PROBLEMAS RESUELTOS DE EXÁMENES DE OPOSICIONES A LOS CUERPOS DE ESTADÍSTICOS DEL ESTADO

ANA CLARA LOPE MARISCAL • JULIO CÉSAR HERNÁNDEZ SÁNCHEZ

PROBLEMAS RESUELTOS DE EXÁMENES DE OPOSICIONES A LOS CUERPOS DE ESTADÍSTICA DEL ESTADO

Julio César Hernández Sánchez y Ana Clara Lope Mariscal

ISBN: 978-84-09-37546-2

Edición: 1ª

Nº de páginas: 550

Actualizado a 25 de abril de 2024

Ilustración de la portada y diseño gráfico:

Tomás Hernández Sánchez

A mis padres y a Isa,
porque sin ellos hubiera sido
imposible preparar las oposiciones

A mis padres

Julio César Hernández Sánchez

Licenciado en Matemáticas

Ingeniero Técnico en Informática de Sistemas

Estadístico Superior del Estado desde 2005

Delegado Provincial del INE en Zamora

Ana Clara Lope Mariscal

Graduada en Economía

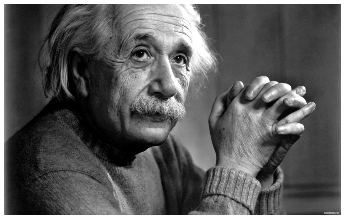
Graduada en Matemáticas - Estadística

Estadístico Superior del Estado desde 2021

I didn't fail the test. I just
found 100 ways to do it wrong.
Benjamin Franklin

Siempre merece la pena un esfuerzo más.

Agradecimientos



Try not to become a man of success but a man of value.

Albert Einstein

Nos gustaría agradecer el esfuerzo de todas las personas que han hecho posible que este libro salga adelante para que pueda ayudar a aquellos opositores ilusionados con este camino. En particular queremos hacer una mención especial a Cristobal Rojas Montoya, que con sus aportaciones ha ayudado a que la versión de este libro haya mejorado, así como a Ignacio González Veiga, Carlos Valero Rodríguez, María Antonia Martínez Luengo y María Dolores Lorca López por su disposición y colaboración en todo aquello que hemos necesitado.

Queremos agradecer también a Miguel Ángel de Castro y a Adolfo Gálvez toda su ayuda y apoyo en este proyecto, que recibieron con entusiasmo y que han impulsado para que pudiera ver la luz, así como a todas las personas del INE que de una u otra forma han colaborado en que esto haya podido ser una realidad.

En particular nos acordamos de nuestras familias, a las que hemos robado mucho tiempo para que esta iniciativa saliera adelante, y que siempre nos apoyan, y a tantos amigos que nos han animado con esta aventura.

Prólogo

Este libro ha nacido para que puedan utilizarlo todos aquellos opositores que estén preparando el acceso a los cuerpos de estadística del estado, aunque por su carácter multidisciplinar puede ser de utilidad en otros ámbitos.

Esperamos que tengáis la bondad de perdonar los errores que pudiéramos haber cometido al elaborarlo.

También nos gustaría agradecer las correcciones, aportaciones e ideas de quien en el uso de este libro pudiera detectar alguna cuestión que fuera susceptible de cambio con el objetivo de mejorar este material. Todo aquello que recibamos será considerado para publicar futuras versiones mejoradas y que puedan seguir siendo útiles para toda la comunidad estadística. Nuestro contacto es el siguiente:

juhersan.zz@gmail.com

anaclopem@gmail.com

Este libro se irá actualizando con los exámenes que cada año vayan apareciendo en las sucesivas convocatorias con la idea de crear una herramienta útil y actual para el opositor/a.

Índice general

Capítulo 1

Introducción

"True science teaches, above all, to doubt and to be ignorant."

Miguel de Unamuno

Actualmente los procesos selectivos a los Cuerpos de Estadísticos del Estado se ajustan a lo dispuesto en un Real Decreto de 2015, de 30 de octubre, por el que se aprueba el texto refundido de la Ley del Estatuto Básico del Empleado Público, y constan de una fase de oposición, tanto para los aspirantes que se presenten por el sistema general de acceso libre, como para los aspirantes que se presenten por la convocatoria de promoción interna. Además el proceso selectivo, para aquellos que superen la fase de oposición, incluye un curso selectivo que se valora y que influye en la puntuación final de los candidatos.

La fase de oposición al Cuerpo Superior de Estadísticos del Estado se estructura en un primer examen de preguntas cortas, seguido de un examen de problemas, uno de idiomas y un examen oral, siendo todos ellos eliminatorios de forma que sólo aquellos aspirantes que superen todos y cada uno de ellos podrán acceder a dicho cuerpo; en las oposiciones al Cuerpo de Diplomados en Estadística del Estado hay dos exámenes de preguntas cortas sobre diversos bloques de contenido y uno de problemas que engloba a todas las áreas de conocimiento del temario. En los últimos años, no obstante, ha variado el orden de algunos de estos exámenes, aunque a día de hoy este es el planteamiento. La incorporación de exámenes de preguntas cortas ha sustituido a pruebas consistentes en redactar uno o varios temas de diversas áreas temáticas, pero lo que ha permanecido constante a lo largo de muchos años, y sigue vigente actualmente son los exámenes de problemas en una u otra fase de cada proceso selectivo.

Nuestra experiencia cuando tuvimos que preparar las oposiciones se cimentó en la utilización de un gran número de textos de diversa índole, tanto de teoría como de problemas, pero dada la diversidad del temario no existen referencias que cubran conjuntamente aspectos de campos tan dispares como demografía, econometría, economía, estadística y muestreo. Es decir, que para profundizar en el conocimiento de cada parte se necesita bibliografía especializada. Y esta situación se traslada de igual forma a la preparación del examen de problemas, teniendo que trabajar para cada apartado en la búsqueda de problemas que pudieran ser objeto de ser seleccionados en estas pruebas. No hay ningún material que aglutine de alguna manera el conocimiento necesario de lo que ha sido susceptible de ser preguntado y tampoco hay material que explique y resuelva los problemas en los diversos procesos de oposición a lo largo de los años. De esta forma, al opositor/a le resulta muy complicado poder tener cierta certeza de cómo se resuelven aquellas cuestiones de ramas tan diversas.

Por este motivo, este libro se enfoca en dichas pruebas, que en el acceso a ambos cuerpos corresponden a uno de los exámenes eliminatorios. En el caso del Cuerpo de Diplomados es el tercer y último examen de la oposición, en el cual se decide el proceso y cobra una importancia vital, y en el caso del Cuerpo Superior suele ser el segundo examen, que es la puerta al examen de idiomas y al oral final.

El objetivo del libro no es otro que el de intentar que los opositores cuenten con un material de ayuda que les acompañe en todo el proceso y en su camino de preparación y que de alguna forma les genere seguridad sobre cómo abordar esta parte de la oposición, puesto que un examen de problemas no está exento de cierta incertidumbre.

Los cuerpos de estadística del estado para los que este material pretende ser de ayuda han sufrido muchos cambios a lo largo de los últimos tres siglos. Los antecedentes se remontan a 1856 cuando se creó la Comisión Estadística General del Reino, que estaba compuesta por un grupo de especialistas en estadística que se dedicaron a supervisar, coordinar y asesorar a los diversos Departamentos de la Administración en el ámbito estadístico.

Los Cuerpos de Estadística en sus orígenes estuvieron muy vinculados a lo que hoy conocemos como Catastro, de forma que había una interrelación entre la topografía catastral, la geografía y la estadística. En 1870 se creó una Dirección General de Estadística Autónoma y se publicó un listado con los funcionarios que llevaron a cabo trabajos estadísticos y topográficos. Pero es en 1873 cuando se vio conveniente crear un Cuerpo de Estadística con diversas categorías administrativas, cómo había que acceder, así como las retribuciones y otros aspectos de relevancia.

Fue en 1902 cuando un Real Decreto del entonces Ministerio de Instrucción Pública y Bellas Artes establecía una separación del Cuerpo de Estadística en dos partes: el Cuerpo Facultativo de Estadística y el Cuerpo Auxiliar de Estadística, fijando las plantillas de ambos.

En los años treinta la estadística tuvo una gran explosión con los avances de la estadística matemática, así como de otros campos como la probabilidad y la inferencia, pero la guerra civil afectó a todo el engranaje estadístico de forma considerable. Se incorporaron, no obstante, máquinas para el procesamiento de información que hasta entonces se hacía de forma manual, contribuyendo a la formación del Censo de Población en 1930.

El Real Decreto de 29 de enero de 1930 del Ministerio de Trabajo y Previsión aprobó un reglamento que establecía las funciones de los Cuerpos Facultativo y Técnico de Ayudantes de Estadística, incluyendo aspectos sobre el ingreso, vacantes, tribunales de oposición etc...

El ingreso en el Cuerpo Facultativo de Estadística se llevaría a cabo por oposición con ciertos requisitos sobre la edad y la formación de los candidatos, teniendo que superar una serie de pruebas sobre álgebra, cálculo, estadística, economía, derecho y geografía. Para el Cuerpo de Ayudantes de Estadística también se estableció un sistema de oposición, pero no se requería titulación académica.

En 1942 se creó el Cuerpo Técnico de Estadística y Colocación que requería para poder formar parte de él también un proceso de concurso-oposición para el que era necesario tener un título universitario. Es en la Ley Estadística de 1945 cuando queda reflejado que los trabajos del INE deberían ser llevados a cabo por dos cuerpos de nueva creación, que fueron el Cuerpo de Estadísticos Facultativos y el Cuerpo de Estadísticos Técnicos. Al primero le correspondería el trabajo de dirección y de investigación y el acceso al mismo sería mediante oposición entre titulados de grado superior. Respecto al segundo, se encargaría de la ejecución técnica de los diversos trabajos estadísticos.

En los años 60 se produjo un gran cambio en los cuerpos de estadística puesto que había que acometer nuevos retos profesionales e institucionales, a la vez que una evolución en la legislación del funcionariado también requería diversas adaptaciones. El INE inicia operaciones importantes en esos años, como la Encuesta de Población Activa(EPA), pero se producen situaciones que restan importancia al papel que la institución tenía y trataba de adquirir. Otros organismos e incluso los propios Ministerios comienzan a elaborar estadísticas de diversa índole al margen del INE, de forma que poco a poco esa labor que se supone debía tener la institución como responsable del sistema estadístico se va debilitando en cierto modo.

Para el acceso al Cuerpo de Estadísticos Facultativos, en los años 60, se requería un título universitario de grado superior, pero a partir del 68 se limitaba fundamentalmente, salvo alguna excepción en alguna convocatoria, a licenciados en matemáticas y en ciencias políticas, económicas y comerciales. El temario fue cambiando con el paso del tiempo, y se introdujeron materias como informática y econometría a las ya existentes de derecho, demografía, economía y estadística. Esta circunstancia hizo que la dificultad en las mismas fuera en aumento puesto que no se adecuaba el temario a ningún plan de las carreras universitarias existentes. La plantilla del Cuerpo era de 200 funcionarios en 1945 y se amplió en 40 plazas en 1973, con la idea a partir de 1974 de dotar de 10 por año al cuerpo, pero muchas quedaron sin cubrir por la dificultad de los conocimientos necesarios para el acceso, así como de las pruebas selectivas.

En cuanto al Cuerpo de Estadísticos Técnicos, inicialmente se requería un título de enseñanza media o asimilado, pero a partir de 1968 había que poseer un título de Bachiller Superior, Perito Mercantil o Maestro de Primaria, a la vez que se actualizaron los contenidos y exámenes de la oposición, que se centrarían en Matemáticas, Estadística, Legislación y Economía fundamentalmente. La plantilla inicial fue de 300 funcionarios y la Ley del 73 la incrementó en 120 plazas, que se convocarían a partir de 1974 de forma gradual.

El inicio del uso de los ordenadores en el INE se remonta al año 1964, en el que se utilizó un equipo IBM con tarjetas perforadoras y 16Kb de memoria y para el que fue necesario formar a un equipo de 14 personas, entre facultativos y técnicos. En la década de los 70 se van sustituyendo esos equipos por otros más modernos y con más capacidad, que van dotando al organismo de experiencia en la evolución de las posibilidades técnicas que van ofreciento estos avances.

En el año 77, mediante Real Decreto, se creó el Cuerpo de Estadísticos Técnicos Diplomados, extinguiéndose así el Cuerpo de Estadísticos Técnicos. Una orden en ese mismo año regularía la integración de un cuerpo en el otro.

A finales de esta década se empezó a gestar la elaboración de una nueva ley de estadística que dejase atrás la ley de 1945. Ya la Constitución del 78 establecía las competencias del Estado y entre ellas incluía la estadística para fines estatales, aunque este último término no estaba suficientemente definido. Por este motivo era necesario elaborar una nueva ley de Estadística que pudiera aclarar y desarrollar esta problemática.

Ya en el año 80 se configuraba en los distintos documentos previos a la nueva ley al INE como punto central y pilar fundamental del Sistema

Estadístico Nacional, de forma que sería el Plan Nacional de Estadística el que reflejara las estadísticas con fines estatales, así como aquellas que fueran de interés para las CCAA.

En la década de los 80 se exigió ser Licenciado, Ingeniero, Arquitecto o estudios equivalentes para poder acceder al Cuerpo de Estadísticos Facultativos y al igual que ocurrió anteriormente el temario se fue adaptando progresivamente a los avances en las distintas disciplinas que lo componían, como informática y econometría.

La publicación de la Ley de la Función Estadística Pública tuvo lugar en el año 89, en la cual se elevaba el rango administrativo del INE, que pasaba a ser Subsecretaría de Estado, y establecía la posibilidad de tener autonomía dentro de la Administración, de forma que pasaba a ser un Organismo Autónomo adscrito al Ministerio de Economía y Hacienda.

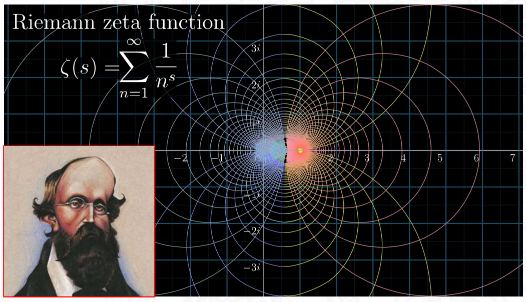
En el año 2001, la Ley de Medidas Fiscales Administrativas y de Orden Social cambia la denominación del Cuerpo de Estadísticos Facultativos, que a partir de ese momento se llamaría Cuerpo Superior de Estadísticos del Estado, e igualmente cambia la del Cuerpo de Estadísticos Técnicos Diplomados por el nombre Cuerpo de Diplomados en Estadística del Estado, que son los nombres actualmente vigentes de los cuerpos sobre los que este libro de problemas hace referencia. Puede consultarse una exposición mucho más extensa y detallada de la historia de los cuerpos de estadística de la Administración General del Estado en referencias como [Celestino, 2010] y [Merediz, 2004].

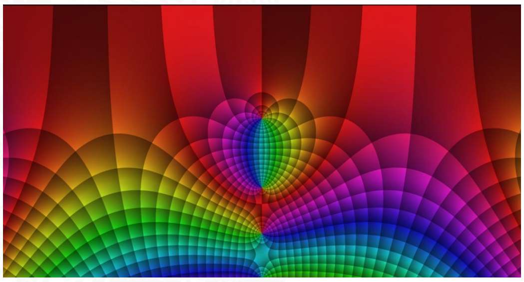
Por último, hemos considerado importante añadir en los anexos de este ejemplar, para cada año, el enunciado de los dos exámenes de problemas (uno del cuerpo de diplomados y el otro del cuerpo superior) tal y como se han presentado a los opositores, teniendo cada capítulo del libro el objetivo de resolver todas o la mayoría de las cuestiones propuestas con el mayor detalle posible. Además iremos actualizando el contenido de esta versión con la resolución de los exámenes que se realicen en futuras convocatorias para que este material se vaya consolidando y sea útil para tantos estudiantes cuyo objetivo es poder formar parte de estos colectivos de la administración.

Capítulo 2

Año 2019

"Todo número entero estrictamente más grande que 1
es la suma de a lo más tres números primos."
—Christian Goldbach, 1742





The Riemann Hypothesis is a mathematical conjecture, first proposed in 1859. In order to prove the Prime Number Theorem, mathematicians needed to study a mathematical object known as the Riemann zeta function. The zeta function was introduced in Riemann's 1859 paper and shown to control the fluctuations of the prime numbers around their "average"behaviour. The zeta function operates on a two-dimensional "number plane" called the complex plane, and associated with it is an infinite set of points known as its "nontrivial zeros"(commonly known as the "zeta zeros" or "Riemann zeros"). The positions of these zeros on the complex plane can be related to an infinite set of wave-like entities which collectively govern the fluctuation of the primes. All of the zeros Riemann was able to calculate lay on a vertical line, and he hypothesised that all of the zeta function's (nontrivial) zeros lie on this "critical line", i.e. the real part of every nontrivial zero of the Riemann zeta function is $\frac{1}{2}$. It remains one of the six unsolved Millennium Prize Problems in mathematics.

$2019 = 3 \cdot 673 = 17^2 + 19^2 + 37^2$

2019 is a polite number. It can be written in 3 ways as a sum of consecutive naturals, for example, $334 + \cdots + 339$

2.1. Soluciones del segundo examen del Cuerpo Superior de Estadísticos del Estado de 2019

Cuestión 1

Sea X una variable aleatoria con ley de probabilidad definida como

$$P_\theta(x) = \left(\frac{\theta}{2}\right)^{|x|} \cdot (1-\theta)^{1-|x|}, \qquad x = -1, 0, 1; \qquad 0 \leq \theta \leq 1$$

Y sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria simple de ella. Resolver razonadamente las siguientes cuestiones:

a) Encontrar el estimador de máxima verosimilitud para θ y obtener la estimación que tendríamos con él si para 20 observaciones de X se obtienen frecuencias $n_1 = 7$, $n_2 = 5$, $n_3 = 8$ para los valores -1, 0 y 1, respectivamente.

b) Encontrar el estimador para θ por el método de los momentos y analizar si es insesgado. Obtener la estimación de θ que propone este estimador para la muestra descrita en el apartado a).

c) Obtener la varianza para ambos estimadores y discutir, para aquel que sea insesgado, si es de mínima varianza

Solución. Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria simple de la variable aleatoria X con ley de probabilidad definida anteriormente, la variable aleatoria X toma los valores -1, 0 y 1 con probabilidades:

$$P_\theta(X=-1) = \left(\frac{\theta}{2}\right)^{1} \cdot (1-\theta)^{1-1} = \frac{\theta}{2}$$

$$P_\theta(X=0)=\left(\frac{\theta}{2}\right)^0\cdot(1-\theta)^{1-0}=1-\theta$$
$$P_\theta(X=1)=\left(\frac{\theta}{2}\right)^1\cdot(1-\theta)^{1-1}=\frac{\theta}{2}$$

a) Para el cálculo del estimador de máxima verosimilitud, $\widehat{\theta}_{MV}$, se comienza considerando el caso general de una muestra de tamaño n para la que se obtienen las frecuencias:

x_i	n_i
-1	n_1
0	n_2
1	n_3

La función de verosimilitud de la muestra es

$$L(X;\theta)=(P(X=-1))^{n_1}\cdot(P(X=0))^{n_2}\cdot(P(X=1))^{n_3}=$$
$$=\left(\frac{\theta}{2}\right)^{n_1}\cdot(1-\theta)^{n_2}\cdot\left(\frac{\theta}{2}\right)^{n_3}=\left(\frac{\theta}{2}\right)^{n_1+n_3}\cdot(1-\theta)^{n_2}$$

y por tanto

$$\ln L(X;\theta)=-(n_1+n_3)\ln 2+(n_1+n_3)\ln\theta+n_2\ln(1-\theta)$$
$$\frac{\partial\ln L(X;\theta)}{\partial\theta}=\frac{n_1+n_3}{\theta}-\frac{n_2}{1-\theta}$$

Así,

$$\frac{\partial\ln L(X;\theta)}{\partial\theta}=0\Leftrightarrow\frac{n_1+n_3}{\theta}=\frac{n_2}{1-\theta}$$
$$\Leftrightarrow(n_1+n_3)(1-\theta)=\theta n_2$$
$$\Leftrightarrow n_1+n_3-(n_1+n_3)\theta=\theta n_2$$
$$\Leftrightarrow n_1+n_3=\theta(n_1+n_2+n_3)$$

Luego

$$\widehat{\theta}_{MV} = \frac{n_1 + n_3}{n_1 + n_2 + n_3} = \frac{n_1 + n_3}{n},$$

es decir, el estimador máximo "verosímil" es la proporción de no ceros en la muestra. En el caso particular de $n = 20$ observaciones para las que se obtienen frecuencias $n_1 = 7$, $n_2 = 5$, $n_3 = 8$, el estimador máximo verosímil sería

$$\widehat{\theta}_{MV} = \frac{n_1 + n_3}{n} = \frac{7+8}{20} = \frac{15}{20} = \frac{3}{4}.$$

b) Para el cálculo del estimador de θ por el método de los momentos, es necesario calcular e igualar los momentos poblacionales a los muestrales. El momento de orden 1 es siempre nulo ya que

$$\alpha_1 = E\left(X\right) = \left(-1\right)\frac{\theta}{2} + 0\left(1-\theta\right) + 1\frac{\theta}{2} = 0,$$

y dado que α_1 no depende de θ, hay que utilizar momentos de orden superior. El momento poblacional de orden 2 viene dado por

$$\alpha_2 = E\left(X^2\right) = \left(-1\right)^2 \cdot \frac{\theta}{2} + 0^2 \cdot \left(1-\theta\right) + 1^2 \cdot \frac{\theta}{2} = \theta.$$

Por otra parte, se calcula el momento muestral de orden 2, que para la muestra dada será:

$$a_2 = \frac{\sum_{i=1}^{n} x_i^2 n_i}{n} = \frac{\left(-1\right)^2 n_1 + 0^2 n_2 + 1^2 n_3}{n} = \frac{n_1 + n_3}{n}.$$

Finalmente, igualando el momento poblacional al momento muestral se obtiene el estimador del método de los momentos:

$$\alpha_2 = a_2 \Leftrightarrow \widehat{\theta}_{MM} = \frac{n_1 + n_3}{n},$$

que coincide con el estimador de máxima verosimilitud. A continuación, se comprueba la insesgadez:

$$E\left(\widehat{\theta}_{MM}\right) = E\left(a_2\right) = \alpha_2 = \theta.$$

Tanto el estimador de máxima verosimilitud como el estimador obtenido por el método de los momentos son insesgados.

c) Visto que ambos son insesgados, se calcula ahora su varianza:

$$Var\left(\widehat{\theta}_{MV}\right) = Var\left(a_2\right) = E\left[a_2^2\right] - \left(E\left[a_2\right]\right)^2 = E\left[a_2^2\right] - \alpha_2^2$$

donde

$$\begin{aligned} E\left[a_2^2\right] =& E\left[\left(\frac{\sum_{i=1}^{n} x_i^2}{n}\right)^2\right] = E\left[\frac{\sum_{i=1}^{n} x_i^4}{n^2} + \frac{\sum_{i\neq j} x_i^2 x_j^2}{n^2}\right] = \\ =& \frac{1}{n}\left[\frac{\sum_{i=1}^{n} E\left(x_i^4\right)}{n} + \frac{\sum_{i\neq j} E\left(x_i^2\right) E\left(x_j^2\right)}{n}\right] = \frac{1}{n}\left[\alpha_4 + \frac{1}{n}\alpha_2^2\binom{n}{2}\right] = \\ =& \frac{\alpha_4}{n} + \frac{n-1}{n}\alpha_2^2, \end{aligned}$$

por tanto,

$$Var(\widehat{\theta}_{MV}) = E\left(a_2^2\right) - \alpha_2^2 = \frac{\alpha_4}{n} + \frac{n-1}{n}\alpha_2^2 - \alpha_2^2 = \frac{\alpha_4 - \alpha_2^2}{n}.$$

Para ver si el estimador tiene varianza mínima se calcula la Cota de Cramer-Rao, que al ser insesgado será $\frac{1}{I(\theta)}$, y sabemos que siempre se cumple que:

$$Var(T) \geq \frac{1}{I\left(\theta\right)}$$

donde la cantidad de información de Fisher, $I\left(\theta\right)$, viene dada por

$$\begin{aligned} I\left(\theta\right) =& -E\left[\frac{\partial^2 \ln f\left(x_1,\ldots .x_n\right)}{\partial\theta^2}\right] = -E\left[\frac{\partial}{\partial\theta}\left(\frac{n_1+n_3}{\theta} - \frac{n_2}{1-\theta}\right)\right] = \\ =& -E\left[\frac{-\left(n_1+n_3\right)}{\theta^2} - \frac{n_2}{\left(1-\theta\right)^2}\right] = E\left[\frac{n_1+n_3}{\theta^2} + \frac{n_2}{\left(1-\theta\right)^2}\right] \end{aligned}$$

Para continuar con el cálculo, hay que tener en cuenta que se dispone de una muestra aleatoria simple $X_1, \ldots, X_n$ y que se ha denotado, por ejemplo, n_1 al "número de valores -1 en la muestra", de forma que el vector (n_1, n_2, n_3) sigue una distribución Multinomial de parámetros n y p es decir

$$(n_1, n_2, n_3) \equiv M(n, p)$$

donde $p = (p_1, p_2, p_3) = \left(\frac{\theta}{2}, 1-\theta, \frac{\theta}{2}\right)$ y así $E[n_1] = n\frac{\theta}{2}$, $E[n_2] = n(1-\theta)$ y $E[n_3] = n\frac{\theta}{2}$. De esta forma,

$$\begin{aligned} I(\theta) =& E\left[\frac{n_1+n_3}{\theta^2} + \frac{n_2}{(1-\theta)^2}\right] = \frac{E(n_1)+E(n_3)}{\theta^2} + \frac{E(n_2)}{(1-\theta)^2} = \\ =& \frac{n\frac{\theta}{2}+n\frac{\theta}{2}}{\theta^2} + \frac{n(1-\theta)}{(1-\theta)^2} = \frac{n}{\theta} + \frac{n}{1-\theta} = \frac{n}{\theta(1-\theta)}. \end{aligned}$$

Así,

$$Var(T) \geq \frac{1}{I(\theta)} = \frac{1}{\frac{n}{\theta(1-\theta)}} = \frac{\theta(1-\theta)}{n} = CCR.$$

Hemos visto que

$$Var\left(\widehat{\theta}_{MM}\right) = \frac{\alpha_4 - \alpha_2^2}{n}$$

donde

$$\begin{aligned} &\alpha_4 = E\left(X^4\right) = (-1)^4\frac{\theta}{2} + 0 + 1^4\frac{\theta}{2} = \theta \\ &\alpha_2 = \theta \Rightarrow \alpha_2^2 = \theta^2; \end{aligned}$$

y por tanto

$$Var\left(\widehat{\theta}_{MM}\right) = \frac{\theta-\theta^2}{n} = \frac{\theta(1-\theta)}{n}.$$

Dado que

$$Var(T) = \frac{1}{I(\theta)} = \frac{\theta(1-\theta)}{n} = CCR,$$

el estimador es eficiente y consecuentemente tiene varianza mínima.

Cuestión 2

En un laboratorio se llevan a cabo ensayos clínicos para comprobar la eficacia de algunos fármacos sobre el control del peso y se sospecha que hay dos de ellos que son equivalentes en cuanto a los efectos que producen. Supongamos que se disponen de las siguientes muestras de pesos medidos en libras.

	Participantes								
	1	2	3	4	5	6	7	8	9
Muestra 1	132	139	126	114	122	132	142	119	126
Muestra 2	124	141	118	116	114	132	145	123	121

Resolver razonadamente si los datos nos sugieren rechazar dicha sospecha a un nivel $\alpha = 0,01$ en las siguientes situaciones:

a) Las muestras 1 y 2, corresponden al peso de las mismas 9 personas adultas, después de administrarle a cada una de ellas los 2 tratamientos. Para ello, se ha diseñado el ensayo de forma que el participante i-ésimo estuviera en idénticas condiciones clínicas (con un peso de partida p_i idéntico) cuando se le administró uno y otro tratamiento.

b) Suponiendo que las muestras 1 y 2 corresponden a 18 personas en una misma situación clínica de partida (con un mismo peso de partida p). Se les divide al azar en dos grupos de 9 de personas cada uno de ellos, y a cada grupo se les administra uno y solo uno de los dos fármacos.

c) Interpretar las discrepancias observadas entre las metodologías de resolución de los apartados a) y b), y de cómo se concretan con las muestras analizadas.

Si fuera necesario, admitir normalidad para la variable peso y homocedasticidad para la variabilidad que presenta por grupo.

Solución.

a) En este primer caso, se trata de un constraste de diferencia de medias para muestras apareadas, puesto que las dos muestras corresponden a las mismas personas tras aplicar los dos tratamientos. No es posible pensar que X e Y son independientes. Por tanto, si queremos ver si los tratamientos son equivalentes, llamamos $d_i = x_i - y_i$ a la diferencia entre las observaciones con uno y otro tratamiento.

Participante	x_i M1	y_i M2	d_i
1	132	124	8
2	139	141	-2
3	126	118	8
4	114	116	-2
5	122	114	8
6	132	132	0
7	142	145	-3
8	119	123	-4
9	126	121	5

El enunciado permite suponer que la variable aleatoria que define la diferencia entre ambos tratamientos es tal que $d \equiv N\left(\mu_d, \sigma_d^2\right)$ con media y varianza desconocidas, y lo que se quiere es contrastar

$$\left\{\begin{array}{l} H_0 : \mu_d = 0 \\ H_1 : \mu_d \neq 0 \end{array}\right. ,$$

para lo cual se utiliza el estadístico de contraste

$$T_{exp} = \frac{\overline{d} - 0}{\frac{\widehat{S}_d}{\sqrt{n}}} \sim t_{n-1}$$

con $\overline{d}$ y $\widehat{S}_d$ la media de las diferencias y la cuasivarianza muestral respectivamente y que toman los siguientes valores:

$$\overline{d} = \frac{8-2+8-2+8+0-3-4+5}{9} = \frac{18}{9} = 2$$

$$\widehat{S}_d^2 = \frac{\sum_{i=1}^{9}\left(d_i - \overline{d}\right)^2}{n-1} = \frac{6^2+4^2+6^2+4^2+6^2+2^2+5^2+6^2+3^2}{8} =$$
$$= \frac{214}{8} = 26,75$$
$$\widehat{S}_d = \sqrt{\widehat{S}_d^2} = 5,172$$

Luego

$$T_{exp} = \frac{2-0}{\frac{5,172}{\sqrt{9}}} = 1,16$$

La región crítica del contraste será:

$$\text{RC} = \{T_{exp} < -t_{n-1,1-\frac{\alpha}{2}}; T_{exp} > t_{n-1,1-\frac{\alpha}{2}}\}$$

y como $\alpha = 0,01 \Rightarrow \frac{\alpha}{2} = 0,005$ y por tanto $t_{n-1;\frac{\alpha}{2}} = t_{8;0{,}995} = 3,355$, entonces no se rechaza la hipótesis nula y la diferencia entre los tratamientos no es significativa.

b) En el segundo caso, las muestras no son apareadas sino que el grupo se divide en dos partes y podemos considerar que son muestras independientes y que a cada una se le suministra un fármaco distinto. Lo que queremos comprobar en este caso es la igualdad de medias entre ambos grupos, para lo cual se efectua una prueba T para la que se suponen varianzas desconocidas pero iguales ($\sigma_1^2 = \sigma_2^2 = \sigma$) ya que el enunciado establece que se puede admitir homocedasticidad para la variabilidad que presenta por grupo. En este caso, el estadístico de contraste sería:

$$T_{exp} = \frac{\left(\overline{Y}_1 - \overline{Y}_2\right) - (\mu_1 - \mu_2)}{\widehat{\sigma}_{\overline{Y}_1 - \overline{Y}_2}} \sim t_{n_1+n_2-2}$$

siendo

$$\widehat{\sigma}_{\overline{Y}_1 - \overline{Y}_2} = \widehat{\sigma}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

y

$$\widehat{\sigma}^2 = \frac{(n_1-1)\widehat{S}^2_{Y_1;n_1-1} + (n_2-1)\widehat{S}^2_{Y_2;n_2-1}}{n_1+n_2-2}$$

donde $\widehat{S}^2_{Y_1;n_1-1}$ y $\widehat{S}^2_{Y_2;n_2-1}$ son las cuasivarianzas muestrales.

El contraste es

$$\begin{cases} H_0 : \mu_1 = \mu_2 \\ H_1 : \mu_1 \neq \mu_2 \end{cases}$$

y la región crítica

$$\text{RC} = \{T_{exp} < -t_{n_1+n_2-2,1-\frac{\alpha}{2}}; T_{exp} > t_{n_1+n_2-2,1-\frac{\alpha}{2}}\}$$

Se calcula el estadístico T:

$$\overline{Y}_1 = \frac{\sum_{i=1}^{9} Y_{1i}}{n_1} = \frac{132+139+\cdots+119+126}{9} = 128$$

$$\overline{Y}_2 = \frac{\sum_{i=1}^{9} Y_{2i}}{n_2} = \frac{124+141+\cdots+123+121}{9} = 126$$

$$\begin{aligned}\widehat{S}^2_{Y_1;n_1-1} &= \frac{\sum_{i=1}^{9}\left(Y_{1i}-\overline{Y}_1\right)^2}{n_1-1} = \\ &= \frac{4^2+11^2+(-2)^2+(-14)^2+(-6)^2+4^2+14^2+(-9)^2+(-2)^2}{8} = \\ &= 83,75\end{aligned}$$

$$\begin{aligned}\widehat{S}^2_{Y_2;n_2-1} &= \frac{\sum_{i=1}^{9}\left(Y_{2i}-\overline{Y}_2\right)^2}{n_2-1} = \\ &= \frac{(-2)^2+15^2+(-8)^2+(-10)^2+(-12)^2+6^2+19^2+(-3)^2+(-5)^2}{8} =\end{aligned}$$

$$= 121$$
$$\widehat{\sigma}^2 = \frac{8 \cdot 83,75 + 8 \cdot 121}{9+9-2} = 102,375$$
$$\widehat{\sigma}_{\overline{Y}_1 - \overline{Y}_2} = \sqrt{102,375}\sqrt{\frac{1}{9}+\frac{1}{9}} = 4,7696$$
$$T_{exp} = \frac{128-126}{4,7696} = 0,4193$$

Como $\alpha = 0,01 \Rightarrow \frac{\alpha}{2} = 0,005$ y por tanto $t_{n_1+n_2-2,1-\frac{\alpha}{2}} = t_{16;0,995} = 2,92078$ de modo que no se rechaza la hipótesis nula de igualdad de medias y puede concluirse que los fármacos son equivalentes.

c) Si las muestras son dependientes, como en el apartado a), se podría pensar que es más probable que las personas con más peso en la primera muestra tengan también más peso en la segunda. En el apartado b), esto no tiene por qué ocurrir puesto que la distribución de los valores de la primera muestra no le indica nada con respecto a la distribución de los valores de la segunda.

La diferencia principal entre ambos métodos es que

$$Var(D) = Var(X-Y) = Var(X) + Var(Y) - 2Cov(X,Y)$$

y este resultado se extiende a las cuasivarianzas, de modo que en la T para muestras apareadas se añade en el denominador la covarianza con signo negativo. Por tanto, si la covarianza entre ambas muestras fuera cero (es decir, que no hubiera relación lineal entre ellas), sería indiferente considerarlas independientes o aparedas. Si la covarianza muestral es distinta de cero, cuanto más negativa sea esta, mayor será el denominador y menor será T haciendo que el test de muestras apareadas sea más conservador que el de muestras independientes. Cuanto más positiva sea la covariaza, ocurrirá lo contrario, esto es, más pequeña será la varianza en el denominador y mayor será el estadístico T siendo el contraste para muestras apareadas menos conservador. La covarianza muestral sería

$$
\begin{aligned}
Cov(Y_1, Y_2) &= \frac{1}{n-1}\left[\sum_{i,j=1}^{n} Y_{1i}Y_{2i} - n\overline{Y}_1\overline{Y}_2\right] = \\
&= \frac{1}{9-1}\left[(132\cdot 124) + (139\cdot 141) + \cdots (126\cdot 121) - 9\cdot 128\cdot 126\right] = \\
&= \frac{145864}{8} - \frac{145152}{8} = \frac{712}{8} = 89
\end{aligned}
$$

De modo que la covarianza es positiva y por tanto, en el caso de muestras apareadas, el denominador del estadístico T es más pequeño, consecuentemente el estadístico T es mayor y por tanto, el contraste para el caso de muestras aparedas, es menos conservador.

Cuestión 3

A) En una población con 5.000 viviendas, determinar el tamaño muestral necesario para que, con un nivel de confianza del 95 % $\left(F^{-1}_{N(0,1)}\left(1-\frac{0{,}05}{2}\right) = 1,96\right)$, la estimación de la proporción de viviendas en alquiler no difiera en más del $0,1$ del valor verdadero. El muestreo se realiza sin reposición. No tenemos información sobre la proporción poblacional P de viviendas en alquiler y por tanto se supone que $P = 1/2$.

B) La población de las 5.000 viviendas se estratifica en 3 grupos. Repartir el tamaño muestral obtenido en el apartado A) entre los estratos aplicando una afijación de mínima varianza o de Neyman y aplicando una afijación proporcional. Para ello se conoce el número de viviendas por estrato (N_h) y la proporción de viviendas de alquiler por estrato (P_h). Comentar los resultados.

N_h	P_h
2500	0,8
2500	0,5
500	0,1

Solución.

A) Se tiene que N $= 5000$, $\alpha = 5\,\%$, $\lambda_\alpha = F^{-1}_{N(0,1)}\left(1-\frac{0{,}05}{2}\right)$ y la proporción poblacional de viviendas en alquiler $P = \dfrac{1}{2}$. El objetivo es encontrar n tal que $\mathrm{P}\left(|\widehat{P}-P| \leq 0,1\right) \geq 0,95$. De esta forma, se halla n con un grado de tolerancia definido por el coeficiente de confianza $1-\alpha = 0,95$ y donde el error máximo admisible es $e = 0,1$. Se tiene que:

$$\mathrm{P}\left(|\widehat{P}-P| \leq e\right) = 1-\alpha \Rightarrow \mathrm{P}\left(-e \leq \widehat{P}-P \leq e\right) = 1-\alpha \Rightarrow$$
$$\Rightarrow \mathrm{P}\left(-\frac{e}{\sigma\left(\widehat{P}\right)} \leq \frac{\widehat{P}-P}{\sigma\left(\widehat{P}\right)} \leq \frac{e}{\sigma\left(\widehat{P}\right)}\right) = 1-\alpha$$

donde $\lambda_\alpha = \dfrac{e}{\sigma\left(\widehat{P}\right)}$ y por tanto $e = \lambda_\alpha \sigma\left(\widehat{P}\right)$. La varianza del estimador de la proporción en este tipo de muestreo es:

$$\begin{aligned}
Var\left(\widehat{P}\right) &= \left(1-\frac{n}{N}\right)\left(\frac{N}{N-1}\right)\frac{PQ}{n}\\
&= \left(\frac{N-n}{N}\right)\left(\frac{N}{N-1}\right)\frac{PQ}{n}\\
&= \left(\frac{N-n}{n}\right)\left(\frac{1}{N-1}\right)PQ\\
&= \left(\frac{N}{n}-1\right)\left(\frac{1}{N-1}\right)PQ
\end{aligned}$$

De forma que

$$e = \lambda_\alpha \sqrt{Var\left(\widehat{P}\right)} \Rightarrow e^2 = \lambda_\alpha^2 \left(\frac{N}{n}-1\right)\left(\frac{1}{N-1}\right)PQ$$

$$\Rightarrow \frac{e^2(N-1)}{\lambda_\alpha^2 \cdot P \cdot Q} = \left(\frac{N}{n} - 1\right)$$
$$\Rightarrow \frac{N}{n} = 1 + \frac{e^2(N-1)}{\lambda_\alpha^2 \cdot P \cdot Q} = 1 + \frac{N-1}{\frac{\lambda_\alpha^2 \cdot P \cdot Q}{e^2}}$$

Denotando $n_0 = \dfrac{\lambda_\alpha^2 \cdot P \cdot Q}{e^2}$,

$$\frac{N}{n} = 1 + \frac{N-1}{n_0} \Leftrightarrow n = \frac{N}{1 + \frac{N-1}{n_0}} = \frac{N \cdot n_0}{n_0 + (N-1)}$$

donde $n_0 = \dfrac{1,96^2 \cdot \frac{1}{2} \cdot \frac{1}{2}}{0,1^2} = 96,04$ y así

$$n = \frac{N \cdot n_0}{n_0 + (N-1)} = \frac{5000 \cdot 96,04}{96,04 + (4999)} = \frac{480200}{5095,04} = 94,248.$$

Por lo que se tomará un tamaño muestral de $n = 95$ viviendas.

B) La población de las 5.000 viviendas se estratifica en 3 grupos, considerando el tamaño muestral $n = 95$ obtenido en el apartado A), se aplica la afijación de mínima varianza o de Neyman. En la tabla del enunciado se proporciona el número de viviendas por estrato, (N_h), y la proporción de viviendas de alquiler por estrato, (P_h). Sin embargo, estos datos no son válidos pues no se cumple que $\displaystyle\sum_{h=1}^{3} N_h = N$. Suponemos que se trata de una errata y proponemos la siguiente división de viviendas por estrato

N_h	P_h
2500	0,8
2000	0,5
500	0,1

de forma que ahora sí $\sum_{h=1}^{3} N_h = 2500 + 2000 + 500 = 5000 = N$.

Para aplicar la afijación de Neyman es necesario calcular

$$\begin{cases} \min Var\left(\overline{x}_{st}\right) = \min \sum_{h=1}^{L} W_h^2 \frac{S_h^2}{n_h}\left(1 - \frac{n_h}{N_h}\right) \\ s.a. \sum_{h=1}^{L} n_h = n \end{cases}$$

Resolviendo dicha minimización se obtiene que

$$n_h = n\frac{N_h S_h}{\sum_{h=1}^{L} N_h S_h}.$$

Para obtener el tamaño muestral de cada uno de los estratos es necesario calcular las cuasidesviaciones típicas poblacionales S_1, S_2 y S_3. La cuasivarianza de cada estrato viene dada por

$$S_h^2 = \frac{N_h}{N_h - 1}\sigma_h^2 = \left(\frac{N_h}{N_h - 1}\right) P_h Q_h$$

de forma que para $h = 1$, $h = 2$ y $h = 3$ respectivamente se tiene:

$$S_1^2 = \frac{2500}{2499} \cdot 0,8 \cdot 0,2 = 0,160064 \Rightarrow S_1 \simeq 0,4$$
$$S_2^2 = \frac{2000}{1999} \cdot 0,5 \cdot 0,5 = 0,250125 \Rightarrow S_2 \simeq 0,5$$
$$S_3^2 = \frac{500}{499} \cdot 0,1 \cdot 0,9 = 0,09018 \Rightarrow S_3 \simeq 0,3$$

Consecuentemente, el denominador del cociente resultado de resolver la minimización es

$$\sum_{h=1}^{3} N_h S_h = 2500 \cdot 0,4 + 2000 \cdot 0,5 + 500 \cdot 0,3 = 2150.$$

Así, para cada uno de los estratos se tiene

$$n_1 = 95\frac{2500 \cdot 0,4}{2150} = 44,186 \simeq 44$$
$$n_2 = 95\frac{2000 \cdot 0,5}{2150} = 44,186 \simeq 44$$
$$n_3 = 95\frac{500 \cdot 0,3}{2150} = 6,627 \simeq 7$$

y la afijación de Neyman es tomar 44 viviendas en el estrato 1, 44 viviendas en el estrato 2 y 7 viviendas en el estrato 3.

Por otra parte, si se aplica una afijación proporcional el tamaño muestral de cada estrato vendrá dado por

$$n_h = kN_h \qquad \text{para } h = 1, 2, 3.$$

Sumando para todos los estratos, debe cumplirse que $n = kN$. Por ser $n = 95$ y $N = 5000$, se tiene que $k = 0,019$. De esta forma, la afijación proporcional determina los siguientes tamaños muestrales:

$$n_1 = 0,019 \cdot 2500 = 47,5 \simeq 47$$

$$n_2 = 0,019 \cdot 2000 = 38$$

$$n_3 = 0,019 \cdot 500 = 9,5 \simeq 9$$

Es decir, 47 viviendas en el estrato 1, 38 viviendas en el estrato 2 y 9 viviendas en el estrato 3. Esta configuración sumaría 94 en lugar de 95 debido al valor de k, puesto que aparecen decimales. El elemento de muestra que falta podría extraerse del estrato 1 o del 3.

Es posible evaluar la diferencia en las varianzas al aplicar la afijación de mínima varianza o de Neyman y al aplicar la afijación proporcional. Siendo $\sigma_h^2 = P_h Q_h$, la varianza de la afijación proporcional para la propoción de clase en muestreo con reposición es:

$$Var_{PROP}(\widehat{P}) = \frac{1}{n} \sum_{h=1}^{L} W_h \sigma_h^2,$$

mientras que la varianza de la afijación de Neyman para la estimación de la propoción de clase en muestreo con reposición es:

$$Var_{NEY}(\widehat{P}) = \frac{1}{n} \left(\sum_{h=1}^{L} W_h \sigma_h \right)^2 .$$

La diferencia entre ambas es:

$$
\begin{aligned}
Var_{PROP}(\widehat{P}) - Var_{NEY}(\widehat{P}) &= \frac{1}{n}\sum_{h=1}^{L} W_h\sigma_h^2 - \frac{1}{n}\left(\sum_{h=1}^{L} W_h\sigma_h\right)^2 = \\
&= \frac{1}{n}\left(\sum_{h=1}^{L} W_h\sigma_h^2 - \left(\sum_{h=1}^{L} W_h\sigma_h\right)^2\right) = \\
&= \frac{1}{n}\sum_{h=1}^{L} W_h\left(\sigma_h - \overline{\sigma}\right)^2 \geq 0
\end{aligned}
$$

donde $\overline{\sigma} = \sum_{h=1}^{L} W_h\sigma_h$. El último paso de la demostración anterior se justifica como sigue:

$$
\sum_{h=1}^{L} W_h\left(\sigma_h - \overline{\sigma}\right)^2 = \sum_{h=1}^{L} W_h\sigma_h^2 - 2\overline{\sigma}\sum_{h=1}^{L} W_h\sigma_h + \overline{\sigma}^2\sum_{h=1}^{L} W_h.
$$

Teniendo en cuenta que $\sum_{h=1}^{L} W_h\sigma_h = \overline{\sigma}$ y que $\sum_{h=1}^{L} W_h = 1$, se tiene que:

$$
\begin{aligned}
\sum_{h=1}^{L} W_h\left(\sigma_h - \overline{\sigma}\right)^2 &= \sum_{h=1}^{L} W_h\sigma_h^2 - 2\overline{\sigma}^2 + \overline{\sigma}^2 = \sum_{h=1}^{L} W_h\sigma_h^2 - \overline{\sigma}^2 = \\
&= \sum_{h=1}^{L} W_h\sigma_h^2 - \left(\sum_{h=1}^{L} W_h\sigma_h\right)^2.
\end{aligned}
$$

La diferencia entre la varianza al aplicar la afijación proporcional y la varianza al aplicar la afijación de Neyman es siempre mayor o igual a cero dándose la igualdad únicamente cuando $\sigma_h = \overline{\sigma}$, $\forall h = 1, \ldots, L$. Por tanto, la diferencia viene dada en función de la distribución de los valores $\sigma_1, \ldots, \sigma_L$. La mayor dispersión de estos valores proporcionará mejores resultados a la afijación de mínima varianza con respecto a la proporcional.

Cuestión 4

Sea una población de 500 empresas con un total de 6.000 de asalariados. Se toma una muestra aleatoria simple sin reemplazamiento de 25 empresas con el objetivo de estimar el total de la cifra de negocios. Se denota por y_i la cifra de negocios de la empresa i, en miles de euros y x_i el número de asalariados de la empresa i. Los datos disponibles son:

$$\sum_{i=1}^{25} x_i = 265; \quad \sum_{i=1}^{25} y_i = 710; \quad \sum_{i=1}^{25} x_i y_i = 9772; \sum_{i=1}^{25} x_i^2 \quad = 4147; \sum_{i=1}^{25} y_i^2 = 25702;$$

Se pide:

a) Estimar el total de la cifra de negocios y su error de muestreo usando el método de la razón.

b) Estimar el total de la cifra de negocios y su error de muestreo usando el método de regresión lineal.

c) Comentar cuál de los dos métodos es más eficiente.

Solución. Del enunciado se sabe que es un muestreo aleatorio simple sin reposición con $N = 500$, $X_{asal} = 6000$ y $n = 25$.

a) Usando el método de la razón:

$$\widehat{Y}_R = \widehat{R} \cdot X = \frac{\sum_{i=1}^{25} y_i}{\sum_{i=1}^{25} x_i} \cdot X = \frac{\overline{y}}{\overline{x}} \cdot X = \frac{710}{265} \cdot 6000 = 16075,47$$

$$\begin{aligned}\widehat{Var}\left(\widehat{Y}_R\right) =& X^2 \cdot \widehat{Var}\left(\widehat{R}\right) = X^2 \frac{1}{\overline{X}^2}\left(\frac{1-f}{n}\right)\left[\widehat{S}_Y^2 + \widehat{R}^2 \widehat{S}_X^2 - 2\widehat{R}\widehat{S}_{XY}\right] = \\ =& \frac{N^2(1-f)}{n(n-1)}\left[\sum_{i=1}^{n} y_i^2 + \widehat{R}^2 \sum_{i=1}^{25} x_i^2 - 2\widehat{R}\sum_{i=1}^{25} x_i y_i\right] = \\ =& \frac{500^2\left(1-\frac{25}{500}\right)}{25(24)}\left[25702 + \left(\frac{710}{265}\right)^2 4147 - 2\left(\frac{710}{265}\right) 9772\right] =\end{aligned}$$

$$=1230040,035.$$

De este modo, el error se muestreo es

$$\sqrt{\widehat{Var}\left(\widehat{Y}_R\right)} = 1109,0717.$$

b) Usando el método de regresión lineal:

$$\widehat{Y}_{reg} = N \cdot \overline{y}_{reg}, \text{ donde } \overline{y}_{reg} = \overline{y} + b_0\left(\overline{X} - \overline{x}\right)$$

siendo

$$\overline{y} = \frac{710}{25} \qquad \overline{x} = \frac{265}{25} \qquad \overline{X} = \frac{6000}{500} = 12$$

y la pendiente de la recta de regresión de Y sobre X se estima por

$$\widehat{b}_0 = \frac{\widehat{S}_{XY}}{\widehat{S}^2_X} = \frac{93,583}{55,75} = 1,768$$

ya que

$$\widehat{S}_{XY} = \frac{1}{n-1}\left[\sum_{i=1}^{25} x_i y_i - n\overline{x}\overline{y}\right] = \frac{1}{24}\cdot 9772 - \frac{25}{24}\cdot\frac{265}{25}\cdot\frac{710}{25} = 93,583$$
$$\widehat{S}^2_X = \frac{1}{n-1}\left[\sum_{i=1}^{25} x_i^2 - n\overline{x}\right] = \frac{1}{24}\left[4147 - 25\left(\frac{265}{25}\right)^2\right] = 55,75$$
$$\widehat{S}^2_Y = \frac{1}{n-1}\left[\sum_{i=1}^{25} y_i^2 - n\overline{y}\right] = \frac{1}{24}\left[25702 - 25\left(\frac{710}{25}\right)^2\right] = 230,75$$

Por tanto,

$$\overline{y}_{reg} = \frac{710}{25} + \frac{93,583}{55,75}\left(12 - \frac{265}{25}\right) = 30,75$$

y así

$$\widehat{Y}_{reg} = N \cdot \overline{y}_{reg} = 500 \cdot 30,75 = 15375.$$

En cuanto a la varianza, se sabe que

$$Var_{min}\left(\overline{y}_{reg}\right) = \frac{1-f}{n}\left(1-\rho^2\right) S^2_Y$$

$$\widehat{Var}_{min}\left(\overline{y}_{reg}\right) = \frac{1-f}{n}\left(1-\widehat{\rho}^2\right)\widehat{S}_Y^2, \text{ con } \widehat{\rho} = \frac{\widehat{S}_{XY}}{\widehat{S}_X\widehat{S}_Y}$$

de modo que

$$\widehat{Var}\left(\widehat{Y}_{reg}\right) = N^2\widehat{Var}\left(\overline{y}_{reg}\right).$$

Sustituyedo se tiene:

$$\widehat{Var}\left(\widehat{Y}_{reg}\right) = 500^2\frac{1-\frac{25}{500}}{25}\left(1-\frac{93,583^2}{55,75\cdot 230,75}\right)230,75 = 699800,1233.$$

De forma que el error se muestreo es

$$\sqrt{\widehat{Var}\left(\widehat{Y}_{reg}\right)} = 836,54.$$

c) Para comparar la precicisón de la estimación usando el método de la razón y la estimación usando el método de regresión, se comparan las expresiones de las varianzas en cada uno de los métodos.

$$Var\left(\widehat{Y}_R\right) = N^2\left(\frac{1-f}{n}\right)\left[S_Y^2+R^2S_X^2-2RS_{XY}\right]$$
$$Var_{min}\left(\widehat{Y}_{reg}\right) = N^2\left(\frac{1-f}{n}\right)\left(1-\rho^2\right)S_Y^2$$

Calculando la diferencia entre ambas varianzas se tiene

$$\begin{aligned}
&Var\left(\widehat{Y}_R\right) - Var_{min}\left(\widehat{Y}_{reg}\right) = \\
&\qquad = N^2\left(\frac{1-f}{n}\right)\left[S_Y^2+R^2S_X^2-2RS_{XY}-\left(1-\rho^2\right)S_Y^2\right] = \\
&\qquad = N^2\left(\frac{1-f}{n}\right)\left[S_Y^2+R^2S_X^2-2RS_{XY}-S_Y^2+\rho^2S_Y^2\right] = \\
&\qquad = N^2\left(\frac{1-f}{n}\right)\left[R^2S_X^2-2RS_{XY}+\rho^2S_Y^2\right] = \\
&\qquad = N^2\left(\frac{1-f}{n}\right)\left(RS_X-\rho S_Y\right)^2 \geq 0.
\end{aligned}$$

La desigualdad siempre es cierta, es decir la varianza del estimador de regresión será siempre menor a la varianza del estimador de razón. Efectivamente, se observa en los resultados obtenidos, que el error de muestreo del estimador utilizando el método de la razón es mayor que el obtenido por el método de regresión.

$$\sqrt{\widehat{Var}\left(\widehat{Y}_R\right)} = 1109,0717 > \sqrt{\widehat{Var}\left(\widehat{Y}_{reg}\right)} = 836,54.$$

La igualdad de precisiones en la estimación por razón y por regresión se produce únicamente si

$$RS_X - \rho S_Y = 0 \Leftrightarrow R = \rho \frac{S_Y}{S_X} = \beta,$$

es decir cuando la recta de regresión de Y sobre X pase por el origen pues si $R = \beta$, la recta de regresión de Y sobre X tiene de ecuación $Y = \beta X + \overline{Y} - \beta \overline{X}$ y la ordenada al origen de la recta de regresión de Y sobre X valdrá $\overline{Y} - \beta \overline{X} = \overline{Y} - R\overline{X} = \overline{Y} - \overline{Y} = 0$.

Cuestión 5

Se dispone de la siguiente información respecto a los agregados de una economía (datos en miles de millones de euros):

Producción de bienes y servicios (a precios básicos)	P.1	2.038
Consumo intermedio (a precios de adquisición)	P.2	1.048
Variación de existencias	P.52	5
Formación bruta de capital fijo	P.51g	249
Gasto en consumo final de los hogares	P.3	608
Gasto en consumo final de las ISFLSH*	P.3	11
Remuneración de los asalariados	D.1	541
Adquisiciones menos cesiones de objetos valiosos	P.53	1
Exportaciones de bienes y servicios	P.6	276
Importaciones de bienes y servicios	P.7	290
Excedente bruto de explotación y renta mixta	B2G+B3G	446
Impuestos sobre producción e importaciones	D.2	112
Subvenciones	D.3	18
Impuestos sobre los productos	D.21	97

*Instituciones sin Fines de Lucro al Servicio de los Hogares

Se pide:

1. Calcule el PIB a precios de mercado (B1*G).
2. Calcule el valor de las subvenciones a los productos (D.31).
3. Calcule la formación bruta de capital (P.5).
4. Calcule el valor del gasto en consumo final de las Administraciones Públicas (P.3).

Solución.

1. PIBpm (B1*G) = RA (D.1) + EBE (B2G+B3G) + Impuestos sobre producción e importaciones (D.2) – Subvenciones (D.3)

   ⇒ PIBpm (B1*G) = 541 + 446 + 112 – 18 = 1.081

2. Utilizando la cuenta de producción se tiene el resultado:

E	Cuenta de producción		R
1.048	Consumos intermedios (precios adq.) (P.2)	Producción de bienes y servicios (a precios básiscos) (P.1)	2.038
		Imps. sobre los productos (D.21)	97
		Subs. a los productos (D.31)	x
1.081	PIB pm (B1*G)		

   El total de empleos debe ser igual al total de recursos de modo que, denotando por x a las Subvenciones a los productos (D.31),

$$1048 + 1081 = 2038 + 97 + x \Leftrightarrow x = 1048 + 1081 - 2038 - 97 = -6$$

   Por tanto, el valor de las Subvenciones a los productos (D.31) es 6.

3. La formación bruta de capital (P.5) comprende:

   (P.51g) la formación bruta de capital fijo;

   (P.52) la variación de existencias;

   (P.53) las adquisiciones menos cesiones de objetos valiosos.

   ⇒ FBC(P.5) = P.52 + P.51g + P.53 = 5 + 249 + 1 = 255.

4. De la siguiente igualdad,

   PIBpm (B1*G) = C(P.3) + FBC (P.5) +X (P.6) – M (P.7), se deduce el valor del Gasto en Consumo Final:

$$\text{C (P.3)} = 1081–255 - 276 + 290 = 840$$

que se descompone en $C = C_{\text{AAPP}} + C_{\text{ISFLSH}} + C_{\text{hogares}}$. Así,

$$C_{\text{AAPP}} = 840–608–11 = 221$$

Cuestión 6

Registre las transacciones descritas a continuación, correspondientes al hogar de los Martínez durante el año de referencia t, en la secuencia de cuentas del sector hogares. Deberá comenzar por la cuenta de producción e ir avanzando especificando la cuenta de explotación, la cuenta de asignación de la renta primaria, la cuenta de distribución secundaria de la renta hasta finalizar con la cuenta de utilización de la renta disponible.

El hogar de los Martínez está compuesto por la pareja formada por Juan y María y sus hijos, Carlota y Miguel.

Durante el año t:

María ha recibido un salario de 2.000 unidades monetarias (u.m). Su empleador ha pagado 20 u.m. en concepto de cotizaciones sociales. María ha pagado 25 u.m. de impuesto sobre la renta y 15 u.m. de cotización social. Ha gastado 100 u.m. en transporte, y comidas fuera del hogar. También se ha comprado una moto por 280 u.m. que ha financiado con un préstamo por el que paga de interés 5 u.m. durante todo el año. El resto de su sueldo se lo cede a Juan que ha sido el responsable este año de llevar las cuentas en el hogar.

Juan ha estado desempleado recibiendo durante el año una prestación por desempleo de 350 u.m. Ha gastado 1100 u.m. en alimentos, bebidas, artículos de droguería y vestido y calzado para toda la familia y 900 u.m. en alquiler de la vivienda.

Carlota ha recibido pagas de sus padres por valor de 30 u.m. que ahorra para poder comprarse una bicicleta en el futuro.

Miguel no tiene empleo fijo pero hace ocasionalmente trabajos de pintura de pequeña entidad para los vecinos, que no declara. Ha cobrado por su producción 800 u.m. pero ha gastado 300 u.m. en pintura y brochas. Ocasionalmente ha llamado a un amigo para que le ayude y le ha pagado 40 u.m. Por otra parte, ha gastado 200 u.m. en bebidas, tabaco y tickets para eventos deportivos.

Solución.

María ha recibido un salario de 2.000 unidades monetarias (u.m), que se registra en recursos en la cuenta de asignación de la renta primaria. Su empleador ha pagado 20 u.m. en concepto de cotizaciones sociales. Por una parte, María recibe estas 20 u.m. como D.12 dentro de la remuneración y, por otra, las cotizaciones sociales pagadas a cargo de los empleadores se contabilizan en la cuenta de distribución secundaria de la renta dentro de D.61 como un empleo para los hogares. En realidad el flujo de este tipo de cotizaciones es de las empresas al Estado, pero en la contabilidad iría del empresario al trabajador primero y posteriormente de este a la Seguridad Social porque es una remuneración suya.

María ha pagado 25 u.m. de impuestos sobre la renta, que se registra como empleos en la cuenta de distribución secundaria de la renta del contribuyente. Además María ha pagado 15 u.m. de cotización social. En el sistema de cuentas, las cotizaciones sociales efectivas a cargo de los hogares se registran en los empleos de la cuenta de distribución secundaria de la renta de los hogares. Podríamos matizar que esas 15 u.m. de cotizaciones sociales se encuentran dentro de la partida D.11 (sueldos y salarios), que después se transfieren a la Seguridad Social.

Por otra parte, María ha gastado 100 u.m. en transporte, y comidas fuera del hogar. Este gasto representa adquisiciones de los hogares sin reembolso que se utilizan para satisfacer directamente las necesidades individuales o colectivas de los miembros del hogar y que dentro del SEC se clasifican como gasto en consumo final. También se ha comprado una moto por 280 u.m. que se registrará como gasto en consumo final dado que se trata de un bien de consumo duradero pero, a su vez, dado que la ha financiado con un préstamo por el que paga de interés 5 u.m. durante todo el año, habrá de hacerse una anotación en los empleos de la cuenta de asignación de la renta primaria. Esto es así pues los intereses (D.41) son las rentas de propiedad que reciben los propietarios de los activos financieros por poner dichos activos financieros a disposición de otra unidad institucional. En este caso es María la que hace uso de dicho activo. No obstante, hay que tener en cuenta que lo que se anota en la columna de empleos en D.41 no son los intereses pagados al banco por un préstamo, sino que deberían ser los intereses descontando el servicio de intermediación financiera que realiza el banco (SIFMI) o margen de beneficio por intermediar, lo que podríamos llamar intereses SEC. Es decir, que esas 5 u.m. que paga de préstamo habría que descomponerlo en la parte de SIFMI que está abonando implícitamente (que se anotaría como gasto en consumo final de los hogares del producto servicios financieros) y la parte restante , que es realmente el interés contable (que se anotaría en la columna de empleos dentro de D.41). El enunciado no especifica nada sobre esta cuestión por lo que sería razonable considerar que SIFMI es igual a cero y las 5 u.m. se registrarían totalmente como un empleo en la cuenta de asignación de la renta primaria.

Juan ha recibido una prestación por desempleo de 350 u.m. Las prestaciones sociales son transferencias a los hogares, en efectivo o en especie, destinadas a aliviar la carga financiera que representa para ellos la cobertura de una serie de riesgos o necesidades y efectuadas por medio de sistemas organizados colectivamente o, fuera de estos sistemas, por unidades de las administraciones públicas y por las ISFLSH. Así, esta se recoge como recurso en la cuenta de distribución secundaria de la renta de los hogares.

En la producción de los hogares pueden distinguirse las siguientes funciones principales: vivienda, alimentación, vestido, cuidado de hijos, adultos y animales de compañía, y trabajo voluntario. Juan ha gastado 1100 u.m. en alimentos, bebidas, artículos de droguería y vestido y calzado para toda la familia que se recoge en empleos en la cuenta de utilización de la renta disponible de los hogares como gasto en consumo final. Además Juan ha gastado 900 u.m. en alquiler de la vivienda que igualmente se registra como gasto en consumo final.

El hecho de que Carlota haya recibido pagas de sus padres por valor de 30 u.m. es un aspecto que no se debería registrar en las cuentas nacionales, puesto que la unidad institucional mínima es el hogar, no la persona. Es decir, las transferencias corrientes de la cuenta de distribución secundaria de la renta se refieren a las que se producen de un hogar a otro hogar, como por ejemplo, las remesas de emigrantes o las de los padres a sus hijos cuando estudian fuera de casa por un tiempo, pero no las que se realizan entre personas que pertenecen a un mismo hogar.

A pesar de que Miguel no declare los trabajos de pintura que hace ocasionalmente, estos han de registrarse en la secuencia de cuentas. Esta actividad se considera como producción de un servicio por un miembro del hogar (en este caso sería un autónomo no dado de alta, pero en el SEC-2010 se registra la economía no observada como infradeclaraciones, actividades ilegales, etc...), por lo que la anotación pertinente sería producción P.1=800 u.m. en la cuenta de producción. El resto de vecinos le están pagando en este caso por la prestación de un servicio (producción), no porque tengan una empresa de pintura y esta persona trabaje en ella ocasionalmente.

Por otra parte, Miguel ha gastado 300 u.m. en pintura y brochas que se consideran herramientas de escaso valor utilizadas para operaciones sencillas siendo considerado este tipo de bienes de consumo duradero un consumo intermedio. Por tanto, se hace una anotación en los empleos de la cuenta de producción por valor de 300 u.m.. Además ha pagado 40 u.m. a un amigo para que le ayude, lo cual se considera como sueldos y salarios, que se registra como remuneración de asalariados en los empleos de la Cuenta de explotación.

Finalmente, Miguel ha gastado 200 u.m. en bebidas, tabaco y tickets para eventos deportivos. Si consideramos que Miguel ha realizado ese gasto en su tiempo libre, sería gasto en consumo final. Pero si consideramos que ha realizado ese gasto con algunos vecinos a cambio de conseguir los trabajos de pintura, se podría considerar como consumos intermedios. Por simplicidad, consideraremos esta partida como gasto en consumo final.

A continuación se recoge la sucesión de cuentas:

E	Cuenta de producción		R
300	Consumos intermedios (P.2)	Producción (P.1)	800
500	VAB (B1.b)		

E	Cuenta de explotación		R
40	Remuneración de asalariados (D.1)	Valor añadido bruto (B1.b)	500
460	EBE+RM (B2.b/B3.b)		

E	Cuenta de asignación de la renta primaria		R
5	Rentas de la propiedad (intereses) (D.41)	EBE+RM (B2.b/B3.b)	460
		Sueldos y salarios (D.11)	2.000
		Cotizaciones sociales a cargo del empresario (D.12)	20
2.475	Saldo de rentas primarias bruto (B5.b)		

E	Cuenta de distribución secundaria de la renta		R
25	Impuestos corrientes pagados (D.5)	Saldo de rentas primarias bruto (B5.b)	2.475
15 + 20	Cotizaciones sociales pagadas (D.61)	Prestaciones sociales distintas de transferencias sociales en especie (D.62)	350
2.765	RBD(B6.b)		

E	Cuenta de utilización de la renta disponible		R
100+280+ +900+200+ +1.100	Gasto en consumo final (P.3)	RBD(B6.b)	2.765
185	Ahorro bruto (B8.b)		

Cuestión 7

Usando datos de distintas Comunidades Autónomas, un investigador desea examinar la relación entre la renta media pagada (RENT) por alquiler en función de los precios medianos de la vivienda (MDHOUSE en miles de euros). Usamos una variable de control adicional que es el porcentaje de la población de la comunidad autónoma que vive en una zona urbana (PCTURBAN). Las columnas de la primera fila de esta tabla indican cuales son las variables "dependientes" en cada regresión; C hace referencia a la denominada constante de regresión. Errores estándar en paréntesis.

	(1) $RENT$	(2) $MDHOUSE$	(3) $MDHOUSE$	(4) $RENT$	(5) $RENT$	(6) $EHAT$
C	125,9 (14,19)	−18,67 (12,00)	7,225 (8,936)	120,7 (12,43)	120,7 (15,71)	−62,85 (26,95)
$PCTURBAN$	0,525 (0,249)	0,182 (0,115)	0,616 (0,131)	0,0815 (0,244)	0,0815 (0,305)	−0,283 (0,258)
$MDHOUSE$	1,521 (0,228)			2,240 (0,268)	2,240 (0,339)	
$FAMINC$		2,731 (0,682)				4,448 (1,532)
$REG2$		−5,095 (4,122)				−6,768 (9,262)
$REG3$		−1,778 (4,073)				4,847 (9,151)
$REG4$		13,41 (4,048)				−18,77 (9,096)
$VHAT$				−1,589 (0,398)		
N	50	50	50	50	50	50
R^2	0.669	0.691	0.317	0.754	0.599	0.226
SCR	20259.6	3767.6	8322.2	15054.0	24565.7	19019.9

a) Las estimaciones de mínimos cuadrados del modelo se encuentran en la columna (1). ¿Es factible que MDHOUSE, el precio mediano de las casas, sea endógeno en esta regresión? Justifique su respuesta

b) Un investigador considera cuatro instrumentos: ingreso mediano de la familia (FAMINC 1,000€) y región del país (REG1, REG2, REG3). A continuación considere las regresiones indicadas en los modelos de las columnas (2) y (3). A partir de aquel, indique cómo contrastar si los instrumentos son débiles.

c) En la columna (4) los residuos mínimo cuadráticos (VHAT) de la regresión en la columna (2) se agregan como un regresor a la regresión básica. Las estimaciones están obtenidas utilizando mínimos cuadrados. ¿Cuál es la utilidad de esta regresión? ¿Qué indica sobre los resultados en (1)?

d) En la columna (5) están las estimaciones VI/ MC2E utilizando los instrumentos enumerados en la parte (b). ¿Qué diferencias observas entre estos resultados y los resultados de los mínimos cuadrados en la columna (1)? Observe que las estimaciones (aunque no los errores estándares) son las mismas en las columnas (4) y (5). ¿Es un error? Justifique sus respuestas.

e) En la columna (6) los residuos de la estimación en (5) son regresados sobre las variables mostradas. ¿Qué información está contenida en estos resultados?

Solución.

a) El modelo propuesto en la columna (1) es:

$$RENT = C + \beta_1 PCTURBAN + \beta_2 MDHOUSE + UHAT.$$

La renta media pagada por alquiler (RENT) está positivamente afectada por los precios medianos de la vivienda (MDHOUSE), por el porcentaje de la población de la comunidad autónoma que vive en una zona urbana (PCTURBAN) y por otras muchas magnitudes macroeconómicas omitidas en la regresión como pueden ser la tasa de desempleo, el tipo de interés, el crecimiento o decrecimiento de la población, etc. Dado que no aparecen representadas en la regresión, será el término de error del modelo, UHAT, el que recoja el efecto de estas variables omitidas y por tanto existirá una correlación positiva entre el término de error UHAT y el precio mediano de la vivienda MDHOUSE. Dicha correlación positiva hace que el coeficiente de MDHOUSE sea inconsistente, sesgado y sobrestimado de forma que la endogeneidad de MDHOUSE es factible.

b) El modelo de la columna (1) contiene una variable potencialmente endógena, MDHOUSE. Para salvar el problema de endogeneidad, un investigador propone, en la columna (2), la siguiente regresión con cuatro instrumentos:

$$MDHOUSE = C + \beta_1 PCTURBAN + \delta_1 FAMINC + \delta_2 REG2 + \\ + \delta_3 REG3 + \delta_4 REG4 + VHAT.$$

Sin embargo, para llevar a cabo la estimación utilizando instrumentos es necesario comprobar la utilidad de estos. Para ello, hay que comprobar la significatividad conjunta de los cuatro instrumentos (FAMINC, REG2, REG3 y REG4) mediante un test F. En él, la hipótesis nula es que los coeficientes de FAMINC, REG2, REG3 y REG4 no son conjuntamente significativos mientras que la hipótesis alternativa es que sí lo son.

$$\begin{cases} H_0: & \delta_1 = \delta_2 = \delta_3 = \delta_4 = 0 \\ H_1: & \text{al menos uno es distinto de } 0 \end{cases}$$

Para llevar a cabo el contraste, se utiliza un estadístico F para el que se necesitan tanto el modelo de la columna (2) como el modelo de la columna (3). Este último, que se denomina modelo restringido (R), se obtiene de suponer que la hipótesis nula es cierta en el modelo de la columna (2), que se denomina modelo no restringido (NR). El modelo restringido de la columna (3) es:

$$MDHOUSE = C + \beta_1 PCTURBAN + XHAT.$$

Sabemos que

$$SCR_{NR} = 3767{,}6 \qquad SCR_R = 8322{,}2$$

donde SCR_{NR} y SCR_R son la suma de cuadrados de residuos del modelo no restringido y restringido, respectivamente. Además,

$$q = 4 \qquad T = 50 \qquad k = 6$$

siendo q el número de variables instrumentos cuya significatividad se quiere contrastar, T el tamaño de la muestra y k el número de parámetros estimados en el modelo sin restringir. Bajo la hipótesis nula de no significatividad,

$$F = \frac{\frac{SCR_R - SCR_{NR}}{q}}{\frac{SCR_{NR}}{(T-k)}} \sim F_{q,T-k;\alpha}$$

Por tanto,

$$F = \frac{\frac{8322{,}2-3767{,}6}{4}}{\frac{3767{,}6}{(50-6)}} = 13,002$$

y el valor proporcionado por las tables para el contraste es

$$F_{q,T-k;\alpha} = F_{4,44;0{,}05} = 3,204.$$

Dado que el valor del estadístico experimental es mayor que el obtenido en las tablas para un nivel de significación del 5 %, se rechaza la hipótesis nula y a este nivel de significación las variables instrumento sí son conjuntamente significativas.

c) La columna (4) se obtiene de añadir los residuos mínimo cuadráticos (VHAT) de la regresión de la columna (2) como un regresor a la regresión inicial. La utilidad de esta regresión es que proporciona una forma de contrastar la supuesta endogeneidad de la variable MDHOUSE mediante un procedimiento de dos etapas sobre el que se basa el test de Hausman. Si no existe endogeneidad en el modelo inicial de la columna (1), tanto MCO como VI son consistentes, pero MCO es más eficiente. Si existe endogeneidad, solamente VI es consistente. Por lo tanto, es importante realizar un contraste de endogeneidad para evitar usar VI cuando no es necesario. Las hipótesis nula y alternativa del test de Hausman son:

$$\begin{cases} H_0: & \text{MDHOUSE es exógena} \\ H_1: & \text{MDHOUSE es endógena} \end{cases}$$

Es necesario contar como mínimo con un instrumento exógeno para llevar a cabo el procedimiento de dos etapas. En este caso, contamos con las variables instrumento FAMINC, REG2, REG3 y REG4. Partiendo del modelo incial la columna (1)

$$RENT = C + \beta_1 PCTURBAN + \beta_2 MDHOUSE + UHAT,$$

en la primera etapa del procedimiento, se regresa la variable cuya endogeneidad se desea contrastar sobre las variables exógenas y los instrumentos:

$$\begin{aligned} MDHOUSE = C + \beta_1 PCTURBAN + \delta_1 FAMINC + \delta_2 REG2 + \\ + \delta_3 REG3 + \delta_4 REG4 + VHAT, \end{aligned}$$

que coincide con la regresión de la columna (2). En caso de exogeneidad de MDHOUSE, dado que suponemos que PCTURBAN, FAMINC, REG2, REG3 y REG4 no están correlacionadas con UHAT, los residuos VHAT tampoco deberían estarlo. En una segunda etapa, se estima por mínimos cuadrados el modelo original agregando el valor ajustado de VHAT a la ecuación:

$$RENT = C + \beta_1 PCTURBAN + \beta_2 MDHOUSE + \beta_3 \widehat{VHAT} + XHAT,$$

que coincide con la columna (4). El objetivo es contrastar la hipótesis nula de que MDHOUSE es exógena. Bajo exogeneidad, el coeficiente de $\widehat{VHAT}$ no debería ser significativo:

$$\begin{cases} H_0: & \beta_3 = 0 \\ H_1: & \beta_3 \neq 0 \end{cases}$$

Rechazar la hipótesis nula sería evidencia en contra de la exogeneidad de MDHOUSE y por tanto en contra de estimar el modelo por MCO. Por tanto, el estadístico de contraste es un estadístico tipo t para la significatividad del coeficiente de $\widehat{VHAT}$. El valor del estadístico de contraste es

$$t_{exp} = \left| \frac{\hat{\beta}_3}{\hat{\sigma}(\hat{\beta}_3)} \right| = \left| \frac{-1,589}{0,398} \right| = |-3,99| = 3,99.$$

Mientras que el valor crítico proporcionado por la tabla de t - Student es: $t_{\{T-k,\ \alpha\}} = t_{\{50-4,\ 0{,}05\}} = t_{\{46,\ 0{,}05\}} = 2,012$. Dado que $3,99 > 2,012$ se rechaza la hipótesis nula de que el coeficiente de $\widehat{VHAT}$ es cero usando un nivel de confianza de 0,05. Se puede concluir que MDHOUSE es endógena. De este modo, comparando los resultados en la columna (1) con los resultados de la columna (4), en la primera de ellas las estimaciones son sesgadas e inconsistentes mientras que en la columna (4) las estimaciones son mejores. Incluso el coeficiente de determinación R^2 ha mejorado significativamente en la columna (4) frente al obtenido en la columna (1). El resultado en la columna (4) también indica que un incremento de una unidad en MDHOUSE contribuye a un incremento de 2,24 unidades en la renta media frente al incremento de 1,521 unidades en la renta en la columna (1).

d) En la columna (5) están las estimaciones VI/MC2E utilizando los instrumentos FAMINC, REG2, REG3 y REG4. Si los instrumentos son exógenos y están correlacionados con la variable para la que hacen de instrumento, es posible obtener estimadores consistentes para el coeficiente de MDHOUSE mediante el llamado estimador de mínimos cuadrados en dos etapas: MC2E. El estimador MC2E regresa, en la primera etapa, la variable MDHOUSE sobre la variable exógena PCTURBAN y los instrumentos FAMINC, REG2, REG3 y REG4:

$$MDHOUSE = C + \beta_1 PCTURBAN + \delta_1 FAMINC + \delta_2 REG2+$$
$$+ \delta_3 REG3 + \delta_4 REG4 + VHAT.$$

Esta primera etapa, se corresponde con la columna (2). La idea es utilizar el componente de MDHOUSE que puede predecirse con PCTURBAN, FAMINC, REG2, REG3 y REG4, es decir $C + \beta_1 PCTURBAN + \delta_1 FAMINC + \delta_2 REG2 + \delta_3 REG3 + \delta_4 REG4$, que estará incorrelacionado con VHAT al ser PCTURBAN, FAMINC, REG2, REG3 y REG4 exógenas. Como los parámetros $\beta_1, \delta_1, \delta_2, \delta_3$ y δ_4 son desconocidos, en esta primera etapa se obtienen estimaciones MCO de dichos parámetros y el valor ajustado de MDHOUSE que se denota por $M\widehat{DHOU}SE$. La segunda etapa es la regresión de RENT sobre $M\widehat{DHOU}SE$ utilizando MCO. Los estimadores de la regresión de la segunda etapa son los estimadores MC2E.

$$RENT = C + \beta_1 PCTURBAN + \beta_2 M\widehat{DHOU}SE + EHAT.$$

Esta segunda etapa, se corresponde con la columna (5).

Se observan dos cambios importantes al comparar las estimaciones de mínimos cuadrados de la columna (1) y la estimación utilizando instrumentos de la columna (5). La primera es que la estimación VI del coeficiente de PCTURBAN es mucho más pequeña que dicha estimación usando mínimos cuadrados. De este modo, el coeficiente de PCTURBAN no es ahora significativo, pues $t_{exp}^{VI/MC2E} = \left|\frac{\hat{\beta}_1}{\hat{\sigma}(\hat{\beta}_1)}\right| = \left|\frac{0,0815}{0,305}\right| = 0,2328$, el valor crítico proporcionado por la tabla de t - Student es: $t_{\{T-k,\ \alpha\}} = t_{\{50-3,\ 0{,}05\}} = t_{\{47,\ 0{,}05\}} = 2,011$ y por tanto $t_{exp}^{VI/MC2E} < t_{\{47,\ 0{,}05\}}$. Mientras que la estimación de mínimos cuadrados, proporciona un valor del estadístico de contraste de $t_{exp}^{MCO} = \left|\frac{\hat{\beta}_1}{\hat{\sigma}(\hat{\beta}_1)}\right| = \left|\frac{0,525}{0,249}\right| = 2,108$ siendo por tanto significativo con $\alpha = 0,05$, ya que $t_{exp}^{MCO} > t_{\{47,\ 0{,}05\}}$. En segundo lugar, en la estimación usando VI, el efecto de MDHOUSE sobre RENT es mayor en magnitud, indicando un mayor efecto que en la primera estimación. Un incremento de una unidad en MDHOUSE en la columna (5) contribuye a un incremento de 2,24 unidades en la renta frente a un incremento de 1,521 en la columna (1). El error estándar del coeficiente en la estimación VI es mayor que el correspondiente error estandar en la estimación de mínimos cuadrados, pero el estadístico de contraste $t_{exp}^{VI/MC2E} = \left|\frac{\hat{\beta}_2}{\hat{\sigma}(\hat{\beta}_2)}\right| = \left|\frac{2,240}{0,339}\right| = 6{,}61$ es muy significativo.

Las estimaciones en las columnas (4) y (5) son las mismas aunque no los errores estándares y esto se debe a que los residuos VHAT de la regresión de mínimos cuadrados de primera etapa están incorrelados con PCTURBAN, porque es una variable explicativa en la primera etapa de la regresión y una propiedad de los residuos de las estimaciones de mínimos cuadrados es que están incorrelados con las variables explicativas del modelo. Del mismo modo, VHAT tampoco está correlacionado con el valor ajustado de MDHOUSE ($\widehat{MDHOUSE}$) que se usa para obtener la estimación VI/MC2E. Si se omite una variable de una regresión en la que está incorrelada con el resto de variables, no hay sesgo por variables omitidas y las estimaciones de mínimos cuadrados permanecen invariadas.

e) En la columna (6) los residuos de la estimación en (5) son regresados sobre las variables PCTURBAN, FAMINC, REG2, REG3 y REG4. Esta regresión auxiliar,

$$\begin{aligned} EHAT = C &+ \beta_1 PCTURBAN + \delta_1 FAMINC + \delta_2 REG2 + \delta_3 REG3 + \\ &+ \delta_4 REG4 + EPSHAT, \end{aligned}$$

(donde *EPSHAT* son nuevos residuos utilizados en la regresión) permite contrastar la utilización de más variables instrumento de las necesarias. Si todos los instrumentos fueran exógenos, sus estimadores tenderían a ser muy parecidos. Por el contrario, si fueran muy diferentes se interpretaría como una evidencia de que alguno o todos ellos no son exógenos, aunque no se podría determinar cuál. El contraste se construye implícitamente sobre esta idea. En la práctica, si los instrumentos son exógenos entonces serían independientes de los residuos de la estimación en (5), EHAT, por lo que si se construye una regresión auxiliar de los errores de la estimación MC2E sobre los instrumentos y las variables exógenas explicativas, los coeficientes asociados a los instrumentos no deberían ser conjuntamente significativos. Por tanto el conjunto de restricciones a contrastar sería

$$\begin{cases} H_0: & \delta_1 = \delta_2 = \delta_3 = \delta_4 = 0 \\ H_1: & \text{al menos uno es distinto de 0} \end{cases}$$

Bajo la hipótesis nula de que todos los instrumentos son exógenos, el estadístico

$$N \cdot R^2 \sim \chi^2_{k-q;\alpha},$$

donde N es el tamaño de la muestra, R^2 es el coeficiente de determinación de la regresión de la columna (6) y la diferencia entre el número de variables instrumentales y el número de variables endógenas en el modelo original son los grados de libertad de la distribución χ^2, es decir, $k - q = 4 - 1 = 3$. Por tanto, suponiendo que la hipótesis nula sea cierta,

$$N \cdot R^2 \sim \chi^2_{3;0,05}.$$

Si se rechaza la hipótesis nula, habrá dudas acerca de la adecuación del conjunto de instrumentos. Uno o más de los instrumentos podría no tener correlación cero con el error. En nuestro caso, $N \cdot R^2 = 50 \cdot 0,226 = 11,3$ y $\chi^2_{4-1=3;0,05} = 7,8147$. Al ser el valor del estadístico de contraste mayor que el valor crítico para los grados de libertad y el nivel de confianza elegidos, se rechaza la hipótesis nula de no significatividad conjunta de los instrumentos y se puede concluir que hay al menos una variable instrumento que es inválida.

Cuestión 8

La tabla siguiente incluye estimaciones del correlograma (fac) y del correlograma parcial (fap) de un conjunto de serie temporales. Las series A, B y C contienen 200 observaciones temporales, mientras que las restantes tienen 300 observaciones. En algunas series se ofrecen estimaciones de la "serie en Niveles" y de la "serie en Diferencias". Teniendo esto en cuenta conteste a las siguientes preguntas:

a) Identifique las series A, B, C, D y E basándose en las FAC y FAP estimadas (explicite los argumentos en los que se basa)

b) Estime, cuando sea posible, los parámetros de los modelos identificados (desarrolle explícitamente sus cálculos y en qué se basa) en el apartado a)

Tabla

Correlogramas (fac) y correlogramas parciales (fap) estimados

	Retardos												Error
	1	2	3	4	5	6	7	8	9	10	11	12	estandár
Serie A													
Correlación (fac)	0.13	-0.05	0.00	-0.01	0.08	0.04	-0.02	0.01	-0.05	-0.08	-0.02	0.06	0.07
Correlación parcial (fap)	0.13	-0.08	0.02	-0.02	0.09	0.01	-0.01	0.02	-0.05	-0.07	-0.01	0.07	0.07
Serie B													
Correlación	0.42	0.03	-0.03	-0.12	0.15	-0.11	-0.06	0.02	0.05	0.07	0.02	-0.10	0.07
Correlación parcial	0.42	-0.18	0.04	-0.14	-0.05	-0.05	-0.01	0.04	-0.00	0.04	-0.06	-0.09	0.07
Serie C													
Correlación	0.58	0.35	0.21	0.13	0.02	-0.03	0.05	0.04	0.08	0.05	-0.08	-0.06	0.07
Correlación parcial	0.58	0.03	-0.01	0.01	-0.09	-0.03	-0.03	0.08	0.04	0.01	-0.12	0.01	0.07
Serie D													
Niveles correlación	0.90	0.84	0.78	0.71	0.66	0.61	0.57	0.53	0.49	0.45	0.42	0.37	0.06
Diferencias	-0.23	0.03	-0.01	-0.01	-0.08	0.03	-0.04	-0.01	0.03	-0.06	0.08	-0.03	0.06
Correlación parcial (niveles)	0.90	0.19	-0.01	-0.03	-0.02	0.06	-0.02	0.01	-0.00	-0.04	0.03	-0.09	0.06
Diferencias	-0.23	-0.02	-0.00	-0.02	-0.10	-0.01	-0.04	-0.03	0.02	-0.06	0.06	-0.01	0.06
Serie E													
Niveles correlación	0.98	0.95	0.92	0.89	0.86	0.84	0.81	0.78	0.75	0.72	0.69	0.66	0.06
Diferencias	0.42	-0.01	0.04	0.07	0.02	-0.08	-0.08	0.05	0.13	0.02	-0.06	0.03	0.06
Correlación parcial (niveles)	0.98	-0.21	0.07	0.01	-0.05	-0.03	0.03	0.01	-0.10	-0.06	-0.01	0.01	0.06
Diferencias	0.42	-0.06	-0.04	-0.10	0.01	-0.03	0.05	0.03	0.02	-0.08	-0.03	-0.07	0.06

Solución.

Antes de comenzar la resolución del ejercicio, es preciso señalar que para contestar a esta pregunta en el examen no sería necesario detallar la misma de la forma que proponemos a continuación, pues no es realista resolver el problema en este sentido. Sin embargo, creemos conveniente aclarar la teoría sobre la que se basa la resolución del ejercicio y situar el mismo en un contexto adecuado que ayude al opositor a entender de forma global el planteamiento.

El objetivo del análisis estocástico de series temporales es inferir a partir de una serie temporal las caracterísitcas de la estructura probabilística subyacente. Se utiliza la metodología de Box - Jenkins que se compone de cuatro etapas: especificación inicial, estimación, chequeo y utilización.

Para determinar la estructura subyacente se consideran casos particulares de procesos estocásticos lineales discretos con estructura conocida: procesos de medias móviles (MA), procesos autorregresivos (AR) y procesos mixtos autorregresivos - medias móviles (ARMA). Además, cuando son series temporales referidas a periodos inferiores al año se consideran modelos estocásticos estacionales. Estos modelos estocásticos estacionales pueden ser de dos tipos: puros y multiplicativos; y al mismo tiempo, cada uno de ellos puede ser estacionario o no. Se denomina s al período estacional, de manera que si los datos son, por ejemplo, trimestrales $s = 4$ y si son mensuales $s = 12$. El modelo estacional puro se caracteriza porque solo existen relaciones entre las observaciones que distan entre sí s o múltiplos de s periodos. Sin embargo, los modelos estacionales puros no van a ser los más frecuentes ya que normalmente no sólo se dan relaciones entre las observaciones que distan s (o múltiplos de s) periodos, sino que lo habitual es que existan relaciones de tipo múltiple. En este caso, se trata de un modelo estacional multiplicativo $SARIMA(P, D, Q)$x(p, d, q). Todas las series que se van a estudiar tienen como punto de partida común este modelo $SARIMA(P, D, Q)$x(p, d, q), cuya estructura es:

$$\left(1 - \Phi_1 L^s - \cdots - \Phi_P L^{Ps}\right)\left(1 - \phi_1 L - \cdots - \phi_p L^p\right)\left(1 - L^s\right)^D \left(1 - L\right)^d y_t =$$
$$= \left(1 - \Theta_1 L^s - \cdots - \Theta_Q L^{Qs}\right)\left(1 - \theta_1 L - \cdots - \theta_q L^q\right) u_t$$

siendo u_t un proceso de ruido blanco con las siguientes propiedades:

$$E(u_t) = 0, \qquad \sigma^2(u_t) = 0, \qquad cov(u_t, u_s) = 0 \ \forall t \neq s.$$

Sin embargo, en este ejercicio vamos eliminar la parte estacional y considerar únicamente la parte regular. Así, consideramos el modelo $ARIMA(p, d, q)$ cuya estructura es:

$$\left(1 - \phi_1 L - \cdots \phi_p L^p\right)\left(1 - L\right)^d y_t = \left(1 - \theta_1 L - \cdots \theta_q L^q\right) u_t,$$

donde u_t es un proceso de ruido blanco. En adelante, u_t representará un proceso de ruido blanco que cumple las propiedades expuestas.

En la etapa de especificación inicial de la metodología Box - Jenkins se determinan el orden de integración de la serie temporal d, el orden del polinomio autorregresivo p y el de medias móviles q. El orden de integración d es el número de diferencias que se requieren para convertir en estacionaria la variable objeto de análisis. En esta primera etapa de especificación se utilizan como principal instrumento la FAC y la FAP de los procesos estacionarios.

Los procesos lineales tienen pautas de autocorrelación teórica características y reconocibles:

- En un proceso de ruido blanco, las funciones de autocorrelación y autocorrelación parcial son nulas
- En un proceso AR (autorregresivo) la función de autocorrelación es infinita y la de autocorrelación parcial es finita (el número de coeficientes significativos indica el orden).
- En un proceso MA (de media móvil) la función de autocorrelación parcial es finita (el número de coeficientes significativos indica el orden) y la de autocorrelación parcial es infinita.
- En un proceso ARMA mixto ambas funciones son infinitas.

Se recoge un resumen de las características de la FAC y de la FAP para los procesos AR, MA y $ARMA$ en la siguiente tabla:

	FAC	FAP
$AR(p)$	Muchos coeficientes no nulos que decrecen con el retardo como mezcla de exponenciales y senoides.	p primeros coeficientes no nulos, el resto cero.
$MA(q)$	q primeros coeficientes no nulos, el resto cero.	Muchos coeficientes no nulos que decrecen con el retardo como mezcla de exponenciales y senoides.
$ARMA(p,q)$	Muchos coeficientes no nulos con decrecimiento hacia cero.	Muchos coeficientes no nulos con decrecimiento hacia cero.

Los valores de la función de autocorrelación (FAC) se corresponden con los coeficientes de autocorrelación ρ_k del proceso estocástico que miden el grado de asociación lineal existente entre dos variables aleatorias del proceso separadas k periodos. Los coeficientes de autocorrelación se calculan a partir de la función de autocovarianzas mediante la expresión:

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \frac{Cov\left(Y_t, Y_{t-k}\right)}{\sqrt{Var(Y_t)Var(Y_{t-k})}} = \frac{Cov\left(Y_t, Y_{t-k}\right)}{\sqrt{Var(Y_t)Var(Y_t)}} = \frac{Cov\left(Y_t, Y_{t-k}\right)}{Var(Y_t)}$$

donde γ_k representa cada una de las autocovarianzas del proceso para $k = 0, 1, 2, \ldots$. Se llama correlograma a la representación mediante un gráfico de barras de la función de autocorrelación de un proceso.

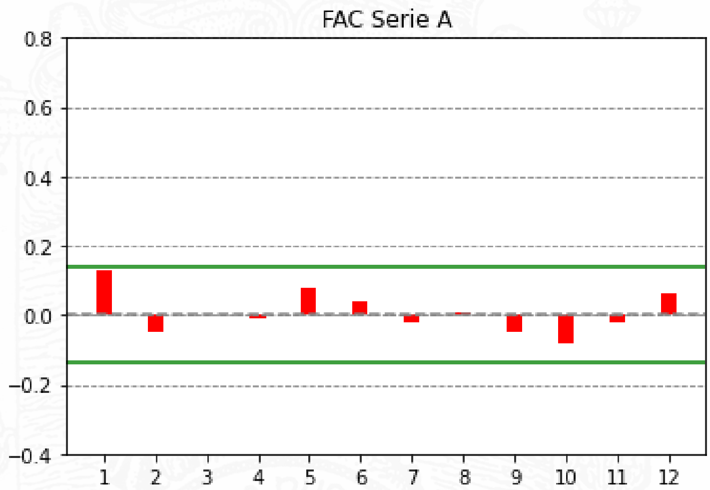
Por otra parte, los valores de la función de autocorrelación parcial (FAP) se corresponden con los coeficientes de autocorrelación parcial ϕ_{kk} del proceso estocástico. El concepto de coeficiente de autocorrelación parcial entre y_t e y_{t-k} es similar al de coeficiente de autocorrelación entre dichos valores de la serie temporal, salvo porque se mide dicha correlación ajustada por el efecto de los retardos intermedios. Es decir, al obtener la correlación entre y_t e y_{t-k} se elimina la influencia que, sobre ambas, tienen $y_{t-1}, y_{t-2}, \ldots, y_{t-j+1}$. La representación mediante un gráfico de barras de la función de autocorrelación parcial se llama correlograma parcial.

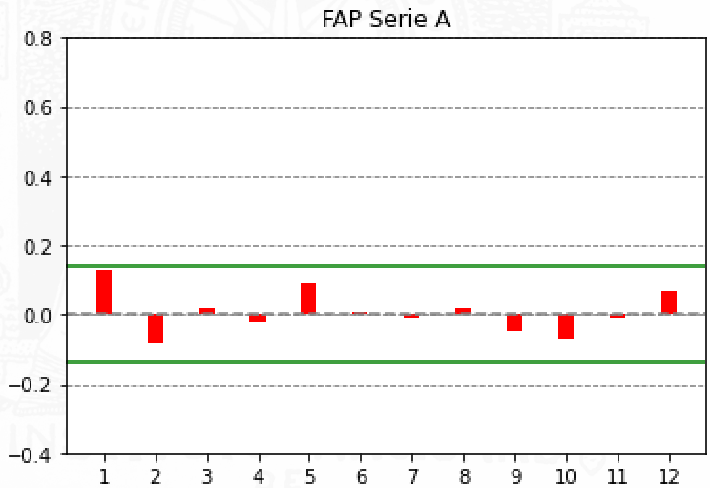
Como ya adelantábamos, la elección del orden se basa en contrastes de significación llevados a cabo con los valores estimados de las funciones de autocorrelación y de autocorrelación parcial. Suponiendo que se identifique una estrucrura $AR(p)$, la distribución de los valores estimados de la FAP es $\widehat{\phi}_{jj} \sim N\left(0, \frac{1}{T}\right)$ para $j > p$. Así, cuando $\widehat{\phi}_{jj}$ pertenece al intervalo $\left(-\frac{1,96}{\sqrt{T}}, \frac{1,96}{\sqrt{T}}\right)$ para todos los valores de j superiores a un cierto p, entonces no se rechaza la hipótesis de que el orden del modelo AR es menor o igual que p.

Analogamente, teniendo en cuenta que un proceso $MA(q)$ tiene función de autocorrelación simple teórica nula para valores $k > q$, se trata de efectuar contrastes de hipótesis de significación estadística acerca de los valores estimados de dicha función. En un proceso de ruido blanco, las estimaciones $\widehat{\rho}_k \sim N\left(0, \frac{1}{T}\right)$, por lo que los retardos serán no significativos si caen dentro del intervalo $\left(-\frac{1,96}{\sqrt{T}}, \frac{1,96}{\sqrt{T}}\right)$. Si ocurre esto para todo $k > q$, el proceso es $MA(q)$. Si ocurre para todo k, la serie temporal es un ruido blanco.

En la segunda etapa de la metodología de Box - Jenkins se estima la estructura propuesta en la etapa de identificación. En la tercera estapa se realiza el chequeo o validación del modelo. Algunos de los criterios de validación más importantes incluyen el análisis de la autocorrelación de los residuos, la identificación de residuos atípicos, la significatividad de los parámetros estimados del modelo (así como sus correlaciones), las condiciones de estacionariedad e invertibilidad y distintos criterios de ajuste (como el error estándar de la regresión, el de Akaike o el de Schwarz). En este ejercicio, la tercera etapa está limitada por la información de la que se dispone. Finalmente, en la cuarta y última etapa de la metodología Box - Jenkins se lleva a cabo la utilización del modelo, por ejemplo, para predecir valores futuros de la serie, para describir las propiedades de un fenómeno económico en cuestión (tendencia, estacionalidad, oscilaciones (cíclicas) estacionarias, ...), etc. La última etapa no se va a implementar en este ejercicio.

SERIE A





En los gráficos de la FAC y de la FAP de la serie A no se percibe, a priori, una estructura determinada y se observa que la FAC y la FAP son muy parecidas. Por otra parte, al disponer de $T = 200$ observaciones, el intervalo de no significatividad es $\left(-\frac{1,96}{\sqrt{200}}, \frac{1,96}{\sqrt{200}}\right) = (-0,13859,\ 0,13859)$. De modo que todos aquellos retardos que se encuentren dentro del intervalo de no significatividad, es decir, aquellos cuyo valor pertenezca a $(-0,13859,\ 0,13859)$, entran dentro de las bandas de confianza, lo que indica que no son distintos de cero. Las bandas de confianza se representan en color verde en el gráfico. Se observa que ninguno de los retardos es significativo pues todos pertenecen al intervalo de no significatividad. Consecuentemente, se identifica la serie como un proceso de ruido blanco y el modelo identificado se puede expresar como

$$y_t = u_t.$$

De forma adicional se podría llevar a cabo la prueba de Ljung y Box, que se utiliza para probar que una serie se distribuye como un proceso de ruido blanco. Siendo las hipótesis de la prueba:

$$\begin{cases} H_0: & y_t \sim RB(0,\sigma^2) \\ H_1: & y_t \textit{ no es ruido blanco} \end{cases}$$

El estadístico de contraste que establecen Ljung y Box es

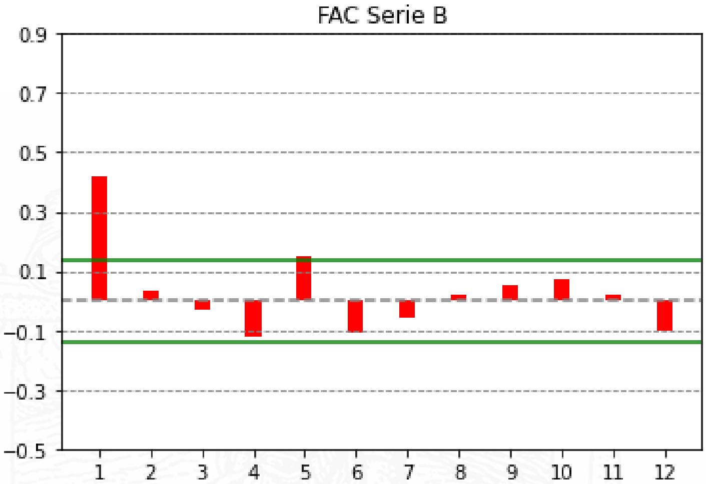
$$Q_{LB} = T(T+2)\sum_{k=1}^{m} \frac{1}{T-k}\widehat{\rho}_k^2 \sim \chi_m^2.$$

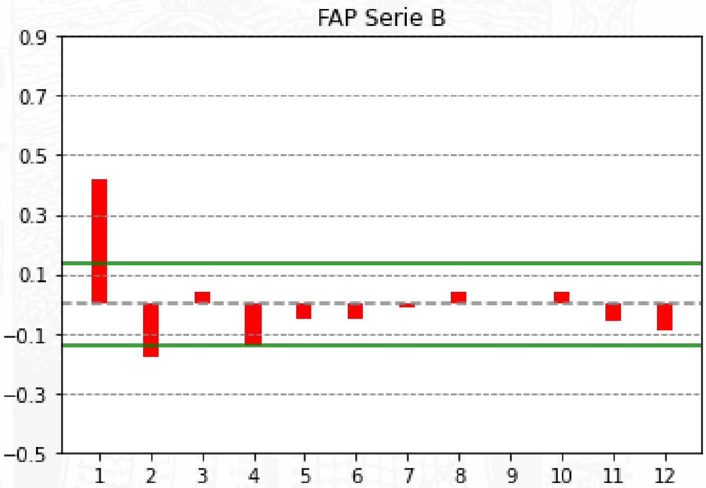
Entonces si $H_0: y_t \sim RB(0,\sigma^2)$ es cierta se cumple que $\widehat{\rho}_k^2 \sim N\left(0, \frac{1}{T}\right),\ k = 1, \ldots, m$ y por tanto $Q_{LB} \overset{H_0}{\sim} \chi_m^2$. En el caso que nos ocupa, el estadístico de contraste es

$$Q_{LB} = 200(200+2)\sum_{k=1}^{1} 2\frac{1}{T-k}\widehat{\rho}_k^2 = 40400 \cdot 0,000209761 = 8,4743$$

y considerando $\alpha = 0,05$, $\chi^2_{12;0,05} = 21,026$. Al ser $Q_{LB} < \chi^2_{12;0,05}$, no se rechaza H_0 y por tanto se trata de un proceso de ruido blanco.

SERIE B





En la serie B, el intervalo de no significatividad de los retardos de ambas gráficas vuelve a ser $\left(-\frac{1,96}{\sqrt{200}}, \frac{1,96}{\sqrt{200}}\right) = (-0,13859,\ 0,13859)$ ya que de nuevo se dispone de $T = 200$ observaciones. De esta forma, en la FAC el primero es el único retardo significativo mientras que en la FAP tanto el primero como el segundo y el cuarto serían significativos. Sin embargo se observa que el valor de estos retardos decrece a medida que aumentan k. Dado que la FAC tiene un único retardo significativo, el primero, mientras que en la FAP parece presentarse una convergencia a cero exponencialmente alternando en signo, se cree que se trata de un modelo $MA(1)$ de la forma

$$y_t = (1 - \theta_1 L)u_t \Leftrightarrow y_t = u_t - \theta_1 u_{t-1}$$

donde θ_1 ha de ser negativo, pues el primer retardo de la FAC y el primer retardo de la FAP son positivos. Dado que se trata de un proceso estacionario en media $E(y_t) = 0$. La estacionariedad en covarianza implica que

$$\gamma_0 = E[y_t - E(y_t)]^2 = E[y_t]^2 = E[u_t - \theta_1 u_{t-1}]^2 = \sigma^2 + \theta_1^2\sigma^2$$

Para retardos superiores a cero se tiene:

$$\begin{aligned}
\gamma_1 &= E\left[y_t y_{t-1}\right] = E\left[\left(u_t - \theta_1 u_{t-1}\right)\left(u_{t-1} - \theta_1 u_{t-2}\right)\right] = \\
&= E\left[u_t u_{t-1} - \theta_1 u_t u_{t-2} - \theta_1 u_{t-1}^2 + \theta_1^2 u_{t-1}u_{t-2}\right] = \\
&= -\theta_1 E\left(u_{t-1}^2\right) = -\theta_1\sigma^2; \\
\gamma_2 &= E\left[y_t y_{t-2}\right] = E\left[\left(u_t - \theta_1 u_{t-1}\right)\left(u_{t-2} - \theta_1 u_{t-3}\right)\right] = 0 \\
\gamma_j &= E\left[y_t y_{t-j}\right] = E\left[\left(u_t - \theta_1 u_{t-1}\right)\left(u_{t-j} - \theta_1 u_{t-j-1}\right)\right] = 0 \ \forall j > 1.
\end{aligned}$$

En resumen, la función de autocovarianzas de un $MA(1)$ será:

$$\gamma_k = \begin{cases} \left(1 + \theta_1^2\right)\sigma^2 & \text{si } k = 0 \\ -\theta_1\sigma^2 & \text{si } k = 1 \\ 0 & \text{si } k > 1 \end{cases}.$$

Consecuentemente, la función de autocorrelación de un proceso $MA(1)$ es:

$$\rho_k = \begin{cases} 1 & \text{si } k = 0 \\ -\frac{\theta_1}{1+\theta_1^2} & \text{si } k = 1 \\ 0 & \text{si } k > 1 \end{cases}.$$

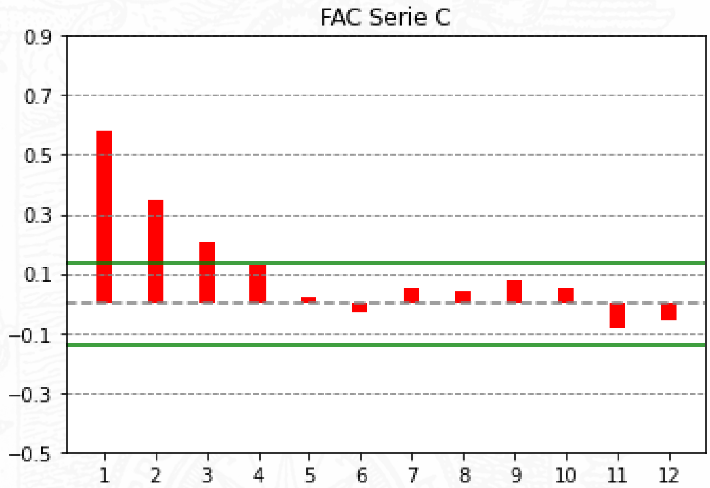
Por tanto al ser $\widehat{\rho_1} = 0,42$ se tiene que

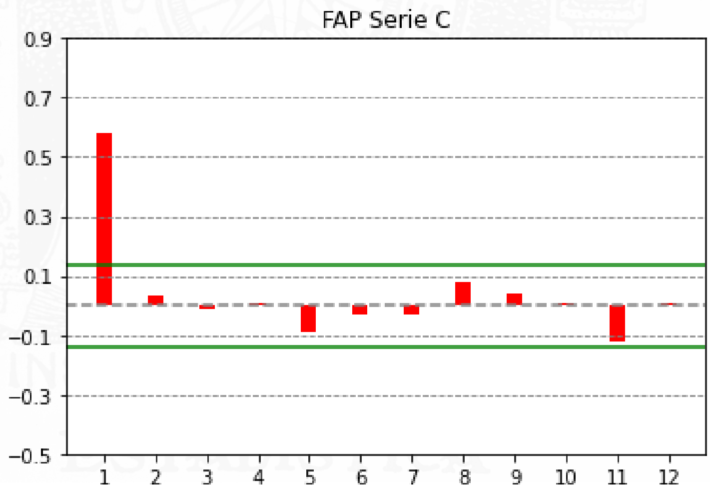
$$0,42 = -\frac{\theta_1}{1 + \theta_1^2} \Rightarrow \theta_1 = -0,544.$$

La otra raíz se descarta porque $|\theta_1| > 1$ y por tanto la serie no sería estacionaria. El modelo estimado para la serie B, estacionaria e invertible, es

$$y_t = u_t + 0,544u_{t-1}.$$

SERIE C





El correlograma es de la serie C presenta un valor relativamente alto de $\widehat{\rho}_1$ seguido de algunos coeficientes $\widehat{\rho}_k$ distintos de cero, pero que tienden a cero a medida que k aumenta. Esto hace pensar que se trata de una serie sin tendencia, que oscila en torno a una media constante, pero cuyas observaciones sucesivas están correlacionadas positivamente, es decir, una serie en la que a una observación por encima de la media, le suele seguir otra o más observaciones por encima de la media (lo mismo, para observaciones por debajo de la media).

Llama la atención el elevado valor del primer retardo tanto en la FAC como en la FAP. Se observa que la FAC muestra un decrecimiento amortiguado hacia cero mientras que en la FAP el único retardo significativo es el primero. Estos argumentos llevan a pensar que podría tratarse de un modelo $AR(1)$. El orden del modelo es 1 ya que de nuevo se dispone de $T = 200$ observaciones y el intervalo de no significatividad es $\left(-\frac{1,96}{\sqrt{200}}, \frac{1,96}{\sqrt{200}}\right) = (-0,13859,\ 0,13859)$. La estructura del modelo $AR(1)$ identificado será:

$$(1 - \phi_1 L)\, y_t = u_t.$$

Reescribiéndolo sin hacer uso del operador retardo L:

$$y_t - \phi_1 y_{t-1} = u_t \Leftrightarrow y_t = \phi_1 y_{t-1} + u_t$$

La esperanza del modelo identificado es:

$$E(y_t) = E(\phi_1 y_{t-1} + u_t) = \phi_1 E(y_{t-1})$$

Para que el proceso sea estacionario la media ha de ser constante y finita:

$$E(y_t) = \phi_1 E(y_t) \Rightarrow (1 - \phi_1)E(y_t) = 0 \Rightarrow E(y_t) = \frac{0}{(1 - \phi_1)} = 0.$$

Así, el proceso será estacionario si $\phi_1 \neq 1$. Además, para que el proceso sea estacionario, la varianza ha de ser constante y finita. Operando se tiene que:

$$E\left[(y_t - E(y_t))^2\right] = E\left[(y_t)^2\right] = E\left[(\phi_1 y_{t-1} + u_t)^2\right] = E(\phi_1^2 y_{t-1}^2 + u_t^2 + 2\phi_1 y_{t-1} u_t) =$$

$$= E(\phi_1^2 y_{t-1}^2) + E(u_t^2) + E(2\phi_1 y_{t-1} u_t) =$$
$$= \phi_1^2 E(y_{t-1}^2) + \sigma^2 + 2\phi_1 E(y_{t-1}u_t) = \phi_1^2 E(y_{t-1}^2) + \sigma^2.$$

Ya que

$$E(y_{t-j}u_t) = \begin{cases} \sigma^2 & \text{para } j = 0 \\ 0 & \text{para } j > 0 \end{cases}.$$

Por tanto,

$$\Rightarrow \gamma_0 = \phi_1^2 \gamma_0 + \sigma^2 \Rightarrow \gamma_0 = \frac{\sigma^2}{1-\phi_1^2}.$$

Siguiendo con el cálculo del resto de coeficientes de covarianzas, para $k > 0$:

$$\begin{aligned}
\gamma_1 &= E[y_t y_{t-1}] = E\left[y_{t-1}\left(\phi_1 y_{t-1} + u_t\right)\right] = \phi_1 \gamma_0 \\
\gamma_2 &= E[y_t y_{t-2}] = E\left[y_{t-2}\left(\phi_1 y_{t-1} + u_t\right)\right] = \phi_1 \gamma_1 = \phi_1^2 \gamma_0 \\
&\dots \\
\gamma_j &= E[y_t y_{t-j}] = E\left[y_{t-j}\left(\phi_1 y_{t-1} + u_t\right)\right] = \phi_1 \gamma_{j-1} = \phi_1^j \gamma_0
\end{aligned}$$

Por tanto la función de autocovarianzas de un proceso $AR(1)$ estacionario es:

$$\gamma_k = \begin{cases} \dfrac{\sigma^2}{1-\phi_1^2} & \text{si } k = 0 \\ \phi_1 \gamma_{k-1} & \text{si } k > 0 \end{cases}$$

Los coeficientes de autocorrelación de un proceso estacionario $AR(1)$ se obtienen de $\rho_k = \dfrac{\gamma_k}{\gamma_0}$ $\forall k$. Es decir, la función de autocorrelación de un proceso estacionario $AR(1)$ es:

$$\rho_k = \begin{cases} 1 & \text{si } k = 0 \\ \phi_1 \rho_{k-1} & \text{si } k > 0 \end{cases}.$$

De esta expresión, se puede deducir el valor de ϕ_1 ya que $\rho_0 = 1$ y $\widehat{\rho}_1 = 0,58$. Por tanto,

$$\widehat{\rho}_1 = \phi_1 \rho_0 \Rightarrow \widehat{\phi}_1 = 0,58.$$

Otra forma de obtener el mismo valor, sería utilizar las ecuaciones de Yule-Walker para el modelo $AR(1)$. En definitiva el modelo estimado será

$$(1 - 0,58L)y_t = u_t,$$

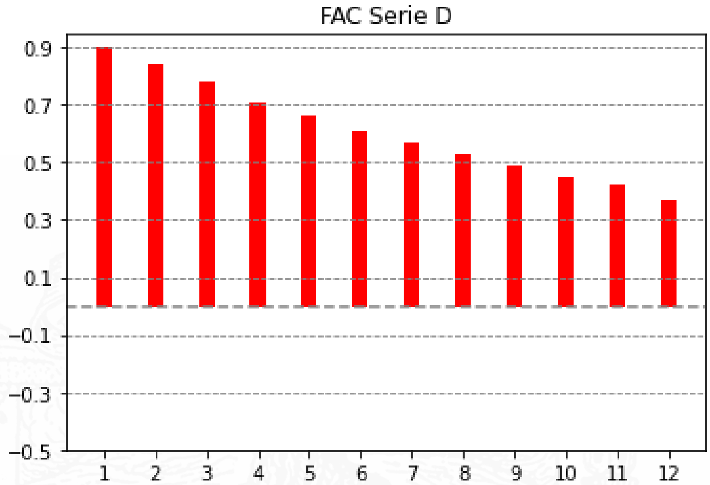
que, sin hacer uso del operador retardo, se puede reescribir como

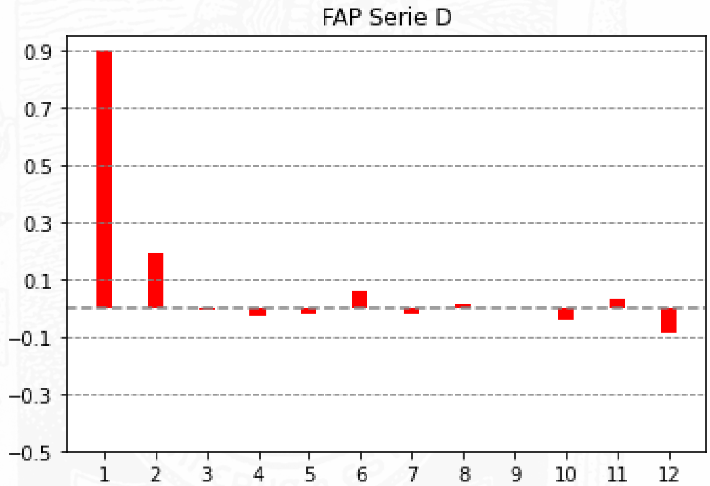
$$y_t - 0,58y_{t-1} = u_t \Leftrightarrow y_t = 0,58y_{t-1} + u_t.$$

Se trata de un modelo estacionario y que además es invertible, pues $|\phi_1| < 1$.

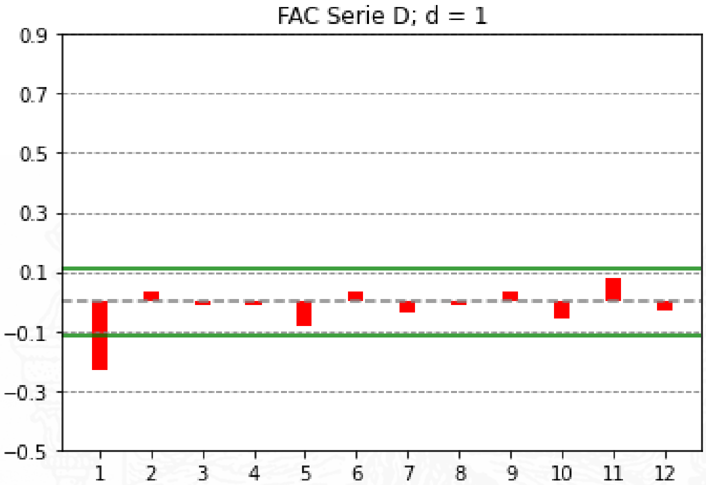
SERIE D

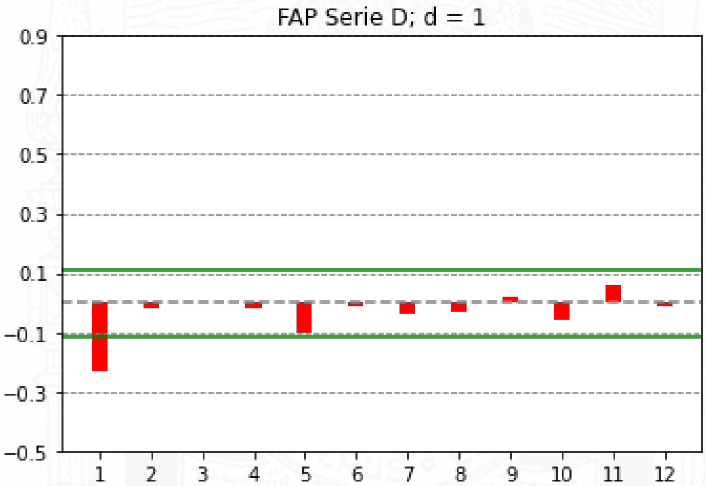
La serie D es claramente no estacionaria debido al comportamiento persistente de la FAC (los valores de $\widehat{\rho}_k$ no decrecen hacia cero rápidamente) y al valor próximo a 1 del primer retardo en la FAP. Este comportamiento persistente es consecuencia de la presencia de una tendencia en la serie, es decir, la serie cambia continuamente de nivel (creciendo o decreciendo). Una observación por encima (o por debajo) de la media general es seguida de muchas observaciones por encima (o por debajo) de la media, las cuales generan el comportamiento persistente en la FAC.





Cuando la serie no es estacionaria, es preciso transformarla: para eliminar la tendencia se toman una o varias diferencias en la serie. En este caso, basta con tomar una única diferencia y se observa, en las figuras que aparecen a continuación, que ahora sí la serie es estacionaria. Tomar una diferencia significa que en el modelo general $ARIMA(p, d, q)$ el valor del parámetro d es $d = 1$.





Dado que para la serie D se dispone de $T = 300$ observaciones, el intervalo de no significatividad sería $\left(-\dfrac{1,96}{\sqrt{300}}, \dfrac{1,96}{\sqrt{300}}\right) = (-0,11316,\ 0,11316)$. Se observa que el primer retardo de la FAC y el primer retardo de la FAP de la serie diferenciada son significativos. Se considera que el modelo presenta una estructura autorregresiva de orden $p = 1$, $AR(1)$, dada la significatividad del primer retardo de la FAP de la serie diferenciada y el comportamiento convergente hacia cero en la FAC de la serie diferenciada alternando el signo de los retardos. La estructura del modelo en este caso será:

$$(1-\phi_1 L)\,(1-L)y_t = u_t$$

Como hemos dicho, se observa que los signos del correlograma (FAC) ya diferenciado se van alternando. Por ejemplo, el valor de $\widehat{\rho}_1$ es negativo pero el valor de $\widehat{\rho}_2$ es positivo. Esto puede ser consecuencia de que las observaciones separadas dos periodos tienden a estar al mismo lado de la media. Se podría tratar por tanto de una serie sin tendencia que oscila en torno a una media constante pero que alterna valores con observaciones sucesivas a diferentes lados de la media general. Como se ha visto en la estimación de la serie anterior, la función de autocorrelación de un proceso $AR(1)$ estacionario es:

$$\rho_k = \begin{cases} 1 & \text{si } k = 0 \\ \phi_1 \rho_{k-1} & \text{si } k > 0 \end{cases}.$$

De esta expresión, se puede deducir el valor de ϕ_1 ya que $\rho_0 = 1$ y $\widehat{\rho}_1 = -0,23$. Por tanto,

$$\widehat{\rho}_1 = \phi_1 \rho_0 \Rightarrow \widehat{\phi}_1 = -0,23$$

y el modelo estimado es

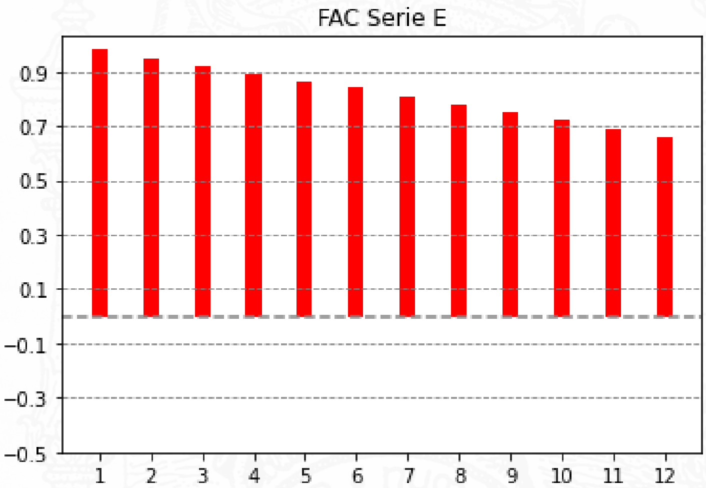
$$(1+0,23L)\,(1-L)y_t = u_t$$

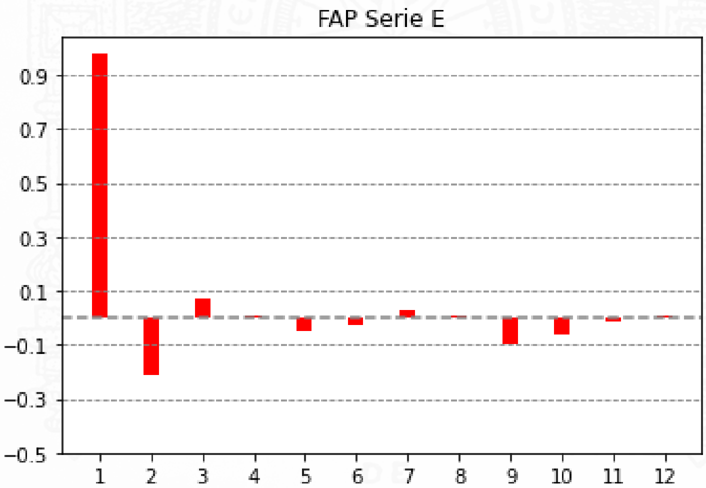
que, sin hacer uso de los operadores retardo, puede reescribirse como

$$\begin{aligned} &(1-L+0,23L-0,23L^2)y_t = u_t \Leftrightarrow \\ \Leftrightarrow\ & y_t - y_{t-1} + 0,23y_{t-1} - 0,23y_{t-2} = u_t \Leftrightarrow \\ \Leftrightarrow\ & y_t - 0,77y_{t-1} - 0,23y_{t-2} = u_t \end{aligned}$$

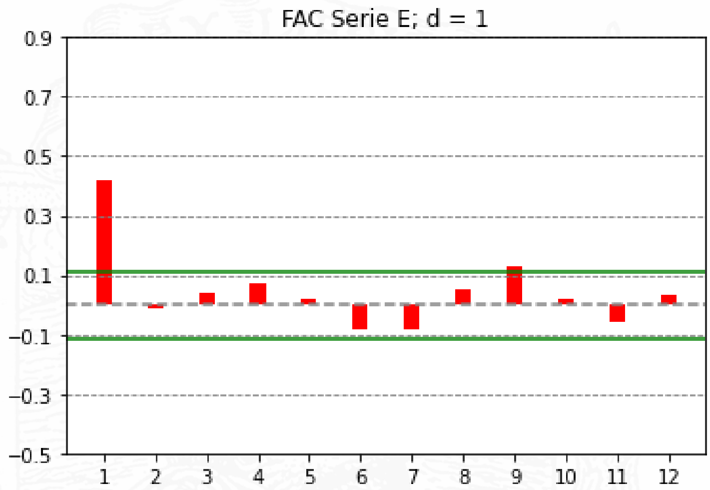
SERIE E

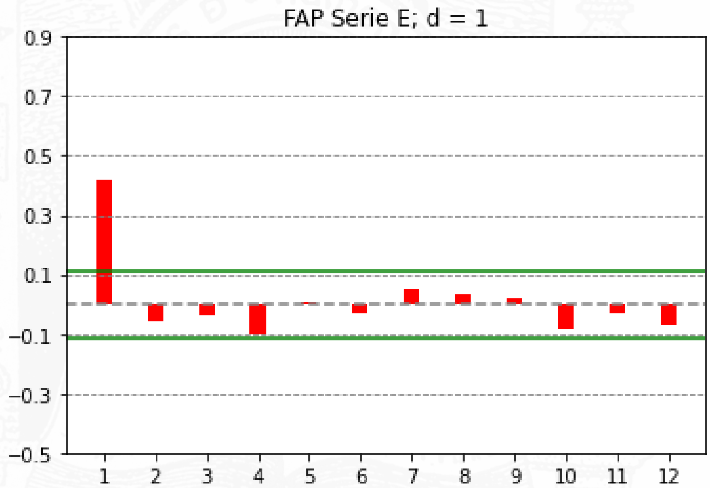
En este último caso, ocurre algo similar al caso anterior. De nuevo es preciso llevar a cabo una diferenciación de la serie no estacionaria, de modo que en el modelo general $ARIMA(p, d, q)$ el valor del parámetro d sería $d = 1$.





Dado que la serie E dispone de $T = 300$ observaciones, el intervalo de no significatividad sería $\left(-\frac{1,96}{\sqrt{300}}, \frac{1,96}{\sqrt{300}}\right) = (-0,11316,\ 0,11316)$. De nuevo tras la diferenciación, como se observa en los gráficos siguientes, el primer retardo de la FAC y el primer retardo de la FAP siguen siendo significativos, señal de que hay cierta estructura que todavía no hemos reflejado en la ecuación del modelo.





Por otra parte, llama la atención la significatividad del retardo 9 en la FAC de la serie E diferenciada. Esta significatividad puede ser, o bien consecuencia de la existencia de un outlier o valor atípico, o bien consecuencia de la existencia de estructura estacional. Dado que no se dispone de datos más allá de los valores de la FAC y de la FAP, no se pueden llevar a cabo contrastes para determinar la verdadera causa de dicha significatividad y suponemos que el motivo de significatividad del noveno retardo es meramente aleatorio. Llegados a este punto, pueden considerarse dos enfoques distintos. El primero de ellos, es pensar que se trata de un modelo $MA(1)$ ya que tanto la FAC como la FAP tienen un único retardo significativo y que la FAP presenta una convergencia a cero alternando en signo. Por tanto, el modelo estímado será de la forma

$$(1-L)\,y_t = (1-\theta_1 L)u_t \Leftrightarrow y_t - y_{t-1} = u_t - \theta_1 u_{t-1}$$

Por el mismo razonamiento que en el apartado B, la función de autocorrelación de un proceso $MA(1)$ es:

$$\rho_k = \begin{cases} 1 & \text{si } k=0 \\ -\frac{\theta_1}{1+\theta_1^2} & \text{si } k=1 \\ 0 & \text{si } k>1 \end{cases}.$$

Por tanto al ser $\widehat{\rho_1} = 0,42$ se tiene que

$$0,42 = -\frac{\theta_1}{1+\theta_1^2} \Rightarrow \theta_1 = -0,544.$$

La otra raíz se descarta porque $|\theta_1| > 1$ y por tanto la serie no sería estacionaria. Por tanto, el modelo estimado por este primer enfoque es

$$y_t - y_{t-1} = u_t + 0,544u_{t-1}.$$

El segundo de los enfoques que se proponen, es considerar que no se perciben de forma clara las características de la posible estructura de media movil $MA(1)$ propuesta. Puede pensarse que la FAC no muestra un decrecimiento amortiguado hacia cero ni que en la FAP se produce una convergencia exponencial a cero y que por tanto no se puede determinar una estructura autorregresiva o una estructura de medias móviles. De esta forma, parece necesario incluir términos tanto autorregresivos como de medias móviles. Es decir, se trata de un modelo mixto autorregresivo - medias móviles. La FAC presenta un pico en el retardo 1, en el que influye tanto la parte autorregresiva como la de medias móviles del modelo y, a continuación presenta un comportamiento amortiguado hacia cero pero sin llegar a anularse. Podría preocupar la significatividad del retardo noveno pero debido a que es un retardo alejado del principio de la serie, se piensa que es consecuencia de las interacciones de la serie y que dejará de ser significativo cuando se lleve a cabo la modelización de la serie. En cuanto a la FAP, esta no se anula como consecuencia de la parte MA a pesar de presentar un comportamiento amortiguado hacia cero y además de nuevo el primero de los retardos es claramente significativo. Vistas estas características, el orden de la parte autorregresiva será tanto $p = 1$ e, igualmente, el orden de la parte de medias móviles será $q = 1$ ya que el primero es el único retardo significativo tanto en la FAC como en la FAP. Por tanto, recordando que se ha tomado una diferencia para hacer que el modelo fuera estacionario, la estructura determinada para el modelo sería el modelo general ARIMA(1,1,1) que se escribe como

$$(1 - \phi_1 L)(1 - L)\, y_t = (1 - \theta_1 L)\, u_t.$$

Para la estimación de los parámetros es necesario calcular las expresiones de la varianza, la función de autocovarianza y las funciones de autocorrelación y de autocorrelación parcial del modelo ARMA(1,1). Cumplida la condición de estacionariedad de la serie, $E(y_t) = E(y_{t-1}) = \mu$. Entonces, el modelo ARMA (1,1) se puede escribir sin hacer uso de los operadores retardos, esto es,

$$(1 - \phi_1 L)\, y_t = (1 - \theta_1 L)\, u_t \Leftrightarrow y_t = \phi_1 y_{t-1} + u_t - \theta_1 u_{t-1}$$

El cálculo de la varianza será:

$$\begin{aligned}\gamma_0 &= E\left[y_t^2\right] = E\left[\left(\phi_1 y_{t-1} + u_t - \theta_1 u_{t-1}\right)^2\right] = \\ &= \phi_1^2 E\left[y_{t-1}^2\right] + E(u_t^2) + \theta_1^2 E\left(u_{t-1}^2\right) + 2\phi_1 E\left(y_{t-1}u_t\right) - 2\phi_1\theta_1 E\left(y_{t-1}u_{t-1}\right) + \\ &- 2\theta_1 E\left(u_t u_{t-1}\right)\end{aligned}$$

donde

$$\begin{aligned}E\left(y_{t-1}u_t\right) &= E\left[\left(\phi_1 y_{t-2} + u_{t-1} - \theta_1 u_{t-2}\right)u_t\right] = 0, \\ E\left(y_{t-1}u_{t-1}\right) &= E\left[\left(\phi_1 y_{t-2} + u_{t-1} - \theta_1 u_{t-2}\right)u_{t-1}\right] = \sigma^2.\end{aligned}$$

En consecuencia,

$$\begin{gathered}\gamma_0 = \phi_1^2\gamma_0 + \sigma^2 + \theta_1^2\sigma^2 - 2\phi_1\theta_1\sigma^2 = \phi_1^2\gamma_0 + \sigma^2\left(1 + \theta_1^2 - 2\phi_1\theta_1\right) \\ \Rightarrow \gamma_0 = \frac{\sigma^2\left(1 + \theta_1^2 - 2\phi_1\theta_1\right)}{1 - \phi_1^2}\end{gathered}$$

El cálculo de la función de autocovarianza, para $j = 1$ será:

$$\begin{aligned}\gamma_1 &= E\left[y_t y_{t-1}\right] = E\left[y_{t-1}\left(\phi_1 y_{t-1} + u_t - \theta_1 u_{t-1}\right)\right] = \phi_1\gamma_0 - \theta_1\sigma^2 = \\ &= \phi_1\frac{\sigma^2\left(1 + \theta_1^2 - 2\phi_1\theta_1\right)}{1 - \phi_1^2} - \theta_1\sigma^2 = \frac{\sigma^2\phi_1\left(1 + \theta_1^2 - 2\phi_1\theta_1\right) - \left(1 - \phi_1^2\right)\theta_1\sigma^2}{1 - \phi_1^2} = \\ &= \frac{\sigma^2\left(\phi_1 + \phi_1\theta_1^2 - 2\phi_1^2\theta_1 - \theta_1 + \theta_1\phi_1^2\right)}{1 - \phi_1^2} = \frac{\sigma^2\left(\phi_1 - \theta_1 + \phi_1\theta_1^2 - \phi_1^2\theta_1\right)}{1 - \phi_1^2} = \\ &= \frac{\sigma^2\left(1 - \phi_1\theta_1\right)\left(\phi_1 - \theta_1\right)}{1 - \phi_1^2}.\end{aligned}$$

Para $j > 1$,

$$\gamma_j = E\left[y_t y_{t-j}\right] = E\left[y_{t-j}\left(\phi_1 y_{t-1} + u_t - \theta_1 u_{t-1}\right)\right] = \phi_1\gamma_{j-1}.$$

Por tanto, la función de autocorrelación se obtiene directamente de la expresión $\rho_j = \frac{\gamma_j}{\gamma_0}$. Esto es,

$$\rho_j = \left\{\begin{array}{cl} \frac{(1-\phi_1\theta_1)(\phi_1-\theta_1)}{1+\theta_1^2-2\phi_1\theta_1} & \text{para } j = 1 \\ \phi_1\rho_{j-1} & \text{para } j > 1 \end{array}\right. .$$

El valor del coeficiente de la parte autorregresiva de obtiene como sigue:

$$\rho_j = \phi_1 \rho_{j-1} \Rightarrow \widehat{\rho}_2 = \widehat{\phi}_1 \widehat{\rho}_1 \Rightarrow \widehat{\phi}_1 = \frac{-0,01}{0,42} = -0,023.$$

Por otro lado, para $j = 1$ se tiene:

$$\rho_1 = \frac{(1 - \phi_1 \theta_1)(\phi_1 - \theta_1)}{1 + \theta_1^2 - 2\phi_1 \theta_1}$$

Por lo que,

$$0,42 = \frac{\left(1 + 0,023\widehat{\theta}_1\right)\left(-0,023 - \widehat{\theta}_1\right)}{1 + \widehat{\theta}_1^2 + 0,046\widehat{\theta}_1}$$

$$0,42 + 0,42\widehat{\theta}_1^2 + 0,01932\widehat{\theta}_1 = -0,023 - 1,000529\widehat{\theta}_1 + 0,023\widehat{\theta}_1^2$$

$$0,397\widehat{\theta}_1^2 + 1,019849\widehat{\theta}_1 + 0,443 = 0$$

De la ecuación anterior, se obtiene $\widehat{\theta}_1 = -0,581$. La otra solución se descarta porque $|\widehat{\theta}_1| > 1$. Por lo que el modelo estimado sería:

$$(1 + 0,023L)(1 - L)\, y_t = (1 + 0,581L)\, u_t,$$

que, sin utilizar operadores retardo, tenemos:

$$y_t - 0,977y_{t-1} - 0,023y_{t-2} = u_t + 0,581u_{t-1}.$$

Recopilando los modelos estimados por cada uno de los enfoques, tenemos:

	Estructura	Modelo estimado
Enfoque 1	$IMA(1,1)$	$y_t - y_{t-1} = u_t + 0,544u_{t-1}$
Enfoque 2	$ARIMA(1,1,1)$	$y_t - 0,977y_{t-1} - 0,023y_{t-2} = u_t + 0,581u_{t-1}$

A pesar de que se observan diferencias debido a que los modelos estimados no son más que aproximaciones, pues el enunciado proporciona poca información, observamos que la parte MA, a la derecha del igual en los dos modelos estimados por sendos enfoques, es muy similar. De manera análoga, en la expresión que aparece a la izquierda del igual, vemos que los valores de los coeficientes que acompañan a y_{t-1} son muy próximos, siendo 1 en el primer enfoque y 0,977 en el segundo. De la misma forma, el valor del coeficiente que acompaña a y_{t-2} en el segundo enfoque es bastante pequeño, no apareciendo y_{t-2} en la modelización del primer enfoque.

Cuestión 9

Dada la siguiente tabla de mortalidad,

Edad	Tasa de mortalidad*	Promedio de años vividos el último año de vida	Riesgo de muerte*	Supervivientes	Defunciones teóricas	Población estacionaria	Tiempo por vivir	Esperanza de vida
4 años	0,103	0,504	0,103	99697,657	10,262		7920017,651	79,440
5 años	0,096	0,413		99687,395	9,575	99681,776	7820325,081	
(...)								
8 años	0,060	0,518	0,060	99671,017	6,022	99668,117	7620968,585	76,461
9 años	0,080	0,480	0,080		7,970	99660,851	7521300,468	75,466
10 años	0,057	0,499	0,057	99657,025		99654,168	7421639,617	74,472

* Riesgos y tasas de mortalidad vienen referidos a 1.000 habitantes

Se pide calcular:

a) El número de supervivientes de 9 años.

b) Población estacionaria de 4 años.

c) Defunciones teóricas de 10 años.

d) Esperanza de vida a los 5 años.

e) Riesgo de muerte a los 5 años.

Solución.

a)

$$l_9 = l_8 - d_8 = 99671,017 - 6,022 = 99665,049 \text{ supervivientes}$$

b)

$$L_4 = l_5 + a_4 \cdot d_4 = 99687,395 + 0,504 \cdot 10,262 = 99692,56705$$

c)

$$q_{10} = \frac{d_{10}}{l_{10}} \Rightarrow d_{10} = q_{10} \cdot l_{10} = \frac{0,057}{1000} \cdot 99657,025 = 5,6804 \text{ defunciones teóricas}$$

d)

$$e_5 = \frac{T_5}{l_5} = \frac{7820325,081}{99687,395} = 78,4484 \text{ años}$$

e)

$$q_5 = \frac{d_5}{l_5} = \frac{9,575}{99687,395} = 0,09605‰$$

Otra forma de calcularlos es mediante la siguiente expresión:

$$q_5 = \frac{m_5}{1 + (1 - a_5)\, m_5} = \frac{0,000096}{1 + (1 - 0,413)\, 0,000096} = 0,09599‰$$

donde la diferencia obtenida respecto al calculo anterior se debe al redondeo en los datos proporcionados por el enunciado.

Cuestión 10

Considere la siguiente matriz de migración de las regiones A, B y C de un país en un año dado t.

	Región de residencia actual				
Región de nacimiento	A	B	C	Fuera de la región	Total
A	10.359	53	74	127	10.613
B	100	8.250	22	79	8.451
C	88	15	9.750	94	9.947
Fuera de la región	250	48	69	-	-
En el extranjero	303	99	108	-	-
Total	11.100	8.465	10.023	-	-

Calcule:

a) El índice de aloctonía de A.

b) La proporción de emigrantes de B.

c) La proporción de inmigrantes de C.

d) La tasa bruta de intercambio entre A y B.

e) La tasa neta de intercambio entre B y C.

f) El índice de compensación entre A y B.

Solución.

a) El índice de aloctonía de A se calcula como el cociente entre el número de no nativos en la región A en el año t y la población en la región A en el momento actual.

$$\acute{I}ndice\ aloctonia_A = \frac{No\ nativos_A}{P_A^t} = \frac{11100 - 10359}{11100} = \frac{741}{11100} = 0,0667.$$

Lo cual significa que el 6,67 % de la población de la región A no es nativa de la misma, sino que inmigró a ella en algún momento de su vida.

b) La proporción de emigrantes de B relaciona el número total de emigrantes de la región y la población en el momento inicial.

$$PE_B = \frac{E_B}{P_B^{t-n}} = \frac{8451 - 8250}{8451} = 0,0237.$$

Lo que significa que el 2,37 % de la población de la región B emigra a lo largo del periodo.

c) La proporción de inmigrantes de C corresponde a la relación entre el número de personas inmigrantes y la población total de dicha región en el momento final del periodo de estudio.

$$PI_C = \frac{I_C}{P_C^t} = \frac{10023 - 9750}{10023} = \frac{273}{10023} = 0,0272.$$

Así, el 2,03 % de la población de la región C se debe a inmigración.

d) La tasa bruta de intercambio entre A y B puede ser definida como

$$TBI = \frac{M_{AB} + M_{BA}}{P_A + P_B} \cdot 1000$$

donde M_{AB} es la población con residencia pasada en A que ahora tiene su residencia en B, M_{BA} es la población con residencia pasada en B que ahora tiene su residencia en A, P_A es la población actual de la región A y P_B sería análogo. Por tanto, la tasa bruta de intercambio entre A y B será

$$TBI = \frac{53 + 100}{11100 + 8465} \cdot 1000 = \frac{153}{19565} \cdot 1000 = 7,82‰$$

e) La tasa neta de intercambio entre B y C se puede definir como

$$TNI = \frac{M_{BC} - M_{CB}}{P_B + P_C} \cdot 1000$$

donde de nuevo M_{BC} es la población con residencia pasada en B que ahora tiene su residencia en C, M_{CB} es la población con residencia pasada en C que ahora tiene su residencia en B y P_B, P_C son las poblaciones actuales de la región B y la región C respectivamente. Así, la tasa neta de intercambio entre B y C será

$$TNI = \frac{22 - 15}{8465 + 10023} \cdot 1000 = \frac{7}{18488} \cdot 1000 = 0,3786‰$$

f) Uno de los índices más comunes para comparar las corrientes migratorias entre dos regiones es el índice de compensación. Sea un espacio definido por las regiones A y B; si se consideran los movimientos migratorios que afectan a la región A, el índice de compensación para esta región se define como la relación entre la migración neta entre las regiones A y B, y la migración total de A, es decir la suma de corrientes migratorias procedentes de A con destino a B y las procedentes de B con destino a A:

$$\textit{Índice de compensación} = \frac{M_{AB} - M_{BA}}{M_{AB} + M_{BA}} = \frac{\textit{Saldo neto}}{\textit{Migración bruta}}$$

donde la defnción de M_{AB} y M_{BA} es equivalente a la que se ha presentado en los casos anteriores. El índice de compensación entre las regiones A y B es

$$\textit{Índice de compensación} = \frac{53-100}{53+100} = \frac{-47}{153} = -0,307$$

Un índice de compensación negativo es señal de que la región B pierde más población de la que recibe de la región A.

2.2. Soluciones del tercer examen del Cuerpo de Diplomados en Estadística del Estado de 2019

Cuestión 1

Los jugadores A y B lanzan un dado cada uno: si el dado de A marca un valor mayor o igual que el de B, comienza a jugar A; en caso contrario, comienza B. El juego consiste en sacar una bola de una urna que contiene una bola roja, una verde y una negra. Si comienza el jugador A y la bola es roja, gana el jugador A. Si la bola no es roja, se devuelve a la urna y saca una bola el jugador B, que gana si la bola es verde o negra; en caso contrario, gana A. Si empieza el jugador B, y la bola es roja o verde, gana el jugador B; en caso contrario, gana A. Se pide calcular las siguientes probabilidades, razonando la respuesta:

a) La probabilidad de que el jugador A inicie el juego. ¿Y de que lo inicie el jugador B?

b) La probabilidad de que el jugador A gane el juego. ¿Y de qué lo gane el jugador B?

c) Si el juego lo ha ganado el jugador B ¿Cuál es la probabilidad de que lo haya iniciado?

Solución.

Se definen A=“puntuación de A al lanzar un dado” y B=“puntuación de B al lanzar un dado”.

a) Puesto que el conjunto de sucesos del experimento aleatorio es finito, podemos utilizar el teorema de la probabilidad total de la siguiente forma:

$$\begin{aligned}P\left(A \geq B\right) =& P(B=1)\cdot P(A\geq 1|B=1) + P(B=2)\cdot P(A\geq 2|B=2)+\\ &+\cdots + P(B=6)\cdot P(A\geq 6|B=6) =\\ =&\frac{1}{6}\cdot 1 + \frac{1}{6}\cdot\frac{5}{6} + \frac{1}{6}\cdot\frac{4}{6} + \frac{1}{6}\cdot\frac{3}{6} + \frac{1}{6}\cdot\frac{2}{6} + \frac{1}{6}\cdot\frac{1}{6} =\\ =&\frac{1}{6}\left(1+\frac{5}{6}+\frac{4}{6}+\frac{3}{6}+\frac{2}{6}+\frac{1}{6}\right) = \frac{1}{6}\left(\frac{7}{2}\right) =\\ =&\frac{7}{12}.\end{aligned}$$

Por tanto:

$$P\left(A_{sale}\right) = P\left(A \geq B\right) = \frac{7}{12};$$
$$P\left(B_{sale}\right) = 1 - P\left(A_{sale}\right) = \frac{5}{12}.$$

b) Se define ahora A_G="A gana el juego", entonces:

$$\begin{aligned}P\left(A_G\right) &= P(A_{sale})P(A_G|A_{sale}) + P(B_{sale})P(A_G|B_{sale}) =\\ &= \frac{7}{12}\cdot\left(\frac{1}{3}+\frac{2}{3}\cdot\frac{1}{3}\right) + \frac{5}{12}\cdot\frac{1}{3} =\\ &= \frac{7}{12}\cdot\frac{5}{9} + \frac{5}{12}\cdot\frac{1}{3} =\\ &= \frac{25}{54}.\end{aligned}$$

Sea B_G="B gana el juego", se tiene:

$$\begin{aligned}P\left(B_G\right) &= P(A_{sale})P(B_G|A_{sale}) + P(B_{sale})P(B_G|B_{sale}) =\\ &= \frac{7}{12}\cdot\left(\frac{2}{3}\cdot\frac{2}{3}\right) + \frac{5}{12}\cdot\frac{2}{3} =\\ &= \frac{7}{12}\cdot\frac{4}{9} + \frac{5}{12}\cdot\frac{2}{3} =\\ &= \frac{29}{54}.\end{aligned}$$

Se obtiene el mismo resultado si se calcula la probabilidad de que B gane el juego como $P(B_G) = 1 - P(A_G)$.

c) La probabilidad de que B haya iniciado el juego dado que lo ha ganado es:

$$P(B_{sale}|B_G) = \frac{P(B_{sale}) \cdot P(B_G|B_{sale})}{P(B_G)} = \frac{\frac{5}{12} \cdot \frac{2}{3}}{\frac{29}{54}} = \frac{15}{29}.$$

Cuestión 2

Una muestra aleatoria simple con reemplazamiento de 900 familias es seleccionada en una ciudad formada por 24.000 familias obteniendo los siguientes resultados:

$$\sum_{i=1}^{900} x_i = 9000 \qquad \sum_{i=1}^{900} x_i^2 = 162819$$

Donde x_i representa el dinero en la cuenta corriente bancaria de la familia i investigada (en miles de euros). Se desea estimar la cantidad media de dinero depositado en las cuentas corrientes bancarias:

a) Obtenga un intervalo de confianza al 95 % para la cantidad que se desea estimar.

b) Se agrupan las familias de la ciudad en dos estratos de 4.000 familias de renta alta y 20.000 de renta baja. Además, se sabe que la cantidad media depositada por las familias de renta alta es nueve veces la de las rentas bajas y que las desviaciones típicas en cada estrato son el doble de las medias correspondientes. ¿Cómo distribuiría la muestra de 900 familias entre los dos estratos de manera que el error de muestreo sea mínimo?

(Nota: Para un valor $\alpha = 0,025$, el valor de la distribución normal que deja esa probabilidad a su derecha es $1,96$)

Solución.

a) Un estimador lineal adecuado para estimar la cantidad media de dinero depositado en las cuentas corrientes bancarias es:

$$\widehat{\overline{X}} = \frac{1}{n}\sum_{i=1}^{n} X_i = \overline{x},$$

que es un estimador insesgado, pues

$$E\left[\widehat{\overline{X}}\right] = E\left[\overline{x}\right] = E\left[\frac{\sum_{i=1}^{n} X_i}{n}\right] = E\left[\frac{\sum_{i=1}^{N} X_i e_i}{n}\right] = \frac{\sum_{i=1}^{N} X_i E\left[e_i\right]}{n} =$$
$$= \frac{\sum_{i=1}^{N} X_i \cdot nP_i}{n} = \frac{\sum_{i=1}^{N} X_i \cdot n\frac{1}{N}}{n} = \frac{1}{N}\frac{n\sum_{i=1}^{N} X_i}{n} =$$
$$= \frac{1}{N}\frac{nX}{n} = \frac{X}{N} = \overline{X}.$$

donde se ha utilizado que $P_i = \frac{1}{N}$ pues se trata de un muestreo con reposición. Además, dada la nota al pie del ejercicio y siendo el tamaño muestral suficientemente grande, se puede suponer que el estimador sigue una distribución normal. Entonces, el intervalo de confianza para la cantidad media de dinero depositado en las cuentas corrientes bancarias es

$$IC\left(\overline{X}\right) = \left[\widehat{\overline{X}} - \lambda_\alpha \widehat{\sigma}\left(\widehat{\overline{X}}\right), \widehat{\overline{X}} + \lambda_\alpha \widehat{\sigma} e\left(\widehat{\overline{X}}\right)\right]$$

donde $\widehat{\sigma}\left(\widehat{\overline{X}}\right)$ es la raíz cuadrada de la estimación de la varianza del estimador y $\lambda_{\alpha=0{,}025}$ es $1{,}96$ según el enunciado. Calculando los valores restantes:

$$\widehat{\overline{X}} = \overline{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{9000}{900} = 10$$

$$\widehat{S}^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}\right] = \frac{1}{899}\left[162819 - \frac{(9000)^2}{900}\right] = 81$$

$$\widehat{Var}\left(\widehat{\overline{X}}\right) = \frac{\widehat{S}^2}{n} = \frac{81}{900} = 0{,}09$$

$$\widehat{\sigma}\left(\widehat{\overline{X}}\right) = \sqrt{\widehat{Var}\left(\widehat{\overline{X}}\right)} = \sqrt{0,09} = 0,3$$

Y por tanto

$$IC\left(\widehat{\overline{X}}\right) = [10 - 1,96 \cdot 0,3\ ,\ 10 + 1,96 \cdot 0,3] = [9,412\ ,\ 10,588]$$

b) Se selecciona una muestra de $n = 900$ familias en una ciudad formada por $N = 24000$ familias. Se forman dos estratos, $h = 1, 2$: el primero de ellos con $N_1 = 4000$ familias de renta alta y el segundo de ellos con $N_2 = 20000$ familias de renta baja. Se sabe que la cantidad media depositada por las familias de renta alta es nueve veces la de las rentas baja, es decir $\overline{x}_1 = 9\overline{x}_2$, y que las desviaciones típicas en cada estrato, σ_1 y σ_2, son el doble de las medias correspondientes a cada uno de los estratos, esto es $\sigma_1 = 2\overline{x}_1$ y $\sigma_2 = 2\overline{x}_2$. La afijación que minimiza el error de muestreo establece que el número de elementos que forman parte de la muestra de cada estrato es

$$n_h = n \cdot \frac{N_h \sigma_h}{\sum_{h=1}^{L} N_h \sigma_h}.$$

Sabemos que

$$\sigma_1 = 2\overline{x}_1 = 2\left(9\overline{x}_2\right) = 18\overline{x}_2$$

$$\sigma_2 = 2\overline{x}_2$$

Por tanto

$$n_1 = n \cdot \frac{N_1 \sigma_1}{\sum_{h=1}^{L} N_h \sigma_h} = n \cdot \frac{N_1 \cdot 18\overline{x}_2}{N_1 \cdot 18\overline{x}_2 + N_2 \cdot 2\overline{x}_2} =$$

$$= 900 \cdot \frac{4000 \cdot 18\overline{x}_2}{4000 \cdot 18\overline{x}_2 + 20000 \cdot 2\overline{x}_2} =$$

$$= 900 \cdot 0,64288 = 578,592$$

$$n_2 = n \cdot \frac{N_2 \sigma_2}{\sum_{h=1}^{L} N_h \sigma_h} = n \cdot \frac{N_2 \cdot 2\overline{x}_2}{N_1 \cdot 18\overline{x}_2 + N_2 \cdot 2\overline{x}_2} =$$

$$= 900 \cdot \frac{20000 \cdot 2\overline{x}_2}{4000 \cdot 18\overline{x}_2 + 20000 \cdot 2\overline{x}_2} =$$
$$= 900 \cdot 0,35711 = 321,407$$

De modo que la distribución de la muestra de 900 familias entre los dos estratos será $n_1 = 579$ y $n_2 = 321$ para que el error de muestreo sea mínimo.

Cuestión 3

Se realiza el experimento de lanzar una moneda tres veces y registrar el número de caras obtenidas. Al repetir 80 veces el experimento se ha obtenido 7 veces ninguna cara, 24 veces una cara, 35 veces dos caras y 14 veces tres caras. Si p es "la probabilidad de obtener una cara al lanzar la moneda", se pide lo siguiente:

a) Obtenga el estimador de p por el método de los momentos.

b) ¿Es insesgado el estimador obtenido en el apartado anterior?

c) Calcule la varianza del estimador obtenido en el apartado (a).

d) Plantee y resuelva el contraste adecuado para estudiar si la moneda está equilibrada.

(Nota: para un valor $\alpha = 0{,}05$, el valor crítico necesario de la distribución del estadístico es $7,81$)

Solución. Se considera el experimento "lanzar una moneda 3 veces" y se define $X =$ "número de caras obtenidas en los 3 lanzamientos" de forma que puede tomar los valores $\{0, 1, 2, 3\}$. Se repite el experimento 80 veces y se obtiene

X	n_i
0	7
1	24
2	35
3	14

Sea $p = P\left(\text{“obtenter cara al lanzar la moneda''}\right)$, la variable X definida como hemos dicho sigue una distribución Binomial de parámetros $n = 3$ y p desconocido, es decir $X \sim B\left(3, p\right)$.

a) Para obtener el estimador por el método de los momentos es necesario igualar los momentos muestrales a los momentos poblacionales. Así el primer momento poblacional es:

$$\alpha_1 = E\left[X\right] = 3p$$

Ahora se iguala al primer momento muestral y se obtiene el estimador de p por el método de los momentos.

$$a_1 = \alpha_1 \Rightarrow \overline{x} = 3p \Rightarrow \widehat{p}_{MM} = \frac{\overline{x}}{3}$$

Utilizando los datos proporcionados en el enunciado, para esta realización del experimento se tiene que el estimador por el método de los momentos es

$$\widehat{p}_{MM} = \frac{\frac{0 \cdot 7 + 1 \cdot 24 + 2 \cdot 35 + 3 \cdot 14}{7+24+35+14}}{3} = \frac{24 + 70 + 42}{3 \cdot 80} = \frac{136}{240} = \frac{17}{30} \simeq 0,566.$$

b) Para ver si es insesgado, se toma la esperanza:

$$E\left[\widehat{p}_{MM}\right] = E\left[\frac{\frac{\sum_{i=1}^{n} x_i}{n}}{3}\right] = \frac{1}{3n}\sum_{i=1}^{n} E\left(x_i\right) = \frac{1}{3n}\sum_{i=1}^{n} 3p = \frac{1}{3n} \cdot 3p \cdot n = p$$

donde se ha utilizado que $E\left[x_i\right] = 3p$ al ser $X \sim B\left(3, p\right)$ y queda demostrado que el estimador es insesgado.

c) Puesto que $X \sim B(3,p)$, $Var(X) = 3p(1-p)$, y la varianza del estimador es:

$$Var(\widehat{p}_{MM}) = Var\left(\frac{\overline{x}}{3}\right) = \frac{1}{9}Var(\overline{x}) = \frac{1}{9}Var\left(\sum_{i=1}^{n}\frac{x_i}{n}\right) = \frac{1}{9n^2}Var\left(\sum_{i=1}^{n}x_i\right) =$$
$$= \frac{1}{9n}Var(x_i) = \frac{1}{9n}3p(1-p) = \frac{p(1-p)}{3n}.$$

d) Finalmente, que la moneda esté equilibrada es equivalente a llevar a cabo un contraste de bondad de ajuste donde las hipótesis a contrastar son

$$\begin{cases} H_0 : p = \frac{1}{2} \\ H_1 : p \neq \frac{1}{2} \end{cases}.$$

Bajo la hipótesis nula $p = \frac{1}{2}$ las frecuencias teóricas se calculan utilizando

$$P(X=0) = \frac{1}{8}, \quad P(X=1) = \frac{3}{8}, \quad P(X=2) = \frac{3}{8}, \quad P(X=3) = \frac{1}{8}.$$

Siendo $n = 80$ y $np_i = n \cdot P(X=i)$, se tiene:

X	np_i
0	10
1	30
2	30
3	10

El estadístico de contraste es

$$\chi^2 = \sum_{i=1}^{n}\frac{(n_i - np_i)^2}{np_i} =$$
$$= \frac{(7-10)^2}{10} + \frac{(24-30)^2}{30} + \frac{(35-30)^2}{30} + \frac{(14-10)^2}{10} =$$
$$= \frac{9}{10} + \frac{36}{30} + \frac{25}{30} + \frac{16}{10} = \frac{68}{15} = 4,5333.$$

El valor crítico de la distribución del estadístico para un valor $\alpha = 0{,}05$ es $7,81$, que es mayor que el valor obtenido para el estadístico de contraste. Por tanto, no se puede rechazar la hipótesis nula con un nivel de confianza del 95 % y se concluye que la moneda está equilibrada.

Cuestión 4

En un proceso de fabricación se analizan dos variables cuantitativas X e Y, obteniéndose los siguientes resultados: (0, 2), (1, 6), (3, 14), (-1, -2) y (2, 10). Calcule:

a) Las distribuciones marginales.

b) La distribución de las frecuencias relativas de X para valores de $Y > 2,5$.

c) El coeficiente de correlación lineal de ambas variables.

d) Los valores de X e Y para las siguientes observaciones: $(-3, y)$ y $(x, 4)$

Solución. Los resultados obtenidos en el proceso de fabricación se recogen en la siguiente tabla:

$X \vert Y$	-2	2	6	10	14
-1	1	0	0	0	0
0	0	1	0	0	0
1	0	0	1	0	0
2	0	0	0	1	0
3	0	0	0	0	1

a) Las distribuciones marginales serán:

X	n_i	f_i
-1	1	$\frac{1}{5}$
0	1	$\frac{1}{5}$
1	1	$\frac{1}{5}$
2	1	$\frac{1}{5}$
3	1	$\frac{1}{5}$

Y	n_j	f_j
-2	1	$\frac{1}{5}$
2	1	$\frac{1}{5}$
6	1	$\frac{1}{5}$
10	1	$\frac{1}{5}$
14	1	$\frac{1}{5}$

b) La distribución de las frecuencias relativas de X para valores de $Y > 2,5$ se calcula usando $f_{i|j} = \frac{f_{ij}}{f_{\cdot j}}$ y teniendo en cuenta que la variable $X|Y > 2,5$ toma los valores $\{-1, 0, 1, 2, 3\}$. Por tanto

$$f_{-1|Y>2,5} = \frac{f_{-1,Y>2,5}}{f_{Y>2,5}} = 0$$
$$f_{0|Y>2,5} = \frac{f_{0,Y>2,5}}{f_{Y>2,5}} = 0$$
$$f_{1|Y>2,5} = \frac{f_{1,Y>2,5}}{f_{Y>2,5}} = \frac{1}{3}$$
$$f_{2|Y>2,5} = \frac{f_{2,Y>2,5}}{f_{Y>2,5}} = \frac{1}{3}$$
$$f_{3|Y>2,5} = \frac{f_{3,Y>2,5}}{f_{Y>2,5}} = \frac{1}{3}$$

c) El coeficiente de correlación lineal será:

$$\rho = \frac{Cov\,(XY)}{\sqrt{Var\,(X)}\sqrt{Var\,(Y)}}$$

donde

$$
\begin{aligned}
Cov\left(XY\right) &= \frac{\sum_{i,j=1}^{n} x_i y_j n_{ij}}{n} - \overline{xy} = \\
&= \frac{0\cdot 2\cdot 1 + 1\cdot 6\cdot 1 + 3\cdot 14\cdot 1 + (-1)\cdot(-2)\cdot 1 + 2\cdot 10\cdot 1}{5} - 1\cdot 6 = \\
&= \frac{70}{5} - 6 = 8 \\
Var\left(X\right) &= \frac{\sum_{i=1}^{n} x_i^2 n_i}{n} - \overline{x}^2 = \\
&= \frac{0^2\cdot 1 + 1^2\cdot 1 + 3^2\cdot 1 + \left(-1\right)^2\cdot 1 + 2^2\cdot 1}{5} - 1^2 = \\
&= \frac{15}{5} - 1 = 2 \\
Var\left(Y\right) &= \frac{\sum_{j=1}^{n} y_j^2 n_j}{n} - \overline{y}^2 = \\
&= \frac{2^2\cdot 1 + 6^2\cdot 1 + 14^2\cdot 1 + \left(-2\right)^2\cdot 1 + 10^2\cdot 1}{5} - 6^2 = \\
&= \frac{340}{5} - 36 = 32
\end{aligned}
$$

siendo

$$
\begin{aligned}
\overline{x} &= \frac{\sum_{i=1}^{n} x_i}{n} = \frac{0+1+3-1+2}{5} = \frac{5}{5} = 1 \\
\overline{y} &= \frac{\sum_{j=1}^{n} y_j}{n} = \frac{2+6+14-2+10}{5} = \frac{30}{5} = 6
\end{aligned}
$$

Por tanto:

$$
\rho = \frac{Cov\left(XY\right)}{\sqrt{Var\left(X\right)}\sqrt{Var\left(Y\right)}} = \frac{8}{\sqrt{2}\sqrt{32}} = \frac{8}{8} = 1
$$

d) Los valores de X e Y para las observaciones dadas se obtienen al calcular la recta de regresión que pasa por esos puntos. Al ser $\rho = 1$, se sabe que hay relación lineal perfecta entre Y y X y entre X e Y. Por tanto, basta con calcular una de las rectas de regresión. La recta de regresión de Y sobre X será:

$$
Y = a + bX
$$

donde

$$b = \frac{Cov\,(X,Y)}{Var\,(X)} = \frac{8}{2} = 4 \qquad\qquad a = \overline{y} - b\overline{x} = 6 - 4 \cdot 1 = 2$$

de modo que la recta de regresión de Y sobre X será:

$$Y = 2 + 4X.$$

Sustituyendo en la ecuación obtenida se tienen los valores buscados para las observaciones propuestas:

$$x = -3 \Rightarrow y = 2 + 4 \cdot (-3) \Leftrightarrow y = -10 \Rightarrow (-3, -10)\,;$$
$$y = 4 \Rightarrow 4 = 2 + 4 \cdot x \Leftrightarrow x = \frac{4-2}{4} = \frac{1}{2} \Rightarrow \left(\frac{1}{2}, 4\right).$$

***Nota**. Se ha visto que el coeficiente de correlación, ρ, es igual a 1 por lo que se puede calcular la recta de regresión se X sobre Y simplemente despejando de la recta de regresión de Y sobre X:*

$$Y = 2 + 4X \Leftrightarrow Y - 2 = 4X \Leftrightarrow X = \frac{Y}{4} - \frac{1}{2}.$$

Cuestión 5

Un ayuntamiento proporcionó los siguientes datos relacionados con la edad de la población extranjera residente en su ciudad en el año 2017.

	Edad mínima	Edad máxima	Edad media	nº extranjeros	nº acum.	amplitud interv.		
	L_{i-1}	L_i	X_i	n_i	N_i	c_i	X_i*n_i	X_i^2*n_i
Menos de 16 años	0	16	8	930	930	16	7.440	59.520
De 16 a 24 años	16	25	20,5	985	1.915	9	20.192,5	413.946,25
De 25 a 44 años	25	45	35	3.840	5.755	20	134.400	4.704.000
De 45 a 64 años	45	65	55	5.345	11.100	20	293.975	16.168.625
De 65 años y más	65	90	77,5	720	11.820	25	55.800	4.324.500
Suma				11.820			511.808	25.670.591

Responda a los siguientes apartados, especificando en cada caso, la medida estadística propuesta, su cálculo y la interpretación del resultado:

a) ¿Cuál es la edad más frecuente entre los residentes extranjeros?

b) ¿Entre qué edades se encuentra la mitad central de los residentes extranjeros?

c) ¿Cuál es la edad máxima que tiene el 50 % de los extranjeros más jóvenes?

d) ¿Cuál es la edad media de los residentes extranjeros? ¿Es representativa de todo el colectivo?

e) Si el coeficiente de asimetría de Fisher de la distribución es 0,367 ¿puede afirmar que la distribución de edades de los residentes extranjeros es simétrica?

Solución. En primer lugar, se observa que se trata de una distribución de tipo III en la que las observaciones están distribuidas en intervalos que además son de distinta amplitud.

a) Para determinar la edad más frecuente entre los residentes se calcula la moda, ya que es una medida de posición central que representa el valor que más se repite, es decir, el que tiene máxima frecuencia. Para el cálculo de la moda en las distribuciones de tipo III hay dos criterios: el primer criterio establece que los valores del intervalo están uniformemente distribuidos. Así, en el intervalo modal, si suponemos que la moda está en un punto del intervalo tal que la distancia a los extremos inferior y superior del mismo son inversamente proporcionales a las frecuencias de los intervalos contiguos, la fórmula para el cálculo de la moda es:

$$Mo = L_{i-1} + \frac{d_{i+1}}{d_{i-1} + d_{i+1}} c_i$$

donde i hace referencia al intervalo modal, que será el de mayor densidad de frecuencia, $d_i = \frac{n_i}{c_i}$, donde c_i es la longitud del intervalo y n_i el número de observaciones en el intervalo. El segundo de los criterios para el cálculo de la moda en las distribuciones de tipo III supone que la moda está en un punto para el que la distancia a los extremos inferior y superior del intervalo son directamente proporcionales a las diferencias entre la frecuencia del intervalo modal y los intervalos contiguos a los extremos. La fórmula para el cálculo de la moda en este caso es:

$$Mo = L_{i-1} + \frac{(d_i - d_{i-1})}{(d_i - d_{i-1}) + (d_i - d_{i+1})} c_i$$

Las densidades de frecuencia serían:

	d_i
Menos de 16 años	$\frac{930}{16} = 58,125$
De 16 a 24 años	$\frac{985}{9} = 109,444$
De 25 a 44 años	$\frac{3840}{20} = 192$
De 45 a 64 años	$\frac{5345}{20} = 267,25$
De 65 y años y más	$\frac{720}{25} = 28,8$

De forma que el intervalo modal es de 45 a 64 años, al ser el intervalo de mayor densidad de frecuencia. Utilizando el primero de los criterios, la edad más frecuente entre los residentes es 47,6087 años:

$$Mo = 45 + \frac{28,8}{192 + 28,8} \cdot 20 = 47,6087.$$

Utilizando el segundo de ellos, la edad más frecuente entre los residentes es 49,7975 años:

$$Mo = 45 + \frac{267,25 - 192}{(267,25 - 192) + (267,25 - 28,8)} \cdot 20 = 49,7975.$$

b) La mitad central de los residentes extranjeros corresponde con el grupo de edad que comprende desde el primer cuartil hasta el tercer cuartil. Los cuartiles se calculan usando la siguiente fórmula:

$$Q_{\frac{r}{q}} = L_{i-1} + \frac{\frac{r}{q}N - N_{i-1}}{n_i} c_i$$

donde i hace referencia al intervalo con una frecuencia acumulada $N_i \geq \frac{r}{q}N$. En el caso del primer cuartil, $\frac{r}{q}N = \frac{25}{100}11820 = 2955$ y el intervalo de interés es de 25 a 44 años. Para el tercer cuartil, $\frac{r}{q}N = \frac{75}{100}11820 = 8865$, por lo que el intervalo en cuestión será de 45 a 64 años.

$$Q_{\frac{25}{100}} = 25 + \frac{\frac{25}{100}11820 - 1915}{3840} \cdot 20 = 30,42$$
$$Q_{\frac{75}{100}} = 45 + \frac{\frac{75}{100}11820 - 5755}{5345} \cdot 20 = 56,64$$

Así, la mitad central de los residentes extranjeros tienen entre 30,42 años y 56,64 años.

c) Para determinar la edad máxima que tiene el 50 % de los extranjeros más jóvenes hay que calcular la mediana, que determina el lugar central de la distribución como un equilibrio de frecuencias dejando el mismo número de frecuencias a ambos lados. La fórmula a aplicar es

$$Me = L_{i-1} + \frac{\frac{N}{2} - N_{i-1}}{n_i} c_i$$

donde ahora i hace referencia al intervalo mediano, que es aquel cuya N_i es la primera superior a $\frac{N}{2}$. Siendo $N = 11820$, $\frac{N}{2} = 5910$ por tanto el primer intervalo cuya N_i supera dicho valor es el intervalo de 45 a 64 años pues $N_i = 11100$ y la mediana será

$$Me = 45 + \frac{5910 - 5755}{5345} \cdot 20 = 45,58.$$

d) La edad media de los residentes extranjeros será

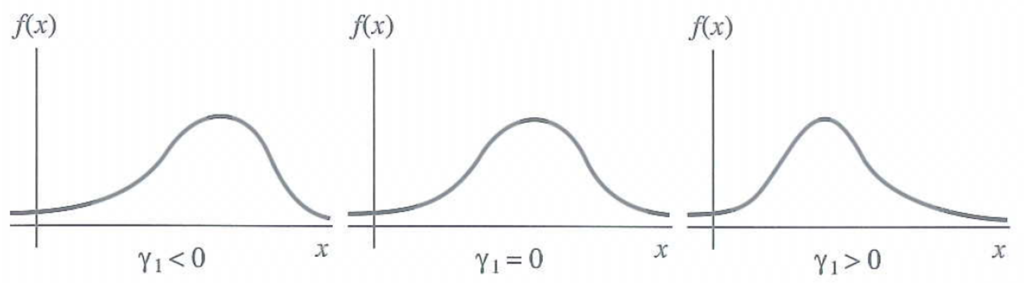
$$\overline{x} = \frac{\sum_{i=1}^{n} x_i n_i}{N} = \frac{511808}{11820} = 43,4$$

donde $N = \sum_{i=1}^{n} n_i$ y la edad media de los residentes extranjeros es 43,4 años.

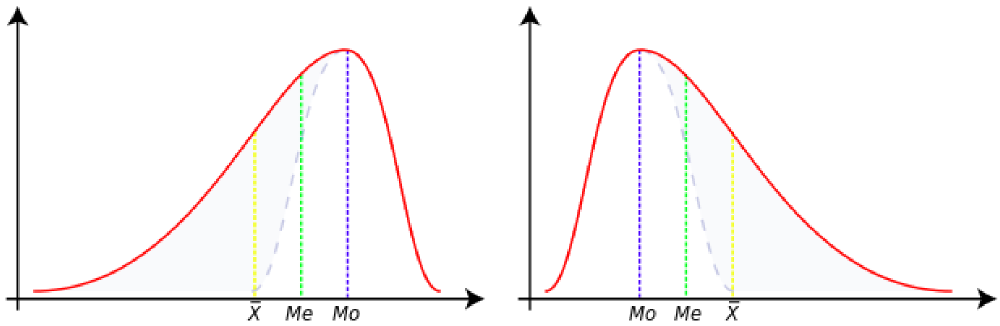
e) El coeficiente de asimetría de Fisher trata de medir y comparar las distancias de las observaciones que están a un lado y otro de la media aritmética para determinar la simetría de la distribución. Se define como:

$$\gamma_1 = \frac{\mu_3}{\sigma^3},$$

siendo μ_3 el momento respecto a la media de orden 3 y σ^3 la desviación típica elevada al cubo. Un coeficiente de asimetría de Fisher positivo es señal de que la distribución es asimétrica por la derecha.



Según el enunciado, para esta distribución $\gamma_1 = 0,367 > 0$ por lo que se trataría de una distribución asimétrica por la derecha. Sin embargo, los resultados obtenidos en los apartados anteriores no son coherentes con este coeficiente de asimetría de Fisher positivo. De hecho, los resultados obtenidos establecen lo contrario: al ser $\overline{x} < Me < Mo$, la distribución presenta asimetría por la izquierda, como se observa en el siguiente gráfico, que correspondería con un coeficiente de asimetría de Fisher negativo.



Si se calcula el coeficiente de asimetría de Fisher para esta distribución utilizando los datos del enunciado, se tiene:

$$\mu_3 = \frac{\sum_{i=1}^{n} (x_i - \overline{x})^3 n_i}{N} = \frac{(8-43,4)^3 \cdot 930 + \cdots + (77,5-43,4)^3 \cdot 720}{11820} =$$
$$= \frac{-17417617}{11820} = -1473,5716$$
$$\sigma^2 = \frac{\sum_{i=1}^{n} (x_i - \overline{x})^2 n_i}{N} = \frac{(8-43,4)^2 \cdot 930 + \cdots + (77,5-43,4)^2 \cdot 720}{11820} =$$
$$= \frac{3509262}{11820} = 296,8918$$
$$\sigma = \sqrt{\sigma^2} = \sqrt{296,8918} = 17,2305$$

Por tanto, el coeficiente de Fisher será

$$\gamma_1 = \frac{\mu_3}{\sigma^3} = \frac{-1473,5716}{(17,2305)^3} = -0,288$$

Este coeficiente de Fisher negativo sí concuerda con los resultados obtenidos en los apartados anteriores y la distribución presenta asimetría por la izquierda.

Cuestión 6

A partir de la información que aparece en la tabla adjunta, calcule, para el total de la economía:

1) La cuenta de bienes y servicios.

2) Las cuentas de producción y explotación.

3) La Renta Nacional Bruta.

4) El Ahorro Nacional Bruto.

5) La capacidad o necesidad de financiación del país. Interprete el resultado.

Gasto en consumo final	900
Producción	1.660
Formación bruta de capital	680
Consumos intermedios	240
Remuneración de asalariados (interior)	740
Remuneración de asalariados pagada al resto del mundo	40
Remuneración de asalariados recibida del resto del mundo	30
Rentas de la propiedad pagadas al resto del mundo	250
Rentas de la propiedad recibidas del resto del mundo	158
Impuestos netos sobre los productos	110
Otros impuestos sobre la producción	354
Otras subvenciones a la producción	116
Transferencias corrientes pagadas al resto del mundo	190
Transferencias corrientes recibidas del resto del mundo	70
Consumo de capital fijo	180
Transferencias de capital recibidas	28
Transferencias de capital pagadas	20
Transferencias sociales en especie	45
Importaciones de bienes y servicios	120
Exportaciones de bienes y servicios	70

Solución.

1) Cuenta de bienes y servicios.

R	Cuenta de bienes y servicios		E
1.660	Producción (P.1)	Consumos intermedios (P.2)	240
110	Imps. netos sobre los productos (D.21 - D.31)	Gasto en consumo final (P.3)	900
		Formación bruta de capital (P.5)	680
120	Importaciones de bienes y servicios (P.7)	Exportaciones de bienes y servicios (P.6)	70
1.890			1.890

2) Cuenta de producción.

E	Cuenta de producción		R
240	Consumos intermedios (P.2)	Producción (P.1)	1.660
		Impuestos netos sobre los productos (D.21-D.31)	110
1.530	PIB pm (B1*g)		
-180	CCF (P.51c)		
1.350	PIN pm (B1*n)		

Cuenta de explotación.

E	Cuenta de explotación		R
740	Remuneración de asalariados interior (D.1)	PIB pm (B1*g)	1.530
110	Impuestos netos sobre los productos (D.21-D.31)		
354	Otros impuestos. sobre la producción (D.29)		
-116	Otras subvenciones a la producción (D.39)		
442	EBE+RM (B.2*g+B.3g)		

3) La renta nacional bruta es el saldo de la cuenta de asignación de la renta primaria.

En esta cuenta, en la columna de recursos, al estar calculando las cuentas para el total de la economía, se podía haber simplificado la misma poniendo el PIB a precios de mercado y añadiendo los agregados relativos a la remuneración de asalariados y las rentas de la propiedad recibidos del resto del mundo.

E	Cuenta de asignación de la renta primaria		R
250	Rentas de la propiedad pagadas al resto del mundo (D.4)	EBE+RM (B.2*g+B.3g)	442
		Remuneración de asalariados (interior) (D.1)	740
40	Remuneración de asalariados pagada al resto del mundo (D.1)	Remuneración de asalariados recibida del resto del mundo (D.1)	30
		Impuestos netos sobre los productos (D.21-D.31)	110
		Otros impuestos. sobre la producción (D.29)	354
		Otras subvenciones a la producción (D.39)	-116
		Rentas de la propiedad recibidas del resto del mundo (D.4)	158
1.428	RNBpm (B.5*g)		

4) El ahorro nacional bruto es el saldo de la cuenta de utilización de la renta disponible, pero previamente hay que completar la cuenta de distribución secundaria de la renta.

E	Cuenta de distribución secundaria de la renta		R
190	Transferencias corrientes pagadas al resto del mundo (D.5)	RNBpm (B.5*g)	1.428
		Transferencias corrientes recibidas del resto del mundo (D.5)	70
1.308	RNBD(B.6*g)		

E	Cuenta de utilización de la renta disponible		R
900	Gasto en consumo final (P.3)	RNBD(B.6*g)	1.308
408	Ahorro bruto (B.8*g)		
-180	-CCF (P.51c)		
228	Ahorro neto (B.8*n)		

5) Finalmente, la capacidad o necesidad de financiación es el saldo de la cuenta de capital que se divide en dos cuentas: la primera de ellas es la cuenta de variaciones del patrimonio neto debidas al ahorro y a las transferencias de capital y la segunda de ellas es la cuenta de adquisiciones de activos no financieros.

Cuenta de Capital.

VA	VPNDAHYTC		VP/PN
		Ahorro neto (B.8*n)	228
		Transferencias de capital recibidas (D.9r)	28
		Transferencias de capital pagadas(D.9p)	-20
236	VPNDAHYTC (B.101)		

Cuenta de Capital.

VA	Adquisición de activos no financieros		VP/PN
680	Formación bruta de capital (P.5)	VPNDAHYTC (B.101)	236
-180	-CCF (P.51c)		
-264	Necesidad de financiación (B.9)		

El país presenta necesidad de financiación frente al resto del mundo por valor de 264 unidades monetarias. Corresponde a la financiación de un déficit del país frente al resto del mundo, es decir, cantidad que el país en cuestión se ve obligado a pedir prestado al resto del mundo.

Cuestión 7

En una economía con tres ramas de actividad, en la que cada una elabora un solo tipo de producto, y en ausencia de impuestos, subvenciones y márgenes de distribución, se conocen los siguientes datos:

1) Datos de la utilización de la producción como consumo intermedio:

    I) De la producción de la rama 1 se destinan 28 millones de euros(M€) a consumo intermedio de la propia rama; 18M€ se adquieren como consumo intermedio por la rama 2; y 23M€ como consumo intermedio por la rama 3.

    II) De la producción de la rama 2 se destinan 15 M€ a consumo intermedio de la propia rama, 22M€ a consumo intermedio de la rama 1 y 15M€ a consumo intermedio de la rama 3.

    III) De la producción de la rama 3 se destinan 19M€ de su producción a consumo intermedio de la propia rama, 20M€ a consumo intermedio de la rama 1 y 28M€ a consumo intermedio de la rama 2.

2) La utilización de factores productivos por cada rama es la siguiente: el factor trabajo utilizado ha sido remunerado con 70M€ por la rama 1, 60M€ por la rama 2 y 50M€ por la rama 3; y el factor capital utilizado, ha sido remunerado con 12M€ por la rama 1, 36M€ por la rama 2 y 29M€ por la rama 3.

3) Las importaciones, clasificadas de acuerdo con los tres grupos de productos que se elaboran en esa economía, han sido: 25M€ del producto 1; 20M€ del producto 2; y 15M€ del producto 3. Las importaciones se destinan exclusivamente a demanda final.

Se pide:

a) Construir la tabla input-output simétrica.

b) Calcular la demanda final para cada producto.

c) Determinar el Valor Añadido Bruto de cada rama.

d) Calcular los coeficientes técnicos de la rama 1. Explique su significado económico.

Solución.

a) En la tabla input-output simétrica que se presenta a continuación, las celdas de color amarillo son los datos proporcionados por el enunciado para completar la tabla. El resto, se deducen de estos como sigue:

- El total de los Consumos Intermedios a precios básicos, (1), se obtiene sumando los Consumos Intermedios de cada una de las ramas verticalmente.
- Dado que en esta economía hay ausencia de impuestos y subvenciones, el total de los consumos intermedios a precios básicos coincide con los Consumos Intermedios a precios de adquisición (2).
- El Valor Añadido Bruto a precios básicos, (3), se obtiene como resultado de la suma de la Remuneración de Asalariados y el Excedente Bruto de Explotación/Renta Mixta.
- Sumando el total de los Consumos Intermedios a precios de adquisición y el Valor Añadido Bruto, se obtiene la Producción a precios básicos, (4).
- La Oferta a precios básicos, (5), se obtiene sumando la Producción a precios básicos y las importaciones.

- La demanda intermedia total, (6), se obtiene sumando la demanda de las ramas horizontalmente.
- El total de empleos, (7), coincide con la Oferta a precios básicos tanto para cada una de las ramas como para el total.
- La demanda final, (8), se obtiene restando al total de empleos la demanda intermedia total. La demanda final es la suma del Gasto en Consumo Final, la Formación Bruta de Capital y las Exportaciones.

	Rama 1	Rama 2	Rama 3	Total demanda intermed. (6)	Demanda final (8)	Total empleos (7)
Rama 1	28	18	23	69	108	177
Rama 2	22	15	15	52	125	177
Rama 3	20	18	19	67	84	151
Total (pb) (1)	70	61	57	188	317	505
Imps. netos sobre prods.	-	-	-	-	-	-
Total CI (p.adq) (2)	70	61	57	188		
VAB (pb) (3)	82	96	79	257		
RA	70	60	50	180		
EBE/RM	12	36	29	77		
Otros imps. netos sobre la producción	-	-	-	-		
Producción (pb) (4)	152	157	136	445		
Importaciones	25	20	15	60		
Oferta (pb) (5)	177	177	151	505		

b) La demanda final para cada producto, (8), se recoge en la penúltima columna de la tabla input-output. Estos valores se obtienen por diferencia entre el total de empleos, (7), (que coincide con la oferta a precios básicos, (5)) y el total de la demanda intermedia, (6), (que se obtiene como suma de las columnas que hacen referencia al consumo intermedio en cada una de las ramas).

c) El Valor Añadido Bruto, (3), de cada rama se obtiene de sumar la utilización de factores productivos por cada rama.

d) La matriz de coeficientes técnicos será:

$$A = \begin{pmatrix} \frac{28}{152} & \frac{18}{157} & \frac{23}{136} \\ \frac{22}{152} & \frac{15}{157} & \frac{15}{136} \\ \frac{20}{152} & \frac{28}{157} & \frac{19}{136} \end{pmatrix} = \begin{pmatrix} 0,184 & 0,115 & 0,169 \\ 0,145 & 0,095 & 0,110 \\ 0,132 & 0,178 & 0,140 \end{pmatrix}$$

donde los valores 152, 157 y 136 de los denominadores son la producción a precios básicos, (4), en las ramas 1, 2 y 3 respectivamente (como se recoge en la tabla input-output). El elemento $a_{11} = 0,184$ representa que existe un consumo intermedio de 0,184 euros en la rama 1 por cada euro que se produce en dicha rama. El elemento $a_{12} = 0,115$ representa que para producir cada euro de producto en la rama 2 se necesitan 0,115 euros de consumos intermedios en productos de la rama 1. Para el resto de los coeficientes la interpretación sería análoga.

Cuestión 8

Se conocen las siguientes operaciones de la balanza de pagos de España, todas ellas realizadas en el mismo periodo (un año t):

1) Se importan textiles de China por 500M€. Se pagan 300M€ al contado y por los 200M€ restantes se recibe un crédito del vendedor.

2) Inversores no residentes adquieren en la bolsa acciones emitidas por empresas residentes por 370M€ . Pagan 150M€ al contado y el resto con un préstamo de un banco nacional a devolver durante el año siguiente.

3) Empresas residentes pagan dividendos a sus accionistas residentes en el extranjero por un importe de 275M€ .

4) Hogares residentes reciben remesas procedentes del extranjero por un importe de 45M€ .

Se pide lo siguiente:

a) Realice las correspondientes anotaciones contables en la Balanza de Pagos, indicando balanza y sub-balanza donde se registran.

b) Indique, como resultado del conjunto de todas las operaciones, cuál es la capacidad o necesidad de financiación para España y si se ha producido una pérdida o una ganancia de reservas.

Solución.

a) 1) La importación de textiles supone una anotación en la cuenta de bienes y servicios de 500M€. Como contrapartida se tiene: una disminución del volumen de reservas de 300M€ por el pago al contado y un aumento de las obligaciones en forma de otras inversiones de 200M€.

2) Se producirá una anotación en la cuenta financiera como inversión en cartera en forma de un aunento de los pasivos frente a los no residentes por valor de 370M€, y como contrapartida se producirá un aumento de las reservas por 150M€ y un aumento de los derechos en forma de otras inversiones por los 220M€ del importe del préstamo.

3) Los dividendos repartidos representan rentas de la propiedad y de la empresa. Por tanto, se realiza una anotación de 275M€ en la cuenta del ingreso primario, siendo la contrapartida la anotación negativa en reservas por el mismo importe.

4) El país recibe efectivo que supone un aumento del volumen de reservas de 45M€ que al no tener contrapartida, implica una anotación por el mismo importe en ingresos en la cuenta de ingreso secundario.

CUENTA CORRIENTE	ingresos	pagos	SALDO = ingresos - pagos
Cuenta de Bienes y Servicios			
textiles importados		500(1)	
↪SALDO Cuenta de Bienes y Servicios			-500
Cuenta del Ingreso Primario			
dividendos		275(3)	
↪SALDO Cuenta del Ingreso Primario			-275
Cuenta del Ingreso Secundario			
remesas recibidas	45(4)		
↪SALDO Cuenta del Ingreso Secundario			45
Saldo Cuenta Corriente			**-730**
CUENTA DE CAPITAL	ingresos	pagos	
	-	-	
Saldo Cuenta de Capital			**0**
SALDO BALANZA CORRIENTE Y DE CAPITAL			**-730**

CUENTA FINANCIERA	variación de activos	variación de pasivos	SALDO = VA - VP
inversión directa			
inversión en cartera		370(2)	
otras inversiones	220(2)	200(1)	
reservas	-300(1);150(2) -275(3); 45(4)		
SALDO CUENTA FINANCIERA			**-730**

b) La capacidad o necesidad de financiación del país coincide con el saldo de la balanza corriente y de capital, que es -730M€. Al ser dicho saldo negativo, el país presenta necesidad de financiación. De la misma forma, se ha producido una pérdida de reservas por valor de -380M€ ya que es la suma de dichas anotaciones:

$$-300 + 150 - 275 + 45 = -380.$$

Cuestión 9

Se muestra en las siguientes tablas la población residente en España, junto con los nacimientos y defunciones producidas en las fechas que se especifican en las tablas:

Año	Población total a 1 de enero
2010	46.500.000
2015	46.000.000

Período	Nacimientos	Defunciones
1 de enero de 2010 a 1 de enero de 2015	2.200.000	2.000.000

A continuación, se proporciona la población y los migrantes por grupos de edad.

Grupo de edad	Población a 1 de enero de 2010	Población a 1 de enero de 2015	Inmigrantes procedentes del extranjero entre el 1 de enero de 2010 y de 2015	Emigrantes con destino al extranjero entre el 1 de enero de 2010 y de 2015
0-15	7.500.000	7.000.000	260.000	300.000
16-64	31.000.000	30.000.000	1.200.000	1.050.000
65-80	8.000.000	9.000.000	100.000	75.000

Calcule:

a) Saldo migratorio y tasa de migración neta para el quinquenio.

b) Índice de atracción para el grupo de edad 16-64.

c) Tasa de migración neta para el grupo de edad 0-15.

d) Edad media a la emigración.

Solución.

a) El saldo migratorio exterior del quinquenio vendría dado por

$$sm = I - E$$

donde I es el total de inmigrantes en España en el periodo en estudio y E es el total de emigrantes en España en dicho periodo. Sin embargo, dado que piden el saldo migratorio sin especificar, este será el resultado de la diferencia entre inmigración y emigración obtenido de la ecuación fundamental o compensadora:

$$P_{t+n} - P_t = N - D + I - E$$

siendo P_{t+n} y P_t la población al final y al principio del periodo en estudio respectivamente y N y D los nacimientos y defunciones respectivamente. Así,

$$SM = I - E = P_{2015} - P_{2010} - N + D =$$

$$= 46000000 - 46500000 - 2200000 + 2000000 =$$
$$= -700000.$$

Este valor debería coincidir con la diferencia entre la suma de inmigrantes y la suma de emigrantes de la tabla anterior del enunciado, y no es así, lo cual hace pensar que puede haber alguna incongruencia en los datos del problema.

Por otra parte, la tasa de migración neta será

$$tmn = i - e$$

donde i y e son las tasas de inmigración y de emigración en España en el quinquenio respectivamente y por tanto:

$$\begin{aligned} tmn &= \left(\frac{I}{n\frac{P^{2010}+P^{2015}}{2}} - \frac{E}{n\frac{P^{2010}+P^{2015}}{2}} \right) \cdot 1000 = \\ &= \left(\frac{-700000}{5 \cdot \frac{46500000+46000000}{2}} \right) \cdot 1000 = \\ &= -3,027‰ \end{aligned}$$

b) Al analizar la inmigración hay que considerar que la población susceptible de migrar es, en teoría, cualquiera que no estuviera viviendo en el área de destino al comienzo del periodo. Calcular dicho denominador presenta dificultades prácticas evidentes, por lo que generalmente se utiliza el mismo denominador que para las tasas de emigración, es decir, la población del lugar de llegada. El indicador así definido se aparta del concepto de tasa, pero sirve para medir la importancia de la inmigración con respecto al crecimiento de la población de acogida. Por tanto, será más correcto denominarlo índice de entrada o índice de atracción. Para el grupo de edad 16-64 años, el índice de atracción se define como:

$$i_{16-64} = \frac{I_{16-64}}{n\frac{P_{16-64}^{2010}+P_{16-64}^{2015}}{2}} \cdot 1000 = \frac{1200000}{5 \cdot \frac{31000000+30000000}{2}} \cdot 1000 = 7,868‰$$

c) La tasa de migración neta para el grupo de edad 0-15 años se calcula de forma análoga a la total:

$$tmn_{0,15} = i_{0,15} - e_{0,15}$$

donde $i_{0,15}$ y $e_{0,15}$ son las tasas de inmigración y de emigración en el grupo de edad 0-15 años en España en el quinquenio estudiado respectivamente y por tanto:

$$i_{0,15} = \frac{I_{0,15}}{n\frac{P_{0,15}^{2010}+P_{0,15}^{2015}}{2}} \cdot 1000 = \frac{260000}{5 \cdot \frac{7500000+7000000}{2}} \cdot 1000 = 7,172‰$$

$$e_{0,15} = \frac{E_{0,15}}{n\frac{P_{0,15}^{2010}+P_{0,15}^{2015}}{2}} \cdot 1000 = \frac{300000}{5 \cdot \frac{7500000+7000000}{2}} \cdot 1000 = 8,275‰$$

$$tmn_{0,15} = i_{0,15} - e_{0,15} = 7,172‰ - 8,275‰ = -1,103‰$$

d) La edad media a la emigración se define como la edad media a la que un individuo saldría de un determinado ámbito con destino al extranjero, en caso de mantener la misma intensidad a la emigración por edad que la observada en el año t en ese ámbito. Se calcula como la media de las edades a las que los individuos emigran con destino al extranjero ponderada por las tasas de emigración por edad expresadas en tanto por uno, es decir, se calcula como:

$$\overline{x}_m = \frac{\sum_x \left(x + \frac{n'}{2}\right) \cdot n' \cdot e_{x,x+n'}}{\sum_x \cdot n' \cdot e_{x,x+n'}}$$

donde x hace referencia al inicio del intervalo de edad, n' a la amplitud del mismo y $e_{x,x+n'}$ es la tasa de emigración para el grupo de edades comprendidas entre x y $x + n'$ años en el ámbito de estudio y en el año t, expresada en tanto por uno.

*(**Aclaración**: se utiliza n' en esta fórmula para diferenciarla de n, utilizada en el cálculo de las tasas de emigración e inmigración pues hacen referencia a distintas amplitudes. Mientras que n se refiere a la amplitud del periodo en estudio, es decir $n = 5$ pues se estudia el periodo que va del 1 de enero de 2010 al 1 de enero de 2015, la segunda de ellas, n', hace referencia a cada una de las amplitudes de los intervalos de edad.)*

La tasa de emigración para cada grupo de edad será

$$e_{x,x+n'} = \frac{E_{x,x+n'}}{n\frac{\left(P^{2010}_{x,x+n'}+P^{2015}_{x,x+n'}\right)}{2}} \cdot 1000$$

para cada uno de los grupos de edad:

$$e_{0,15} = \frac{300000}{5\frac{(7500000+7000000)}{2}} \cdot 1000 = 0,008275 \cdot 1000 = 8,275‰$$
$$e_{16,64} = \frac{1050000}{5\frac{(31000000+30000000)}{2}} \cdot 1000 = 0,006885 \cdot 1000 = 6,885‰$$
$$e_{65,80} = \frac{75000}{5\frac{(8000000+9000000)}{2}} \cdot 1000 = 0,001764 \cdot 1000 = 1,764‰$$

Así,

$$\begin{aligned}\overline{x}_m &= \frac{\left(0+\frac{16}{2}\right)\cdot 16 \cdot 0,008275 + \left(16+\frac{49}{2}\right)\cdot 49 \cdot 0,006885 + \left(65+\frac{16}{2}\right)\cdot 16 \cdot 0,001764}{16 \cdot 0,008275 + 49 \cdot 0,006885 + 16 \cdot 0,001764} \\ &= \frac{16,7828345}{0,497989} = 33{,}70121\end{aligned}$$

por lo que la edad media a la emigración es aproximadamente 34 años.

Cuestión 10

Complete todos los datos del siguiente extracto de tabla de mortalidad abreviada. (Es suficiente redondear a dos decimales, excepto en las tasas específicas de mortalidad).

Edad	Tasas específicas de mortalidad	Promedio de años vividos el último año de vida	Riesgo o probabilidad de muerte	Supervivientes		Población estacionaria	Tiempo por vivir	Esperanza de vida
0		0,123939	0,002645464	100.000				
1		0,481492	0,000223966					
5		0,500314	0,000155676				8.209.014,8	

Solución. En la primera columna se recogerá la tasa específica de mortalidad a la edad x, ${}_nm_x$, que representa el número de individuos de la cohorte ficticia que fallecen con edad cumplida x por tiempo de exposición al riesgo de muerte de los individuos de dicha generación. Es decir, se trata del cociente entre el número de defunciones de individuos con edad cumplida x y el tiempo total (medido en años) vivido por los individuos de la cohorte con dicha edad: ${}_nm_x = \frac{{}_nd_x}{{}_nL_x}$. A partir de la relación entre población estacionaria y función de supervivencia y la propia definición de tasa específica de mortalidad en cada edad, se deriva una aproximación clásica entre tasa de mortalidad y riesgo de muerte a cada edad x: ${}_nm_x = \frac{{}_nq_x}{n\cdot(1-(1-a_x)_nq_x)}$, donde n es la amplitud del intervalo de edad.

La segunda columna recoge el promedio de años vivios, a_x, que representa el tiempo promedio vivido con edad cumplida x por aquellos individuos que mueren con dicha edad.

La tercera columna recoge el riesgo o probabilidad de muerte con edad cumplida x, ${}_nq_x$, que es la probabilidad de un individuo perteneciente a la cohorte ficticia inicial que sobrevive hasta cumplir x años, de morir con dicha edad, es decir, el cociente entre el número de ocurrencias del fenómeno (defunciones teóricas a la edad x) y el total de casos posibles (población sometida al riesgo del fenómeno, es decir, los supervivientes a la edad x); ${}_nq_x = \frac{{}_nd_x}{l_x}$.

Los supervivientes a la edad x en la cuarta columna, l_x, son el número de individuos de la cohorte inicial que llegan con vida a la edad x.

En la quinta columna, suponemos que se recogen las defunciones teóricas con edad x, ${}_nd_x$, es decir el número de defunciones de la cohorte inicial que tienen lugar en individuos de edad cumplida x: ${}_nd_x = l_x - l_{x+n}$.

La siguiente columna recoge la población estacionaria a la edad x, ${}_nL_x$ que es el tiempo total vivido (medido en años) por los individuos de la generación ficticia con edad cumplida x: ${}_nL_x = n\cdot(l_{x+n} + a_x \cdot {}_nd_x)$ donde n es la amplitud del intervalo de edad.

El tiempo por vivir, T_x, es el tiempo que le queda por vivir a la generación desde los x años de edad hasta su extinción: $T_x = T_{x+n} + {}_nL_x$, siendo $T_w = L_w$.

Por último, la esperanza de vida a la edad x, e_x es el número medio de años que a un individuo de edad x perteneciente a la cohorte ficticia inicial le restaría por vivir, es decir, el cociente entre el tiempo total (medido en años) que le resta por vivir a partir de cumplir x años de edad a los individuos de la generación ficticia hasta su completa extinción y el número de supervivientes de la misma a la edad x de forma que se calcula con el cociente $e_x = \frac{T_x}{l_x}$.

Utilizando la información anterior y redondeando a dos decimales excepto en las tasas específicas de mortalidad, como se indica en el enunciado, se completa la tabla:

$$ {}_1d_0 = {}_1q_0 \cdot l_0 = 0,002645464 \cdot 100000 = 264,55 $$

$$ l_1 = l_0 - {}_1d_0 = 100000 - 264,55 = 99735,45 $$

$$ {}_4d_1 = {}_4q_1 \cdot l_1 = 0,000223966 \cdot 99735,45 = 22,34 $$

$$ l_5 = l_1 - {}_4d_1 = 99735,45 - 22,34 = 99713,12 $$

$$ {}_5d_5 = {}_5q_5 \cdot l_5 = 0,000155676 \cdot 99713,12 = 15,52 $$

$$ {}_1m_0 = \frac{{}_1q_0}{1\cdot(1-(1-a_0){}_1q_0)} = \frac{0,002645464}{1\cdot(1-(1-0,123939)0,002645464)} = 0,0026516 $$

$$ {}_4m_1 = \frac{{}_4q_1}{4\cdot(1-(1-a_1){}_4q_1)} = \frac{0,000223966}{4\cdot(1-(1-0,481492)0,000223966)} = 0,00005599 $$

$$_5m_5 = \frac{_5q_5}{5 \cdot (1-(1-a_5)_5q_5)} = \frac{0,000155676}{5 \cdot (1-(1-0,500314)0,000155676)} = 0,00003113$$

$$_1L_0 = 1 \cdot (l_1 + a_0 \cdot d_0) = 1 \cdot (99735,45 + 0,123939 \cdot 264,55) = 99768,24$$

$$_4L_1 = 4 \cdot (l_5 + a_1 \cdot d_1) = 4 \cdot (99713,12 + 0,481492 \cdot 22,34) = 398895,49$$

$$_5L_5 = n \cdot (l_{x+1} + a_5 \cdot d_5) = 5 \cdot ((99713,12 - 15,52) + 0,500314 \cdot 15,52) = 498526,80$$

donde se ha supuesto que el próximo intervalo será 10 y así la amplitud es 5

$$T_1 = T_5 + {}_4L_1 = 8209014,80 + 398895,49 = 8607910,29$$

$$T_0 = T_1 + {}_1L_0 = 8607910,29 + 99768,24 = 8707678,53$$

$$e_0 = \frac{T_0}{l_0} = \frac{8707678,53}{100000,00} = 87,08$$

$$e_1 = \frac{T_1}{l_1} = \frac{8607910,29}{99735,45} = 86,31$$

$$e_5 = \frac{T_5}{l_5} = \frac{8209014,80}{99713,12} = 82,33$$

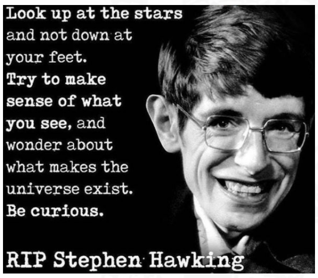
	$_nm_x$	a_x	$_nq_x$	l_x	d_x	$_nL_x$	T_x	e_x
0	0,0026516	0,123939	0,002645464	100000,00	264,55	99768,24	8707678,53	87,08
1	0,00005599	0,481492	0,000223966	99735,45	22,34	398895,49	8607910,29	86,31
5	0,00003113	0,500314	0,000155676	99713,12	15,52	498526,80	8209014,80	82,33
				99697,59				

Capítulo 3

Año 2018

"Do not worry too much about your difficulty in mathematics,
I can assure you that mine are still greater."

—Albert Einstein



Nació en la ciudad de Oxford en Reino Unido, el 8 de enero de 1942 y fue un científico británico que se dedicó al estudio de los agujeros negros y de la relatividad en general. Obtuvo el Doctorado en 1966 en Física Teórica y Cosmología. En 1963 le diagnosticaron esclerosis lateral amiotrófica, o mejor conocida como el Mal de Lou Gehrig, una patología degenerativa y su vida poco a poco empezó a presentar limitaciones, pero su trabajo intelectual no se detuvo, empezó a acrecentarse de manera considerable, convirtiéndose en poco tiempo en uno de los científicos e investigadores del mundo con más popularidad después de Einstein. Esta enfermedad fue progresiva, sus extremidades inferiores empezaron a paralizarse, luego el resto de su cuerpo y desafortunadamente una neumonía también le afectó a tal punto, que tuvieron que practicarle una traqueotomía y de ahí en adelante, perdió la capacidad para hablar. Dependía de un sintetizador y de una silla de ruedas, las 24 horas del día. Murió el 14 de marzo de 2018, a los 76 años de edad.

$2018 = 2 \cdot 1009 = 7^2 + 8^2 + \cdots + 18^2$

2018 is a deficient number. It is larger than the sum of its divisors (1012).

3.1. Soluciones del segundo examen del Cuerpo Superior de Estadísticos del Estado de 2018

Cuestión 1

De una distribución $N(100, \sigma)$, se toman dos muestras aleatorias simples independientes entre sí de tamaño 4 y 5.
En la muestra 1: $\bar{x} = 100,4; \sum_{i=1}^{4}(x_i - \bar{x})^2 = 15,26$.
En la muestra 2: $\bar{y} = 99,6; \sum_{i=1}^{5}(y_i - \bar{y})^2 = 10,72$.
Calcular la $P(\bar{X} - \bar{Y} \geq 2)$, donde $\bar{X}$ e $\bar{Y}$ son las medias muestrales, indicando la distribución muestral necesaria para su cálculo.

Tabla 3.1: Observaciones muestrales

Muestra 1	Muestra 2
98	97,8
103,4	101,3
100,5	97,9
99,7	100,7
	100,3

Solución.

Se trata de dos muestras independientes con varianzas poblacionales desconocidas pero iguales. Dado que $\mu_1 - \mu_2 = 0$, se tiene:

$$\frac{\bar{x} - \bar{y}}{\sqrt{\frac{1}{n} + \frac{1}{m}}\sqrt{\frac{(n-1)S_{c1}^2 + (m-1)S_{c2}^2}{n+m-2}}} \sim t_{n+m-2}$$

Por tanto:

$$P(\bar{X}-\bar{Y} \geq 2) = P\left(\frac{\bar{x}-\bar{y}}{\sqrt{\frac{1}{n}+\frac{1}{m}}\sqrt{\frac{(n-1)S_{c1}^2+(m-1)S_{c2}^2}{n+m-2}}} \geq \frac{2}{\sqrt{\frac{1}{n}+\frac{1}{m}}\sqrt{\frac{(n-1)S_{c1}^2+(m-1)S_{c2}^2}{n+m-2}}}\right)$$

Es decir,

$$P\left(t_{n+m-2} \geq \frac{2}{\sqrt{\frac{1}{n}+\frac{1}{m}}\sqrt{\frac{(n-1)S_{c1}^2+(m-1)S_{c2}^2}{n+m-2}}}\right)$$

Dado que $(n-1)S_{c1}^2 = 15,26$ y $(m-1)S_{c2}^2 = 10,72$,

$$P(\bar{X}-\bar{Y} \geq 2) = P\left(t_7 \geq \frac{2}{\sqrt{\frac{9}{20}}\sqrt{\frac{25,98}{7}}}\right) = P\left(t_7 \geq 1,5475\right) = 0,08618$$

valor que se obtiene interpolando para los diferentes valores tabulados en la distribución t de Student con 7 grados de libertad, que es la distribución muestral. En concreto utilizando $P(t_7 \geq 1,41492) = 0,1$ y $P(t_7 \geq 1,89458) = 0,05$ obtenemos la solución anterior.

Cuestión 2

Contrastar con un nivel de significación $2,5\,\%$ que la duración de un determinado tipo de bombillas eléctricas es una variable aleatoria con función de densidad:

$$f(x) = \frac{1}{\theta} exp\left(-\frac{1}{\theta}x\right) \qquad\qquad x > 0$$

con $\theta = 200$ horas, teniendo en cuenta que en una muestra aleatoria de 75 bombillas probadas hasta fundirse, se han observado las duraciones recogidas en la tabla 3.2.

Tabla 3.2: Observaciones muestrales.

Duración	Número de bombillas
Hasta 200 horas	40
De 200 a 300 horas	15
De 300 a 400 horas	8
De 400 a 500 horas	6
Más de 500 horas	6
	N=75

Solución.

Siendo $X \equiv$ "duración de la bombilla hasta fundirse" y teniendo en cuenta que $F(x) = \int_0^x f(x)dx$,

$$F(x) = \begin{cases} \int_0^x \frac{1}{\theta} e^{-\frac{x}{\theta}} = \left(-e^{\frac{x}{\theta}}\right)_0^x = -e^{\frac{x}{\theta}} - (-e^0) = 1 - e^{-\frac{x}{\theta}} & \text{si} \quad x > 0 \\ \\ 0 & \text{si} \quad x < 0 \end{cases}$$

los valores teóricos proporcionados por la distribución serían:

$$P(X \leq 200) = F(200) = 1 - exp\left(-\tfrac{1}{200}200\right) = 0,632$$
$$P(200 < X \leq 300) = F(300) - F(200) = 0,1447$$
$$P(300 < X \leq 400) = F(400) - F(300) = 0,0877$$
$$P(400 < X \leq 500) = F(500) - F(400) = 0,0532$$
$$P(X \geq 500) = 1 - F(500) = 0,08208$$

Las frecuencias se obtienen multiplicando estas probabilidades por 75, y como la penúltima frecuencia esperada es menor que 5, las dos últimas frecuencias se agrupan.

Frecuencias observadas	40	15	8	6	6
Frecuencias esperadas	47,4	10,853	6,578	3,99	6,156

Frecuencias observadas	40	15	8	12
Frecuencias esperadas	47,4	10,853	6,578	10,146

El estadístico de contraste D sería:

$$D = \sum_{i=1}^{k} \frac{(n_i - np_i^0)^2}{np_i^0} = \frac{(40-47,4)^2}{47,4} + \cdots + \frac{(12-10,146)^2}{10,146} = 3,379$$

El número de grados de libertad del contraste es k-1=3. Así, $D \sim \chi^2_{k-1;\alpha} \equiv \chi^2_{3;0,025} = 9,35$. Al ser, $D = 3,379 < \chi^2_{3;0,025}$, no se rechaza la hipótesis nula ($H_0 : F = F_0$, con F_0 la función de distribución de una exponencial con parámetro $\frac{1}{\theta}$) y puede decirse que la duración de las bombillas es una variable aleatoria con función de densidad exponencial de parámetro $\frac{1}{\theta} = 200$.

Cuestión 3

De una población de N = 500 hogares, se obtiene una muestra aleatoria simple (sin reposición) de tamaño n = 50. En cada hogar de la muestra se mide el gasto en alimentación (Y) y el gasto total (X).

Tabla 3.3: Datos en miles de euros

$\sum_{i=1}^{50} Y_i = 213$	$\sum_{i=1}^{50} X_i = 452$
$\sum_{i=1}^{50} Y_i^2 = 964$	$\sum_{i=1}^{50} X_i^2 = 4552$
$\sum_{i=1}^{50} Y_i X_i = 2060$	

Se pide:

A) Estimar el porcentaje de gasto en alimentación y su error de muestreo.

B) Estimar el gasto total en alimentación usando el método de la razón y su error de muestreo. Para ello se conoce que el gasto total de la población es de 5000.

Solución.

A) El estimador del porcentaje de gasto en alimentación será

$$\widehat{R} = \frac{\widehat{Y}}{\widehat{X}} = \frac{\overline{y}}{\overline{x}} = \frac{\sum_{i=1}^{50} Y_i}{\sum_{i=1}^{50} X_i} = \frac{213}{452} = 0,4712 \simeq 47\,\%$$

y su error de muestreo estimado:

$$\begin{aligned}
\widehat{Var}(\widehat{R}) &= \frac{1}{\overline{X}^2}\frac{(1-f)}{n}\left(\widehat{S}_Y^2 + \widehat{R}^2\widehat{S}_X^2 - 2\widehat{R}\widehat{S}_{XY}\right) = \\
&= \frac{1}{\overline{X}^2}\frac{\left(1-\frac{n}{N}\right)}{n(n-1)}\left(\sum_{i=1}^{50} Y_i^2 + \widehat{R}^2\sum_{i=1}^{50} X_i^2 - 2\widehat{R}\sum_{i=1}^{50} X_iY_i\right) = \\
&= \frac{1}{\overline{X}^2}\frac{\left(1-\frac{50}{500}\right)}{50\cdot 49}\left(964 + \left(\frac{213}{452}\right)^2 \cdot 4552 - 2\left(\frac{213}{452}\right)2060\right)
\end{aligned}$$

Para calcular el error de muestreo necesitamos conocer el valor del gasto medio de la población, es decir, de $\overline{X}$, que según el apartado segundo valdría $\overline{X} = \frac{5000}{500} = 10$. Si utilizaramos este dato se obtendría que:

$$\begin{aligned}
\widehat{Var}(\widehat{R}) &= 0,000003673 \cdot 33,3406 = 0,00012246 \\
\widehat{\sigma}(\widehat{R}) &= \sqrt{0,00012246} = 0,01106
\end{aligned}$$

Sin ese dato, que puede ser discutible si en este apartado puede utilizarse, no se podría calcular el error de muestreo, aunque una aproximación al mismo se calcularía utilizando el gasto medio muestral, es decir, $\overline{x} = \frac{452}{50} = 9,04$

$$\begin{aligned}
\widehat{Var}(\widehat{R}) &\simeq 0,000004495 \cdot 33,3406 = 0,00014986 \\
\widehat{\sigma}(\widehat{R}) &\simeq \sqrt{0,00014986} = 0,01224
\end{aligned}$$

B) En cuanto a la estimación del gasto total en alimentación, usando el método de la razón y teniendo en cuenta que el gasto total de la población es de 5000, se tiene:

$$\widehat{Y}_R = \widehat{R} \cdot X = 0,4712 \cdot 5000 = 2356,19$$

Consecuentemente,

$$\widehat{Var}(\widehat{Y}_R) = X^2 \cdot \widehat{Var}(\widehat{R}) = 5000^2 \cdot 0,00012246 = 3061,5$$

de modo que su error de muestreo será:

$$\widehat{\sigma}(\widehat{Y}_R) = \sqrt{3061,5} = 55,33$$

Cuestión 4

Se quiere estimar la proporción de pinos que hay en una zona forestal. Para ello se divide la zona en 20 conglomerados de tamaño diferentes M_i, $i = 1, \ldots, 20$ conociendo que el total de árboles es $M = \sum_{i=1}^{20} M_i = 1000$. Se utiliza un muestreo de conglomerados con submuestreo donde en ambas etapas el procedimiento de selección es con probabilidades iguales sin reposición. En la primera etapa se seleccionan 4 conglomerados y en la segunda se aplica una fracción de muestreo $f_{2i} = \frac{10}{M_i}$. Los valores del tamaño de los conglomerados muestrales y el número de pinos obtenido en cada uno de ellos vienen en la tabla 3.4.

Tabla 3.4: Tamaño muestral y número de pinos en cada conglomerado

Número de árboles = M_i	Número de pinos
60	8
80	7
50	6
30	4

Se pide:

A) Una estimación insesgada de la proporción de pinos y su error de muestreo.

B) Una estimación de la proporción de pinos utilizando el estimador de la razón al tamaño y su error de muestreo.

C) Comentar las ventajas e inconvenientes del estimador B) respecto al A).

Solución.

a) Del enunciado se tiene:

$$N = 20 \qquad M = 1000 \qquad n = 4 \qquad f_{2i} = \frac{10}{M_i} \Rightarrow m_i = 10 \qquad \forall i = 1, \cdots, 4$$

Una estimación insesgada de la proporción de pinos vendría dada por:

$$\widehat{P} = \frac{\widehat{A}}{M} = \frac{1}{M}\sum_{i=1}^{n}\frac{M_i\widehat{P}_i}{\pi_i} = \frac{1}{M}\frac{N}{n}\sum_{i=1}^{n}M_i\widehat{P}_i$$

Puesto que $\widehat{P}_i = \frac{n_i}{m_i}$, donde n_i es el número de pinos en cada conglomerado, para cada uno de los conglomerados se tiene

$$\widehat{P}_1 = \frac{8}{10} \qquad \widehat{P}_2 = \frac{7}{10} \qquad \widehat{P}_3 = \frac{6}{10} \qquad \widehat{P}_4 = \frac{4}{10}$$

y así la estimación insesgada de la proporción de pinos es:

$$\widehat{P} = \frac{1}{M}\frac{N}{n}\sum_{i=1}^{4}M_i\widehat{P}_i = \frac{1}{1000}\frac{20}{4}\left(60\frac{8}{10} + 80\frac{7}{10} + 50\frac{6}{10} + 30\frac{4}{10}\right) = 0,73.$$

La estimación de la varianza de la proporción ha de hacerse teniendo en cuenta que se trata de muestreo bietápico sin reposición en las dos etapas con probabilidades iguales para conglomerados de distinto tamaño. Así:

$$\widehat{Var}(\widehat{P}) =$$

$$= \frac{1}{M^2}\left[\frac{N^2(1-f_1)}{n}\frac{\sum_i^n\left(M_i\widehat{P}_i - \frac{1}{n}\sum_i^n M_i\widehat{P}_i\right)^2}{n-1} + \frac{N}{n}\sum_i^n M_i^2\left(1-f_{2i}\right)\frac{\widehat{P}_i\widehat{Q}_i}{m_i-1}\right]$$

$$= \frac{1}{1000^2}\left[\frac{20^2(1-\frac{4}{20})}{4}\frac{1155}{3} + \frac{20}{4}\frac{760}{3}\right] = 0,032$$

Y el error de muestreo será

$$\widehat{\sigma}(\widehat{P}) = \widehat{\sigma}(\widehat{P}) = \sqrt{\widehat{Var}(\widehat{P})} = \sqrt{0,032} = 0,179$$

b) Para el cálculo de la estimación de la proporción de pinos utilizando el método de la razón se utiliza

$$\widehat{P}_R = \frac{\sum_{i=1}^4 A_i}{\sum_{i=1}^4 M_i} = \frac{\sum_{i=1}^4 M_i\widehat{P}_i}{\sum_{i=1}^4 M_i} = \frac{60\frac{8}{10} + 80\frac{7}{10} + 50\frac{6}{10} + 30\frac{4}{10}}{60+80+50+30} = \frac{146}{220} = 0,663$$

donde se añade el subíndice R para hacer referencia al cálculo de la estimación utilizando el método de la razón. En cuanto a su error de muestreo, se puede aproximar mediante la expresión:

$$\begin{aligned}
\widehat{Var}(\widehat{P}_R) \approx & \frac{1}{(\bar{M})^2}\left[\frac{1}{n}\left(1-f_1\right)\frac{1}{n-1}\left(\sum_{i=1}^{4}\left(\frac{M_i\cdot n_i}{m_i}-\widehat{P}_R\cdot M_i\right)^2\right)+\right.\\
& \left.+\frac{1}{nN}\sum_{i=1}^{4}\frac{M_i\left(M_i-m_i\right)\widehat{P}_i\left(1-\widehat{P}_i\right)}{m_i-1}\right]=\\
& =\frac{1}{\frac{M^2}{N^2}}\left[\frac{1}{n}\left(1-f_1\right)\frac{1}{n-1}\left(\sum_{i=1}^{4}\left(\frac{M_i\cdot n_i}{m_i}-\widehat{P}_R\cdot M_i\right)^2\right)+\right.\\
& \left.+\frac{1}{nN}\frac{1}{m-1}\sum_{i=1}^{4}M_i\left(M_i-m_i\right)\widehat{P}_i\left(1-\widehat{P}_i\right)\right]
\end{aligned}$$

donde,

$$\begin{aligned}
\sum_{i=1}^{4}\left(\frac{M_i\cdot n_i}{m_i}-\widehat{P}_R\cdot M_i\right)^2 = & \left(\frac{60\cdot 8}{10}-\frac{146}{220}\cdot 60\right)^2+\left(\frac{80\cdot 7}{10}-\frac{146}{220}\cdot 80\right)^2+\\
& +\left(\frac{50\cdot 6}{10}-\frac{146}{220}\cdot 50\right)^2+\left(\frac{30\cdot 4}{10}-\frac{146}{220}\cdot 30\right)^2=\\
& =\left(48-\frac{438}{11}\right)^2+\left(56-\frac{584}{11}\right)^2+\left(30-\frac{365}{11}\right)^2+\\
& +\left(12-\frac{219}{11}\right)^2=\\
& =\frac{17918}{121}
\end{aligned}$$

$$\begin{aligned}
\sum_{i=1}^{4}M_i\left(M_i-m_i\right)\widehat{P}_i\left(1-\widehat{P}_i\right) & = 60\cdot 50\cdot\frac{8}{10}\frac{2}{10}+80\cdot 70\cdot\frac{7}{10}\cdot\frac{3}{10}+\\
& +50\cdot 40\cdot\frac{6}{10}\cdot\frac{4}{10}+30\cdot 20\cdot\frac{4}{10}\cdot\frac{6}{10}=\\
& =480+1176+480+144=2280.
\end{aligned}$$

De esta forma,

$$\widehat{Var}(\widehat{P}_R) = \frac{20^2}{1000^2}\left[\frac{1}{4}\cdot\left(1-\frac{4}{20}\right)\cdot\frac{1}{3}\cdot\left(\frac{17918}{121}\right)+\frac{1}{4\cdot 20}\cdot\frac{1}{9}\cdot 2280\right] =$$
$$= 0,005216.$$

Y el error de muestreo será

$$\widehat{\sigma}(\widehat{P}_R) = \widehat{em}(\widehat{P}_R) = \sqrt{\widehat{var}(\widehat{P}_R)} = \sqrt{0,005216} = 0,0722.$$

c) El cálculo de este estimador $\widehat{P}_R$ no requiere el conocimiento del número de conglomerados total ni el número total de elementos de la población, sino solo la información obtenida de la muestra, pero suele ser sesgado, aunque dicho sesgo puede ser reducido incrementando el tamaño muestral. El estimador $\widehat{P}$ del primer apartado es insesgado, aunque suele ser menos preciso que el de la razón al tamaño.

El denominador del estimador $\widehat{P}_R$ es $\sum_{i=1}^{4} M_i$, mientras que en el de las dos etapas $\widehat{P}$ aparece en el denominador $M \cdot \frac{n}{N}$ y de ahí su diferencia. En el segundo caso tiene en cuenta tantos elementos como la proporción de conglomerados muestreados por el total de elementos de la población, y en el primero ajusta el denominador por el total de elementos de los conglomerados muestrados concretamente.

En general, también será más eficiente el estimador de razón, puesto que hace posible corregir los desequilibrios de la muestra de unidades primarias. El estimador insesgado suele ser más ineficiente porque es muy sensible a los valores del tamaño de cada conglomerado.

Cuestión 5

Se dispone de la siguiente información respecto a los agregados de oferta, demanda y rentas de una economía en miles de millones de euros (entre guiones el código de las operaciones y saldos en SEC2010):

Tabla 3.5: Agregados de oferta, demanda y rentas en miles de millones de euros

Producción de bienes y servicios (a precios básicos) - P. 1 -	2.000
Formación bruta de capital - P.5 -	250
Subvenciones sobre los productos - D.31 -	8
Gasto en consumo final - P. 3 -	900
Importaciones de bienes y servicios - P. 7 -	360
Exportaciones de bienes y servicios - P. 6 -	400
Formación bruta de capital fijo - P.51g -	230
Excedente de explotación bruto y Renta mixta bruta - B.2g + B.3g -	500
Impuestos sobre la producción y las importaciones - D.2 -	130
Consumo intermedio (precio de adquisición) - P.2 -	1.000
Subvenciones - D.3 -	20

Se pide:

1. Calcule el valor añadido a precios básicos.

2. Calcule el PIB a precios de mercado (PIBpm).

3. Calcule el valor de la variación de existencias. Suponga para ello que las adquisiciones menos cesiones de objetos valiosos son nulas.

4. Calcule la remuneración de los asalariados.

Solución.

$$VAB_{pb} = Producción_{pb} - CI_{padq} = 2000 - 1000 = 1000$$

$$PIB_{pm} = GCF + FBC + X - M = 900 + 250 + 400 - 360 = 1190$$

$$FBC = FBCF + \Delta existencias \Leftrightarrow \Delta existencias = FBC - FBCF = 250 - 230 = 20$$

$$RA = PIB_{pm} - EBE/RM - \textit{ impuestos netos produccion e importaciones } =$$

$$= 1190 - 500 - (130 - 20) = 580$$

Cuestión 6

El PIBpm de una determinada economía en el año t ha sido de 1100 (miles de millones de euros). Se conoce también que la tasa de variación de este agregado a precios corrientes entre el año $t-1$ y t fue de 4,9 % mientras que la correspondiente tasa de variación anual en volumen fue del 3,4 %. Además de la cuenta del resto del mundo se tiene la siguiente información del año $t-1$ (en miles de millones de euros):

Tabla 3.6: Agregados del resto del mundo en el año $t-1$ (en miles de millones de euros)

Remuneración de asalariados recibida del resto del mundo	20
Remuneración de asalariados pagada al resto del mundo	8
Impuestos sobre la producción y las importaciones pagadas al resto del mundo	10
Subvenciones recibidas del reto del mundo	50
Rentas de la propiedad recibidas del resto del mundo	50
Rentas de la propiedad pagadas al resto del mundo	70

Se pide:

1. Calcule el PIBpm y la Renta Nacional Bruta (RNB) del año t–1.
2. Calcule la variación entre t–1 y t del deflactor implícito del PIBpm.
3. ¿Considera que según los principios recogidos en el SEC 2010 tiene sentido el cálculo de la RNB en términos reales? Razone su respuesta.

Solución.

1.)

$$TVI(PIB)_{t-1}^{t} = \frac{PIB_t}{PIB_{t-1}} - 1 \Leftrightarrow PIB_{t-1} = \frac{1100}{1,049} = 1048,61 \text{ miles de millones de euros}$$

$$\begin{aligned} RNB_{t-1} \quad &= PIB_{t-1} + \text{rentas primarias reales a cobrar del resto del mundo} \\ &\qquad - \text{rentas primarias reales a pagar al resto del mundo} \end{aligned}$$

Por tanto:

$$RNB_{t-1} = 1048,61 + (20 + 50) - (8 + 70 + 10 - 50) = 1080,61$$

Es decir, en el año $t-1$ el PIB fue de 1048,61 miles de millones de euros y la RNB de 1080,61 miles de millones de euros.

2.) Por otra parte, sabemos que:

$$\text{Índice(volumen)}_{t-1}^{t} = (\text{TV(volumen)}_{t-1}^{t} + 1) \cdot 100$$
$$\text{Índice(valor)}_{t-1}^{t} = (\text{TV(valor)}_{t-1}^{t} + 1) \cdot 100$$

y como el enunciado nos proporciona las tasas de variación en volumen y a precios corrientes, puesto que se cumple que :

$$\text{Índice de valor} = \text{índice de precios} \cdot \text{índice de volumen}$$
$$104{,}9 = \text{índice de precios} \cdot 103{,}4$$
$$\text{Índice de precios} = \tfrac{1{,}049}{1{,}034} = 101,45$$

Esto es, la tasa de variación entre t–1 y t del deflactor implícito del PIB es 1,45 %.

3.) Dado que la RNB es un saldo contable que incluye distintas operaciones de distribución, resulta difícil desglosar directamente los valores corrientes en componentes de precio y volumen. Si bien, esta descomposición podría realizarse de forma indirecta utilizando los flujos de las operaciones correspondientes.

En concreto, los flujos de renta pueden medirse en términos reales únicamente si se selecciona una cesta de bienes y servicios a cuya adquisición se destina habitualmente la renta y se utiliza el índice de precios de esta cesta como deflactor de las rentas corrientes. La selección es siempre arbitraria, en el sentido de que la renta rara vez se destina de forma específica a las adquisiciones efectuadas durante el periodo en cuestión. Una parte de la misma puede ahorrarse para realizar adquisiciones en periodos posteriores o, por el contrario, las adquisiciones del periodo pueden estar financiadas parcialmente por ahorros de periodos precedentes.

La renta nacional bruta real se obtiene como:

El producto interior bruto en volumen

+ las ganancias o pérdidas resultantes de las variaciones de la relación real de intercambio (es decir, la renta real total de los residentes se ve influida también por la tasa a la que las exportaciones pueden comerciarse frente a las importaciones del resto del mundo);

O, equivalentemente, la renta nacional bruta real es igual a:

La renta interior bruta real

+ las rentas primarias reales a cobrar del resto del mundo

- las rentas primarias reales a pagar al resto del mundo.

De este modo, para poder expresar los diversos agregados de la renta nacional en términos reales, se debería deflactar los ingresos y pagos de rentas primarias y transferencias del/al resto del mundo con un índice del gasto final interior bruto. Sin embargo, dado que este deflactor no es fácil de encontrar es muy complejo el cálculo de la RNB en términos reales. En 2019 se ha intentado armonizar entre los Estados miembros, mediante el Reglamento 2019/516 del Parlamento Europeo y del Consejo, la armonización de la renta nacional bruta a precios de mercado.

Cuestión 7

¿Cuál es la relación entre crimen y castigo? Esta importante pregunta fue estudiada mediante un panel de datos de Carolina del Norte. Las secciones transversales son 90 condados, y los datos son anuales para los años 1981-1987. En estos modelos la tasa de delincuencia pretende ser explicada a partir de variables como el efecto disuasivo del sistema legal, los salarios en el sector privado (que representan el retorno a las actividades legales). Los autores comentan que puede haber heterogeneidad entre los condados (características no observadas de cada condado).

En este marco analítico, considere un modelo en el cual la tasa de criminalidad (y) es una función de la probabilidad de detención (X_1), probabilidad de ser un convicto (preso) (X_2), la probabilidad de una pena de prisión (X_3), el promedio de las penas de prisión (X_4), y el salario semanal promedio en el sector manufacturero (X_5). Importante: en **todos los casos se utilizan los logaritmos de las variables**.

(1) Indique los signos esperados en un modelo regresión lineal múltiple

(2) Una analista de datos propone estimar el modelo indicado en (1) mediante mínimos cuadrados ordinarios. Los resultados son los siguientes (errores estándar en paréntesis)

$$\widehat{y} = \underset{(0{,}3654)}{-6{,}0861} \underset{(0{,}043)}{-0{,}6566}\, X_1 \underset{(0{,}0277)}{-0{,}4466}\, X_2 \underset{(0{,}0727)}{+0{,}2082}\, X_3 \underset{(0{,}0606)}{-0{,}0586}\, X_4 \underset{(0{,}0619)}{+0{,}2921}\, X_5$$

Analice los signos de los coeficientes estimados y su significación (al 95 % utilizando para ello la distribución normal). ¿Son como esperaba? Interprete el coeficiente de X_1.

(3) Otra económetra sin embargo considera oportuno estimar el modelo (1) usando un estimador de efectos fijos. El estimador de efectos fijos arroja el siguiente modelo estimado

$$\widehat{y} = \underset{(0{,}3236)}{-3{,}2288} \underset{(0{,}0376)}{-0{,}2313}X_1 \underset{(0{,}0222)}{-0{,}1378}X_2 \underset{(0{,}0393)}{-0{,}1431}X_3 \underset{(0{,}0310)}{+0{,}0183}X_4 \underset{(0{,}0553)}{-0{,}1666}X_5$$

¿Los coeficientes coinciden ahora con lo que esperaba? Interpreta el coefiiente en X_1 y compáralo con la estimación en (2). ¿Qué concluye sobre el efecto disuasivo de la probabilidad de arresto? Por último, interpreta el coeficiente en X_4. ¿Qué conclusión tienes sobre la gravedad del castigo como elemento disuasivo?

(4) Tras estimar los dos modelos, argumente cuáles han podido ser los motivos que la económetra ha tenido para proponer un modelo de efectos fijos. Justifique cuál de los dos modelos le parece más adecuado para estimar la relación entre crimen y castigo.

(5) ¿Puede haber endogeneidad en el modelo? Justifique la respuesta en términos del modelo propuesto.

Solución.

Dado que en todos los casos se utilizan los logaritmos de las variables, se trata de un modelo log - log que presenta la siguiente estructura:

$$\begin{aligned}\log\left(tasa\ crim\right) = \beta_0 &+ \beta_1 \log\left(prob\ detención\right) + \beta_2 \log\left(prob\ convicto\right)\\ &+ \beta_3 \log\left(prob\ pena\ prisión\right) + \beta_4 \log\left(promedio\ penas\ prisión\right)\\ &+ \beta_5 \log\left(salario\ promedio\ sect.\ manufact\right)\end{aligned}$$

que, siguiendo la notación del ejercicio, se reescribe como:

$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5.$$

Cada uno de los coeficientes β_i representan la elasticidad de Y respecto de cada una de las variables X_i, es decir, el cambio porcentual de Y ante un pequeño cambio porcentual en X_i, es decir:

$$\beta_i = \frac{\frac{\Delta Y}{Y}}{\frac{\Delta X_i}{X_i}}$$

(1) El intercepto β_0 es la tasa de criminalidad media que se predice para cualquier condado siempre que son nulos la probabilidad de detención, la de

ser un convicto, la de una pena de prisión, el promedio de las penas de prisión y el salario semanal promedio. Sin embargo, es poco probable que todas estas variables tomen el valor cero ya que 0 está fuera del rango de muchas de ellas, por tanto, la interpretación de este coeficiente de forma teórica en la práctica es más difuso. Por otra parte, los signos esperados para cada una de las variables son:

$\beta_1 < 0$: si aumenta porcentualmente la probabilidad de detención, debería implicar que disminuye porcentualmente la tasa de criminalidad

$\beta_2 < 0$: la tasa de criminalidad debería ser menor si hay mayor probabilidad de ser preso (por simplicidad entendemos en este contexto que mayor probabilidad significa un aumento porcentual de dicha probabilidad)

$\beta_3 < 0$: menor tasa de criminalidad a medida que la probabilidad de que la pena de prisión sea mayor

$\beta_4 < 0$: menor tasa de criminalidad si el promedio de penas de prisión es mayor

$\beta_5 < 0$: menor tasa de criminalidad si el salario aumenta

(2) Suponemos normalidad y consideramos $\alpha = 0,05$, $\lambda_{\frac{\alpha}{2}} = 1,96$. De forma que en el contraste de significación individual

$$\begin{cases} H_0 : \beta_j = 0 \\ H_1 : \beta_j \neq 0 \end{cases},$$

la región de aceptación será $(-1,96, 1,96)$. A la vista de los valores para los estadísticos obtenidos a continuación, se rechaza la hipótesis nula para todos los coeficientes a excepción del coeficiente de la variable X_4. Por tanto todos los coeficientes son individualmente significativos a la hora de explicar la variabilidad de la tasa de criminalidad menos el coeficiente del promedio de penas de prisión.

$$\left|\frac{\hat{\beta}_0}{\hat{\sigma}(\hat{\beta}_0)}\right| = \left|\frac{-6,0861}{0,3654}\right| = |-16,65| > |1,96|$$

$$\left|\frac{\hat{\beta}_1}{\hat{\sigma}(\hat{\beta}_1)}\right| = \left|\frac{-0,6566}{0,0403}\right| = |-16,29| > |1,96|$$

$$\left|\frac{\hat{\beta}_2}{\hat{\sigma}(\hat{\beta}_2)}\right| = \left|\frac{-0,4466}{0,0277}\right| = |-16,12| > |1,96|$$

$$\left|\frac{\hat{\beta}_3}{\hat{\sigma}(\hat{\beta}_3)}\right| = \left|\frac{0,2082}{0,0727}\right| = |2,863| > |1,96|$$

$$\left|\frac{\hat{\beta}_4}{\hat{\sigma}(\hat{\beta}_4)}\right| = \left|\frac{-0,0586}{0,0606}\right| = |0,9669| < |1,96|$$

$$\left|\frac{\hat{\beta}_5}{\hat{\sigma}(\hat{\beta}_5)}\right| = \left|\frac{0,2921}{0,0619}\right| = |4,7189| > |1,96|$$

La variable X_3 tiene un signo contrario a la intuición, pues se supone que una mayor probabilidad de pena de prisión debería hacer decaer la tasa de criminalidad. Ocurre lo mismo con la variable X_5 pues cabría esperar que un mayor salario en el sector manufacturero funcionaría como incentivo para aumentar el empleo y disminuir la criminalidad.

En cuanto a la interpretación del coeficiente de X_1 podemos decir que si la probabilidad de detención aumenta un 1 %, la tasa de criminalidad disminuirá un 0,65 %.

(3) En este caso, todos los coeficientes coinciden con el signo esperado a excepción del coeficiente de X_4, según el cuál, un aumento del 1 % en el promedio de penas de prisión supondría que la tasa de criminalidad aumentase un 0,018 %. En cuanto a la interpretación del coeficiente de X_1, si la probabilidad de detención aumenta un 1 %, la tasa de criminalidad disminuirá un 0,23 %, es decir, el efecto ahora es menor que en el caso anterior, que

era casi el triple que en este caso. Esto lleva a pensar que parte del efecto que antes se atribuía a la probabilidad de arresto era en realidad debido a efectos inobservables de los condados. A la vista de los resultados se puede concluir que en efecto el aumento de la probabilidad de arresto supone una disminución de la tasa de criminalidad. De nuevo, a excepción del coeficiente del promedio de penas de prisión, todos los coeficientes son individualmente significativos a la hora de explicar la variabilidad de la tasa de criminalidad.

Si seguimos suponiendo normalidad tenemos:

$$\left|\frac{\hat{\beta}_0}{\hat{\sigma}(\hat{\beta}_0)}\right| = |-9,97| > |1,96|$$

$$\left|\frac{\hat{\beta}_1}{\hat{\sigma}(\hat{\beta}_1)}\right| = |-6,15| > |1,96|$$

$$\left|\frac{\hat{\beta}_2}{\hat{\sigma}(\hat{\beta}_2)}\right| = |-6,20| > |1,96|$$

$$\left|\frac{\hat{\beta}_3}{\hat{\sigma}(\hat{\beta}_3)}\right| = |3,641| > |1,96|$$

$$\left|\frac{\hat{\beta}_4}{\hat{\sigma}(\hat{\beta}_4)}\right| = |0,5903| < |1,96|$$

$$\left|\frac{\hat{\beta}_5}{\hat{\sigma}(\hat{\beta}_5)}\right| = |3,012| > |1,96|$$

En cuanto a la interpretación del coeficiente de X_4, tiene un signo contrario a la intuición como hemos comentado, pues cabría esperar que si en un determinado condado el promedio de penas de prisión es mayor, la tasa de criminalidad sea menor. Sin embargo, el modelo proporciona el siguiente resultado: un aumento de un 1 % en el promedio de las penas de prisión supondrá un aumento de la criminalidad del 0,018 %. Sin embargo, este coeficiente no es significativo de forma individual por lo que, a la vista del modelo, no deberíamos centrar el análisis en esta variable y sería más oportuno incluso analizar el planteamiento sin ella porque parece que la gravedad del castigo no es un elemento disuasivo.

(4) Como ya se adelantaba en el enunciado, los autores sospechaban que podía haber heterogeneidad entre los condados de forma que habría que incluir en el modelo un efecto fijo no observable específico a cada condado. Dado que en el segundo modelo todos los signos esperados coinciden con los obtenidos excepto el cuarto y que este no es significativo, la segunda especificación parece más acertada.

(5) En el caso concreto del modelo que se propone tiene sentido pensar que la relación entre las variables explicativas y la variable Y es bidireccional, salvo quizá la variable X_5, puesto que a medida que aumenta el salario lo lógico es que sí tuviera una influencia en la disminución de la tasa de criminalidad pero no al revés, es decir, en ese caso no parece bidireccional. Las demás variables sí parecen tener una relación con la endógena en ambas direcciones. Por ejemplo, para X_1, probabilidad detención: una variación en la probabilidad de detención sí debería influir en la tasa de criminalidad y una variación de la tasa de criminalidad también debería reflejarse en la probabilidad de detención. Por tanto, esto confirma nuestra intuición de que sí que podría haber endogeneidad en el modelo.

Formalmente, que alguna de las variables X_j pueda ser endógena significa que está correlacionada con el término de error, y esto podría ocurrir debido a la omisión de variables relevantes, a errores de medida, simultaneidad, etc...

La existencia de este tipo de variables invalida los estimadores MCO de los parámetros del modelo, que serían inconsistentes, y quizá por eso esa otra económetra ha considerado la estimación con efectos fijos, que de alguna forma constituyen una herramienta de control de la endogeneidad que se resuelve con regresiones en diferencias, entre otras posibles soluciones. La heterogeneidad entre condados puede ser una causa de esa endogeneidad y por tanto parece preferible el uso de un modelo con efectos fijos frente al propuesto en primer lugar. Esto podría indicar que se rechazaría la hipótesis nula del contraste de ausencia de correlaciones

$$H_0 : E\left(\alpha_i | x_{i1}, \ldots, x_{iT}, z_i\right) = 0$$

por lo que en efecto podría haber endogeneidad.

No obstante, aun estimando con efectos fijos uno de los coeficientes no parece responder a lo que sería más lógico, por lo que todavía el modelo seguramente puede mejorarse.

Cuestión 8

Una variable macroeconómica Y_t se modeliza con $Y_t = 0,01 + 0,5\epsilon_{t-1} + 0,1\epsilon_{t-2} + \epsilon_t$ donde $\{\epsilon_t\}$ es i.i.d. con media cero y varianza σ^2.

(1) Calcula la esperanza y varianza incondicionada de Y_t.

(2) Calcula la autocovarianza de primer y segundo orden de Y_t.

(3) ¿Cómo es la autocovarianza para retardos superiores a 2?

(4) A partir de la información anterior, ¿puede concluirse que el proceso es estacionario débil? Justifique la respuesta.

(5) ¿Es un proceso invertible? Justifique la respuesta.

(6) ¿Cuál es la esperanza condicionada de Y_{t+1} dada toda la información disponible en el periodo t?

Solución.

(1) Puesto que la variable presenta una estructura correspondiente a un modelo $MA(2)$, la esperanza incondicionada de Y_t será

$$E[Y_t] = E[0,01 + 0,5\epsilon_{t-1} + 0,1\epsilon_{t-2} + \epsilon_t] = 0,01 = \mu,$$

mientras que la varianza incondicionada vendrá dada por

$$\begin{aligned}Var(Y_t) =& E[Y_t - \mu]^2 = E(0,01 + 0,5\epsilon_{t-1} + 0,1\epsilon_{t-2} + \epsilon_t - 0,01)^2 = \\ =& E(0,5\epsilon_{t-1} + 0,1\epsilon_{t-2} + \epsilon_t)^2 = \\ =& 0,5^2E(\epsilon_{t-1}^2) + 0,1^2E(\epsilon_{t-2}^2) + E(\epsilon_t^2) + 2\cdot 0,5\cdot 0,1\cdot E(\epsilon_{t-1}\epsilon_{t-2}) + \\ &+ 2\cdot 0,5\cdot E(\epsilon_{t-1}\epsilon_t) + 2\cdot 0,1\cdot E(\epsilon_{t-2}\epsilon_t) = \\ =& 0,5^2\sigma^2 + 0,1^2\sigma^2 + \sigma^2 = 1,26\sigma^2\end{aligned}$$

pues

$$E\left(\epsilon_t\epsilon_{t-j}\right) = 0 \qquad \forall j > 0$$

(2) De nuevo, teniendo en cuenta que $E\left(\epsilon_t\epsilon_{t-j}\right) = 0\ \forall j > 0$, la autocovarianza de primer orden es

$$\begin{aligned}\gamma_1 &= E[(Y_t - \mu)(Y_{t-1} - \mu)] = \\ &= E[(0,5\epsilon_{t-1} + 0,1\epsilon_{t-2} + \epsilon_t)(0,5\epsilon_{t-2} + 0,1\epsilon_{t-3} + \epsilon_{t-1})] = \\ &= 0,5E(\epsilon_{t-1}^2) + 0,1\cdot 0,5\cdot E(\epsilon_{t-2}^2) = 0,5\sigma^2 + 0,05\cdot\sigma^2 = 0,55\sigma^2\end{aligned}$$

y la autocovarianza de segundo orden

$$\begin{aligned}\gamma_2 &= E[(Y_t - \mu)(Y_{t-2} - \mu)] \\ &= E[(0,5\epsilon_{t-1} + 0,1\epsilon_{t-2} + \epsilon_t)(0,5\epsilon_{t-3} + 0,1\epsilon_{t-4} + \epsilon_{t-2})] = \\ &= 0,1E\left(\epsilon_{t-2}^2\right) = 0,1\sigma^2\end{aligned}$$

(3) Para todo $j > 2$, la autocovarianza es nula pues

$$\gamma_j = E[(Y_t - \mu)(Y_{t-j} - \mu)] =$$

$$= E[(0,5\epsilon_{t-1} + 0,1\epsilon_{t-2} + \epsilon_t)(0,5\epsilon_{t-j-1} + 0,1\epsilon_{t-j-2} + \epsilon_{t-j})] = 0$$

(4) Se dice que un porceso estocástico es estacionario en sentido débil si cumple las siguientes tres propiedades:

1. Las esperanzas matemáticas de las variables aleatorias no dependen del tiempo, es decir, son constantes:

$$E[Y_t] = E[Y_{t+m}] \ \forall \ m$$

2. Las varianzas de las variables aleatorias no dependen del tiempo y son finitas, es decir:

$$Var(Y_t) = Var(Y_{t+m}) < \infty \qquad\qquad \forall m$$

3. Las covarianzas entre dos periodos de tiempo distintos solamente dependen del lapso de tiempo trasncurrido entre estos dos períodos. Es decir:

$$cov(Y_t, Y_{t'}) = cov(Y_{t+m}, Y_{t'+m}) \qquad\qquad \forall m$$

A la vista de los resultados obtenidos en los apartados anteriores, se puede concluir que el proceso es estacionario débil pues:

1. La esperanza de la variable aleatoria es constante.

$$E[Y_t] = 0,01 = \mu \qquad\qquad \forall m$$

2. La varianza no dependen del tiempo y es finita.

$$Var(Y_t) = 1,26\sigma^2 < \infty \qquad\qquad \forall m$$

3. La covarianza entre dos periodos de tiempo distintos solamente dependen del lapso de tiempo trasncurrido entre ellos.

$$
\begin{aligned}
cov(Y_t, Y_{t-1}) &= 0,55\sigma^2 & \text{si } j = 1 \\
cov(Y_t, Y_{t-2}) &= 0,1\sigma^2 & \text{si } j = 2 \\
cov(Y_t, Y_{t-j}) &= 0 & \forall j > 2
\end{aligned}
$$

(5) La variable presenta una estructura correspondiente a un modelo $MA(2)$

$$y_t = \mu + \theta_1\epsilon_{t-1} + \theta_2\epsilon_{t-2} + \epsilon_t$$

que puede escribirse como

$$y_t = \mu + \theta(L)\epsilon_t$$

donde

$$\theta(L) = 1 - \theta_1 L - \theta_2 L^2$$

Como culaquier modelo de medias móviles, el modelo $MA(2)$ será siempre estacionario mientras que para que se trate de un proceso invertible deberá cumplirse que las raíces de la ecuación

$$\theta(L) = 1 - \theta_1 L - \theta_2 L^2 = 0$$

caigan fuera del círculo unidad. Estas raíces podrán ser reales o complejas. En cualquiera de los dos casos, el que las raíces caigan fuera del círculo unitario implica la siguientes restricciones en los valores de los parámetros θ_1 y θ_2:

$$\theta_2 + \theta_1 < 1$$

$$\theta_2 - \theta_1 < 1$$

$$|\theta_2| < 1$$

En el caso que nos ocupa $\theta_1 = -0,5$ y $\theta_2 = -0,1$ por lo que el proceso es invertible al ser:

$$\theta_2 + \theta_1 = -0,5 + (-0,1) = -0,6 < 1$$
$$\theta_2 - \theta_1 = -0,5 - (-0,1) = -0,4 < 1$$
$$|\theta_2| = |0,1| < 1$$

(6) La predicción es uno de los objetivos del análisis de series temporales. Se trata de predecir valores futuros de una variable con el menor error posible. La información disponible hasta el periodo t es $Y_1, \ldots, Y_t, \epsilon_1, \ldots, \epsilon_t$ y deseamos obtener la predicción de Y_{t+1}. A ese conjunto de información hasta el periodo t lo denotamos por I_t y al predictor óptimo en el periodo $t+l$ en general lo escribimos como $\hat{Y}_t(l)$, utilizando toda la información disponible hasta t.

Se puede demostrar que $\hat{Y}_t(l) = E(Y_{t+l}|I_t)$, por lo que teniendo el modelo $Y_t = 0,01{+}0,5\epsilon_{t-1}{+}0,1\epsilon_{t-2}{+}\epsilon_t$, para calcular esa esperanza condicionada que se nos pide, que es precisamente la predicción en el momento $t+1$ utilizamos el supuesto de que las perturbaciones son conocidas para los periodos de la muestra y tienen carácter de ruido blanco para períodos posteriores, es decir:

$$E(\epsilon_{t+j}|I_t) = \begin{cases} \epsilon_{t+j} \text{ para } j \leq 0 \\ E(\epsilon_{t+j}) = 0 \text{ para } j > 0 \end{cases}$$

Luego $\hat{Y}_t(1)$ se calcula de la siguiente forma:

$$\hat{Y}_t(1) = E[Y_{t+1}|I_t] = E[0,01 + 0,5\epsilon_t + 0,1\epsilon_{t-1} + \epsilon_{t+1}|I_t] =$$
$$= 0,01 + 0,5E[\epsilon_t|I_t] + 0,1E[\epsilon_{t-1}|I_t] + E[\epsilon_{t+1}|I_t]$$

El error ϵ_{t+1} es independiente de lo que ha pasado hasta t, su historia pasada no ayuda a predecir el futuro de forma que la esperanza condicional es idéntica a la esperanza incondicional e igual a cero, esto es,

$$E[\epsilon_{t+1}|I_t] = E[\epsilon_{t+1}] = 0$$

Así:

$$\hat{Y}_t(1) = E[Y_{t+1}|I_t] = 0,01 + 0,5\epsilon_t + 0,1\epsilon_{t-1}$$

Cuestión 9

Dada la siguiente tabla de mortalidad:

	Tasa de mortalidad	Promedio de años vividos el último año de vida	Riesgo de muerte	Supervivientes	Defunciones teóricas	Población estacionaria	Tiempo por vivir	Esperanza de vida
0 años		0,123939	2,645464					
1 año		0,481492	0,223966				8209014,8	

Se pide:

a) Calcular los supervivientes a 0 años.

b) Defunciones teóricas a 0 años.

c) Población estacionaria a 0 años.

d) Tasa de mortalidad a 0 años.

e) Calcular los supervivientes a 1 años.

f) Defunciones teóricas a 1 años.

g) Población estacionaria a 1 años.

h) Tasa de mortalidad a 1 años.

i) Esperanza de vida para las personas de 1 año.

j) Tiempo por vivir para las personas de 0 años.

k) Esperanza de vida para las personas de 0 años

Solución.

a)

$$l_0 = 100000$$

b)

$$q_0 = \frac{d_0}{l_0} \Leftrightarrow d_0 = q_0 \cdot l_0 \Leftrightarrow d_0 = \frac{2,645462}{1000} \cdot 100000 = 264,5464$$

c)

$$l_1 = l_0 - d_0 = 100000 - 264,5464 = 99735,4536$$

$$L_0 = l_1 + a_0 d_0 = 99735,4536 + 0,123939 \cdot 264,5465 = 99768,24122$$

d)

$$m_0 = \frac{d_0}{L_0} = \frac{264,5464}{99768,24122} = 2,6516‰$$

Otra posibilidad sería utilizar:

$$m_0 = \frac{q_0}{1 + (1 - a_0)q_0}$$

e)

$$l_1 = l_0 - d_0 = 100000 - 264,5464 = 99735,4536$$

f)

$$d_1 = q_1 \cdot l_1 = \frac{0,223966}{1000} \cdot 99735,4536 = 22,337$$

g)

$$l_2 = l_1 - d_1 = 99735,4536 - 22,337 = 99713,1162$$

$$L_1 = l_2 + a_1 d_1 = 99713,1162 + 0,481492 \cdot 22,337 = 99723,87$$

h)

$$m_1 = \frac{d_1}{L_1} = \frac{22,337}{99723,87} = 0,22398‰$$

i)

$$e_1 = \frac{T_1}{l_1} = \frac{8209014,8}{99735,4536} = 82,3078$$

j)

$$T_0 = T_1 + L_0 = 8209014,8 + 99768,24122 = 8308783,041$$

k)

$$e_0 = \frac{T_0}{l_0} = \frac{8308783,041}{100000} = 83,08783$$

Cuestión 10

A partir de la información de la Tabla A, y sabiendo que la población total a 1 de enero de 2015 era de 46450,0 miles de personas mientras que a 1 de enero de 2016 era de 46440,0 miles de personas, se pide calcular:

a) La Tasa Bruta de Natalidad (TBN^{2015})

b) La Tasa General o Global de Fecundidad (TGF^{2015})

c) Las Tasas Específicas de Fecundidad por edad de la madre (TEF_x)

d) El Índice Sintético de Fecundidad (ISF^{2015})

e) La edad media a la maternidad en 2015.

Tabla 3.7: Tabla A

Grupos de edad	Mujeres residentes (miles)		Nacimientos por edad de las madres, año 2015 (miles)
	1.enero.2015	1.enero.2016	
15-19	1.045,3	1.060,2	8,2
20-24	1.137,8	1.117,0	29,8
25-29	1.318,5	1.279,8	75,2
30-34	1.627,1	1.549,0	148,8
35-39	1.938,5	1.897,6	125,5
40-44	1.907,2	1.926,2	30,6
45-49	1.829,7	1.838,4	2,0
Total	10.804,1	10.668,2	420,1

Solución.

La población media a 1 de julio de 2015 es 46445,0 miles de personas teniendo en cuenta que la población total era de 46450,0 miles de personas a 1 de enero de 2015 y de 46440,0 miles de personas a 1 de enero de 2016. La tasa bruta de natalidad es el cociente entre el número total de nacimientos y la población media:

$$TBN^{2015} = \frac{N^{2015}}{P^{01-07-15}} \cdot 1000 = \frac{420,1}{46445,0} \cdot 1000 = 9,045‰.$$

Por otra parte, la Tasa Global de Fecundidad será el cociente entre el número total de nacimientos y la población media de mujeres en edad fértil, es decir,

$$TGF^{2015} = \frac{N^{2015}}{M_{15-49}^{01-07-15}} \cdot 1000 = \frac{420,1}{\frac{10804,1+10668,2}{2}} \cdot 1000 = \frac{420,1}{10736,15} \cdot 1000 = 39,12‰.$$

En cuanto a las Tasas Específicas de Fecundidad por edad de la madre serán el cociente entre el número de nacimientos de cada grupo de edad y la población media de ese grupo determinado. Para ello se calculan primero las poblaciones medias de cada grupo de edad y a continuación el cociente $TEF_x^{2015} = N_x^{2015}/M_x^{01-07-15}$.

$$M_{15-19}^{01-07-15} = \frac{1045,3+1060,2}{2} = 1052,75 \Rightarrow TEF_{15-19}^{2015} = 7,789‰$$

$$M_{20-24}^{01-07-15} = 1127,4 \Rightarrow TEF_{20-24}^{2015} = 26,43‰$$

$$M_{25-29}^{01-07-15} = 1299,15 \Rightarrow TEF_{25-29}^{2015} = 57,88‰$$

$$M_{30-34}^{01-07-15} = 1588,05 \Rightarrow TEF_{30-34}^{2015} = 93,69‰$$

$$M_{35-39}^{01-07-15} = 1918,05 \Rightarrow TEF_{35-39}^{2015} = 65,43‰$$

$$M_{40-44}^{01-07-15} = 1916,7 \Rightarrow TEF_{40-44}^{2015} = 15,96‰$$

$$M_{45-49}^{01-07-15} = 1834,05 \Rightarrow TEF_{45-49}^{2015} = 1,0904‰$$

El Índice Sintético de Fecundidad será la suma de las tasas específicas de fecundidad en tanto por uno. Al estar manejando grupos de edad quinquenales, cada mujer “vive” cinco años dentro de cada uno de los intervalos de edad, y por tanto hay que multiplicar por 5 para obtener el número medio de hijos que tendría una mujer a lo largo de los 35 años de vida fértil:

$$\begin{aligned} ICF^{2015} &= 5\sum_{x=15}^{49} f_x^{2015} = \\ &= 5\left(0,007789 + 0,02643 + 0,05788 + \cdots + 0,01596 + 0,0010904\right) = \\ &= 5 \cdot 0,2682 = 1,341. \end{aligned}$$

En término medio, cada mujer tendrá 1,34 hijos a lo largo de su vida fértil.

Para el cáclulo de la edad media a la maternidad se utiliza

$$EMM^{2015} = \frac{\sum_{x=15}^{49}\left(x+\frac{n}{2}\right) f_x^{2015}}{\sum_{x=15}^{49} f_x^{2015}}$$

donde x hace referencia al extremo izquierdo del intervalo de cada uno de los grupos de edad y n hace referencia a la amplitud de dichos intervalos. Por tanto, la edad media a la maternidad será

$$\begin{aligned} EMM^{2015} &= \frac{\left(15+\frac{5}{2}\right)0,007789 + \left(20+\frac{5}{2}\right)0,02643 + \cdots + \left(45+\frac{5}{2}\right)0,0010904}{0,2682} = \\ &= \frac{8,53899}{0,2682} = 31,8 \end{aligned}$$

Por tanto la edad media a la maternidad es 31,8 años.

3.2. Soluciones del tercer examen del Cuerpo de Diplomados en Estadística del Estado de 2018

Cuestión 1

Sea $(X_1, X_2, \ldots X_n)$ una muestra aleatoria simple de una variable aleatoria X discreta que toma los valores $\{-1, 0, 1\}$ con las siguientes probabilidades:

$$P(X=-1) = \frac{1-\theta}{2} \qquad P(X=0) = \frac{\theta+\lambda}{2} \qquad P(X=1) = \frac{1-\lambda}{2}$$

$$0 < \theta < 1; \qquad 0 < \lambda < 1$$

Calcule, razonando la respuesta:

a) Los estimadores de θ y de λ por el método de los momentos.

b) ¿Es insesgado el estimador de θ obtenido en el apartado anterior?

c) Análogamente ¿es insesgado el estimador de λ?

Solución.

a) El procedimiento establecido por el método de los momentos consiste en igualar los momentos poblacionales a los momentos muestrales. El primer momento poblacional es

$$E[X] = -1 \cdot \frac{1-\theta}{2} + 0 \cdot \frac{\theta+\lambda}{2} + 1 \cdot \frac{1-\lambda}{2} = \frac{\theta-1}{2} + \frac{1-\lambda}{2} = \frac{\theta-\lambda}{2},$$

que, por tanto, hay que igualarlo al primer momento muestral:

$$\overline{x} = \frac{\theta-\lambda}{2} \Leftrightarrow \theta - \lambda = 2\overline{x} \Leftrightarrow \widehat{\theta} = 2\overline{x} + \lambda$$

Por otra parte, el momento poblacional de segundo orden es

$$E[X^2] = (-1)^2 \cdot \frac{1-\theta}{2} + 1^2 \cdot \frac{1-\lambda}{2} = \frac{2-\theta-\lambda}{2} = 1 - \frac{\theta+\lambda}{2}$$

que, de forma análoga al caso anterior, se iguala al momento muestral de segundo orden:

$$\frac{1}{n}\sum_{i=1}^{n} x_i^2 = 1 - \frac{\theta+\lambda}{2}$$

Equivalentemente,

$$\frac{\theta+\lambda}{2} = 1 - \frac{1}{n}\sum_{i=1}^{n} x_i^2$$

donde sustituyendo el resultado obtenido en el paso anterior, $\widehat{\theta} = 2\overline{x} + \lambda$, se tiene

$$\frac{2\overline{x}+\lambda+\lambda}{2} = 1 - \frac{1}{n}\sum_{i=1}^{n} x_i^2 \Leftrightarrow 2\overline{x}+2\lambda = 2 - \frac{2}{n}\sum_{i=1}^{n} x_i^2 \Leftrightarrow 2\lambda = 2 - 2\overline{x} - \frac{2}{n}\sum_{i=1}^{n} x_i^2$$

Es decir:

$$\widehat{\lambda} = 1 - \overline{x} - \frac{1}{n}\sum_{i=1}^{n} x_i^2$$

Y como $\widehat{\theta} = 2\overline{x} + \lambda$,

$$\widehat{\theta} = 2\overline{x} + 1 - \overline{x} - \frac{1}{n}\sum_{i=1}^{n} x_i^2 = 1 + \overline{x} - \frac{1}{n}\sum_{i=1}^{n} x_i^2$$

b) Utilizando que $\overline{x} = \frac{\theta-\lambda}{2}$ y $\frac{\theta+\lambda}{2} = 1 - \frac{1}{n}\sum_{i=1}^{n} x_i^2$, se calcula la esperanza del estiamdor de θ obtenido en el apartado anterior:

$$\begin{aligned} E\left[\widehat{\theta}\right] &= E\left[1 + \overline{x} - \frac{1}{n}\sum_{i=1}^{n} x_i^2\right] = 1 + \frac{n\frac{\theta-\lambda}{2}}{n} - \frac{n\left(1-\frac{\theta+\lambda}{2}\right)}{n} = \\ &= 1 + \frac{\theta-\lambda}{2} - \left(1 - \frac{\theta+\lambda}{2}\right) = 1 + \frac{\theta-\lambda}{2} - 1 + \frac{\theta+\lambda}{2} = \\ &= \frac{\theta-\lambda}{2} + \frac{\theta+\lambda}{2} = \frac{\theta}{2} - \frac{\lambda}{2} + \frac{\theta}{2} + \frac{\lambda}{2} = \theta. \end{aligned}$$

Dado que $E\left[\widehat{\theta}\right] = \theta$, se concluye que el estimador de θ proporcionado por el método de los momentos es insesgado.

c) Siguiendo el mismo razonamiento, se demuestra que el estimador del parámetro λ obtenido utilizando el método de los momentos es insesgado:

$$E\left[\widehat{\lambda}\right] = E\left[1 - \overline{x} - \frac{1}{n}\sum_{i=1}^{n} x_i^2\right] = 1 - \frac{n\frac{\theta-\lambda}{2}}{n} - \frac{n\left(1 - \frac{\theta+\lambda}{2}\right)}{n} = \lambda.$$

Cuestión 2

Sea (X, Y) una variable aleatoria bidimensional con función de densidad:

$$f(x) = \begin{cases} c(x-y) & \text{si } 0 < x < 2, -x < y < x \\ \\ 0 & \text{en el resto de los casos} \end{cases}$$

Calcular:

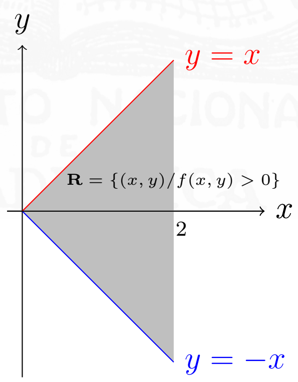
a) El valor de c para que $f(x, y)$ sea función de densidad.

b) Las funciones de densidad marginales de X e Y.

c) ¿Son X e Y variables aleatorias independientes? Razone la respuesta.

Solución.

a) El recinto sobre el que está definida la variable aleatoria bidimensional es el siguiente:



Para que sea función de densidad tiene que ocurrir que la integral sobre dicho recinto sea uno, es decir,

$$1 = \int_0^2 \int_{-x}^{x} c(x-y)dydx = c\int_0^2 \left(xy - \frac{y^2}{2}\right)_{-x}^{x} dy =$$
$$= c\int_0^2 x^2 - \frac{x^2}{2} - \left(-x^2 - \frac{x^2}{2}\right) dx = c\int_0^2 2x^2 dx = 2c\left(\frac{x^3}{3}\right)_0^2 =$$
$$= 2c\frac{8}{3} = \frac{16}{3}c$$

Luego necesariamente

$$c = \frac{3}{16}$$

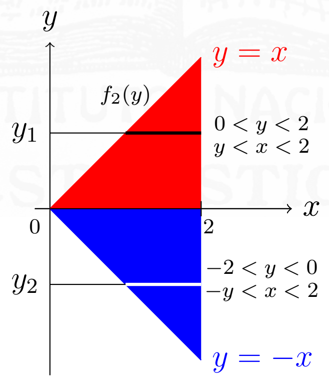
b) Dada la forma del recinto en que está definida la variable (X, Y), la función de densidad marginal de la variable Y está definida en dos partes:

- Para $0 < y < 2$:

$$f_2(y) = \int_y^2 \frac{3}{16}(x-y)dx = \frac{3}{16}\left(\frac{x^2}{2} - yx\right)_y^2 =$$
$$= \frac{3}{16}\left(2 - 2y - \left(\frac{y^2}{2} - y^2\right)\right) = \frac{3}{16}\left(2 - 2y + \frac{y^2}{2}\right)$$

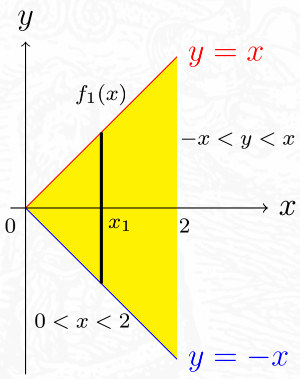
- Para $-2 < y \leq 0$:

$$f_2(y) = \int_{-y}^2 \frac{3}{16}(x-y)dx = \frac{3}{16}\left(\frac{x^2}{2} - yx\right)_{-y}^2 =$$
$$= \frac{3}{16}\left(2 - 2y - \left(\frac{y^2}{2} + y^2\right)\right) = \frac{3}{16}\left(2 - 2y - \frac{3y^2}{2}\right)$$



La función de densidad marginal de la variable X viene dada por

$$f_1(x) = \begin{cases} \int_{-x}^{x} \frac{3}{16}(x-y)dy = \frac{3x^2}{8} & \text{si } 0 < x < 2 \\ \\ 0 & \text{en el resto de los casos} \end{cases}$$



c) Para que fueran independientes tendría que ocurrir que $f_1(x) \cdot f_2(y) = f(x, y)$, sin embargo, por ejemplo, para $0 < x < 2$, $0 < y < 2$:

$$\frac{3x^2}{8} \cdot \frac{3}{16}\left(2 - 2y + \frac{y^2}{2}\right) \neq \frac{3}{16}(x - y)$$

por lo tanto no son independientes.

Cuestión 3

En una población dividida en dos estratos, se desea investigar el valor medio de una determinada característica cuantitativa y para ello se propone un muestreo estratificado considerando una función del coste de los trabajos de campo del tipo:

$$\sum_{h=1}^{2} c_h n_h$$

donde h indica el estrato; c_h el coste por unidad encuestada en el estrato h; n_h el número total de unidades encuestadas en el estrato h. La información de la que se dispone de los estratos se recoge en la tabla 3.8.

Tabla 3.8: Información por estratos.

Estrato	W_h	S_h	c_h
1	0,4	10	4 euros
2	0,6	20	9 euros

Donde W_h y S_h representan el peso y la raíz de la cuasivarianza poblacional de la característica estudiada, respectivamente, en el estrato h. Si se desea minimizar el coste de la encuesta para un valor dado de la varianza del estimador de la media, se pide:

a) Hallar la relación entre los tamaños muestrales en los estratos, n_1 y n_2.

b) Hallar el coste total de los trabajos de campo para un tamaño total de la muestra de $n = 300$.

c) Despreciando la fracción de muestreo en los estratos, calcular el error de muestreo para $n = 300$.

Solución.

a) Dado que el objetivo es minimizar el coste de la encuesta dado un valor de la varianza, hay que utilizar la afijación óptima. La afijación óptima determina el valor del tamaño muestral de cada estrato como:

$$n_h = n\frac{W_h S_h/\sqrt{c_h}}{W_1 S_1/\sqrt{c_1} + W_2 S_2/\sqrt{c_2}}.$$

Este se deduce de la minimización de la función de los multiplicadores de Lagrange. En este caso, dicha función es

$$\Phi = c_1 n_1 + c_2 n_2 + \lambda \left(\sum_{h=1}^{2} W_h^2 \left(1 - f_h\right) \frac{S_h^2}{n_h} - Var(\overline{x}_{st}) \right)$$

De esta forma, los tamaños muestrales en los estratos considerando el valor proporcionado por la afijación óptima, serán

$$n_1 = n\frac{W_1S_1/\sqrt{c_1}}{W_1S_1/\sqrt{c_1} + W_2S_2/\sqrt{c_2}}$$
$$n_2 = n\frac{W_2S_2/\sqrt{c_2}}{W_1S_1/\sqrt{c_1} + W_2S_2/\sqrt{c_2}}$$

y la relación entre los tamaños muestrales es:

$$\frac{n_1}{n_2} = \frac{n\frac{W_1S_1/\sqrt{c_1}}{W_1S_1/\sqrt{c_1}+W_2S_2/\sqrt{c_2}}}{n\frac{W_2S_2/\sqrt{c_2}}{W_1S_1/\sqrt{c_1}+W_2S_2/\sqrt{c_2}}} = \frac{W_1S_1\sqrt{c_2}}{W_2S_2\sqrt{c_1}}.$$

Es decir,

$$n_1 = n_2\frac{W_1S_1\sqrt{c_2}}{W_2S_2\sqrt{c_1}}$$

Con los datos del problema:

$$n_1 = n_2\frac{0,4\cdot 10\sqrt{9}}{0,6\cdot 20\sqrt{4}} = n_2\frac{4\sqrt{9}}{6\cdot 2\sqrt{4}} = n_2\frac{12}{24} = \frac{n_2}{2}$$

El tamaño muestral en el estrato 2 será el doble que en el estrato 1.

b) Para $n = 300$, $n_2 = 300\frac{0,6\cdot 20/3}{0,4\cdot 10/2+0,6\cdot 20/3} = 200$ y $n_1 = 200\frac{0,4\cdot 10/2}{0,6\cdot 20/3} = 100$ luego el coste total será:

$$C = \sum_{h=1}^{2} c_h n_h = 4\cdot 100 + 9\cdot 200 = 2200$$

c) Despreciando la fracción de muestreo, el error de muestreo sería:

$$Var(\overline{x}_{st}) = \sum_{h=1}^{L} W_h^2(1-f_h)\frac{S_h^2}{n_h} \approx \sum_{h=1}^{L} W_h^2\frac{S_h^2}{n_h}$$
$$Var(\overline{x}_{st}) \simeq W_1^2\frac{S_1^2}{n_1} + W_2^2\frac{S_2^2}{n_2} = 0,4^2\cdot\frac{10^2}{100} + 0,6^2\cdot\frac{20^2}{200} = 0,88$$
$$\sigma\left(\overline{x}_{st}\right) = \sqrt{0,88} = 0,938.$$

Cuestión 4

La siguiente tabla presenta datos de edades sobre un conjunto de 24 individuos seleccionados de dos poblaciones distintas.

P1	38	10	10	60	20	38	5	40	40	90	40	40	40	50	50	50	50	10	60	70	80	80	90	40
P2	40	30	30	30	30	60	35	35	35	40	98	40	40	40	20	40	40	45	45	50	35	60	60	40

Se pide:

1) Representar gráficamente ambas distribuciones de datos.

2) Comparar los grados de dispersión y simetría de ambas distribuciones.

Solución.

1) Para la representación gráfica se obtienen en primer lugar las tablas estadísticas de distribuciones de frecuencias de cada una de las pobalciones (Tablas 3.9 y 3.10 respectivamente).

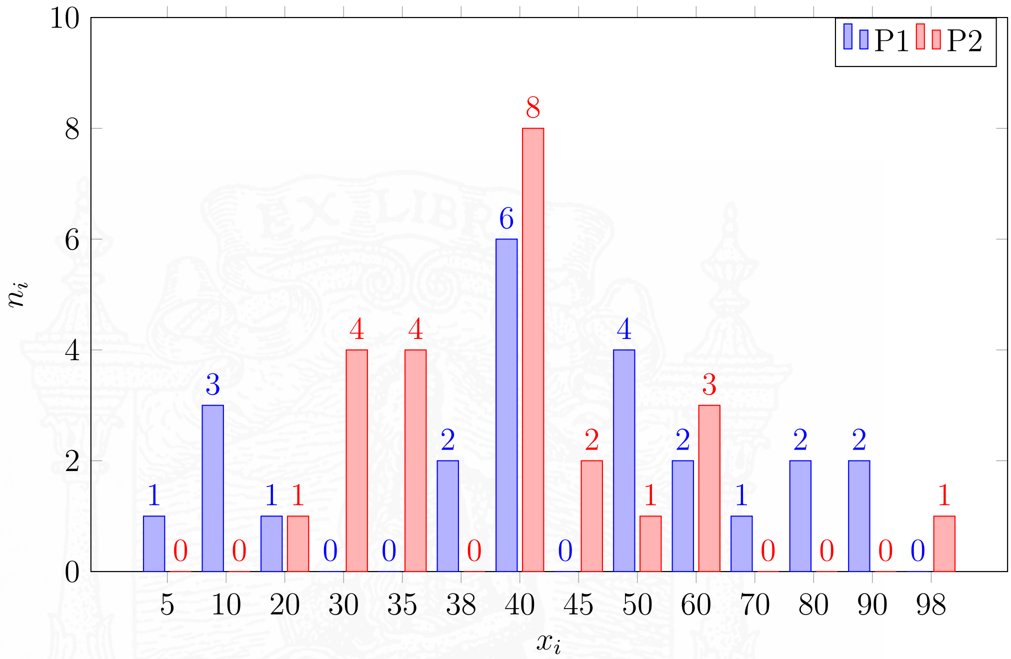
Tabla 3.9: Distribución de P1

x_i	5	10	20	38	40	50	60	70	80	90
n_i	1	3	1	2	6	4	2	1	2	2

Tabla 3.10: Distribución de P2

x_i	20	30	35	40	45	50	60	98
n_i	1	4	4	8	2	1	3	1

Dichas frecuencias se representan utilizando un diagrama de barras, que consiste en utilizar barras de altura n_i sobre la abscisa x_i. En el diagrama que aparece a continuación, se utiliza el color azul para la población P1 y el color rojo para la población P2.



2) Para comparar los grados de dispersión y simetría se calculan diferentes medidas:

- Recorrido, rango o campo de variación: $R = \max x_i - \min x_i$.

  P1: $R = 90 - 5 = 85$

  P2: $R = 98 - 20 = 78$

  El recorrido de **P1** mayor que recorrido de **P2**, luego **P1** presenta mayor dispersión.

- Coeficiente de apertura: $A = \frac{\max x_i}{\min x_i}$.

  P1: $A = \frac{90}{5} = 18$

  P2: $A = \frac{98}{20} = 4,9$

  La apertura en es **P1** mucho mayor que la apertura en **P2**, reflejo de mayor dispersión en **P1**.

- Varianza: $S^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2 n_i}{N}$.

  P1: $S^2 = \frac{13804,625}{24} = 575,19$

P2: $S^2 = \frac{5487{,}40}{24} = 228,08$

La varianza de **P1** es de nuevo mucho mayor que la varianza de **P2**, símbolo de la mayor dispersión de **P1**.

Una vez calculada la varianza, se puede obtener la desviación típica: $S = \sqrt{S}$, que de forma análoga es una medida de dispersión muy frecuentemente utilizada.

P1: $S = \sqrt{575,19} = 23,98$

P2: $S = \sqrt{228,08} = 15,10$

- Coeficiente de variación de Pearson: $CV = \frac{S}{\overline{x}}$.

  P1: $CV = \frac{23{,}98}{45{,}875} = 0,5227$

  P2: $CV = \frac{15{,}10}{42{,}416} = 0,3560$

  Una vez más, el mayor coeficiente de variación de Pearson en **P1** confirma la mayor dispersión existente en los datos de **P1**.

- Índice de asimetría de Fisher: $g_1 = \frac{m_3}{S^3}$ donde $m_3 = \frac{\sum_{i=1}^{n}(x_i-\overline{x})^3 n_i}{N}$.

  P1: $g_1 = \frac{1872{,}32}{(23{,}98)^3} = 0,1357$

  P2: $g_1 = \frac{6993{,}24}{(15{,}10)^3} = 2,0303$

  Cuanto más próximos son estos valores a 0 significa que más simétricas son las distribuciones. Se observa que **P1** presenta mayor grado de simetría que **P2**.

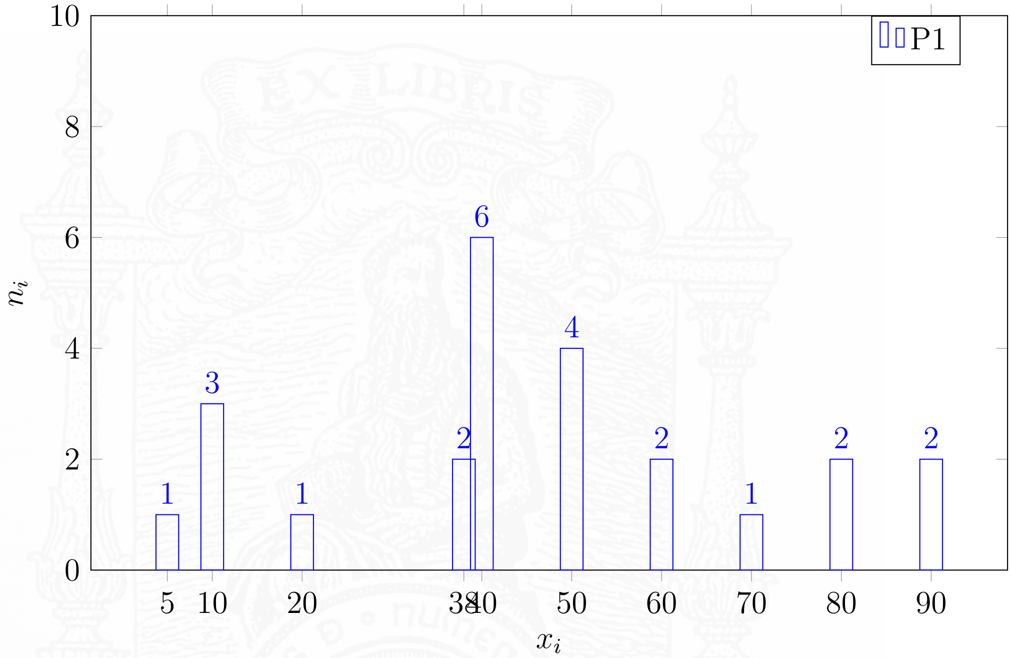
- Índice de asimetría de Pearson: $A_p = \frac{\overline{x}-Mo}{S}$

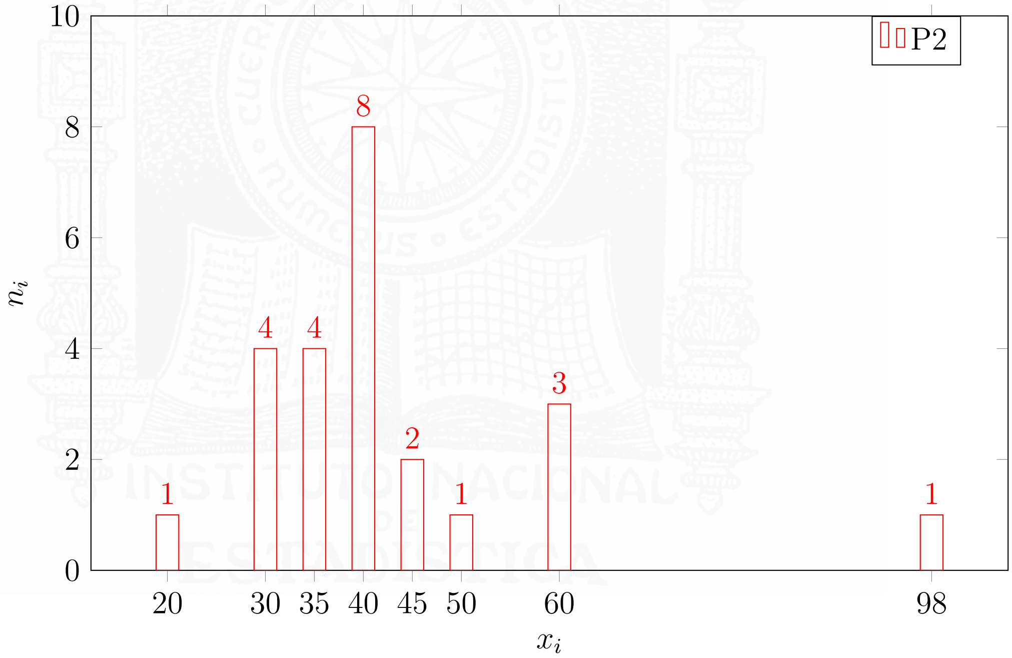
  P1: $A_p = \frac{45{,}875-40}{23{,}98} = 0,2449$, donde la moda es 40 por ser la observación con mayor frecuencia.

  P2: $A_p = \frac{42{,}4166-40}{15{,}10} = 0,4910$, donde la moda es 40 por el mismo motivo .

  Igual que ocurre con el Índice de asimetría de Fisher, cuanto más próximos son estos valores a 0, más simétricas son las distribuciones.

De nuevo, el Índice de asimetría de Pearson refleja mayor simetría de la población **P1** respecto a **P2**.





Cuestión 5

En la tabla 3.11 se recogen los datos de la evolución de los precios de un producto entre 2013 y 2017, así como el valor del IPC de esa economía. A partir de los mismos, se pide:

a) Obtener la serie del IPC para todo el periodo señalado, tomando como año base 2013=100.

b) Calcular la serie de precios de venta del artículo, expresados en euros de 2013.

c) Calcular las variaciones anuales experimentadas en el precio en términos corrientes del artículo y compare su evolución respecto a la variación de los precios en términos constantes.

d) Calcular el precio de venta del artículo previsto para el año 2020 suponiendo que la variación experimentada en el año 2016-2017 se mantuviera constante en los siguientes años.

Tabla 3.11: Datos de la evolución de los precios y del IPC entre 2013 y 2017

Año	Precio venta	IPC (media anual)
2013	114,23	100,859
2014	119,87	100,707
2015	125,03	100,203
2016	129,31	100,000
2017	135,12	101,956

Solución.

a) Para el cálculo de la serie del IPC tomando como año base 2013, se divide para cada año el valor del índice entre el valor del índice en el periodo base (que es 100,859) y se multiplica por 100. Se recoge en la primera columna de la tabla 3.12.

b) El cálculo de la serie de precios de venta del artículo expresados en euros de 2013, que se recoge en la segunda columna de la tabla 3.12, se divide el valor del precio de venta en cada año entre el índice en base 2013 calculado en la primera columna.

Tabla 3.12: Solución apartados a y b.

Año	IPC base 2013 (apartado a)	Precio venta en euros de 2013 (apartado b)
2013	$100,859 \cdot \frac{100}{100,859} = 100$	$114,23 \cdot \frac{100}{100} = 114,23$
2014	$100,707 \cdot \frac{100}{100,859} = 99,84$	$119,87 \cdot \frac{100}{99,84} = 120,06$
2015	$100,203 \cdot \frac{100}{100,859} = 99,34$	$125,03 \cdot \frac{100}{99,34} = 125,86$
2016	$100,00 \cdot \frac{100}{100,859} = 99,14$	$129,31 \cdot \frac{100}{99,14} = 130,43$
2017	$101,956 \cdot \frac{100}{100,859} = 101,08$	$135,12 \cdot \frac{100}{101,08} = 133,66$

c) Por último, en la tabla 3.13 calculamos las variaciones anuales del precio en términos corrientes y constantes.

Tabla 3.13: Solución apartado c.

Año	En términos corrientes	En términos constantes
TVI_{2013}^{2014}	$\left(\frac{119{,}87}{114{,}23}-1\right)\cdot 100 = 4,93\,\%$	$\left(\frac{120{,}06}{114{,}23}-1\right)\cdot 100 = 5,10\,\%$
TVI_{2014}^{2015}	$\left(\frac{125{,}03}{119{,}87}-1\right)\cdot 100 = 4,30\,\%$	$\left(\frac{125{,}86}{120{,}06}-1\right)\cdot 100 = 4,83\,\%$
TVI_{2015}^{2016}	$\left(\frac{129{,}31}{125{,}03}-1\right)\cdot 100 = 3,42\,\%$	$\left(\frac{130{,}43}{125{,}86}-1\right)\cdot 100 = 3,63\,\%$
TVI_{2016}^{2017}	$\left(\frac{135{,}12}{129{,}31}-1\right)\cdot 100 = 4,49\,\%$	$\left(\frac{133{,}66}{130{,}43}-1\right)\cdot 100 = 2,48\,\%$

d) Para calcular el precio de venta del artículo previsto para el año 2020 suponiendo que la variación experimentada en el año 2016 - 2017 se mantuviera constante en los siguientes años, hay que aplicar la tasa de variación de 2016 - 2017 a cada uno de los años hasta llegar a 2020. La tasa de variación de 2016 - 2017 $\left(\text{TVI}_{2016}^{2017}\right)$ es de 4,49 %. Para obtener el precio de venta en 2018, hay que aumentar el precio de venta del año 2017 (que es 135,12) un 4,49 %. Al precio resultante, habría que aplicarle de nuevo un aumento del 4,49 % para obtener el precio de venta del año 2019. Una vez más habría que añadir otro 4,49 % hasta conseguir el precio de venta de 2020. Es decir:

$$P_{2020} = P_{2017}\cdot(1,0449)^3 = 135,12\cdot(1,0449)^3 = 154,15.$$

Por tanto la tasa de variación entre 2017 y 2020 será

$$TVP_{2017}^{2020} = \frac{154,15-135,12}{135,12} = 0,1408.$$

Es decir, el aumento total es de un 14,08 %.

Cuestión 6

Con los datos de la tabla 3.14 relativos al sector Hogares:

Tabla 3.14: Datos del sector Hogares (S.14) (Cifras en millones de euros)

Producción	206
Formación bruta de capital fijo	80
Adquisiciones menos cesiones de activos no financieros no producidos	2
Consumos intermedios	80
Remuneración de asalariados pagada	38
Remuneración de asalariados recibida	530
Impuestos netos sobre la producción y las importaciones	6
Impuestos corrientes	86
Gasto en consumo final	632
Cotizaciones sociales netas	150
Prestaciones sociales	190
Transferencias corrientes netas	-8
Transferencias de capital netas	-2
Transferencias sociales en especie	134
Variación de existencias	2
Adquisiciones menos cesiones de objetos valiosos	1
Ajuste por la variación de los derechos por pensiones	-14
Rentas de la propiedad netas	40

Calcule la sucesión completa de cuentas del sector, además, razone las siguientes preguntas:

1) ¿Se puede calcular el Producto Interior Bruto de este sector?

2) En el sector Hogares ¿se puede distinguir, a nivel teórico, la renta mixta del excedente bruto de explotación?

3) ¿Qué significado económico tiene la operación D.8 – Ajuste por la variación de los derechos por pensiones?

4) ¿Cuál es el consumo final colectivo del sector?

5) ¿Qué significado económico tiene una capacidad de financiación negativa?

Solución.

Se calcula la sucesión completa de cuentas del sector hogares:

E	Cuenta de producción		R
80	Consumos intermedios (P.2)	Producción (P.1)	206
126	Valor añadido Bruto (B1.b)		

E	Cuenta de explotación		R
38	Remuneración de asalariados (D.1)		
6	Impuestos netos sobre la producción e importaciones (D.2–D.3)	Valor añadido bruto (B1.b)	126
82	EBE/RM (B2.b/B3.b)		

E	Cuenta de asignación de la renta primaria		R
		EBE/RM (B2.b/B3.b)	82
		RA (D.1)	530
		Rentas de la propiedad netas (D.4)	40
652	Saldo de rentas primarias (B5.b)		

E	Cuenta de distribución secundaria de la renta		R
		Saldo de rentas primarias (B5.b)	652
86	Imp's corrientes renta (D.5)	Prestaciones sociales distintas de	
150	Cot's sociales netas (D.61)	Transf's sociales en especie (D.62)	190
		Otras transf's corrientes (D.7)	–8
598	RBD (B6.b)		

E	Cuenta de redistribución de la renta en especie		R
		RBD (B6.b)	598
		Trasferencias sociales	134
		en especie (D.63)	
732	RBD Ajustada (B7.b)		

E	Cuenta de utilización de la renta disponible		R
		RBD (B6.b)	598
632	GCF	Ajuste por la variación de los	-14
		derechos por pensiones (D.8)	
-48	Ahorro bruto (B8.b)		

E	Cuenta de utilización de la renta disponible ajustada		R
632 + 134	CFE	RBDAjustada (B7.b)	732
		Ajuste por la variación de los derechos por pensiones (D.8)	-14
-48	Ahorro bruto (B8.b)		

VA	Cuenta de Capital. VPNDAHYTC		VP//PN
		Ahorro neto (B8.n)	-48 - CCF
		Transferencias de capital netas	-2
(-50 - CCF)	VPNDAHYTC		

VA	Cuenta de Capital. Adquisición de activos no financieros		VP//PN
83	FBC(= 80+2+1)	VPNDAHYTC	–50 – CCF
– CCF	- Consumo de capital fijo		
2	Adquisiciones de activos no financieros no producidos		
– 135	Necesidad de financiación (B.9)		

1) El saldo de la cuenta de producción es el valor añadido bruto (VAB) si trabajamos con sectores, es decir, en este caso sería el VAB de los hogares. Pero para el total de economía (incluyendo impuestos netos sobre los productos) hablamos de producto interior bruto (PIB), y faltaría conocer tanto los VAB del resto de sectores como los impuestos netos sobre los productos. Luego no es posible.

2) Sí, la renta mixta es el beneficio derivado de la remuneración del trabajo por cuenta propia realizado por el hogar. Por otra parte, además del

carácter de sueldos y los salarios, se consideran los beneficios por el trabajo realizado como empresario. Esta renta, que no es ni estrictamente una remuneración ni solamente beneficios se conoce como «renta mixta».

3) El ajuste por la variación de los derechos por pensiones (D.8) representa el ajuste necesario para hacer aparecer en el ahorro de los hogares las variaciones de los derechos por pensiones sobre los que los hogares tienen un derecho definido. La variación de los derechos por pensiones obedece a las cotizaciones y prestaciones registradas en la cuenta de distribución secundaria de la renta.

4) El consumo final colectivo es nulo pues en el sector de los hogares, todo el consumo final es individual.

5) La capacidad (+)/necesidad (–) de financiación es igual al saldo de la cuenta de capital (B.9N) en las cuentas del SEC. El término “capacidad/necesidad de financiación” es una especie de simplificación terminológica. Cuando la variable es positiva (es decir, que muestra una capacidad de financiación), debe denominarse capacidad de financiación (+). Cuando es negativa (es decir, que muestra una necesidad de endeudamiento), debe denominarse necesidad de financiación (–). Consecuentemente, la capacidad de financiación corresponde a un excedente que se presta, y la necesidad de financiación corresponde a la financiación de un déficit. En este caso concreto que nos ocupa, el significado económico de la capacidad de financiación negativa corresponde al importe que el sector hogares se ve obligado a pedir prestado a otros sectores.

Cuestión 7

Se conoce la siguiente información del sector S12 “Instituciones financieras” extraída de la Contabilidad Nacional, (cifras en millones de euros). Se pide:

1) Defina lo que son los servicios de intermediación financiera medidos indirectamente (SIFMI) según el SEC- 2010 y en qué rúbrica/s estarían incluidos en la cuenta de producción.

2) Calcule la cuenta de explotación y su saldo.

3) Calcule la cuenta de asignación de la renta primaria y su saldo.

4) Calcule la renta disponible bruta.

5) Sabiendo que el ahorro bruto del sector es 25.510 millones de euros, calcule el gasto el consumo final y la capacidad o necesidad de financiación del sector.

Tabla 3.15: Información relativa al sector S12 "Instituciones financieras" extraída de la Contabilidad Nacional (cifras en millones de euros).

Valor añadido bruto	37.798
Impuestos corrientes sobre los beneficios	4.798
Formación bruta de capital	5.514
Remuneración de asalariados	19.944
Cotizaciones sociales netas recibidas	7.514
Prestaciones sociales pagadas	8.879
Rentas de la propiedad recibidas	76.499
Rentas de la propiedad pagadas	61.076
Impuestos sobre la producción y las importaciones	3.826
Subvenciones a la producción recibidas	95
Transferencias de capital a cobrar	998
Transferencias de capital a pagar	1.651
Consumo de capital fijo	4.415
Transferencias corrientes recibidas	27.530
Transferencias corrientes pagadas	26.768
Adquisiciones menos cesiones de activos no producidos	-25

Solución.

1) En la intermediación financiera hay unidades que disponen de fondos que no necesitan de forma inmediata y hay unidades cuyos fondos son insuficientes para responder a sus necesidades. Las instituciones financieras aceptan depósitos de las primeras para prestarlos a las segundas utilizando para ello un mecanismo mediante el cual cada una de las dos partes paga una comisión al banco por el servicio prestado: la unidad que presta fondos acepta recibir un tipo de interés inferior al tipo «de referencia» y la unidad que

toma prestados los fondos acepta pagar un tipo de interés superior al tipo «de referencia». La diferencia entre el tipo de interés pagado a los bancos por los prestatarios y el tipo de interés efectivamente pagado a los depositarios es un importe por SIFMI. El SIFMI total es la suma de las dos tarifas implícitas pagadas por el prestamista y el prestatario. El PIB se ve afectado por el volumen de SIFMI asignados al consumo final, las exportaciones y las importaciones.

2)

E	Cuenta de explotación		R
19.944	Remuneración de asalariados	Valor añadido bruto	37.798
3.826	Impuestos sobre la producción e importaciones		
-95	Subvenciones a la producción recibidas		
14.123	EBE/RM		

3)

E	Cuenta de asignación de la renta primaria		R
61.079	Rentas de la propiedad pagadas	EBE/RM	14.123
		Rentas de la propiedad recibidas	76.499
29.546	Saldo de rentas primarias		

4)

E	Cuenta de distribución secundaria de la renta		R
4.798	Impuestos corrientes sobre los beneficios	Saldo de rentas primarias	29.546
8.879	Prestaciones sociales pagadas	Cotizaciones sociales netas recibidas	7.514
26.768	Transferencias corrientes pagadas	Transferencias corrientes recibidas	27.530
24.145	Renta Disponible Bruta		

5)

E	Cuenta de utilización de la renta disponible		R
0	GCF	Renta Disponible Bruta	24.145
–1.365	Ajuste por la variación de los derechos por pensiones		
25.510	Ahorro Bruto		

VA	Cuenta de Capital. VPNDAHYTC		VP // PN
		Ahorro bruto	25.510
		– CCF	-4.415
		Transferencias de capital a cobrar	998
		– Transferencias de capital a pagar	-1.651
20.442	VPNDAHYTC		

VA	Cuenta de Capital. Adquisición de activos no financieros		VP // PN
5.514	FBC	VPNDAHYTC	20.442
– 4.415	– CCF		
–25	Adquisiciones de activos no financieros no producidos		
19.368	Capacidad de financiación		

Cuestión 8

Indique cómo se registraría cada una de las siguientes operaciones en la Balanza de Pagos española, comprobando que la balanza queda equilibrada. ¿Cuál sería la capacidad o necesidad de financiación del país?

a) El Gobierno español envía una ayuda de 100.000€ para la construcción de un hospital en un país en vías de desarrollo.

b) Inmigrantes residentes en España envían remesas a sus países de origen por importe de 20.000€.

c) Una fábrica textil española importa materias primas por valor de 30.000€ (FOB), pagando la mitad al contado y la otra mitad mediante un crédito a 4 años concedido por los proveedores no residentes.

d) Una empresa española de instrumentos musicales exporta guitarras por valor de 60.000€, de los cuales 3.000€ corresponden a los fletes y seguros prestados por empresas residentes y pagados al contado por la empresa importadora no residente.

e) Un grupo inversor chino compra una empresa española de videojuegos por valor de 400.000€.

f) Una compañía española compra el 5 % de una empresa de telefonía móvil noruega, por importe de 80.000€.

Solución.

CUENTA CORRIENTE	ingresos	pagos	
Cuenta de Bienes y Servicios			
textiles importados		30.000(c)	
guitarras exportadas	57.000(d)		
fletes	3.000(d)		
↪SALDO Cuenta de Bienes y Servicios			30.000
Cuenta del Ingreso Primario			
↪SALDO Cuenta del Ingreso Primario			0
Cuenta del Ingreso Secundario			
remesas exportadas		20.000(b)	
↪SALDO Cuenta del Ingreso Secundario			- 20.000
Saldo Cuenta Corriente			**10.000**
CUENTA DE CAPITAL	ingresos	pagos	
construcción hospital		100.000(a)	
Saldo Cuenta de Capital			**-100.000**
SALDO BALANZA CORRIENTE Y DE CAPITAL			**-90.000**

CUENTA FINANCIERA	variación de activos	variación de pasivos	
inversión directa		400.000(e)	
inversión en cartera	80.000(f)		
otra inversión		15.000(c)	
reservas	-100.000(a);-20.000(b); -15.000(c);57.000(d); 3.000(d);400.000(e); -80.000(f)		
SALDO CUENTA FINANCIERA			**-90.000**

La capacidad o necesidad de financiación del país coincide con el saldo de la balanza corriente y de capital:

$$10.000 + (-100.000) = -90.000.$$

Al ser dicho saldo negativo, el país presenta necesidad de financiación.

Cuestión 9

La siguiente tabla se refiere a la población de un país contabilizada a 1 de julio.

Edad	Población residente a 1 de Julio	Nacimientos por edad de la madre	Defunciones
0	551	0	4
1-9	5.461	0	3
10-19	5.187	22	1
20-29	5.945	230	0
30-39	6.376	135	2
40-49	5.880	18	9
50 y más	10.600	0	230

Calcule:

1. La tasa bruta de natalidad, sabiendo que 5 nacimientos fueron de madres no residentes, todas ellas de más de 30 años.

2. La tasa específica de mortalidad de menores de 1 año.

3. La tasa de mortalidad infantil.

4. La tasa específica de fecundidad de las mujeres de 20 a 29 años, dada una razón de masculinidad del 103 para ese intervalo.

Solución.

1. La tasa Bruta de Natalidad se define como el cociente entre el total de nacimientos registrados durante un cierto año de mujeres residentes de un determinado ámbito y la población media de ese ámbito en dicho periodo. En general se expresa por mil habitantes.

$$\begin{aligned} TBN^t =& \frac{N^t}{P^t_{med}} = \\ =& \frac{22+230+135+18-5}{551+5461+5187+5945+6376+5880+10600} \cdot 1000 = \\ =& \frac{400}{40000} \cdot 1000 = 10‰ \end{aligned}$$

2. La tasa específica de mortalidad de menores de 1 año se calcula como el cociente entre el total de defunciones de residentes menores de un año registradas durante un cierto año y la población media de menores de un año. En general se expresa por mil habitantes.

$$TEM^t_{<1} = \frac{D^t_{<1}}{P^t_{<1}} \cdot 1000 = \frac{4}{551} \cdot 1000 = 7,259‰$$

3. La tasa de mortalidad infantil se calcula como el cociente entre el total de defunciones de residentes menores de un año registradas durante un cierto año y el total de nacidos vivos de madre residente en ese año determinado. En este caso, sobre el total de nacidos vivos es necesario restar 5 nacimientos que fueron de madres no residentes. En general se expresa por mil nacimientos.

$$TMI^t = \frac{D^t_{<1}}{NV^t} \cdot 1000 = \frac{4}{22+230+135+18-5} \cdot 1000 = \frac{4}{400} \cdot 1000 = 10‰$$

4. La tasa específica de fecundidad de las mujeres de 20 a 29 años será el cociente entre el número de nacimientos de mujeres residentes en un determinado ámbito durante un cierto año y el total de mujeres de dicho colectivo poblacional en ese año y ese ámbito determinados. En general se expresa por mil mujeres.

$$TEF^t_{20-29} = \frac{N^t_{20-29}}{M^t_{20-29}} \cdot 1000$$

Para su cálculo es necesario conocer el número de mujeres de entre 20 y 29 años. Se tiene que el total de la población de dicho rango es 5945 y que la razón de masculinidad del 103. Utilizando la definición de razón de masculinidad, se tiene:

$$Razón\ masc^t = \frac{Población^t_{Hombres}}{Población^t_{Mujeres}} = \frac{Población^t_{Hombres}}{Población^t_{total} - Población^t_{Hombres}}$$

$$103 = \frac{Pob^t_{Homb}}{Pob^t_{Total} - Pob^t_{Homb}} \cdot 100$$

$$1,03 \left(Pob^t_{Total} - Pob^t_{Homb}\right) = Pob^t_{Homb} \Rightarrow 1,03 Pob^t_{Total} = 2,03 Pob^t_{Homb}$$

Resolviendo:

$$Pob^t_{Total} = \frac{2,03}{1,03} Pob^t_{Homb} \Rightarrow Pob^t_{Homb} = \frac{5945}{2,03/1,03} = 3016,42.$$

Consecuentemente,

$$Pob^t_{Muj} = Pob^t_{total} - Pob^t_{Homb} = 5945 - 3016,42 = 2928,57.$$

Finalmente, utilizando el dato de la población femenina de entre 20 y 29 años, se calcula la tasa especifícia de fecundidad.

$$TEF^t_{20-29} = \frac{230}{2928,57} \cdot 1000 = 78,53‰$$

Cuestión 10

En un país con dos regiones, A y B, se ha construido la siguiente matriz de migraciones (datos en número de personas), a partir de la información de los censos de 2001 y 2011.

Tabla 3.16: Matriz de migraciones en las regiones A y B.

Territorio de residencia en 2001	Territorio de residencia en 2011		
	Región A	Región B	Total (A+B)
Región A	5.400.000	300.000	5.700.000
Región B	250.000	7.500.000	7.750.000
En el extranjero	80.000	100.000	180.000
No aplicable (*)	130.000	180.000	310.000
Total	5.860.000	8.080.000	13.940.000

(*): "No aplicable" recoge la población nacida en el periodo intercensal.

A partir de esa matriz, calcule:

a) El número total de emigrantes durante el periodo 2001-2011 y el número total de inmigrantes durante el periodo 2001-2011.

b) La proporción de emigración y la tasa de emigración de la región B durante el periodo 2001-2011.

c) El índice de atracción de la región B durante el periodo 2001-2011.

d) El saldo migratorio y las tasas de migración bruta y neta de la región B durante el periodo.

Solución.

a) El número total de emigrantes es la suma de los emigrantes de la región A y los emigrantes de la región B. Así,

$$M_{AB} + M_{BA} = 300000 + 250000 = 550000$$

El número total de inmigrantes es la suma de los inmigrantes nacionales e internacionales, esto es:

$$M_{AB} + M_{BA} + extranjeros_A + extranjeros_B =$$

$$300000 + 250000 + 80000 + 100000 = 730000$$

b) La proporción de emigración de la región B durante el perido 2001 - 2011 es la relación entre el número de emigrantes de la región B y la población de origen al inicio del periodo. Por otra parte, la tasa de emigración relaciona el número de emigrantes del área B en el periodo 2001 - 2011 con la población media de dicho área durante dicho periodo. Serán respectivamente:

$$PE_B = \frac{E_B}{P_B^{t-n}} = \frac{250000}{7750000} = 0,032$$

$$e_B = \frac{E_B}{n\frac{P_B^{t-n}+P_B^t}{2}} \cdot 1000 = \frac{250000}{10\frac{7750000+8080000}{2}} \cdot 1000 = 3,15‰$$

c) El índice de atracción de la región B durante el periodo 2001-2011 relaciona el número de inmigrantes de la región B en el periodo estudiado con la población media de dicho área durante dicho periodo. El número de inmigrantes de la región B se calcula como la suma de los inmigrantes nacionales e internacionales en la región en estudio. Los inmigrantes nacionales son aquellos cuyo territorio de residencia en 2001 era la región A mientras que en 2011 es la región B. Por otra parte, los inmigranes internacionales son aquellos cuyo terriorio de residencia en 2001 era el extranjero pero en 2011 viven en la región B. Esto es:

$$I_B = M_{AB} + extranjeros_B = 300000 + 100000 = 400000$$

Consecuentemente, el índice de atracción de la región B será:

$$i_B = \frac{I_B}{n\frac{P_B^{t-n}+P_B^t}{2}} \cdot 1000 = \frac{400000}{10\frac{7750000+8080000}{2}} \cdot 1000 = 5,05‰$$

d) El saldo migratorio y las tasas de migración bruta y neta de la región B durante el periodo serían:

$$sm_B = I_B - E_B = 400000 - 250000 = 150000$$

$$tmb_B = i_B + e_B = 5,05 + 3,15 = 8,2‰$$

$$tmn_B = i_B - e_B = 5,05 - 3,15 = 1,9‰$$

Capítulo 4

Año 2017

"All mathematicians share a sense of amazement over the infinite depth and mysterious beauty and usefulness of mathematics."

—Martin Gardner

Maryam Mirzakhani

Se graduó en Matemáticas en 1999 en la Universidad de Tecnología Sharif de Teherán. En 2004 se doctoró en la Universidad de Harvard. Desarrolló su carrera en los campos del espacio de Teichmüller, la geometría hiperbólica, la teoría ergódica y la geometría simpléctica. Tras hacer su tesis en la Universidad de Harvard, trabajó como investigadora en el Instituto Clay de Matemáticas y en la Universidad de Princeton.

Fue investigadora en la Universidad de Stanford (EE. UU.). Sus estudios abarcan impactantes y originales investigaciones sobre geometría y sistemas dinámicos. Su trabajo en superficies de Riemann y sus modelos espaciales conectan varias disciplinas matemáticas muy diversas e influyen en todas ellas. Fue profesora de matemáticas en la Universidad de Stanford desde septiembre de 2008 hasta su fallecimiento el 15 de julio de 2017, tras luchar 4 años contra un cáncer de mama.

En 2014 se convirtió en la primera mujer galardonada con la Medalla Fields.

$2017 = 1008 + 1009$, so it is a polite number.

2017 can be written as a sum of positive squares in only one way, i.e.,

$2017 = 1936 + 81 = 44^2 + 9^2$

4.1. Soluciones del cuarto examen del Cuerpo Superior de Estadísticos del Estado de 2017

Cuestión 1

Dos muestras independientes, ambas de tamaño 7, de sendas distribuciones poblacionales normales de varianza común σ^2 arrojan medias muestrales de 4.8 y 5.4, y varianzas muestrales de 8.38 y 7.62, respectivamente. Encontrar la expresión que permitiría obtener un intervalo de confianza para $\mu_1 - \mu_2$ al nivel 0.95.

Solución.

Teniendo en cuenta que se trata de un caso de intervalo de confianza para diferencia de medias con poblaciones normales con varianza desconocidas pero iguales el estadístico:

$$\frac{\bar{x} - \bar{y} - (\mu_1 - \mu_2)}{\sqrt{\frac{1}{n} + \frac{1}{m}}\sqrt{\frac{nS_1^2 + mS_2^2}{n+m-2}}} \sim t_{n+m-2}$$

y el intervalo de confianza para $\mu_1 - \mu_2$ vendrá dado por:

$$\mathrm{IC}_{\mu_1-\mu_2}^{1-\alpha} = \left[\bar{x} - \bar{y} \pm t_{n+m-2;\frac{\alpha}{2}}\sqrt{\frac{1}{n} + \frac{1}{m}}\sqrt{\frac{nS_1^2 + mS_2^2}{n+m-2}}\right]$$

siendo S_i^2 las varianzas muestrales de cada muestra para $i = 1, 2$. Del enunciado se tiene:

$$n = m = 7$$

$$\overline{x} = 4,8; S_1^2 = 8,38$$

$$\overline{y} = 5,4; S_2^2 = 7,62$$

de forma que

$$\mathrm{IC}^{1-\alpha}_{\mu_1-\mu_2} = \left[-0,6 - \frac{2\sqrt{6}}{3} t_{12;\frac{0,05}{2}}, -0,6 + \frac{2\sqrt{6}}{3} t_{12;\frac{0,05}{2}}\right]$$

$$\mathrm{IC}^{1-\alpha}_{\mu_1-\mu_2} = [-4,158, 2,958]$$

Cuestión 2

Se supone que el número de erratas por página de un libro sigue una distribución de Poisson. Elegidas al azar 95 páginas se obtuvo que había 0, 1, 2, 3, 4, 5 en 40, 30, 15, 7, 1 y 0 páginas respectivamente. ¿Contiene la muestra evidencias estadísticamente significativas para rechazar dicho supuesto? Justifica la respuesta.

Solución.

Sea $X \equiv$ "número de erratas por página". Se quiere comprobar si X sigue una distribución de Poisson, es decir, que $X \sim P(\lambda)$. Dado que la esperanza de la variable es λ, la media de la muestra de valores de la variable puede tomarse como una aproximación a λ. Además se trata de su estimador máximo verosímil y el que se obtiene por el método de los momentos. Así,

$$\widehat{\lambda} = \frac{0 \cdot 40 + 1 \cdot 30 + 2 \cdot 15 + 3 \cdot 7 + 4 \cdot 1 + 5 \cdot 0}{95} = 0,89$$

Por tanto,

$$P(X=0) = \frac{e^{-\lambda}\lambda^0}{0!} = e^{-\lambda} = e^{-0,89} = 0,4106$$

$$P(X=1) = \frac{e^{-\lambda}\lambda^1}{1!} = e^{-0,89} \cdot 0,89 = 0,3654$$

$$P(X=2) = 0,1626$$

$$P(X=3) = 0,0482$$

$$P(X=4) = 0,0107$$

$$P(X=5) = 0,00191$$

Tabla 4.1: Frecuencias teóricas y frecuencias observadas

Nº de erratas	0	1	2	3	4	5
Frecuencia teórica $n \cdot p_i^0$	39,01	34,72	15,45	4,58	1,019	0,181
Frecuencia observada n_i	40	30	15	7	1	0

Dado que en las últimas 3 clases $n \cdot p_i^0 < 5$, se acumulan de forma que:

Tabla 4.2: Frecuencias agrupadas

Nº de erratas	0	1	2	3
Frecuencia teórica $n \cdot p_i^0$	39,01	34,72	15,45	5,79
Frecuencia observada n_i	40	30	15	8

Se utiliza un contraste χ^2 de bondad de ajuste, el cuál estudia la hipótesis nula $H_0 : F_X = F_0$, con F_0 la función de distribución de una Poisson con parámetro λ y se tiene:

$$\chi^2 = \sum_{i=1}^{4} \frac{(n_i - n \cdot p_i^0)^2}{n \cdot p_i^0} = \frac{(40 - 39,01)^2}{39,01} + \cdots + \frac{(8 - 5,79)^2}{5,79} = 1,523$$

El número de grados de libertad del estadístico de contrastes será $\chi^2_{k-q-1} = \chi^2_{4-1-1} = \chi^2_2$, donde $q = 1$ pues se ha estimado el parámetro de la distribución de Poisson a partir de la muestra. El valor proporcionado por la tabla de Chi Cuadrado para el nivel de significación habitual $\alpha = 0,05$ es 5,9915. Dado que este valor es mayor que el obtenido a partir de la muestra $(1,523)$ no se rechaza que el número de erratas siga una distribución de Poisson.

Cuestión 3

Sea una población de 3245 empresas dedicadas al sector de comercio estratificadas en 3 grupos de tamaño según el número de asalariados. Los datos disponibles están recogidos en la tabla 4.3.

Tabla 4.3: Datos disponibles

Estrato	N_h	X_h	S_h
1 (entre 0 y 9 asalariados)	3200	3400	1,5
2 (entre 10 y 49 asalariados)	40	650	10,2
3 (de 50 o más asalariados)	5	580	150

Donde

N_h: número de empresas en el estrato h

X_h: total de asalariados en el estrato h

S_h: cuasivarianza poblacional de X en el estrato h

El tamaño muestral es de 50 empresas y el muestro es aleatorio simple sin reposición. Se pide:

a) Calcular el tamaño muestral por estrato, usando la afijación de mínima varianza o de Neyman.

b) Calcular la varianza y el coeficiente de variación para el estimador insesgado del total de asalariados.

Solución.

Según la afijación de mínima varianza el tamaño de cada estrato será:

$$n_h = n\frac{N_h S_h}{\sum_{h=1}^{L} N_h S_h}$$

Así se obtiene: $n_1 = 40$, $n_2 = 4$ y $n_3 = 6$. Sin embargo, se observa que el número de unidades a seleccionar para la muestra en el tercer estrato es superior al número de unidades de dicho estrato. Ante esta circunstancia, seleccionamos para la muestra las 5 unidades del tercer estrato, es decir, $n_3 = 5$, y las unidades del tercer estrato son autorrepresentadas. El resto de 45 unidades restantes de la muestra han de repartirse mediante la afijación de mínima varianza de nuevo:

$$n_1 = 45\frac{N_1S_1}{N_1S_1 + N_2S_2} = 41,47 \simeq 41$$
$$n_2 = 45\frac{N_2S_2}{N_1S_1 + N_2S_2} = 3,52 \simeq 4$$

Luego la afijación de mínima varianza proporciona los siguientes tamaños muestrales para cada estrato: $n_1 = 41$, $n_2 = 4$ y $n_3 = 5$. Dado que en el último estrato se toma la población completa, la aportación de este estrato a la varianza será nula. Así, siendo $f_h = \frac{n_h}{N_h}$,

$$\begin{aligned} Var\left(\widehat{X}_{st}\right) &= \sum_{h=1}^{L} N_h^2 \left(1 - f_h\right) \frac{S_h^2}{n_h} \\ &= 3200^2 \left(1 - \frac{41}{3200}\right) \frac{1,5^2}{41} + 40^2 \left(1 - \frac{4}{40}\right) \frac{10,2^2}{4} + 0 = \\ &= 554751,219 + 37454,4 + 0 = 592205,619 \end{aligned}$$

Por tanto, el error de muestreo será

$$\sqrt{Var\left(\widehat{X}_{st}\right)} = \sqrt{592205,619} = 769,549$$

Por otra parte, siendo $X = \sum_{h=1}^{3} X_h = 3400 + 650 + 580 = 4630$, el coeficiente de variación será

$$\mathrm{CV}\left(X_{st}\right) = \frac{\sqrt{Var\left(\widehat{X}_{st}\right)}}{X} = \frac{769,549}{4630} = 0,166$$

Cuestión 4

Se quiere conocer la superficie dedicada a la plantación de pinos de una región. Dicha región, con un total de 50.000km^2, se divide en 100 áreas o conglomerados. Se extrae una muestra de 5 conglomerados con reemplazamiento y con probabilidades proporcionales a sus superficies. Las proporciones de superficie dedicadas a la plantación de pinos, en cada uno de los conglomerados de la muestra se recogen en la tabla 4.4. Se pide calcular un estimador insesgado de la superficie total dedicada a la plantación de pinos y su error de muestreo relativo o coeficiente de variación.

Tabla 4.4: Proporciones de superficie dedicadas a la plantación de pinos.

Conglomerado	**1**	**2**	**3**	**4**	**5**
Proporción	**0,05**	**0,2**	**0,1**	**0,15**	**0,25**

Solución.

Se tienen como datos: $N = 100$, $M = 50000\ km^2$ y $n = 5$. Se pide calcular un estimador insesgado de la superficie total dedicada a la plantación de pinos, $\widehat{X}$. Dado que se trata de un muestreo de conglomerados con reposición y probabilidades proporcionales a los tamaños, el estimador vendrá dado por:

$$\widehat{X} = \sum_{i=1}^{n} \frac{X_i}{nP_i} = \sum_{i=1}^{n} \frac{X_i}{n\frac{M_i}{M}} = \frac{M}{n}\sum_{i=1}^{n} \frac{X_i}{M_i} = \frac{M}{n}\sum_{i=1}^{n} \overline{X_i} = M\overline{\overline{x}} =$$
$$= 50000\frac{0,05+0,2+0,1+0,15+0,25}{5} = 50000 \cdot 0,15 = 7500\ km^2$$

El coeficiente de variación será:

$$\widehat{CV}\left(\widehat{X}\right) = \frac{\sqrt{\widehat{Var}\left(\widehat{X}\right)}}{\widehat{X}}$$

donde

$$\widehat{Var}\left(\widehat{X}\right) = \frac{1}{n\left(n-1\right)}\sum_{i=1}^{n}\left(\frac{X_i}{P_i} - \widehat{X}\right)^2 = \frac{1}{n\left(n-1\right)}\sum_{i=1}^{n}\left(\frac{X_i}{M_i/M} - \widehat{X}\right)^2 =$$

$$= \frac{1}{n(n-1)} \sum_{i=1}^{n} \left(M\overline{X}_i - \widehat{X} \right)^2 = \frac{M^2}{n(n-1)} \sum_{i=1}^{n} \left(\overline{X}_i - \overline{\overline{x}} \right)^2 =$$
$$= \frac{50000^2}{5(5-1)} \left[(0,05-0,15)^2 + (0,2-0,15)^2 + \cdots + (0,25-0,15)^2 \right]$$
$$= \frac{50000^2}{20} 0,025 = 3125000.$$

De modo que

$$\widehat{CV}\left(\widehat{X}\right) = \frac{\sqrt{3125000}}{7500} = 0,2357 \simeq 23,6\,\%$$

Cuestión 5

En una determinada economía, para la rama de actividad A, se tiene la información recogida en la tabla 4.5. Por otra parte el PIBp.m. a precios corrientes de esa economía en el año t fue de 1.118.000 millones de euros; la tasa de variación anual de este agregado a precios corrientes entre t y t+1 fue de 4,0 % mientras que su tasa de variación anual en volumen fue del 3,1 %. Se pide:

a) Calcule la tasa de variación del valor añadido (pb) a precios corrientes de la rama de actividad A entre t y t+1. Obtenga asimismo la tasa de variación en volumen del valor añadido (pb) entre t y t+1. Nota: p.b. (precio básico), p.a. (precio de adquisición), p.m. (precios de mercado)

b) Calcule el valor del PIB (pm) en t+1 a precios corrientes y la variación anual entre t y t+1 del deflactor implícito del PIB (pm).

Tabla 4.5: Información relativa a la rama de actividad A

	año t	año t+1
Producción (p.b.) a precios corrientes (millones de euros)	450.000	482.000
Consumos intermedios (p.a.) a precios corrientes (millones de euros)	275.000	292.000
Índice de precios de la producción de la rama (en %)	100	105
Índice de precios de los consumos intermedios de la rama (en %)	100	103

Solución.

El valor añadido bruto se registra a precios básicos y es igual a la producción valorada a precios básicos menos los consumos intermedios valorados a precios de adquisición, por tanto:

$$VAB_A^t = P_{pb}^t - CI_{pa}^t = 450000 - 275000 = 175000 \text{ millones de €}$$

$$VAB_A^{t+1} = P_{pb}^{t+1} - CI_{pa}^{t+1} = 482000 - 292000 = 190000 \text{ millones de €}.$$

De forma que la tasa de variación a precios corrientes será

$$TVI\,(VAB_A)_t^{t+1} = \frac{VAB_{A,t+1} - VAB_{A,t}}{VAB_{A,t}} = \frac{190000}{175000} - 1 = 0,0857 \simeq 8,6\,\%$$

Para obtener la tasa de variación en volumen, deflactamos por los índices de precios correspondientes a cada agregado:

Tabla 4.6: Datos para calcular el $TVI\,(VAB_{A,volumen})_t^{t+1}$ en volumen.

	t	t+1
$\text{Producción}_{Volumen} = \frac{\textit{producción}}{\textit{Índice Precios Producción}}$	450000	$\frac{482000}{1{,}05} = 459047,62$
$\text{CI}_{Volumen} = \frac{CI}{\textit{Índice Precios CI}}$	275000	$\frac{292000}{1{,}03} = 283495,15$

de modo que:

$$VAB_{A,valor}^{t+1} = 459047,62 - 283495,15 = 175552,47 \text{ millones de €.}$$

$$TVI\,(VAB_{A,volumen})_t^{t+1} = \frac{VAB_{A,t+1} - VAB_{A,t}}{VAB_{A,t}} = \frac{175552,5}{175000} - 1 =$$

$$= 0,00315 \simeq 0,31\,\%$$

Para el cálculo del valor del PIB (pm) en t+1 a precios corrientes y la variación anual entre t y t+1 del deflactor implícito del PIB (pm), se sabe que $PIB_t = 1118000$ y además $TVI(PIB)_t^{t+1} = 4\,\%$ donde

$$TVI(PIB)_t^{t+1} = \frac{PIB_{t+1} - PIB_t}{PIB_t}$$

por lo que

$$0,04 = \frac{PIB_{t+1} - PIB_t}{PIB_t} \Leftrightarrow 1,04 \cdot PIB_t = PIB_{t+1}.$$

Así,

$$PIB_{t+1} = 1,04 \cdot 1118000 = 1162720$$

Por otra parte, se sabe que $TVI(PIBvolumen)_t^{t+1} = 3,1\,\%$ donde:

$$TVI(PIBvolumen)_t^{t+1} = \frac{\frac{PIB_{t+1}}{IPC_{t,t+1}} - \frac{PIB_t}{IPC_{base}}}{\frac{PIB_t}{IPC_{base}}}$$

de modo que

$$0,031 = \frac{\frac{PIB_{t+1}}{IPC_{t,t+1}} - \frac{PIB_t}{100}}{\frac{PIB_t}{100}} \Leftrightarrow \frac{PIB_{t+1}}{IPC_{t,t+1}} = 1,031 \cdot PIB_t.$$

Se tiene pues:

$$
\begin{aligned}
IPC_{t,t+1} &= \frac{PIB_{t+1}}{1,031 \cdot PIB_t} = \frac{1162720}{1,031 \cdot 1118000} = \frac{1,04 \cdot 1118000}{1,031 \cdot 1118000} = \\
&= \frac{1,04}{1,031} = 1,008720 \simeq 101
\end{aligned}
$$

Cuestión 6

Se dispone de la información respecto a los agregados de oferta y rentas de una economía en miles de millones de euros, recogidos en la tabla 4.7.

Tabla 4.7: Información de los agregados de oferta y rentas (miles de millones de euros)

	Código SEC 2010	
Producción de bienes y servicios (a precios básicos)	P.1	2.000.000
Consumo intermedio (a precios de adquisición)	P.2	1.020.000
Impuestos sobre la producción y las importaciones	D2	112.000
Impuestos sobre los productos	D.21	90.000
Subvenciones	D.3	18.400
Subvenciones sobre los productos	D.31	6.000
Excedente de explotación de renta mixta	B.2.G+B.3.G	445.000

Por otra parte, de la cuenta del resto del mundo se obtienen los datos (en millones de euros) recogidos en la tabla 4.8.

Tabla 4.8: Información de los agregados de la cuenta del resto del mundo (miles de millones de euros)

	Cod. SEC 2010	
Remuneración de asalariados recibida del resto del mundo	D.1	1.100
Remuneración de asalariados pagada al resto del mundo	D.1	300
Impuestos sobre la producción e importaciones pagados al resto del mundo	D.2	1.500
Subvenciones recibidas del resto del mundo	D.3	6.000
Otras transferencias corrientes recibidas del resto del mundo	D.7	2.000
Rentas de la propiedad recibidas del resto del mundo	D.4	45.000
Rentas de la propiedad pagadas al resto del mundo	D.4	66.000
Otras transferencias corrientes pagadas al resto del mundo	D.7	1.500

Calcule a partir de esta información, el valor añadido total a precios básicos, el PIB a precios de mercado (PIBpm), la remuneración de los asalariados y la renta nacional bruta de la economía.

Solución.

$$VAB_{pb} = \textit{Producción}_{pb} - CI_{padq} = 2000000 - 1020000 = 980000 \text{ millones de €}$$

$$\begin{aligned} PIB_{pm} &= VAB_{pb} + \textit{impuestos netos sobre los productos} = \\ &= VAB_{pb} + \text{D.21} - \text{D.31} = 980000 + 90000 - 6000 = \\ &= 1064000 \text{ millones de €} \end{aligned}$$

$$\begin{aligned} PIB_{pm} &= RA_{interior} + EBE/RM + \textit{impuestos netos sobre la producción y} \\ &\quad \textit{las importaciones} = RA_{interior} + EBE/RM + \text{D.2} - \text{D.3} \end{aligned}$$

$$\Rightarrow RA_{interior} = PIB_{pm} - EBE/RM - (\text{D.2} - \text{D.3})$$

$$\Rightarrow RA_{interior} = 1064000 - 445000 - (112000 - 18400)$$

$$\Rightarrow RA_{interior} = 525400 \text{ millones de €}$$

$$RNB = PIB_{pm} + rentas\ de\ la\ propiedad\ con\ el\ RM\ netas$$

$$= 1064000 - (300 + 66000 + 1500) + (1100 + 45000 + 6000) =$$

$$= 1064000 - 67800 + 52100 = 1048300 \text{ millones de €}$$

En el cálculo de la RNB, al tener datos de la cuenta del resto del mundo, la interpretación es que, por ejemplo, para D.1, que es la Remuneración de asalariados pagada al resto del mundo, desde el punto de vista de nuestra economía, es un agregado que paga nuestra economía, y por tanto se resta al PIB_{pm}. Por otra parte, conceptos como el D.3, que son las Subvenciones recibidas del resto del mundo, desde el punto de vista de nuestra economía, es una cantidad que nuestra economía recibe del resto del mundo, por lo tanto se suma al PIB_{pm}.

Cuestión 7

En un Estado (E) de nuestro interés acontece que los accidentes de tráfico son la causa principal de muerte para jóvenes entre 5 y 32 años. A través de varias políticas de gasto, el gobierno central hace un tiempo promulgó una ley de uso obligatorio del cinturón de seguridad, y así reducir el número de muertes y de lesiones graves. Su departamento de análisis econométrico ubicado en el INE está interesado en examinar cómo de efectiva ha sido esta ley. Para ello tiene disponible un conjunto de datos de panel de 50 unidades político administrativas del país (provincias por ejemplo) para los años 1983 – 1997. Su conjunto de datos incluye las siguientes variables:

- *fatalidades* es el número de muertes por miles de millas de tráfico
- CS_{uso} es el nivel de uso del cinturón de seguridad
- *velo90* es una dummy = 1 si límite de velocidad es de 90 kms por hora, =0 al contrario
- *velo120* es una dummy = 1 si límite de velocidad es de 120 kms por hora, =0 al contrario
- *ba08* es una dummy = 1 si límite de alcohol en la sangre es $\leq$ 0,08 %, =0 al contrario
- *beber - edad18* es una dummy = 1 si tiene 18 años de edad para beber, =0 al contrario
- *ing* es el ingreso per cápita
- *edad* media de edad
- *provincia* es un conjunto de dummies de provincias
- *años* es un conjunto de dummies de años

La siguiente tabla contiene los resultados de varias regresiones (MCO agrupados, MCO con efectos fijos de estado, MC generalizados con efectos aleatorios de estado y MCO con efectos fijos de estado y año). Las siguientes preguntas se basan en estos resultados.

	1	2	3	4
CS_uso	4,07 (1,22)	−5,77 (1,215)	−4,50 (1,12)	−3,72 (1,13)
$velo90$	0,148 (0,403)	−0,425 (0,334)	−0,341 (0,337)	−0,783 (0,424)
$velo120$	2,40 (0,511)	1,23 (0,329)	1,34 (0,328)	0,804 (0,340)
$ba08$	−1,92 (0,445)	−1,38 (0,373)	−1,36 (0,367)	−0,822 (0,352)
$edad - beber18$	0,079 (0,876)	0,745 (0,507)	0,767 (0,510)	−1,13 (0,535)
$ln(ing)$	−18,1 (0,931)	−13,5 (1,42)	−12,6 (1,14)	6,26 (3,86)
$edad$	−0,007 (0,109)	0,979 (0,382)	0,232 (0,239)	1,32 (0,383)
$constante$	196,5 (8,22)		137,9 (8,92)	
Efectos provincia	No	FE	RE	FE
Efectos años	No	No	No	FE
R^2	0.544	0.874	0.683	0.897

FE (Efectos fijos); RE(Efectos aleatorios)

1. Centrándonos en los resultados de la regresión MCO de datos agrupados (fusionados) en la columna 1,

   a) ¿la regresión estimada sugiere que un mayor uso del cinturón de seguridad reduce significativamente las muertes?

   b) Indique si este resultado tiene sentido. Si es así, explica por qué. Si no, explica qué piensas que está pasando aquí.

2. Las columnas 2 y 3 contemplan las estimaciones por efectos fijos de provincia, efectos aleatorios de provincia.

   a) ¿Qué conclusión puede sacarse con respecto al impacto de nivel de uso del cinturón de seguridad en las muertes cuando agregamos los efectos fijos?

3. Los resultados de la prueba de Hausman para los modelos de las columnas 3 y 4 son los siguientes.

	Coeficientes (b) fijos	(B) .	(b-B) diferencia	$(\text{diag(var\_(b)-var\_(B))})^{-1}$ S.E.
CB_uso	-5.774782	-4.503969	-1.270813	0.2698476
velo90	-0.4250387	-0.3405939	-0.0844448	0.0644951
velo120	1.23329	1.335134	-0.1018444	0.0209829
ba08	-1.377456	-1.364296	-0.0131597	0.0648635
edad-beber18	0.7453195	0.766994	-0.0216745	.
ln(ing)	-13.5144	-12.61544	-0.8989608	0.8380231
edad	0.9786802	0.2318357	0.7468445	0.2976956

b=consistente bajo H_0 y H_a;
B=inconsistente bajo H_a, eficiente bajo H_0;
Test: H_0: discrepancias de coeficiente no sistemáticas

$$\begin{aligned} chi2(7) &= (b-B)'[(var(b)-var(B)^{-1}](b-B) \\ &= 26{,}62 \\ Prob > chi2 &= 0{,}0004 \end{aligned}$$

Indica qué conclusión principal puedes sacar respecto de la información suministrada de cara a plantear el modelo más adecuado.

4. El modelo en la columna 4 agrega efectos fijos anuales al modelo de la columna 2. Una prueba F de la significación conjunta de estas dummies de 14 años nos da el estadístico F de 8,85. ¿Qué concluye acerca de la significación conjunta de las dummies del año? ¿Los resultados con respecto al impacto del uso del cinturón de seguridad cambian ahora que hemos agregado EF de los años?

Solución.

En este ejercicio se trabaja con varios ajustes con diferentes modelos para un conjunto de datos de sección cruzada de los que se disponen observaciones en un conjunto de años. El primero de los modelos agrupa todas las observaciones y se estima una regresión sin tener en cuenta la naturaleza de corte transversal y de series de tiempo de los datos. Esto se conoce como MCO agrupados y el modelo general en un caso así sería:

$$y_{it} = \mu + \beta_1 x_{1it} + \cdots + \beta_k x_{kit} + u_{it} = \mu + \boldsymbol{x}_{it}^T \boldsymbol{\beta} + u_{it}$$

, donde i es la i-ésima provincia, $i = 1, ..., N$, y t denota el año de observación, $t = 1, ..., T$, de manera que en nuestro caso la estimación del mismo ha sido:

$$\begin{aligned}
\widehat{fatalidades}_{it} &= 196,5 + 4,07CS_uso_{it} + 0,148velo90_{it} + 2,40velo120_{it} + \\
&- 1,92ba08_{it} + 0,079\textit{edad-beber18}_{it} - 18,1ln(ing)_{it} - \\
&- 0,007edad_{it} \\
&\forall i = 1, \ldots, 50;\ \forall t = 1, \ldots, 15.
\end{aligned}$$

En este modelo, los coeficientes de regresión son iguales para todas las provincias, es decir, que tendrían un comportamiento idéntico, que es un supuesto seguramente difícil de sostener. También estamos suponiendo que el término de error $u_{it} \sim iid(0, \sigma_u^2)$ tiene media cero y varianza constante. El problema que puede tener este planteamiento es que no distingue entre provincias ni indica si la variable respuesta a través del tiempo es la misma o no se comporta igual. Es decir, se oculta de alguna forma la posible heterogeneidad que puede existir entre dichas zonas, y por tanto es posible que el término de error se correlacione con algunas de las variables regresoras del modelo. Si esto fuera así, los coeficientes estimados en la ecuación pueden estar sesgados y ser inconsistentes.

Para tener en cuenta este tipo de efectos no observables se puede utilizar la estrategia de utilizar un modelo de mínimos cuadrados de efectos fijos de provincia en el que se utilizan variables dicótomas. Este modelo sería el de la columna segunda:

$$\begin{aligned}
\widehat{fatalidades}_{it} &= -5,77CS_uso_{it} - 0,425velo90_{it} + 1,23velo120_{it} + \\
&- 1,38ba08_{it} + 0,745\textit{edad-beber18}_{it} - 13,5ln(ing)_{it} +
\end{aligned}$$

$$
\begin{aligned}
&+ 0,979 edad_{it} + \widehat{\alpha}_1 provincia1_i + \widehat{\alpha}_2 provincia2_i + \cdots + \\
&+ \widehat{\alpha}_{50} provincia50_i \\
&\forall i = 1, \ldots, 50;\ \forall t = 1, \ldots, 15.
\end{aligned}
$$

donde $\widehat{\alpha}_i$ es una constante distinta para cada provincia. La expresión general de este modelo sería:

$$y_{it} = \mu + \beta_1 x_{1it} + \cdots + \beta_k x_{kit} + \alpha_i + u_{it} = \mu + \boldsymbol{x}_{it}^T \boldsymbol{\beta} + \alpha_i + u_{it}$$

siendo $i = 1, ..., N$, y $t = 1, ..., T$, y el supuesto fundamental que se hace es que $E\left[u_{it}|\alpha_i, x_{1it}, \ldots, x_{kit}\right] = 0$, los regresores deben estar incorrelados con u_{it}.

El término efectos fijos se corresponde con el hecho de que el término constante puede ser diferente para cada provincia, aunque dicho intercepto no varía con el tiempo. Para conseguir este efecto se utilizan lo que se llaman variables dicótomas y se utilizan tantas como provincias porque no hay término independiente en el modelo. En nuestro caso, por ejemplo, $provincia1_i$ tomaría el valor 1 para la provincia 1 y el valor 0 en otro caso, de manera que el coeficiente estimado $\widehat{\alpha}_1$ sería el coeficiente asociado a la provincia número 1.

También podríamos permitir un efecto tiempo si pensamos que como consecuencia del tiempo y de ciertos avances en las tecnología o en la legislación puede afectar el comportamiento de nuestra variable dependiente. Tales efectos también se podrían tener en cuenta con facilidad introduciendo variables dicótomas temporales, una para cada año del periodo en estudio. En el modelo de la columna 4 tendríamos esta situación, sin término independiente y por tanto teniendo que utilizar 15 variables binarias. Esto haría que haya que estimar más coeficientes y se consumen grados de libertad, quedando el modelo estimado siguiente:

$$\begin{aligned}
\widehat{fatalidades}_{it} = {} & -3,72 CS_uso_{it} - 0,783 velo90_{it} + 0,804 velo120_{it} + \\
& - 0,822 ba08_{it} - 1,13 edad\text{-}beber18_{it} - 6,26 ln(ing)_{it} + \\
& + 1,32 edad_{it} + \widehat{\alpha}_1 provincia1_i + \widehat{\alpha}_2 provincia2_i + \cdots + \\
& + \widehat{\alpha}_{50} provincia50_i + \widehat{\gamma}_1 año1_t + \widehat{\gamma}_2 año2_t + \cdots + \widehat{\gamma}_{15} año15_t \\
& \forall i = 1, \ldots, 50;\ \forall t = 1, \ldots, 15.
\end{aligned}$$

donde $\widehat{\alpha}_i$ es una constante distinta para cada provincia y $\widehat{\gamma}_t$ es una constante distinta para cada año.

El problema que tiene este tipo de modelos de efectos fijos es que si se introducen muchas variables dicótomas puede haber problemas de falta de grados de libertad, es decir, que no haya observaciones suficientes para realizar el análisis, y además pueden presentarse también problemas de multicolinealidad. A veces también puede ocurrir que estos modelos no identifiquen el efecto de aquellas variables que no se ven modificadas con el paso del tiempo. Es conveniente también prestar atención al término de error u_{it}, que suponemos que cumple los supuestos clásicos, es decir, que sigue una normal de media cero y varianza constante, pero quizá deba modificarse la suposición sobre el mismo. Puede suponerse que la varianza del error es constante para cada provincia, o bien que es heteroscedástica. También podría suponerse que para cada provincia no existe autocorrelación con el tiempo, y hay más opciones, pero el análisis se complica y vamos a analizar algunas de ellas con los diversos modelos que se proponen.

Estas propuestas de efectos fijos quizá podrían reflejar una falta de conocimiento de funcionamiento del modelo real al tener que introducir tantas variables dicótomas, de manera que nos podríamos preguntar si se podría expresar todo esto mediante el término de perturbación, y eso es lo que hace precisamente el modelo de efectos aleatorios, que es el que aparece en la columna 3 del resumen de modelos anterior.

La modelización general de efectos aleatorios se basa en partir de la ecuación:

$$y_{it} = \mu + \boldsymbol{x}_{it}^T \boldsymbol{\beta} + \alpha_i + u_{it} \ , \forall i = 1, \ldots, N; \ \forall t = 1, \ldots, T.$$

En lugar de considerar fija a α_i vamos a suponer que es una variable aleatoria con un valor medio igual a α, sin el subíndice i, es decir, $\alpha_i = \alpha + \epsilon_i$, donde ϵ_i es un error aleatorio con media cero y varianza σ_ϵ^2. Luego el modelo sería:

$$y_{it} = \mu + \boldsymbol{x}_{it}^T \boldsymbol{\beta} + \alpha + w_{it} = \alpha_\mu + \boldsymbol{x}_{it}^T \boldsymbol{\beta} + w_{it} \ , \forall i = 1, \ldots, N; \ \forall t = 1, \ldots, T.$$

$$w_{it} = \epsilon_i + u_{it}$$

Ahora el término de error tiene dos componentes, uno relativo al corte transversal específico de cada provincia, y otro que varía tanto en el tiempo como para cada provincia. Los supuestos de este tipo de modelo son:

$$\epsilon_i \sim N(0, \sigma_\epsilon^2) \ , \ u_{it} \sim N(0, \sigma_u^2)$$

$$E(\epsilon_i u_{it}) = 0; \ E(\epsilon_i \epsilon_j) = 0 \ i \neq j$$

$$E(u_{it} u_{is}) = E(u_{ij} u_{ij}) = E(u_{it} u_{js}) = 0 \ i \neq j; t \neq s$$

, es decir, los errores individuales no están correlacionados entre ellos y no están autocorrelacionados en las unidades de series de tiempo ni en las de corte transversal. Es crucial saber que w_{it} no está correlacionada con ninguna variable explicativa del modelo. En el caso que nos ocupa el modelo estimado con efectos aleatorios de provincia sería:

$$\begin{aligned} \widehat{fatalidades}_{it} = {} & 137,9 - 4,50 CS\_uso_{it} - 0,341 velo90_{it} + \\ & + 1,34 velo120_{it} + -1,36 ba08_{it} + 0,767 \textit{edad-beber18}_{it} + \\ & - 12,6 ln(ing)_{it} + 0,232 edad_{it} \end{aligned}$$

$\forall i = 1, \ldots, 50;\ \forall t = 1, \ldots, 15.$

La diferencia principal entre el modelo de efectos fijos y aleatorios es que en el primero de ellos cada provincia tiene un valor fijo de término independiente, pero en efectos aleatorios el intercepto común representaría el valor medio de todos esos términos independientes de cada provincia, y el componente de error ϵ_i sería la desviación aleatoria de ese término individual respecto a este valor medio. De esta forma tenemos que $E(w_{it}) = 0$ y $Var(w_{it}) = \sigma_\epsilon^2 + \sigma_u^2$, es decir, el término de error es homoscedástico. Se puede demostrar que los términos de error de una provincia o unidad en dos puntos de tiempo están correlacionados, $\rho = corr(w_{it}, w_{is}) = \frac{\sigma_\epsilon^2}{\sigma_\epsilon^2+\sigma_u^2}$, $t \neq s$, luego el método de estimación más adecuado en este caso es el método de mínimos cuadrados generalizados.

El estimador de efectos fijos permite estimar el modelo bajo supuestos menos restrictivos, puesto que permite que haya correlación entre los regresores y los efectos individuales, y permite estimar el modelo incluso si los regresores son endógenos, aunque suele ser menos eficiente al no identificar los coeficientes de regresores que no varíen en el tiempo. Efectos aleatorios es más eficiente si se cumplen supuestos adicionales a los de efectos fijos, pero puede ser inconsistente.

Una vez que hemos aclarado los condicionantes de cada modelo estimado podemos discutir sobre la idoneidad de la utilización de uno u otro.

1. a) En el caso que nos ocupa, la regresión estimada no sugiere que un mayor uso del cinturón de seguridad reduzca significativamente las muertes, sino que un aumento del uso de una unidad en el nivel de uso del cinturón de seguridad, aumentará la muertes en 4,07 unidades por cada 1000 millas. Hay que comentar que realizando un test de la t de Student para comprobar si el coeficiente de esta

variable es significativo es claro que sí que lo es, puesto que la $t_{exp} = \frac{4{,}07}{1{,}22} = 3,36$, que es mayor que el valor crítico de la t, que sería 1,962 y por esta razón podemos afirmar lo anterior.

b) Ya que el uso del cinturón debería disminuir el número de muertes en lugar de aumentarlas, el resultado no parece tener mucho sentido. Esto puede deberse a que existan factores no observables que afecten al modelo. También puede ocurrir que existan variables omitidas, o bien porque estos factores no observables supongan un problema de autocorrelación en el modelo.

2. En las columnas 2 y 3 se observa que el modelo especificado con efectos aleatorios tiene un peor ajuste que el modelo con efectos fijos, es decir, existen efectos individuales que están correlacionados con la provincia de origen de los conductores. Cuando se agregan los efectos fijos, si aumenta el uso del cinturón de seguridad 1 una unidad, las muertes disminuirán 5,77 unidades por miles de millas de tráfico. Si se tienen en cuenta efectos aleatorios, las muertes disminuyen 4,5 unidades por miles de millas de tráfico. La variable del uso del cinturón tiene el signo que sería esperable en ambos casos y se puede comprobar del mismo modo que en el apartado anterior, que utilizando el test de la t el coeficiente es significativo en ambos casos, rechazando por tanto que sea nulo, y con el valor del R^2 intuímos qué modelo se ajusta mejor, pero sería conveniente efectuar la prueba de Hausman para decantarnos por efectos fijos o aleatorios.

3. *A pesar de que el enunciado menciona las columnas 3 y 4, los resultados del test de Hausman proporcionados son relativos a las columnas 2 y 3, por tanto suponemos que se trata de una errata en el enunciado.*

El test de Hausman es un contraste de robustez frente a la eficiencia de los estimadores, que permite analizar si el modelo adecuado para nuestros datos es el modelo de efectos fijos o el modelo de efectos aleatorios. La hipótesis nula del contraste es que las variables explicativas y los efectos individuales no están correlacionados, es decir, que se cumplen los supuestos del modelo de efectos aleatorios, mientras que la hipótesis alternativa será que sí existe correlación. Es decir,

$$\begin{cases} H_0: & E\left[\alpha_i | X_{i1}, X_{i2}, \cdots X_{iT}\right] = E\left[\alpha_i\right] = 0 \\ H_1: & E\left[\alpha_i | X_{i1}, X_{i2}, \cdots X_{iT}\right] \neq 0 \end{cases}$$

El contraste compara los coeficientes estimables de los regresores que varían con el tiempo, y el estadístico de contraste mide la distancia entre ambas estimaciones, de manera que si toma un valor elevado se rechaza H_0:

$$H = \left(\hat{\beta}_{EF} - \hat{\beta}_{EA}\right)^t \left[Var\left(\hat{\beta}_{EF}\right) - Var\left(\hat{\beta}_{EA}\right)\right]^{-1} \left(\hat{\beta}_{EF} - \hat{\beta}_{EA}\right) \underset{H_0}{\sim} \chi^2_{(k)},$$

siendo $Var(\hat{\beta}_{EF})$ la matriz de varianzas y covarianzas del modelo estimado con efectos fijos y $Var(\hat{\beta}_{EA})$ la matriz de varianzas y covarianzas del modelo estimado con efectos aleatorios.

Bajo la hipótesis nula, la estimación del modelo de efectos aleatorios será consistente y eficiente mientras que la estimación del modelo de efectos fijos será únicamente consistente. En cambio, bajo la hipótesis alternativa, la estimación del modelo de efectos fijos será consistente pero no eficiente y la estimación del modelo de efectos aleatorios no será ni siquiera consistente.

Si la diferencia observada entre ambos es escasa, se tiene evidencia a favor de la hipótesis nula. Que la diferencia observada sea significativa sugiere que es preferible utilizar los estimadores de efectos fijos. En el caso que nos ocupa, se puede observar que el p-valor es menor al 5 % de significación habitual, por lo cual se rechaza la hipótesis nula. Es decir, el estimador de efectos aleatorios es inconsistente y por lo tanto, es preferible utilizar el estimador de efectos fijos.

4. Finalmente, siendo el valor del estadístico F experimental 8,85 y teniendo en cuenta que el estadístico para contrastar la significación conjunta de los parámetros del modelo proporcionado por la tabla de F de Snedecor para $F_{k,NT-k-1}$ tendrá un valor pequeño, se puede concluir que se rechazará la hipótesis nula de no significación por lo que es posible afirmar que las variables binarias temporales son conjuntamente significativas y pertenecen al modelo. Al agregar los años utilizando un modelo de efectos fijos el coeficiente de determinación de R^2 es más alto. Esto indica un mejor ajuste como consecuencia de la utilización de un mayor número de variables explicativas en el modelo, aunque ese aumento por sí solo no justifica que el modelo sea más adecuado y sería necesario analizar más parámetros del ajuste. En cuanto al impacto del uso del cinturón, el número de personas que mueren si se utiliza el cinturón es menor.

Cuestión 8

Está interesado en estimar una demanda de pescado vendido en el Mercado de Luarca, para ello usted va al mercado y recoge el precio diario y la cantidad vendida de 97 días consecutivos (Dado que el mercado está cerrado los fines de semana, recopila datos de lunes a viernes). Específicamente tiene datos para las siguientes variables:

- totqty - la cantidad total de pescado vendido ese día

- avgprc - el precio promedio del pescado vendido ese día
- lun - una dummy = 1 por si el día es lunes
- mar - una dummy = 1 por si el día es martes
- mie - una dummy = 1 por si el día es miércoles
- jue - una dummy = 1 por si el día es juesves
- wave2 - la altura máxima promedio de olas durante dos días anteriores a los datos de precio y cantidad
- wave3 - la altura media promedio de ola durante tres y cuatro días anteriores a los datos de precio y cantidad

Nota: aunque usemos subíndices de tiempo a lo largo de esta pregunta, no utilizaremos métodos de series temporales. También mantendremos la suposición de que todos los errores son homocedásticos.

1. Suponga que la ecuación de demanda se puede escribir para cada periodo de tiempo como

   $$ln(totqty_t) = \beta_0 + \beta_1 ln(avgprc_t) + \beta_2 lun_t + \beta_3 mar_t + \beta_4 mie_t + \beta_5 jue_t + u_t$$

   , es decir, la demanda puede variar durante días de la semana. ¿Por qué no es apropiado usar MCO para estimar esta ecuación de demanda? ¿Qué información adicional necesitamos para tener estimadores consistentes de los parámetros de ecuación de demanda?

2. Las variables wave2$_t$ y wave3$_t$ son medidas de las alturas de las olas oceánicas en los últimos días. ¿Qué dos suposiciones necesitamos para usar wave2$_t$ y wave3$_t$ como instrumentos para ln(avgprc$_t$) al estimar la

ecuación de la demanda? Asegúrese de analizar como estas suposiciones están relacionadas con las ecuaciones de demanda y oferta.

3. La primera etapa de una regresión de mínimos cuadrados de dos etapas nos da los siguientes resultados:

$$ln(avgprc) = \underset{(0{,}14)}{-1,02} \underset{(0{,}114)}{-0,012}\, lun_t \underset{(0{,}1119)}{-0,0090}\, mar_t \underset{(0{,}112)}{+0,051}\, mie_t \underset{(0{,}111)}{+0,124}\, jue_t$$

$$\underset{(0{,}021)}{+0,094}\, wave2_t \underset{(0{,}020)}{+0,053}\, wave3_t$$

$R^2 = 0,165$

y la prueba de la significación conjunta de wave2_t y wave3_t nos da un estadístico F: F − stat = 19,1, mientras que la prueba de significación conjunta de las dummies de día de la semana da F − stat = 0,53. ¿Son wave2_t y wave3_t individualmente significativos al nivel del 1 %? ¿Qué revelan los resultados de la regresión de esta primera etapa sobre nuestros instrumentos?

4. La estimación de la ecuación de demanda por mínimos cuadrados en dos etapas nos da los siguientes resultados:

$$ln(totqty_t) = \underset{(0{,}18)}{8,16} \underset{(0{,}327)}{-0,816}\, ln(avgprc_t) \underset{(0{,}229)}{-0,307}\, lun_t \underset{(0{,}226)}{-0,685}\, mar_t$$

$$\underset{(0{,}224)}{+0,521}\, mie_t \underset{(0{,}225)}{+0,095}\, jue_t$$

donde los errores estándar entre paréntesis son los correctos (es decir, toman en cuenta el procedimiento de dos etapas). ¿Cuál es la interpretación del coeficiente en ln(avgprc$_t$)? ¿Su magnitud parece razonable? Construya un intervalo de confianza del 95 % para este coeficiente.

5. Dado que tenemos dos instrumentos y una variable endógena, la ecuación de demanda está sobre identificada. La prueba de sobre identificación de restricciones nos da un estadístico F de 0,013. ¿Qué concluye?

6. Dado que la ecuación de la oferta (no especificada) evidentemente depende de las variables de olas(wave), ¿qué dos suposiciones deberíamos hacer para estimar la elasticidad del precio de la oferta?

7. Aquí están los resultados de la estimación de una posible ecuación de oferta para esta industria por mínimos cuadrados en dos etapas:

$$ln(totqty_t) = \underset{(2,23)}{10,82} \underset{(2,24)}{+2,13}\, ln(avgprc_t) \underset{(0,212)}{-0,267}\, wave2_t \underset{(0,139)}{-0,169}\, wave3_t$$

¿Cuál es la interpretación del coeficiente de ln(avgprc$_t$)? ¿Es estadísticamente significativo al nivel de 10 %?

Solución.

1. La variable relativa al precio medio del pescado se podría considerar en este contexto como una variable endógena, puesto que seguramente la cantidad demandada o vendida depende del precio y viceversa. En esta situación nos encontramos ante un problema de ecuaciones simultáneas en el que tratamos de estudiar las ecuaciones de demanda y de oferta para determinar cuál es el equilibro de una forma conjunta. Así, si se estimase directamente la ecuación proporcionada por MCO, al ser uno de los regresores endógeno, los estimadores serían sesgados e inconsistentes pues no se cumple uno de los requisitos para que los estimadores calculados con este método tengan buenas propiedades, que es que $Cov\,(X_t, u_t) \neq 0$, y por tanto existe información relevante para el modelo que no se está teniendo en cuenta. Tendríamos que intentar considerar factores específicos o variables específicas que determinen la oferta y que no tengan influencia sobre la demanda, como veremos posteriormente que podrían ser las alturas de las olas, actuando como variables instrumentales. Sería necesario por tanto que apareciera al

menos una variable exógena en la ecuación de la oferta que no sea relevante en la estimación de la ecuación de demanda para poder obtener estimadores consistentes de los parámetros.

2. Los dos supuestos necesarios para usar wave2_t y wave3_t como instrumentos son: por una parte, ausencia de correlación entre las variables instrumentales y el término de error del modelo, y por otra parte, existencia de correlación entre wave2_t y wave3_t y las variables explicativas a las que sustituirán. El primero de estos supuestos permite asegurar que, definiendo $\boldsymbol{z}$ como el vector que recoge las variables instrumentales, $Cov(\boldsymbol{z}, \boldsymbol{u}) = 0$ y es esta ausencia de correlación con el término de error la que garantiza la consistencia que es lo importante en este tipo de problemas. El segundo de los supuestos es que $Cov(\boldsymbol{z}, \boldsymbol{x}) \neq 0$, es decir, que existe correlación entre los instrumentos y la o las variables que generan el problema, y esta condición hace que el estimador de variables instrumentales sea más eficiente, aunque el primero de los supuestos es más relevante. Este primer supuesto implicaría que para que wave2_t y wave3_t sean instrumentos válidos para $ln(avgprc_t)$ necesitamos que se puedan excluir de la ecuación de demanda. Esto puede no ser del todo razonable, y que la altura de las olas se determinen en parte por el clima o el tiempo y la demanda en un mercado local de pescado podría depender de la propia demanda. La segunda suposición se traduce en que al menos una de las dos variables wave2_t o wave3_t aparezcan en la ecuación de oferta. En el apartado siguiente hay una evidencia indirecta de esta cuestión, puesto que ambas son conjuntamente significativas en la expresión de $ln(avgprc_t)$. Así se cumplirían las condiciones para que la estimación de la ecuación de demanda sea consistente.

3. El método de mínimos cuadrados en dos etapas (MC2E) proporciona estimadores insesgados y consistentes de los parámetros en situaciones en las que alguna de las variables explicativas es endógena, de manera que esté correlacionada con el término de error o bien hay una omisión de variables explicativas exógenas y una estimación por mínimos cuadrados no sería adecuada por las malas propiedades de los estimadores. Disponemos pues de más de 1 variable exógena externa al modelo, que en nuestro caso son wave2_t y wave3_t (que serían los posibles instrumentos) y buscamos el mejor instrumento posible para $ln(avgprc_t)$, que será una combinación lineal de todas las variables exógenas que aparecen en el modelo y además estas dos adicionales, es decir,

$$ln(avgprc_t) = \beta_0 + \beta_1 lun_t + \beta_2 mar_t + \beta_3 mier_t + \beta_4 jue_t + \\ + \beta_5 wave2_t + \beta_6 wave3_t + v_t$$

En la primera etapa se estima el modelo por MCO, y según los datos del ejercicio este modelo es el siguiente:

$$\widehat{ln(avgprc_t)} = \underset{(0{,}14)}{-1,02} \underset{(0{,}114)}{-0,012}\, lun_t \underset{(0{,}1119)}{-0,0090}\, mar_t \underset{(0{,}112)}{+0,051}\, mie_t \underset{(0{,}111)}{+0,124}\, jue_t$$

$$\underset{(0{,}021)}{+0,094}\, wave2_t \underset{(0{,}020)}{+0,053}\, wave3_t$$

$R^2 = 0,165$

Lo que se suele hacer es plantear un contraste de identificación para determinar si los instrumentos están o no correlacionados con la variable endógena $ln(avgprc_t)$, es decir:

$$\begin{cases} H_0: & \beta_5 = \beta_6 = 0 \\ H_1: & \beta_5 \neq 0 \; ; \; \beta_6 \neq 0 \end{cases}$$

En el estudio de la significación conjunta de wave2_t y wave3_t se observa que el estadístico experimental es 19,1 mientras que el estadístico

$F_{(2,97-6-1=90;0,01)} = 4,849$. Por tanto, conjuntamente sí son claramente significativos, lo cuál ya indicaría que pueden ser instrumentos válidos. Además, los contrastes de significación individual de wave2_t y wave3_t se basan en el estadístico:

$$\frac{\hat{\beta}_{wavei_t}}{\hat{\sigma}(\hat{\beta}_{wavei_t})} \sim t_{97-7=90}$$

Por lo tanto, calculando:

$$\frac{\hat{\beta}_{wave2_t}}{\hat{\sigma}(\hat{\beta}_{wave2_t})} = 4,4761 \qquad\qquad \frac{\hat{\beta}_{wave3_t}}{\hat{\sigma}(\hat{\beta}_{wave3_t})} = 2,65$$

y siendo $t_{(90;0,005)}$=2,6316, para ambas variables el estadístico muestral es mayor que el obtenido de las tablas, aunque los valores del estadístico experimental son muy próximos al estadístico teórico en uno de los instrumentos, por tanto también son individualmente significativos y no podemos rechazar la hipótesis nula de que sean nulos.

Por otra parte, en la estimación anterior, las variables binarias relativas a los días de la semana no son conjuntamente significativas, puesto que el p-valor asociado a una $F_{(4,97-6-1=90)}$ = 0,53 es 0,71, por tanto, esto significa que mientras algunas de estas variables tienen significado en la ecuación de demanda, estos efectos se difuminan y no afectan al precio de equilibrio una vez que las variables wave2_t y wave3_t están presentes en la ecuación. Si bien es cierto que es poco probable que la altura de las olas de días previos tenga alguna influencia sobre el deseo de los consumidores de comprar pescado, esta estimación sí que sugiere que estas dos variables podrían ser instrumentos válidos, aunque seguramente con una de ellas sería suficiente.

4. En la segunda etapa de la estimación por MC2E lo que se hace es sustituir en la ecuación de partida:

$ln(totqty_t) = \beta_0 + \beta_1 ln(avgprc_t) + \beta_2 lun_t + \beta_3 mar_t + \beta_4 mie_t + \beta_5 jue_t + u_t$ la variable endógena $ln(avgprc_t)$ por sus valores ajustados en el paso anterior, es decir, por $ln(\widehat{avgprc_t})$ de tal forma que el modelo sería: $ln(totqty_t) = \beta_0 + \beta_1 ln(\widehat{avgprc_t}) + \beta_2 lun_t + \beta_3 mar_t + \beta_4 mie_t + \beta_5 jue_t + u_t$ y estimando por MCO se obtienen los llamados estimadores de Mínimos Cuadrados Bietápicos o en 2 etapas, que serán consistentes y asintóticamente normales, siempre que los instrumentos sean válidos. Según el enunciado la estimación de este modelo es:

$$ln(totqty_t) = \underset{(0{,}18)}{8,16} \underset{(0{,}327)}{-0,816}\, ln(\widehat{avgprc_t}) \underset{(0{,}229)}{-0,307}\, lun_t \underset{(0{,}226)}{-0,685}\, mar_t$$

$$\underset{(0{,}224)}{+0,521}\, mie_t \underset{(0{,}225)}{+0,095}\, jue_t$$

La interpretación del coeficiente de $ln(avgprc_t)$, que sí que parece razonable, es que un aumento de un 1 % en el precio promedio del pescado vendido ese día supondrá una caída del 0,816 % en la cantidad total de pescado vendido ese día. (O equivalentemente, un aumento de un 10 % en el precio supone una bajada de un 8,16 % en la cantidad pescado vendido ese día). El significado que tiene es que la caída en la demanda es algo menor que el incremento en el precio, y este coeficiente es significativo de forma individual por lo que esta interpretación es coherente. El intervalo de confianza para este coeficiente, vendrá dado por

$$IC^{1-\alpha}_{\widehat{\beta}_1} = \left[\widehat{\beta}_1 - t_{n-k;\frac{\alpha}{2}}\sqrt{Var\left(\hat{\beta}_1\right)}; \widehat{\beta}_1 + t_{n-k;\frac{\alpha}{2}}\sqrt{Var\left(\hat{\beta}_1\right)}\right]$$

donde $t_{n-k;\frac{\alpha}{2}} \simeq 1,96$ y por tanto el intervalo de confianza es:

$$IC^{1-\alpha}_{\widehat{\beta}_1} = [-0,816 - 1,96 \cdot 0,327; -0,816 + 1,96 \cdot 0,327] =$$
$$= [-1,45692; -0,17508]\,.$$

En general, si los instrumentos no son válidos, por incumplir alguna de las condiciones que se les requiere, los estimadores por MC2E serán inconsistentes y peores que los de MCO, que tienen en todo caso mínima varianza aunque fueran también inconsistentes.

5. Dado que se dispone de un solo regresor endógeno, $\ln(\text{avgprc}_t)$, y de dos instrumentos, wave2_t y wave3_t, el modelo está sobreidentificado y podemos estudiar si todos los instrumentos están incorrelacionados con el término de error, siendo por tanto válidos, mediante el llamado contraste J.

   Lo que se hace es plantear una regresión de los residuos de MC2E, que son los $\hat{u}_t$, sobre todas las variables exógenas presentes en el modelo y sobre las externas también:

$$\widehat{u}_t = \delta_0 + \delta_1 lun_t + \delta_2 mar_t + \delta_3 mie_t + \delta_4 jue_t + \delta_5 wave2_t + \delta_6 wave3_t + \epsilon_t$$

   y el contraste es:

$$\begin{cases} H_0: & \delta_5 = \delta_6 = 0 \\ H_1: & \delta_5 \neq 0 \text{ y/o } \delta_6 \neq 0 \end{cases}$$

   donde δ_5 y δ_6 son los coeficientes de la regresión. Es decir, la hipótesis nula sería que todos los instrumentos están incorrelacionados con el error frente a la alternativa de que alguno sí que lo estaría. El estadístico para el contraste de significación sería $J = m \cdot F \sim \chi^2_q$ donde m es el número de instrumentos y q es el grado de sobreidentificación, es decir, el número de instrumentos, que en nuestro caso son 2, menos el número de variables explicativas endógenas, que en nuestro caso es una, así que $q = 1$. El enunciado nos dice que F tiene un valor de 0,013 y el valor crítico de una χ^2_1 al 95 % de confianza es 3,84. Como el valor empírico $J = 2 \cdot 0,013 = 0,026$ es menor que el crítico, no rechazamos la hipótesis nula de no correlación de los instrumentos con el término de error del modelo, es decir, se concluye la exogeneidad de los instrumentos utilizados y por tanto no hay evidencia en contra de la validez de los mismos.

Este contraste tiene el inconveniente de que si se rechaza la hipótesis nula, es decir, si se concluye que algún instrumento no es válido no especifica cuál es en caso de haber varios.

6. Para estimar la elasticidad del precio de la oferta tendríamos que asumir que las variables binarias de los días de la semana no aparezcan en la ecuación de la oferta, pero sí que estén en la ecuación de demanda. Con los datos del apartado tercero se muestran evidencias de que existen efectos producidos por el día de la semana en la ecuación de demanda. Esta afirmación no la podemos confirmar en el caso de la ecuación de oferta, puesto que no la conocemos.

   La ecuación de oferta podría estimarse consistentemente si creemos que la cantidad ofrecida de pescado no está afectada por el precio, sino que está determinada por cuestiones exógenas al mercado, como pueden ser factores como la altura de las olas, las horas de luz diarias, la climatología, etc... En este caso lo que tendríamos es un modelo con dos ecuaciones con una estructura conocida con el nombre de recursiva.

7. El modelo de oferta estimado por MC2E es:

$$ln(totqty_t) = \underset{(2,23)}{10,82} \underset{(2,24)}{+2,13}\, ln(avgprc_t) \underset{(0,212)}{-0,267}\, wave2_t \underset{(0,139)}{-0,169}\, wave3_t$$

   La interpretación del coeficiente de ln(avgprc$_t$) es que un aumento de un 1 % en el precio promedio del pescado vendido ese día supondrá una subida del 2,13 % en la cantidad total de pescado vendido ese día, afirmación que no parece muy razonable en principio. Para contrastar su significatividad calculamos:

$$\frac{\widehat{\beta}_{ln(avgprc_t)}}{\hat{\sigma}(\widehat{\beta}_{ln(avgprc_t)})} \sim t_{97-4=93}$$

$$\left|\frac{2,13}{2,24}\right| = 0,95 < 1,66 \simeq t_{93;0,05}$$

Por tanto, de forma individual no es significativo, y por este motivo la interpretación anterior pierde su validez y de alguna forma nos hace pensar que la pesca en este ámbito probablemente se produzca teniendo en cuenta circunstancias climáticas que son exógenas al mercado, de forma que la cantidad de pescado que se ofrece no depende del precio, es decir, la oferta sería inelástica al precio, estando determinada por diversos factores exógenos. Ahora bien, la cantidad que se vende sí que depende del precio a través de la ecuación de demanda.

Cuestión 9

Dada la siguiente tabla de mortalidad:

	Tasa de mortalidad	Promedio de años vividos el último año de vida	Riesgo de muerte	Supervivientes	Defunciones teóricas	Población estacionaria	Tiempo por vivir	Esperanza de vida
98 años	296,19	0,4758	256,38	6.263,77	1.605,93	5.421,93	16.422,50	2,622
99 años		0,4688			1.223,66			
100 y más años		2,0362						

Se pide:

a) Calcular los supervivientes a las edades de 99 y 100 y más años.

b) Calcular el riesgo de muerte a las edades de 99 y 100 y más años.

c) Calcular la población estacionaria a las edades de 99 y 100 y más años.

d) Calcular el tiempo por vivir a las edades de 99 y 100 y más años.

e) Calcular la esperanza de vida a las edades de 99 y 100 y más años.

f) Calcular la tasa de mortalidad a las edades de 99 y 100 y más años.

Solución.

Según la metodología establecida:

$$q_{100} = 1$$
$$e_{100} = a_{100} = 2,0362$$

Los supervivientes se obtienen restando a los del año anterior las defunciones teóricas del mismo año:

$$l_{99} = l_{98} - d_{98} = 6263,77 - 1605,93 = 4657,84$$

El total del tiempo vivido por los individuos de la generación será:

$$T_{98} = T_{99} + L_{98} \Leftrightarrow T_{99} = T_{98} - L_{98} = 16422,50 - 5421,93 = 11000,57$$

Del cociente entre estas cantidades se deriva la esperanza de vida:

$$e_{99} = \frac{T_{99}}{l_{99}} = \frac{11000,57}{4657,84} = 2,361.$$

El riesgo de muerte se obtiene como el cociente entre las defunciones teóricas y el número de supervivientes:

$$q_{99} = \frac{d_{99}}{l_{99}} = \frac{1223,66}{4657,84} = 262,71‰$$

y con ella se calcula la tasa de mortalidad:

$$q_{99} = \frac{m_{99}}{1 + (1 - a_{99})m_{99}} \Leftrightarrow m_{99} = \frac{q_{99}}{1 - q_{99}(1 - a_{99})}$$
$$m_{99} = \frac{262,71}{1 - 262,71(1 - 0,4688)} = 305,31‰$$

Finalmente, la población estacionaria se obtiene como:

$$m_{99} = \frac{d_{99}}{L_{99}} \Leftrightarrow L_{99} = \frac{d_{99}}{m_{99}} = \frac{1223,66}{305,31} = 4007,92.$$

De la población estacionaria se obtienen los supervivientes de 100 y más años:

$$L_{99} = l_{100} + a_{99}d_{99} \Leftrightarrow l_{100} = L_{99} - a_{99}d_{99} = 4007,92 - 0,4688 \cdot 1223,66 = 3434,27$$

que según la metodología establecida coincide con las defunciones:

$$d_{100} = l_{100} = 3434,27$$

En cuanto al tiempo que le queda por vivir a la generación:

$$T_{99} = T_{100} + L_{99} \Leftrightarrow T_{100} = T_{99} - L_{99} = 11000,57 - 4007,92 = 6992,65$$

que coincide con la población estacionaria:

$$L_{100} = T_{100} = 6992,65$$

Por último, la tasa de mortalidad será:

$$m_{100} = \frac{d_{100}}{L_{100}} = \frac{3434,27}{6992,65} = 491,12‰$$

	Tasa de mortalidad	Promedio de años vividos el último año de vida	Riesgo de muerte	Supervivientes	Defunciones teóricas	Población estacionaria	Tiempo por vivir	Esperanza de vida
98 años	296,19	0,4758	256,38	6.263,77	1.605,93	5.421,93	16.422,50	2,622
99 años	305,31	0,4688	262,71	4.657,84	1.223,66	4.007,92	11.000,57	2,361
100 y más años	491,12	2,0362	1000	3.434,27	3.434,27	6.992,65	6.992,65	2,0362

Cuestión 10

Sea un país con dos regiones, A y B, y considere la siguiente matriz migratoria 4.9 por regiones en el periodo 2001 - 2011.

Tabla 4.9

Región de residencia 2001	Región de residencia 2011		
	Región A	Región B	Total
Región A	1244,4	14,4	1258,8
Región B	18,2	889,4	907,6
En el extranjero	72,2	99,3	171,5
No aplicable	130,1	123,2	253,5
Total	1464,9	1126,3	2591,2

Nota: "No aplicable" recoge la población nacida en el periodo intercensal 2001 - 2011.

Se pide:

a) El número total de emigrantes durante el periodo, la proporción de emigración y la tasa de emigración en las regiones A y B.

b) El número total de inmigrantes durante el periodo, la proporción de inmigración y el índice de atracción en las regiones A y B.

c) El saldo migratorio y las tasas de migración brutas y netas durante el periodo en las regiones A y B.

d) El índice de efectividad migratoria y el índice de migración diferencial en las regiones A y B.

Solución.

En primer lugar se calculan los resultados para la Región A y a continuación se procede análogamente para la Región B.

- **Región A.**

  El número de emigrantes de una región se calcula como la diferencia entre la población al inicio del periodo objeto de estudio y la población que reside en la región tanto al inicio como al final del periodo.

$$E_A = P_A^{t-n} - M_{AA} = 1258,8 - 1244,4 = 14,4$$

La proporción de emigración de cada región durante el perido 2001 - 2011 es la relación entre el número de emigrantes de dicha región y la población de origen en la región al inicio del periodo.

$$PE_A = \frac{E_A}{P_A^{t-n}} = \frac{14,4}{1258,8} \cdot 100 = 1,14\,\%$$

La tasa de emigración relaciona el número de emigrantes de la región en el periodo 2001 - 2011 con la población media de dicha región durante dicho periodo.

$$e_A = \frac{E_A}{10\left(\frac{P_A^{t-n}+P_A^t}{2}\right)} \cdot 1000 = \frac{14,4}{10\left(\frac{1258,8+1464,9}{2}\right)} \cdot 1000 = 1,057‰$$

El número de inmigrantes se obtiene restando a la población de la región al final de periodo el número de personas que residen en la región tanto al inicio como al final del periodo y la población nacida en el periodo intercensal, es decir, la población "no aplicable".

$$I_A = P_A^t - M_{AA} - \text{no aplicable}_A = 1464,9 - 1244,4 - 130,1 = 90,4$$

La proporción de inmigrantes relaciona el número de inmigrantes de la región la población total de la misma al final del periodo.

$$PI_A = \frac{I_A}{P_A^t} = \frac{90,4}{1464,9} \cdot 100 = 6,171\,\%$$

El índice de atracción de cada región durante el periodo 2001-2011 relaciona el número de inmigrantes de dicha región en el periodo estudiado con la población media de la región durante el periodo.

$$i_A = \frac{I_A}{10\left(\frac{P_A^{t-n}+P_A^t}{2}\right)} \cdot 1000 = \frac{90,4}{10\left(\frac{1258,8+1464,9}{2}\right)} \cdot 1000 = 6,637‰$$

El saldo migratorio es la diferencia entre el total de inmigrantes y el total de emigrantes.

$$SM_A = I_A - E_A = 76$$

La tasa de migración bruta es la suma de las tasa de inmigración y emigración.

$$tmb_A = i_A + e_A = 7,694‰$$

La tasa de migración neta es la diferencia entre las tasa de inmigración y emigración.

$$tmn_A = i_A - e_A = 5,58‰$$

El índice de efectividad es un indicador que se define para una región como la relación entre la migración neta de dicha región y la migración total de la misma.

$$I^A_{efectividad} = \frac{I_A - E_A}{I_A + E_A} = 0,7251$$

El índice de migración diferencial permite comparar la migración entre diferentes sub–poblaciones. Se calcula utilizando

$$d_A = \frac{E_A \cdot P}{E \cdot P_A} - 1 = \frac{14,4 \cdot 2591,2}{32,6 \cdot 1464,9} - 1 = -0,218$$

donde

$$P = \Sigma_x P^t_x = P^t_A + P^t_B = 1464,9 + 1126,3 = 2591,2$$

$$E = \Sigma_x E_x = E_A + E_B = 14,4 + 18,2 = 32,6$$

siendo P el tamaño de la población total (suma del tamaño de cada uno de los grupos) y E la emigración en la población total (suma de emigración en cada grupo). El índice de migración diferencial negativo indica que la tendencia a migrar en la región A es menor que la tendencia a migrar de la población total.

- **Región B.**

$$E_B = 907,6 - 889,4 = 18,2$$

$$PE_B = \frac{E_B}{P_B^{t-n}} = \frac{18,2}{907,6} \cdot 100 = 2,005\,\%$$

$$e_B = \frac{E_B}{10\left(\frac{P_B^{t-n}+P_B^t}{2}\right)} \cdot 1000 = \frac{18,2}{10\left(\frac{907,6+1126,3}{2}\right)} \cdot 1000 = 1,789‰$$

$$I_B = 1126,4 - 889,4 - 123,3 = 113,7$$

$$PI_B = \frac{I_B}{P_B^t} = \frac{113,7}{1126,3} \cdot 100 = 10,09\,\%$$

$$i_B = \frac{E_B}{10\left(\frac{P_B^{t-n}+P_B^t}{2}\right)} 1000 = \frac{113,7}{10\left(\frac{907,6+1126,3}{2}\right)} 1000 = 11,18‰$$

$$SM_B = I_B - E_B = 95,5$$

$$tmb_B = i_B + e_B = 12,969‰$$

$$tmn_B = i_B - e_B = 9,391‰$$

$$I^B_{efectividad} = \frac{I_B - E_B}{I_B + E_B} = 0,724$$

$$d_B = \frac{E_B \cdot P}{E \cdot P_B} - 1 = -0,2844$$

donde de nuevo, $P = \Sigma_x P_x = 25951,2$ y $E = \Sigma_x E_x = 32,6$. El índice de migración diferencial positivo indica que la tendencia a migrar en la región B es mayor que la tendencia a migrar de la población total.

4.2. Soluciones del tercer Examen del Cuerpo de Diplomados en Estadística del Estado de 2017

Cuestión 1

La probabilidad de que una sucursal de un Banco reciba un cheque sin fondos es del 1 %.

a) Si una sucursal en una hora recibe 20 cheques, ¿cuál es la probabilidad de que reciba algún cheque sin fondos en una hora?

b) Si la media del valor de los cheques sin fondos es de 580€ y la sucursal trabaja 6 horas diarias. Calcular la cantidad total esperada de euros en un día, correspondiente a los cheques sin fondo.

c) El banco dispone de 12 sucursales en una determinada ciudad, ¿cuál es la probabilidad de que al menos 4 de las sucursales reciban algún cheque sin fondos en 1 hora?

d) Por un error informático los sistemas de una sucursal del Banco detectan sólo la mitad de los cheques sin fondos que reciben ¿cuál es la probabilidad de que esta sucursal reciba un cheque sin fondos después de ser procesado por sus sistemas informáticos?

Solución.

a) Se define $X \equiv$ "cheque sin fondos", $X \sim Bernoulli(p = 0{,}01)$ pues con una probabilidad de 0,01 X toma el valor 1 si la sucursal recibe un cheque sin fondos y con probabilidad 0,99 toma el valor 0 si la sucursal no recibe un cheque sin fondos.

Sea $Y \equiv$ "número de cheques sin fondos en una hora", $Y = \sum_{i=1}^{20} X_i$. Dada la propiedad reproductiva de la distribución de Bernoulli se cumple que $Y \sim B(n = 20, p = 0{,}01)$. Por tanto, la probabilidad de recibir algún cheque sin fondos en una hora, dado que la sucursal recibe 20 cheques a la hora, es:

$$P(Y > 0) = 1 - P(Y = 0) = 1 - (0{,}99)^{20} = 1 - 0{,}8179 = 0{,}182.$$

b) El número esperado de cheques sin fondos en una hora es

$$E[Y] = np = 20 \cdot 0{,}01 = 0{,}2.$$

Siendo $Z = 6Y$ el número de cheques sin fondos que recibe a lo largo del día, el número esperado de cheques sin fondos a lo largo del día será

$$E[Z] = 6E[Y] = 6 \cdot 0{,}2 = 1{,}2.$$

Siendo la media del valor de estos 580€, la cantidad total esperada de euros correspondiente a cheques sin fondos en un día será

$$1,2 \cdot 580 = 696€.$$

c) La probabilidad de que una sucursal reciba algún cheque sin fondos en una hora es un suceso de probabilidad 0,182. Si se define la variable aleatoria $W \equiv$"número de sucursales que reciben algún cheque sin fondos en una hora", se tiene que $W \sim B(12; 0,182)$. La probabilidad de que al menos 4 reciban algún cheque sin fondos en una hora será

$$\begin{aligned}P(W \geq 4) = 1 - P(W \leq 3) = \\ &= 1 - \binom{12}{0} 0{,}182^0 \cdot 0{,}8179^{12} - \binom{12}{1} 0{,}182^1 \cdot 0{,}8179^{11} \\ &- \binom{12}{2} 0{,}182^2 \cdot 0{,}8179^{10} - \binom{12}{3} 0{,}182^3 \cdot 0{,}8179^9 \simeq\end{aligned}$$

$$\simeq 1 - 0{,}0896 - 0{,}2393 - 0{,}2928 - 0{,}2172 =$$
$$= 0,1611$$

d) La sucursal trabaja 6 horas al día luego recibe $6 \cdot 20 = 120$ cheques en un día. Sea entonces D ≡"número de cheques sin fondos en un día", $D \sim B(120; 0,01)$. La probabilidad de recibir un cheque sin fondos en un día antes de que los sistemas informáticos traten de detectarlo es por tanto

$$P\left(D=1\right) = \binom{120}{1} 0,01^1 0,99^{119} = 0,3628.$$

Los sistemas de una sucursal detectan sólo la mitad de los cheques sin fondos que reciben, es decir,

$$\frac{1}{2} = P\left(\text{detectado sin fondos/cheque sin fondos}\right)$$

S e definen D_{SF} = "detectado sin fondos", CF = "con fondos" y SF = "sin fondos", y se tiene que

$$P(SF \cap D_{SF}) = P(D_{SF}/SF) \cdot P(SF) = \frac{1}{2} \cdot 0{,}01 = 0,005.$$

Suponiendo que si un cheque que tiene fondos nunca es rechazado por el sistema, se tiene que $P(CF \cap D_{SF}) = 0$ y por tanto

$$P(D_{SF}) = P(CF) \cdot P(D_{SF}/CF) + P(D_{SF}/SF) \cdot P(SF) =$$
$$= 0{,}99 \cdot 0 + \frac{1}{2} \cdot 0{,}01 =$$
$$= 0,005.$$

Así, la variable aleatoria

SF_{ND} ≡"número de cheques sin fondos recibidos en un día y detectados" sigue una distribución $SF_{ND} \sim B(120; 0,005)$ y la probabilidad de que esta sucursal reciba un cheque sin fondos después de ser procesados todos los que recibe en un día es

$$P\left(SF_{ND} = 1\right) = \binom{120}{1} 0,005^1 0,995^{119} = 0,3304.$$

Cuestión 2

Sea (X, Y) una variable aleatoria bidimensional con función de densidad conjunta:

$$f(x,y) = \begin{cases} e^{-(x+y)} & \text{si } x > 0, y > 0 \\ \\ 0 & \text{en el resto de los casos} \end{cases}$$

a) Calcular las funciones de densidad marginales

b) ¿Son independientes X e Y?

c) Calcular la función de distribución conjunta.

d) Calcular la matriz de varianzas - covarianzas de la variable (W, Z) con $W = X + Y$ y $Z = X - Y$.

Solución.

a) La función de densidad marginal de la variable Y viene dada por

$$f(y) = \int_0^\infty e^{-(x+y)} dx = e^{-y} \int_0^\infty e^{-x} dx = e^{-y}\left(-e^{-x}\right)_0^\infty = e^{-y},\ y > 0$$

Operando de forma análoga, la función de densidad marginal de la variable X viene dada por

$$f(x) = \int_0^\infty e^{-(x+y)} dy = e^{-x},\ x > 0$$

Se tiene entonces que ambas variables son exponenciales de parámetro $\lambda = 1$.

b) Dado que
$$e^{-x}e^{-y} = e^{-(x+y)},$$
se cumple la condición de independencia $f(x) \cdot f(y) = f(x, y)$ y por tanto X e Y son independientes.

c) Las funciones de distribución de cada una de las variables marginales son
$$F(x) = 1 - e^{-x},\ x > 0$$
$$F(y) = 1 - e^{-y},\ y > 0$$
y, dado que son independientes, $F(x) \cdot F(y) = F(x, y)$ de modo que la función de distribución conjunta será
$$F(x,y) = \begin{cases} (1 - e^{-x})(1 - e^{-y}) & \text{si } x > 0, y > 0 \\ \\ 0 & \text{en otro caso} \end{cases}$$

d) Finalmente, por ser $X \sim Exp(\lambda = 1)$ e $Y \sim Exp(\lambda = 1)$ y visto que son independientes, $Var(X) = Var(Y) = 1$ y $Cov(X, Y) = 0$. Por tanto, la matriz de varianzas y covarianzas de la variable (X, Y) será
$$\Sigma_{XY} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$
Dada la variable (W, Z) con $W = X + Y$ y $Z = X - Y$, la matriz de varianzas - covarianzas será:
$$\Sigma_{WZ} = \begin{pmatrix} Var(W) & Cov(W, Z) \\ Cov(W, Z) & Var(Z) \end{pmatrix}$$
donde
$$Var(W) = Var(X + Y) = Var(X) + Var(Y) = 1 + 1 = 2$$

$$Var(Z) = Var(X - Y) = Var(X) + Var(Y) = 1 + 1 = 2$$

$$\begin{aligned} Cov(W, Z) &= Cov(X + Y, X - Y) = \\ &= Cov(X, X) - Cov(X, Y) + Cov(Y, X) - Cov(Y, Y) = \\ &= Var(X, X) - Var(Y, Y) = 1 - 1 = 0 \end{aligned}$$

Luego

$$\Sigma_{WZ} = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$$

Cuestión 3

Una empresa láctea, que dispone de 100 plantas de producción distribuidas por toda la geografía, desea estimar la cantidad de leche envasada al mes. Para ello selecciona una muestra, con probabilidades iguales y con reposición, de 8 plantas en las que examina todas las máquinas para envasar. El resultado de la muestra es el siguiente:

Planta	Nº de máquinas	Cantidad total de leche envasada por las máquinas (miles de litros)
1	20	20
2	40	25
3	30	22
4	40	24
5	20	20
6	70	30
7	70	28
8	50	25

Se pide:

a) Proporcione una estimación insesgada del total de leche envasada y una estimación de su varianza.

b) Si se conoce que la empresa láctea tiene 4000 máquinas en total, obtenga una estimación del total de leche envasada por un método diferente al del apartado anterior.

Solución.

a) El estimador del total poblacional dado por

$$\widehat{X}_{HH} = \sum_{i=1}^{n} \frac{X_i}{n \cdot P_i}$$

es un estimador insesgado del total de leche envasada al mes. Se tiene que

$$N = 100 \qquad n = 8 \qquad P_i = \frac{1}{N}$$

y por tanto

$$\widehat{X}_{HH} = \sum_{i=1}^{n} \frac{X_i}{\frac{n}{N}} = \sum_{i=1}^{8} \frac{20 + 25 + 22 \cdots + 25}{\frac{8}{100}} = 2425.$$

Una estimación de su varianza será:

$$\begin{aligned}
\widehat{Var}(\widehat{X}_{HH}) &= \frac{1}{n(n-1)} \sum_{i=1}^{n} \left(\frac{X_i}{\frac{1}{N}} - \widehat{X}_{HH} \right)^2 \\
&= \frac{1}{8 \cdot 7} \left[(100 \cdot 20 - 2425)^2 + \cdots + (100 \cdot 25 - 2425)^2 \right] \\
&= \frac{895000}{56} = 15982,142
\end{aligned}$$

b) Una estimación del total de leche envasada por un método diferente al del apartado anterior se podría obtener utilizando el método de la razón. Así:

$$\widehat{X}_R = M\widehat{R} = M\bar{\bar{x}} = M \frac{\sum_{i=1}^{8} X_i}{\sum_{i=1}^{8} M_i} = 4000 \cdot \frac{20 + 40 + \cdots + 25}{20 + 25 + \cdots + 50} =$$
$$= 4000 \cdot \frac{194}{340} = 2282,35.$$

Cuestión 4

Dada la siguiente distribución bidimensional de frecuencias:

$X \backslash Y$	0	1	2
0	15	15	10
1	5	20	5
2	10	5	15

Se pide calcular:

a) La media aritmética y la varianza de la variable $X - Y$

b) La recta de regresión de Y sobre X y la varianza residual de Y

Solución.

a) En la tabla 4.10 se recoge la distribución de la variable diferencia. A partir de ella, se calculan la media aritmética y la varianza.

Tabla 4.10: Distribución de la variable $Z = X - Y$.

$Z = X - Y$	n_i
-2	10
-1	20
0	50
1	10
2	10
	100

$$\overline{z} = \sum_{i=1}^{5} z_i \frac{n_i}{N} = \frac{(-2) \cdot 10 + (-1) \cdot 20 + \cdots + 2 \cdot 10}{100} = -0,1$$

$$S_Z^2 = \sum_{i=1}^{3} z_i^2 \frac{n_i}{N} - \overline{z}^2 = \frac{(-2)^2 \cdot 10 + (-1)^2 \cdot 20 + \cdots + 2^2 \cdot 10}{100} - (-0,1)^2 = 1,09$$

b) Para el cálculo de la recta de regresión de Y sobre X, es necesario conocer los siguientes valores relativos a las variables X e Y:

$$y = \overline{y} + \frac{S_{XY}}{S_X^2}(x - \overline{x})$$

Se completa la tabla relativa a la distribución bidimensional de frecuencias con una fila y una columna para las distribuciones marginales:

$X \backslash Y$	0	1	2	n_i
0	15	15	10	40
1	5	20	5	30
2	10	5	15	30
n_j	30	40	30	N = 100

A partir de la infromación anterior,

$$\overline{x} = \sum_{i=1}^{3} x_i \frac{n_i}{N} = \frac{0 \cdot 40 + 1 \cdot 30 + 2 \cdot 30}{100} = 0,9$$

$$\overline{y} = \sum_{j=1}^{3} y_j \frac{n_j}{N} = \frac{0 \cdot 30 + 1 \cdot 40 + 2 \cdot 30}{100} = 1$$

$$S_{XY} = \sum_{i=1}^{3} \sum_{j=1}^{3} x_i y_j \frac{n_{ij}}{N} - \overline{x}\overline{y} = $$

$$= \frac{0 \cdot 0 \cdot 15 + 0 \cdot 1 \cdot 15 + \cdots + 2 \cdot 2 \cdot 15}{100} - 0,9 \cdot 1 = 0,1$$

$$S_X^2 = \sum_{i=1}^{3} x_i^2 \frac{n_i}{N} - \overline{x}^2 = \frac{0^2 \cdot 40 + 1^2 \cdot 30 + 2^2 \cdot 30}{100} - 0,9^2 = 0,69$$

y se tiene la recta de regresión

$$y = 1 + \frac{0,1}{0,69}(x - 0,9) \Leftrightarrow y = 0,869 + 0,145x$$

Para el cálculo de la varianza residual de Y se utiliza

$$S^2_{r_Y} = \left(1 - r^2\right) S^2_Y$$

donde

$$S^2_Y = \sum_{j=1}^{3} y_j^2 \frac{n_j}{N} - \overline{y}^2 = \frac{0^2 \cdot 30 + 1^2 \cdot 40 + 2^2 \cdot 30}{100} - 1^2 = 0,6$$

$$r^2 = \frac{S^2_{XY}}{S^2_X S^2_Y} = \frac{(0,1)^2}{0,69 \cdot 0,6} = \frac{10}{414}$$

y por tanto, la varianza residual de Y será

$$S^2_{r_Y} = \left(1 - \frac{10}{414}\right) 0,6 = 0,5855.$$

Cuestión 5

El importe medio y el número de hipotecas constituidas sobre fincas rústicas, viviendas y otras fincas, inscritas en los registros de la propiedad en el periodo 2015-2017 se recogen en la tabla 4.11. A partir de esos datos, se pide:

a) Calcular el índice de Laspeyres del Importe medio de las hipotecas entre 2016 y 2017, tomando como año base 2015.

b) Determinar la variación interanual del índice de Laspeyres entre 2016 y 2015.

c) ¿Cuál es la repercusión de la componente individual "Viviendas" en la variación del índice de Laspeyres calculado anteriormente, entre 2015 y 2016?

d) Calcular la participación de las viviendas en el cambio total experimentado en el índice general de 2016.

Tabla 4.11: Importe medio y número de hipotecas (2015-2017)

Hipotecas constituidas	Fincas rústicas		Viviendas		Otras fincas	
	Importe medio (miles €)	Número (miles)	Importe medio (miles €)	Número (miles)	Importe medio (miles €)	Número (miles)
2015	154,5	18,7	105,9	246,7	214,5	102,5
2016	165,7	16,7	109,8	282,7	195,3	101,4
2017	152,5	16,5	116,7	310,1	185,6	106,5

Solución.

a) El índice de Laspeyres del importe medio de las hipotecas entre 2016 y 2017, tomando como año base 2015, será

$$\begin{aligned}
L_{p_{15}}^{16-17} &= \frac{\sum_{i=1}^{3} \frac{p_{i17}}{p_{i16}} p_{i15} q_{i15}}{\sum_{i=1}^{3} p_{i15} q_{i15}} \cdot 100 = \\
&= \frac{\frac{152,5}{165,7} 154,5 \cdot 18,7 + \frac{116,7}{109,8} 105,9 \cdot 246,7 + \frac{185,6}{195,3} 214,5 \cdot 102,5}{154,5 \cdot 18,7 + 105,9 \cdot 246,7 + 214,5 \cdot 102,5} \cdot 100 = \\
&= \frac{51320,543}{51000,93} \cdot 100 = 1,0062 \cdot 100 = 100,62
\end{aligned}$$

b) Por otra parte, el índice de Laspeyres de 2016 con base en 2015 es:

$$\begin{aligned}
L_{p_{15}}^{16} &= \frac{\sum_{i=1}^{3} p_{i16} q_{i15}}{\sum_{i=1}^{3} p_{i15} q_{i15}} \cdot 100 = \\
&= \frac{165,7 \cdot 18,7 + 109,8 \cdot 246,7 + 195,3 \cdot 102,5}{51000,93} \cdot 100 = \\
&= \frac{50204,5}{51000,93} \cdot 100 = 0,98438401 \cdot 100 = 98,438401
\end{aligned}$$

y la variación interanual del índice de Laspeyres entre 2016 y 2015 es

$$TVI_{15}^{16} = \frac{98,438401}{100} - 1 = -0,01562 = -1,562\,\%$$

c) La repercusión de la componente individual “Viviendas” en la variación del índice de Laspeyres calculado entre 2015 y 2016, se calcula utilizando como ponderación $w_i = p_{i0}q_{i0}$ ya que

$$L_{p_0}^t = \frac{\sum_{i=1}^n p_{it}q_{i0}}{\sum_{i=1}^n p_{i0}q_{i0}} = \frac{\sum_{i=1}^n \frac{p_{it}}{p_{i0}} p_{i0}q_{i0}}{\sum_{i=1}^n p_{i0}q_{i0}} = \frac{\sum_{i=1}^n I_i w_i}{\sum_{i=1}^n w_i}.$$

Así,

$$R_{viviendas} = \frac{\Delta I_{viviendas} w_{viviendas}}{\sum_{i=1}^3 w_i} = \frac{(103,68 - 100)(246,7 \cdot 105,9)}{\sum_{i=1}^3 w_i} =$$
$$= \frac{3,68 \cdot 26125,53}{26125,53 + 2889,15 + 21986,25} = \frac{3,68 \cdot 26125,53}{51000,93} = 1,886$$

Por tanto, dicha repercusión en la variación del índice de Laspeyres es 1,866, es decir, mide cuánto variaría el índice general si únicamente variase el precio de las viviendas manteniendo el resto de productos el precio sin variación. Siguiendo cálculos análogos, recogidos en la tabla 4.12, se obtiene la repercusión del resto de componentes.

Tabla 4.12: Repercusión de variación de cada componente en el índice general

	Índice 2015 (I_i)	Ponderación (w_i)	Índice 2016 $(I_i + \Delta I_i)$	$R_i = \frac{\Delta I_i w_i}{\sum_{i=1}^3 w_i}$
Viviendas	100	$246,7 \cdot 105,9 =$ 26125,53	$\frac{109,8}{105,9} \cdot 100 =$ 103,68	$\frac{3,68 \cdot 26125,53}{51000,93} =$ 1,886
Fincas rústicas	100	$18,7 \cdot 154,5 =$ 22889,15	$\frac{165,7}{154,5} \cdot 100 =$ 107,25	$\frac{7,25 \cdot 2889,15}{51000,93} =$ 0,411
Otras fincas	100	$102,5 \cdot 214,5 =$ 21986,25	$\frac{195,3}{214,5} \cdot 100 =$ 91,05	$\frac{-8,95 \cdot 21986,25}{51000,93} =$ $-3,858$
	$L_{p15}^{15} =$ 100	$\sum_{i=1}^3 w_i =$ 51000,93	$L_{p15}^{16} =$ 98,43	$\sum_{i=1}^3 R_i =$ $-1,562$

El índice en base 2015 aumentó su valor en 1,886 puntos al pasar del año 2015 al 2016 a causa de la subida de precios de las viviendas, y también aumentó su valor en 0,411 puntos a causa de la subida de precios de las fincas rústicas, pero disminuyó en cambio 3,858 puntos a causa de la bajada de precios de las otras fincas. La bajada de las otras fincas ha compensado la subida de las viviendas y las fincas rústicas de forma que globalmente ha hecho que el índice disminuya y su variación sea negativa. En este caso, una parte de la variación absoluta del índice, que es -1,562 se debe a subidas (1,886+0,411=2,297) y otra parte se debe a bajadas (-3,858), así que en total hubo un movimiento de 6,155.

d) La participación de un componente, de forma general, trata de medir dentro de la variación total de un índice, qué porcentaje es debido a dicho componente.

La participación de las "Viviendas" en el cambio total experimentado en el índice general de 2016 será:

$$P_{15}^{15\to 16}(viviendas) = \frac{R_{viviendas}}{\sum_{i=1}^{3} R_i} = \frac{1,886}{1,886 + 0,411 - 3,858} = -1,2074$$

Si calculamos la participación de las fincas rústicas en la variación del índice obtenemos -0,2631 y la de las otras fincas 2,4699. En cuanto a la participación, por tratarse de un porcentaje o una proporción de la tasa de variación, tiene una interpretación clara si todas las repercusiones son positivas o todas son negativas, pero se difumina y se complica cuando como en este caso existen repercusiones de distinto signo.

En el caso de que un índice construido con una serie de componentes tuviera una variación nula, lo que ocurriría es que no se podrían calcular proporciones con respecto a la variación (no se puede dividir por 0), y en ese caso se podría concluir simplemente que como no hubo variación, no hubo participaciones en esa variación. Sin embargo, aunque no hubiera variación en conjunto, sí que hubo una subida (debida a ciertos artículos) compensada con una bajada (debido a otros), y puede ser interesante ver qué proporción de la subida/bajada o movimiento general se debe a cada bien. Por este motivo se puede hablar de la participación en la subida absoluta de un índice, en la bajada absoluta de un índice o en el movimiento general del índice.

La participación de las "Viviendas" en el movimiento general del índice entre 2015 y 2016, se calcula dividiendo la repercusión absoluta del componente en cuestión entre la suma de las repercusiones absolutas sin signo, que vale 6,155:

$$P_{15}^{\updownarrow 15\to 16}(viviendas) = \frac{R_{viviendas}}{\sum_{i=1}^{3}|R_i|} = \frac{1,886}{1,886+0,411+3,858} = 0,3064$$

, es decir, que el 30,64 % del movimiento general se debe a las viviendas. Al ser positivo, esa parte del movimiento es una subida. De igual forma se puede calcular la participación de las fincas rústicas y otras fincas en el movimiento general del índice, resultando respectivamente:

$$P_{15}^{\updownarrow 15\to 16}(\text{fincas rústicas}) = \frac{R_{\text{fincas rústicas}}}{\sum_{i=1}^{3}|R_i|} = \frac{0,411}{1,886+0,411+3,858} = 0,0667$$

$$P_{15}^{\updownarrow 15\to 16}(\text{otras fincas}) = \frac{R_{\text{otras fincas}}}{\sum_{i=1}^{3}|R_i|} = \frac{-3,858}{1,886+0,411+3,858} = -0,6268$$

y la interpretación es que el 6,67 % del movimiento general se debe a las fincas rústicas y esta parte es una subida y el 62,68 % del movimiento se debe a las otras fincas, y al ser negativa, esta parte del movimiento corresponde a una bajada.

La participación de las "Viviendas" en la subida absoluta del índice sería:

$$P_{15}^{\uparrow 15\to 16}(viviendas) = \frac{1,886}{1,886+0,411} = 0,8210$$

, es decir, el 82,10 %, puesto que de los tres componentes hay dos de ellos que tienen repercusiones positivas y el tercero la tiene negativa.

La participación de las "Viviendas" en la bajada absoluta del índice sería:

$$P_{15}^{\downarrow 15\to 16}(viviendas) = \frac{1,886}{3,858} = 0,4888$$

, es decir, el 48,88 %.

Cuestión 6

A partir de la información recogida en la tabla 4.13, extraída de la Contabilidad Nacional Anual del año 2015, correspondiente al sector de Administraciones Públicas, se pide:

a) Calcular la producción del sector.

b) Calcular el excedente neto de explotación.

c) Calcular el consumo final efectivo.

d) Sabiendo que el ahorro bruto del sector es -28.617 millones de euros, calcule la capacidad o necesidad de financiación del sector.

Tabla 4.13: Datos del sector de administraciones públicas (S.13) *(Cifras en millones de euros)*

Formación bruta de capital	26.965
Remuneración de asalariados	119.356
Impuestos netos sobre la producción y las importaciones pagados	476
Impuestos netos sobre la producción y las importaciones recibidos	127.499
Transferencias de capital a cobrar	10.330
Transferencias de capital a pagar	10.826
Consumo de capital fijo	27.628
Gasto en consumo final individual	119.628
Gasto en consumo final colectivo	89.295
Transferencias sociales en especie: producción adquirida en el mercado	28.202
Consumos intermedios	57.474
Adquisiciones menos cesiones de activos no producidos	926
Préstamos recibidos	1.250
Valores representativos de deuda emitidos	55
Participaciones en el capital y en fondos de inversión	180

Solución.

a) Primeramente detallamos las cuentas de explotación y de producción para calcular la producción de este sector.

E	Cuenta de explotación		R
119.356	Remuneración de asalariados	Valor añadido bruto	147.460
476	Impuestos netos sobre la producción e importaciones		
27.628	EBE $\equiv$CCF		

E	Cuenta de producción		R
57.474	Consumos intermedios	Producción	204.934
147.460	Valor añadido Bruto (B1.b)		

b) En el sector de las Administraciones Públicas el Excedente Bruto de Explotación y el Consumo de Capital Fijo coinciden, por tanto el Excedente Neto de Explotación será nulo.

c) El Consumo Final Efectivo coincide con el Gasto en Consumo Final colectivo, es decir, 89.295.

d) Ddado que el Ahorro Bruto del sector es -28.617, su capacidad o necesidad de financiación (D.9), proporcionada por las Cuentas de Acumulación, será:

$$\text{D.9} + \text{FBC - CCF} + \text{Adq acts no fin} =$$
$$= \text{AB - CCF} + \text{transf. de capital a cobrar - transf. de capital a pagar}$$
$$\text{D.9} + 26.965 - 27.628 + 926 =$$
$$= -\ 28.617 - 27.628 + 10.330 - 10.826$$
$$\text{D.9} = -\ 57.004$$

Esto es, hay una necesidad de financiación de 57.004 millones de euros.

Cuestión 7

En una economía con dos ramas de actividad, se dispone del extracto de la tabla input-output simétrica: recogido en la tabla 4.14 (Cifras en millones de euros).

Tabla 4.14: Extracto tabla input-output simétrica

	Consumos intermedios		Gasto en consumo final (GCF)	Formación bruta de capital (FBC)	Exportaciones
	Rama 1	Rama 2			
Producto 1	30	10	50	10	20
Producto 2	10	40	20	20	10

Además, se proporcionan los siguientes datos para completar la tabla:
- La producción de la rama 1 (producto 1) es de 100 millones de euros y la de la rama 2 (producto 2) es de 50 millones de euros.
- La remuneración de asalariados ha sido de 45 millones de euros en la rama 1 y 50 millones de euros en la rama 2.
- Los impuestos netos sobre la producción pagados por la rama 1 ascienden a 5 millones de euros, mientras que en el caso de la rama 2 ascienden a 10 millones.
- No existen impuestos sobre los productos ni márgenes de distribución (comercio y transporte) en esta economía.
- Todas las exportaciones se consideran de servicios de mercado. No existen exportaciones de bienes en esta economía.

Se pide:

a) Calcular las importaciones realizadas de cada producto (a excepción de las exportaciones e importaciones, suponga que todas las demás operaciones del Resto del Mundo se consideran nulas).

b) Calcular el Excedente Bruto de Explotación de la rama 2 y explique su significado económico.

c) Calcular el Producto Interior Bruto de la economía a partir de la tabla, por las tres vías (demanda, oferta, rentas).

d) Calcular la matriz de coeficientes técnicos. Interprete el resultado obtenido en la primera fila de esa matriz.

Solución.

a) La tabla input-output simétrica completa se recoge en la tabla 4.15, donde se aprecia que las importaciones de cada rama son 20 y 50 millones de euros respectivamente.

b) El Excedente Bruto de Explotación de la rama 2 es -60. El hecho de que su signo sea negativo implica que hay "déficit" bruto de explotación, es decir, la suma de los consumos intermedios, la remuneración de asalariados y los impuestos son mayores que la producción.

c) El cálculo del PIB por las tres vías se tiene:

$$\begin{aligned} PIB_{\text{producción}} &= VAB + \textit{impuestos netos sobre los productos} = \\ &= 60 + 0 = 60 \ \textit{millones de €} \\ PIB_{\text{renta}} &= RA + EBE/RM + \textit{impuestos netos sobre la producción} = \\ &= 95 - 50 + 15 = 60 \ \textit{millones de €} \\ PIB_{\text{gasto}} &= GCF + FBC + X - M = \end{aligned}$$

$$= 70 + 30 + 30 - 70 = 60 \ millones \ de \ €$$

d) La matriz de coeficientes técnicos será:

$$A = \begin{pmatrix} \frac{30}{100} & \frac{10}{50} \\ \frac{10}{100} & \frac{40}{50} \end{pmatrix} = \begin{pmatrix} 0,3 & 0,2 \\ 0,1 & 0,8 \end{pmatrix}$$

donde los valores 100 y 50 de los denominadores son la producción a precios básicos en las ramas 1 y 2 respectivamente (como se recoge en la tabla 4.15). El elemento $a_{11} = 0,3$ significa que existen unos consumos intermedios de 0,3 euros en la rama 1 por cada euro que se produce en la rama 1. El elemento $a_{12} = 0,2$ indica que para producir cada euro de producto en la rama 2 se necesitan 0,2 euros de consumos intermedios en productos de la rama 1.

Tabla 4.15: Tabla input-output simétrica completa.

T.D.I.(Total Demanda Intermedia); T.D.F.(Total Demanda Final)

Las celdas sombreadas en color amarillo son los datos proporcionados por el enunciado.

	Rama 1	Rama 2	T.D.I.	GCF	FBC	X	T.D.F.	Total empleos
Producto 1	30	10	40	50	10	20	80	120 (40+80)
Producto 2	10	40	50	20	20	10	50	100 (50+50)
Total (p.b.)	40	50	90	70	30	30	130	220 (90+130)
Imps. netos productos	-	-	-	-	-	-	-	-
Total (p.adq.)	40	50	90	70	30	30	130	220 (90+130)
CI (p.adq.)	40	50	90					
RA	45	50	95					
Otros imps. netos sobre producción	5	10	15					
EBE/RM	10	-60	-50					
VAB (p.b.)	60	0	60					
Producción (p.b.)	100	50	150					
Importaciones	20	50	70					
Oferta (p.b.)	120	100	220					

Cuestión 8

Con los datos en millones de euros de la Contabilidad Nacional de España para el año 2010 recogidos en la tabla 4.16, calcule:

a) El Excedente de Explotación / Renta Mixta Bruto.

b) La Renta Nacional Bruta disponible a precios de mercado.

c) El Ahorro Nacional Neto.

d) La capacidad o necesidad de financiación del país. Interprete el resultado.

Tabla 4.16: Datos de la Contabilidad Nacional de España para el año 2010

Producto interior bruto a precios de mercado	1.080.913
Consumo de capital fijo	182.025
Formación bruta de capital	254.549
Adquisiciones menos cesiones de activos no financieros no producidos	-119
Remuneración de asalariados (interior)	541.475
Rentas primarias netas del exterior	-15.155
Transferencias de capital netas con el exterior	5.732
Transferencias corrientes netas con el exterior	-12.718
Impuestos netos sobre la producción y las importaciones	93.559
Consumo final efectivo	840.470

Solución.

- EBE/RM = PIB – RA(interior) – impuestos netos sobre la producción y las importaciones = 1.080.913 – 541.475 – 93.559 = 445.879 millones de €.

- RNB = PIB + rentas primarias netas del exterior = 1.080.913 – 15.155

= 1.065.758 ⇒ RNBD = RNB + transferencias corrientes netas con el exterior = 1.065.758 – 12.718 = 1.053.040 millones de €.

- ANB = RNBD – CFE = 1.053.040 – 840.470 = 212.570 millones de €. ⇒ ANN = ANB – CCF = 212.570 – 182.025 = 30.545 millones de €.
- Capacidad o necesidad de financiación = ANN + transferencias de capital netas con el exterior – (FBC – CCF + ActNoProd's) = 30.545 +5.732 – (254.549 – 182.025 – 119) = – 36.128 millones de €. Al tener signo negativo significa que la economía necesita financiación, es decir, tiene que recurrir al endeudamiento.

Cuestión 9

Los datos de la tabla 4.17 han sido tomados de un Censo de Población de referencia 31/12/1991. La estadística de Movimiento Natural de Población proporciona la tabla 4.18 sobre defunciones, relativa a la generación de 1990. Suponiendo que el efecto migratorio es nulo, calcular el número de efectivos de la generación de 1990 y cuántos de éstos sobrevivieron a su segundo aniversario.

Tabla 4.17: Población 31/12/1991

Año de nacimiento	Población
1989	59.984
1990	60.120
1991	60.850

Tabla 4.18: Defunciones generación 1990

Edad	Año	Defunciones
0	1990	104
0	1991	95
1	1991	70
1	1992	45
2	1992	27
2	1993	20

Solución.

Los efectivos de la generación de 1990 son:

$$60120 + 70 + 95 + 104 = 60389$$

puesto que se suman a la población existente a 31/12/1991 las personas de la generación que han fallecido hasta ese momento.

Los efectivos de la generación de 1990 que sobreviven a su segundo aniversario son:

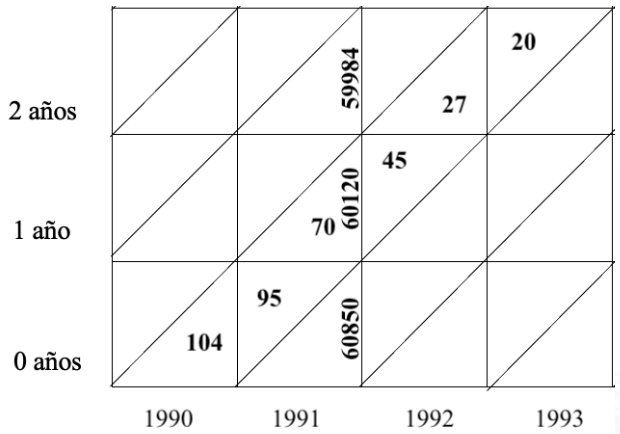


Figura 4.1: Representación en un diagrama de Lexis de los datos de defunciones y población.

$$60120 - (45 + 27 + 20) = 60028.$$

Cuestión 10

La tabla 4.19 presenta, por grupos de edad y sexo, datos correspondientes a la población de un país a 1 de Enero del año 2018 (en millones de habitantes). Complete la tabla 4.19, conocidos los siguientes indicadores:

- Índice de envejecimiento de la población (ambos sexos) = 121,6 %
- Ratio de masculinidad de la población mayor de 64 años = 76,5 %
- Tasa de dependencia de la población masculina menor de 16 años = 25,0 %
- Tasa de dependencia de la población mayor de 64 años (ambos sexos) = 29,7 %

(Nota: es suficiente con que los resultados se aproximen con un solo decimal).

Tabla 4.19: Datos de población a 1 de Enero del año 2018. En millones de habitantes.

Edad	Hombres	Mujeres	Ambos Sexos
0-15	?	?	?
16-64	?	15,1	?
65 y más años	3,9	?	?
Total	?	?	46,7

Solución. Se sabe que

$$\textit{ratio masculinidad}_{>64} = \frac{P_{>64}^{hombres}}{P_{>64}^{mujeres}} = 0,765,$$

luego

$$P_{>64}^{mujeres} = \frac{P_{>64}^{hombres}}{\textit{ratio masculinidad}_{>64}} = \frac{3,9}{0,765} = \underset{(1)}{5,09}$$

Por otra parte,

$$I_{envejecimineto}^{ambos\ sexos} = \frac{P_{>64}^{ambos\ sexos}}{P_{<16}^{ambos\ sexos}} = 1,216$$

donde

$$P_{>64}^{ambos\ sexos} = 3,9 + \underset{(1)}{5,09} \simeq \underset{(2)}{9},$$

por tanto,

$$P_{<16}^{ambos\ sexos} = \frac{P_{>64}^{ambos\ sexos}}{I_{envejecimineto}^{ambos\ sexos}} = \frac{9}{1,216} = \underset{(3)}{7,4}$$

Siendo

$$\textit{tasa de dependencia}_{>64}^{ambos\ sexos} = \frac{P_{>64}^{ambos\ sexos}}{P_{16-64}^{ambos\ sexos}} = 0,297,$$

entonces

$$P_{16-64}^{ambos\ sexos} = \frac{P_{>64}^{ambos\ sexos}}{\textit{tasa de dependencia}_{>64}^{ambos\ sexos}} = \frac{9,0}{0,297} = \underset{(4)}{30,3}.$$

Finalmente, dado que

$$\textit{tasa de dependencia}_{<16}^{hombres} = \frac{P_{<16}^{hombres}}{P_{16-64}^{hombres}} = 0,25,$$

será

$$P_{<16}^{hombres} = \textit{tasa de dependencia}_{<16}^{hombres} \cdot P_{16-64}^{hombres} = 0,25 \cdot \underset{(5)}{15,2} = \underset{(6)}{3,8}$$

donde

$$P_{16-64}^{hombres} = P_{16-64}^{ambos\ sexos} - P_{16-64}^{mujeres} = \underset{(4)}{30,3} - 15,1 = \underset{(5)}{15,2}.$$

Mediante diferencias se obtiene el resto de resultados recogidos en la tabla 4.20.

Tabla 4.20: Datos de población a 1 de Enero del año 2018. Los números entre paréntesis corresponden al orden en el que se ha obtenido ese dato con la información disponible.

Edad	Hombres	Mujeres	Ambos Sexos
0-15	$\underset{(6)}{3,8}$	$\underset{(7)}{3,6}$	$\underset{(3)}{7,4}$
16-64	$\underset{(5)}{15,2}$	15,1	$\underset{(4)}{30,3}$
65 y más años	3,9	$\underset{(1)}{5,1}$	$\underset{(2)}{9}$
Total	$\underset{(8)}{22,9}$	$\underset{(9)}{23,8}$	46,7

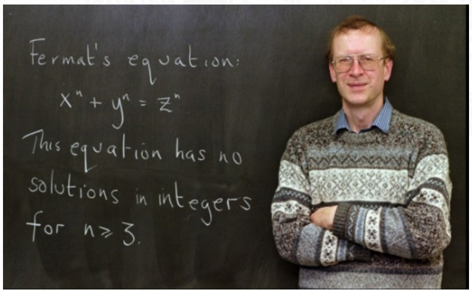
Capítulo 5

Año 2016

"Mathemathics is the queen of the science."

—Carl Friedrich Gauss

Matemático inglés, nacido en Cambridge el 11 de abril de 1953. En 1971 Wiles estudió en el Merton College de Oxford. En 1980 leyó su tésis doctoral, tras tres años como profesor asistente de Benjamin Pierce en la Universidad de Harvard. En 1987 supo que Ken Ribet había demostrado que el último teorema de Fermat era una consecuencia de la conjetura sobre curvas elípticas que Shimura y Taniyama habían enunciado años antes. En ese momento dedicó todos sus esfuerzos a la demostración de la conjetura de Shimura-Taniyama cuya consecuencia sería la prueba del enunciado de Fermat. Tras siete años de intenso trabajo, el primer intento de demostración lo presentó en una serie de lecturas realizadas en el Instituto Isaac Newton de Cambridge el 23 de Junio de 1993. Tras la lectura final, Wiles anunció que había probado el último teorema de Fermat, pero tras la publicación de los resultados y el exhaustivo análisis por parte de la comunidad matemática terminaron por encontrarse algunos errores e incosistencias.
Ante esta situación Andrew Wiles se negó a continuar con su trabajo, pero gracias a R. Taylor, un estudiante que le alentó para que continuara el camino que había empezado e intentase resolver los errores que contenía su anterior demostración, finalmente el 19 de Septiembre de 1994 solventó definitivamente los problemas y llegó a una demostración definitiva



Andrew Wiles

$2016 = 3^3 + 4^3 + \cdots + 9^3 = 6!!(6!! - 6) = 2^5 \cdot 3^2 \cdot 7 = \binom{64}{2}$

It is an evil number, because the sum of its binary digits is even.

5.1. Soluciones del segundo examen del Cuerpo Superior de Estadísticos del Estado de 2016

Cuestión 1

1. Sea una población de M=40.000 individuos distribuida en N=10.000 hogares. Notamos por:

   M_i = tamaño del hogar i-ésimo, (medido en número de individuos)

   A_i = número de mujeres en el hogar i-ésimo

   A = total de mujeres en la población

   Se obtiene una muestra aleatoria simple (sin reemplazamiento) de n=20 hogares y se encuesta a todos los individuos de cada hogar seleccionado. Se pide:

   a) Una estimación insesgada de A y su error de muestreo.

   b) Una estimación de la eficiencia relativa del muestreo por conglomerados respecto a la del muestreo aleatorio simple.

   c) Una estimación de A y su error de muestreo, aplicando el estimador de razón al tamaño.

   d) Comentar las ventajas e inconvenientes del estimador c) respecto al a).

   Los datos necesarios para los cálculos son los siguientes:

   $$\sum_{i=1}^{n} M_i = 69; \quad \sum_{i=1}^{n} A_i = 35; \quad \sum_{i=1}^{n} M_i^2 = 269; \quad \sum_{i=1}^{n} A_i^2 = 81; \quad \sum_{i=1}^{n} M_i \cdot A_i = 142$$

2. Los individuos de la población del apartado anterior se distribuyen en 200 secciones censales de 50 hogares cada una. Se quiere estimar el total

de mujeres (A) en la población y para ello, se obtiene una muestra de 4 secciones censales, de las que se obtiene a su vez una submuestra de 5 hogares de cada sección censal seleccionada. El muestreo se realiza con reposición y probabilidades iguales en ambas etapas. Calcular una estimación insesgada de A y su error de muestreo.

Los datos necesarios para los cálculos son los siguientes:

Secciones censales	1	2	3	4
Número de mujeres en cada sección censal	6	8	11	10

Solución.

1.a) Se tiene un población formada por $M = 40000$ individuos y se considera que estos están divididos en $N = 10000$ hogares. Dado que no se dispone de información para el tamaño de cada uno de los hogares, se considera que los conglomerados no varían mucho en tamaño de modo que se supone que cada uno de los hogares estará formado por un número similar de individuos y se consideran los tamaños M_i de cada uno de los hogares como la media de los tamaños de los conglomerados:

$$\overline{M} = \frac{\sum_{i=1}^{N} M_i}{N} = \frac{40000}{10000} = 4.$$

Por tanto se utilizarán las fórmulas habituales para conglomerados con el mismo tamaño $\overline{M} = 4$. Sabiendo que se obtiene una muestra aleatoria simple (sin reemplazamiento) de $n = 20$ hogares, un estimador insesgado del total de clase A, que representa el total de mujeres en la población, vendrá dado por:

$$\widehat{A}_{\scriptscriptstyle MCU} = \frac{N}{n} \sum_{i=1}^{n} \sum_{j=1}^{\overline{M}} A_{ij} = \frac{N}{n} \sum_{i=1}^{n} A_i = \frac{10000}{20} 35 = 17500.$$

, donde el subíndice MCU significa muestreo de conglomerados unietápico o sin submuestreo.

Para el cálculo de la estimación insesgada de su error de muestreo se obtiene de la raíz cuadrada de la estimación insesgada de la varianza de A_{MCU}. Esta se calcula como:

$$\begin{aligned}
\widehat{Var}_{MCU}\left(\widehat{A}_{MCU}\right) &= N^2\overline{M}^2\widehat{Var}\left(\widehat{P}\right) = \\
&= N^2\overline{M}^2(1-f)\frac{\widehat{S}_b^2}{n\overline{M}} = \\
&= \frac{N^2\overline{M}^2(1-f)}{n\overline{M}}\frac{\overline{M}}{n-1}\sum_{i=1}^{n}\left(P_i-\overline{P}\right)^2 = \\
&= \frac{N^2\overline{M}^2(1-f)}{n(n-1)}\sum_{i=1}^{n}\left(P_i-\overline{P}\right)^2 .
\end{aligned}$$

Siendo

$$P_i = \frac{\sum_{j=1}^{\overline{M}} A_{ij}}{\overline{M}} = \frac{A_i}{\overline{M}},$$

$$\overline{P} = \frac{1}{n}\sum_{i=1}^{n} P_i = \frac{1}{n\overline{M}}\sum_{i=1}^{n}\sum_{j=1}^{\overline{M}} A_{ij} = \frac{1}{n\overline{M}}\sum_{i=1}^{n} A_i \Leftrightarrow \overline{M}\cdot\overline{P} = \frac{1}{n}\sum_{i=1}^{n} A_i$$

se tiene

$$\begin{aligned}
\sum_{i=1}^{n}\left(P_i-\overline{P}\right)^2 &= \sum_{i=1}^{n}\left(\frac{A_i}{\overline{M}}-\frac{1}{n\overline{M}}\sum_{i=1}^{n}A_i\right)^2 = \\
&= \left[\frac{1}{\overline{M}}\right]^2\sum_{i=1}^{n}\left(A_i-\overline{M}\cdot\overline{P}\right)^2 = \\
&= \left[\frac{1}{\overline{M}}\right]^2\sum_{i=1}^{n}\left(A_i^2+\overline{M}^2\overline{P}^2-2\overline{M}\cdot\overline{P}A_i\right) = \\
&= \left[\frac{1}{\overline{M}}\right]^2\left(\sum_{i=1}^{n}A_i^2+n\overline{M}^2\cdot\overline{P}^2-2\overline{M}\cdot\overline{P}\sum_{i=1}^{n}A_i\right) = \\
&= \left[\frac{1}{\overline{M}}\right]^2\left(\sum_{i=1}^{n}A_i^2-n\overline{M}^2\cdot\overline{P}^2\right)
\end{aligned}$$

utilizando en el último paso que

$$\sum_{i=1}^{n} A_i = \sum_{i=1}^{n}\sum_{j=1}^{\overline{M}} A_{ij} = n\overline{M} \cdot \overline{P}.$$

Sustituyendo en la expresión del cálculo de la estimación de la varianza del número de mujeres en la población estimado se tiene:

$$\begin{aligned}
\widehat{Var}_{_{MCU}}\left(\widehat{A}_{_{MCU}}\right) &= \frac{N^2\overline{M}^2(1-f)}{n(n-1)}\left[\frac{1}{\overline{M}}\right]^2\left(\sum_{i=1}^{n} A_i^2 - n\overline{M}^2\overline{P}^2\right) = \\
&= \frac{N^2(1-f)}{n(n-1)}\left(\sum_{i=1}^{n} A_i^2 - n\overline{M}^2\overline{P}^2\right).
\end{aligned}$$

que con los datos del problema será:

$$\begin{aligned}
\widehat{Var}_{_{MCU}}\left(\widehat{A}_{_{MCU}}\right) &= \frac{10000^2\left(1-\frac{20}{10000}\right)}{20(19)}\left(81 - 20\cdot 4^2\cdot\left(\frac{35}{80}\right)^2\right) = \\
&= \frac{10000^2\frac{499}{5000}}{380}\left(81 - \frac{245}{4}\right) = \\
&= \frac{10000^2\cdot 499}{190000}\left(\frac{79}{4}\right) = \\
&= 5186973,684.
\end{aligned}$$

Así, la estimación insesgada del error de muestreo del total de la población estimado será:

$$\widehat{\sigma}_{_{MCU}}\left(\widehat{A}_{_{MCU}}\right) = \sqrt{\widehat{Var}_{_{MCU}}\left(\widehat{A}_{_{MCU}}\right)} = \sqrt{5186973,684} = 2277,49.$$

1.b) Para llevar a cabo la comparación de la eficiencia relativa del muestreo de conglomerados y el muestreo aleatorio simple una posibilidad es calcular la varianza en el caso de realizar muestreo aleatorio simple y compararla con la varianza obtenida en el apartado anterior. Esta comparación es posible ya que ambos estimadores son insesgados. Para un muestreo aleatorio simple el estimador de la proporción de mujeres se calcula como:

$\widehat{P}_{Mujeres} = \frac{\sum_{i=1}^{n} A_i}{\sum_{i=1}^{n} M_i} = \frac{35}{69} = 0,507$, por lo que el estimador del total de mujeres será:

$$\hat{A}_{_{MAS}} = N \cdot \overline{M} \cdot \widehat{P}_{Mujeres} = 10000 \cdot 4 \cdot 0,507 = 20289,85$$

y la varianza sería con la fórmula habitual, pero interpretando de forma correcta los datos de la misma según la información que nos proporciona el enunciado:

$$\begin{aligned}
\widehat{Var}_{_{MAS}}(\hat{A}_{_{MAS}}) &= N'^2 \left(1 - \frac{n^{'}}{N'}\right) \frac{\hat{P}\hat{Q}}{n' - 1} = \\
&= N^2 \overline{M}^2 \left(1 - \frac{\sum_{i=1}^{n} M_i}{N\overline{M}}\right) \frac{\hat{P}\hat{Q}}{\sum_{i=1}^{n} M_i - 1} = \\
&= 10^8 \cdot 4^2 \cdot \left(1 - \frac{69}{4 \cdot 10000}\right) \frac{0,507 \cdot 0,493}{68} = \\
&= 40000 \cdot 39931 \cdot \frac{0,25}{68} = \\
&= 5871054,93
\end{aligned}$$

La eficiencia relativa del muestreo por conglomerados respecto al aleatorio simple se calcula como el cociente entre sus varianzas:

$$eff(\hat{A}_{_{MCU}}, \hat{A}_{_{MAS}}) = \frac{\widehat{Var}_{_{MAS}}(\hat{A}_{_{MAS}})}{\widehat{Var}_{_{MCU}}(\hat{A}_{_{MCU}})} = \frac{5871054,93}{5186973,684} = 1,13188$$

por lo que como es mayor que uno significa que el estimador del MAS es menos eficiente que el de conglomerados. En concreto la varianza en el caso de Muestreo Aleatorio Simple es un 13,19 % mayor.

Otra forma de estudiar la eficiencia relativa del muestreo por conglomerados respecto a la del muestreo aleatorio simple es llevando a cabo el cálculo del coeficiente de correlación intraconglomerados. Este se define como el coeficiente de correlación lineal entre todos los pares de valores posibles de la variable en estudio medidos sobre unidades dentro del mismo conglomerado. Es decir, es una medida de la homogeneidad en el interior de los conglomerados de forma que el muestreo por conglomerados será preferido al muestreo aleatorio simple cuando el coeficiente de homogeneidad intraconglomerados sea lo más pequeño posible, ya que lo ideal es la heterogeneidad dentro de los conglomerados. La estimación de este coeficiente, $\widehat{\delta}$, viene dada por:

$$\widehat{\delta} = \frac{\widehat{S}_b^2 - \widehat{S}_0^2}{\left(\overline{M} - 1\right)\widehat{S}_0^2}$$

donde

$$\widehat{S}_0^2 = \frac{(N-1)\,\widehat{S}_b^2 + \left(N\overline{M} - N\right)\widehat{S}_W^2}{N\overline{M} - 1}$$

$$\widehat{S}_b^2 = \frac{\overline{M}}{n-1}\sum_{i=1}^{n}\left(P_i - \overline{P}\right)^2$$

$$\widehat{S}_W^2 = \frac{\sum_{i=1}^{n}\overline{M}P_i\left(1-P_i\right)}{n\left(\overline{M}-1\right)}$$

La varianza del muestreo monoetápico de conglomerados se puede expresar en función de la varianza del muestreo aleatorio simple y el coeficiente de correlación intraconglomerado:

$$\widehat{Var}_{MCU}\left(\widehat{P}\right) = \widehat{Var}_{MAS}\left(\widehat{P}\right)\left[1 + \left(\overline{M}-1\right)\widehat{\delta}\right]$$

y la comparación de la eficiencia entre muestreo monoetápico de conglomerados y muestreo aleatorio simple podría resumirse como sigue:

- si $\widehat{\delta} > 0 \Rightarrow$ conglomerados peor que aleatorio simple
- si $\widehat{\delta} = 0 \Rightarrow$ conglomerados igual que aleatorio simple
- si $\widehat{\delta} < 0 \Rightarrow$ conglomerados mejor que aleatorio simple

En el caso que nos ocupa,

$$\widehat{S}_b^2 = \frac{\overline{M}}{n-1}\sum_{i=1}^{n}\left(P_i - \overline{P}\right)^2 = \frac{\overline{M}}{n-1}\left(\frac{1}{\overline{M}}\right)^2\left[\sum_{i=1}^{n}A_i^2 - n\overline{M}^2\overline{P}^2\right] =$$

$$= \frac{1}{(n-1)\overline{M}}\left[\sum_{i=1}^{n} A_i^2 - n\overline{M}^2\overline{P}^2\right] =$$

$$= \frac{1}{(n-1)\overline{M}}\left[\sum_{i=1}^{n} A_i^2 - n\overline{M}^2\left(\frac{1}{n\overline{M}}\sum_{i=1}^{n} A_i\right)^2\right] =$$

$$= \frac{1}{19\cdot 4}\left[81 - 20\cdot 4^2\cdot\left(\frac{1}{20\cdot 4}\cdot 35\right)^2\right] = 0,2599$$

$$\widehat{S}_W^2 = \frac{\sum_{i=1}^{n}\overline{M}P_i(1-P_i)}{n\left(\overline{M}-1\right)} = \frac{\overline{M}}{n\left(\overline{M}-1\right)}\sum_{i=1}^{n}\frac{A_i}{\overline{M}}\left(1-\frac{A_i}{\overline{M}}\right) =$$

$$= \frac{\overline{M}}{n\left(\overline{M}-1\right)}\frac{1}{\overline{M}^2}\sum_{i=1}^{n} A_i\left(\overline{M}-A_i\right) = \frac{1}{n\left(\overline{M}-1\right)\overline{M}}\left[\overline{M}\sum_{i=1}^{n} A_i - \sum_{i=1}^{n} A_i^2\right] =$$

$$= \frac{1}{20\cdot 4\cdot(4-1)}\left[4\cdot 35 - 81\right] = 0,2458$$

$$\widehat{S}_0^2 = \frac{(N-1)\widehat{S}_b^2 + \left(N\overline{M}-N\right)\widehat{S}_W^2}{N\overline{M}-1} =$$

$$= \frac{(10000-1)\,0,2599 + (10000\cdot 4 - 10000)\,0,2458}{10000\cdot 4 - 1} =$$

$$= 0,2493$$

Y por tanto, la estimación del coeficiente de correlación intraconglomerados es:

$$\widehat{\delta} = \frac{\widehat{S}_b^2 - \widehat{S}_0^2}{\left(\overline{M}-1\right)\widehat{S}_0^2} = \frac{0,2599 - 0,2493}{(4-1)\,0,2493} = 0,014.$$

Debería ser $\widehat{\delta} < 0$, para confirmar que el muestreo por conglomerados es preferible al muestreo aleatorio simple. No obstante es un valor debido a las aproximaciones muy cercano a cero por lo que no parece que sea ni coherente ni concluyente.

1.c) Considerando que los conglomerados son de distinto tamaño, para estimar el total de mujeres de la población se puede utilizar el estimador de la razón de A. Este estimador suele utilizarse cuando los tamaños de los conglomerados varían mucho de unos a otros. Del enunciado se tiene que M=40000, N=10000, n=20 y el muestreo es sin reposición.

$$\hat{A}_R = M\hat{R} = M\hat{P} = M\frac{\sum_{i=1}^{n} A_i}{\sum_{i=1}^{n} M_i} = 40000\frac{35}{69} = 20289,85$$

Para estimar la varianza del total sin reposición se utiliza el estimador de la varianza del estimador de razón:

$$\widehat{Var}(\hat{P}) = \frac{1-f}{n\overline{M}^2}\left(\hat{S}_A^2 + \hat{R}^2\hat{S}_M^2 - 2\hat{R}\hat{S}_{AM}^2\right) = \frac{1-f}{n\overline{M}^2}\sum_{i=1}^{n}\frac{\left(A_i - \hat{R}M_i\right)^2}{n-1} =$$

$$= \frac{(1-f)}{n\left(\frac{M}{N}\right)^2(n-1)}\left(\sum_{i=1}^{n} A_i^2 + \hat{R}^2\sum_{i=1}^{n} M_i^2 - 2\hat{R}\sum_{i=1}^{n} M_iA_i\right) =$$

$$= \frac{1-\frac{20}{10000}}{20\left(\frac{40000}{10000}\right)^2(20-1)}\left(81 + \left(\frac{35}{69}\right)^2\cdot 269 - 2\frac{35}{69}142\right) = 0,00101$$

$\widehat{Var}(\hat{A}_R) = M^2\widehat{Var}(\hat{P}) = 40000^2 \cdot 0,00101 = 1616610,17$

$\hat{\sigma}(\hat{A}_R) = \sqrt{\widehat{Var}(\hat{A}_R)} = 1271,46.$

1.d) La estimación de razón al tamaño es uno de los métodos más eficaces si no se puede estratificar. Normalmente el total por conglomerado está relacionado de manera proporcional con el tamaño, así que la estimación de razón está justificada. No obstante, aunque en general el método de razón tiende a tener mejores resultados, hay que tener ciertas precauciones. Para construir el estimador $\hat{A}_R$ hay que conocer los tamaños de cada uno de los conglomerados muestreados o al menos la suma de ellos, además de conocer el tamaño de la población y es un estimador sesgado. Este sesgo suele ser pequeño en general, y si los conglomerados son exactamente iguales, este estimador coincide con el $\hat{A}_{MCU}$, puesto que :

$$\hat{A}_R = M\hat{P} = M\frac{\sum_{i=1}^{n} A_i}{\sum_{i=1}^{n} M_i} = M\frac{\sum_{i=1}^{n} A_i}{n\cdot\overline{M}} = \frac{N}{n}\sum_{i=1}^{n} A_i = \hat{A}_{MCU}$$

el cual es insesgado y no requiere conocer el tamaño de la población, sino únicamente el número de conglomerados.

El estimador $\hat{A}_{MCU}$ aunque no tiene sesgo puede ser poco preciso debido a que no tiene en cuenta las ponderaciones dadas por el tamaño de los conglomerados M_i. Cuando las proporciones de cada conglomerado varían poco y los tamaños de los mismos sí que varían considerablemente, ocurriendo por tanto que $A_i = M_i \cdot P_i$ tiene mucha variabilidad para $i = 1, \cdots, N$, hace que su varianza sea grande, puesto que este término aparece en su expresión: $\widehat{Var}_{MCU}\left(\widehat{A}_{MCU}\right) = \frac{N^2(1-f)}{n(n-1)}\left(\sum_{i=1}^n A_i^2 - n\overline{M}^2\overline{P}^2\right)$.

2) Se trata de un muestreo de conglomerados bietápico con conglomerados de igual tamaño y submuestreo de igual tamaño (es decir, 5 hogares en cada sección censal). Además ambas etapas son con reposición y probabilidades iguales. Los datos de los que se dispone son:

$$N = 200; \qquad \overline{M} = 50; \qquad n = 4; \qquad \overline{m} = 5; \qquad f_{2i} = \frac{\overline{m}}{\overline{M}} = \frac{1}{10}.$$

La estimación insesgada de A se obtiene de la siguiente forma:

$$\widehat{A} = \sum_{i=1}^{n} \frac{\hat{A}_i}{n \cdot p_{1i}} = \sum_{i=1}^{n} \frac{\hat{A}_i}{\frac{n}{N}} = \frac{N}{n}\sum_{i=1}^{n} \hat{A}_i = \frac{N}{n}\sum_{i=1}^{n}\sum_{j=1}^{\overline{m}} \frac{A_{ij}}{\overline{m} \cdot p_{2j}} = \frac{N}{n}\sum_{i=1}^{n}\sum_{j=1}^{\overline{m}} \frac{A_{ij}}{\frac{\overline{m}}{\overline{M}}} =$$

$$= \frac{N \cdot \overline{M}}{n \cdot \overline{m}} \sum_{i=1}^{n}\sum_{j=1}^{\overline{m}} A_{ij} = N \cdot \overline{M} \cdot \overline{P} = \frac{200 \cdot 50}{4 \cdot 5}(6 + 8 + 11 + 10) = 17500$$

Para el cálculo de la estimación insesgada de su error de muestreo se utiliza el método de conglomerados últimos, que tiene en cuenta que

$$\widehat{Var}\left(\widehat{A}\right) = \frac{\sum_{i=1}^{n}\left(\widehat{\theta}_i - \widehat{\theta}\right)^2}{n\,(n-1)} = \frac{\sum_{i=1}^{n}\left(\widehat{A}_i - \widehat{A}\right)^2}{n\,(n-1)} = \frac{\sum_{i=1}^{n}\left(N\overline{M}\widehat{P}_i - N \cdot \overline{M} \cdot \overline{P}\right)^2}{n\,(n-1)} =$$

$$= N^2\overline{M}^2 \frac{\sum_{i=1}^{n}\left(\widehat{P}_i - \overline{P}\right)^2}{n\,(n-1)}$$

donde

$$\overline{P} = \frac{1}{n}\sum_{i=1}^{n}\widehat{P}_i = \frac{6+8+11+10}{4\cdot 5} = 1,75$$

y por tanto

$$\widehat{Var}\left(\widehat{A}\right) =$$

$$= 200^2 \cdot 50^2 \frac{\left(\frac{6}{5}-0,175\right)^2 + \left(\frac{8}{5}-0,175\right)^2 + \left(\frac{11}{5}-0,175\right)^2 + \left(\frac{10}{5}-0,175\right)^2}{4\cdot 3} =$$

$$= 10^8 \frac{0,3025+0,0225+0,2025+0,0625}{12} =$$

$$= 10^8 \frac{0,59}{12} = 4916666,6667.$$

La estimación del error de muestreo es consecuentemente

$$\widehat{\sigma}(\widehat{A}) = \sqrt{\widehat{Var}(\widehat{A})} = \sqrt{4916666,6667} = 2217,3557.$$

Cuestión 2

En la tabla 5.1 se muestran datos correspondientes a operaciones y saldos de las cuentas del sector AAPP (S.13) de una economía para un año t.

a) Calcular para el sector S.13:

1. El saldo de rentas primarias bruto y neto.

2. La renta disponible ajustada bruta.

3. El gasto en consumo final de las AAPP.

4. Sabiendo que el concepto de déficit o superávit público es equivalente al concepto de capacidad o necesidad de financiación de las Administraciones Públicas y que el PIB a precios corrientes de esta economía en el año t ha sido de 1070 mm €, calcule el déficit público como porcentaje del PIB para el citado año.

b) El gasto en consumo final de las AAPP a precios corrientes en el año t-1 fue de 194 mm€, y la tasa de variación interanual (t/t-1) del deflactor de este agregado fue del 1 %,

1. Calcule la tasa de variación interanual del gasto en consumo final de las AAPP a precios corrientes.

2. Indique qué parte de esta variación sería atribuible a variaciones de volumen.

3. Calcule el gasto en consumo final de las AAPP del año t valorado a precios del año t-1.

Tabla 5.1: Ejercicio 2. Unidades en miles de millones de euros (mm €)

Nota: para esta economía se considera que las adquisiciones menos cesiones de activos no financieros no producidos son despreciables en el año t

Formación bruta de capital (P.5)	25
Consumo de capital fijo (P.51c)	26
Consumo final colectivo efectivo (P.42)	80
Información sobre impuestos y subvenciones:	
Impuestos sobre la producción y las importaciones(recursos) (D.2)	125
Subvenciones (empleos) (D.3)	10
Impuestos corrientes netos sobre la renta, patrimonio, etc. D.5	100
Información sobre las rentas de la propiedad D.4:	
Intereses recibidos D.41	4
Intereses pagados D.41	40
Otras rentas de la propiedad netas recibidas D.42-D.45	3
Cotizaciones sociales netas (D.61)	130
Prestaciones sociales distintas de transferencias sociales en especie D.62 -	180
Transferencias sociales en especie (D.63) -	120
Información sobre transferencias:	
Otras transferencias corrientes netas pagadas D.7	10
Transferencias de capital netas a pagar D.9	1
Excedente de explotación bruto	30

c) Sabiendo que en el año t el valor a precios corrientes del gasto en consumo final de las AAPP en los dos primeros trimestres fue de 47 y 52 mm€ respectivamente , y que la tasa intertrimestral del 3^{er} trimestre fue del 2 % calcule:

1. El valor del agregado a precios corrientes en el 3^{er} trimestre.

2. El valor del agregado a precios corrientes en el 4^o trimestre. 3. Indique a qué tasa interanual aproximada evolucionó en el 4^o trimestre del año t, sabiendo que las tasas interanuales correspondientes a los tres primeros fueron 2,5 %, 3 % y 3,5 %, respectivamente.

Solución.

a.1) El enunciado de esta pregunta parece tener dos datos que no son coherentes, puesto que en el sector AAPP debe ser CCF = EBE y ambos datos no son coincidentes. No obstante, los utilizaremos como nos los dan, pero se podrían obtener resultados inconsistentes.

E	II.1.2.Cuenta de asignación de la renta primaria		R
10	Subvenciones a producción e importaciones	EBE/RMB (B2.b/B3.b)	30
		RA (D.1)	0
40	Intereses rentas propiedad pagados	Impuestos sb producción e import's	125
		Intereses rentas prop recibidas	4
		Rentas de la propiedad recibidas	3
112	Saldo de rentas primarias bruto		

Saldo de rentas primarias neto = Saldo de rentas primarias bruto - CCF = 112-26=86

a.2)

E	II.2.Cuenta de distribución secundaria de la renta		R
10	Otras transferencias corrientes netas pagadas	Saldo de rentas primarias bruto	112
180	Prestaciones sociales distintas de transferencias sociales en especie	Impuestos corrientes netos sobre la renta, patrimonio, etc...	100
		Cotizaciones sociales netas	130
152	RBD		

E	II.3.Cuenta de redistribución de la renta en especie		R
120	Transferencias sociales en especie	RBD	152
32	RBD Ajustada		

a.3) El GCF de las AAPP es la suma del gasto en consumo final individual de las AAPP y el consumo final colectivo efectivo de las AAPP, es decir, las transferencias sociales en especie y el consumo colectivo de las AAPP. Por tanto:

$$GCF_{AAPP} = GCF_{individual} + GCF_{colectivo} =$$

$$= \text{transferencias sociales en especie} + GCF_{colectivo} = 120 + 80 = 200$$

a.4)

E	II.4.1.Cuenta de utilización de la renta disponible		R
200	GCF	RBD	152
-48	Ahorro Bruto		
-26	-CCF		
-74	Ahorro Neto		

Podemos obtener estos saldos de otra forma:

E	II.4.1.Cuenta de utilización de la renta disponible ajustada		R
80	Consumo Final Efectivo	RBD Ajustada	32
-48	Ahorro Bruto		
-26	-CCF		
-74	Ahorro Neto		

VA	III.1.1.Cuenta de capital. VPNDAHYTK		VP
		ANN	-74
		Transferencias de capital a pagar	-1
-75	VPNDAHYTK		

Además FNC = FBC - CCF = 25-26 = -1

VA	III.1.1.Cuenta de capital. VPNDAHYTK		VP
-1	FNC	VPNDAHYTK	-75
0	Adquisiciones menos cesiones de activos no producidos		
-74	Capacidad o necesidad de Financiación		

Sabiendo que el PIB_{pm} en el año t ha sido de 1070 mm€, el déficit público como porcentaje del PIB será: $\frac{74}{1070} \cdot 100 = 6,91\,\%$

b.1)

Tenemos que $GCF_t^{\text{corrientes}} = 200$, según los datos del apartado anterior y que $GCF_{t-1}^{\text{corrientes}} = 194$. Por tanto, la tasa de variación interanual del gasto en consumo final de las AAPP a precios corrientes es:

$$TVI(GCF)_{t-1}^{t} = \frac{200 - 194}{194} = 0,0309 \approx 3,09\,\%$$

b.2)

Índice de valor = índice de precios · índice de volumen $\Leftrightarrow$

índice de volumen $= \frac{\text{índice de valor}}{\text{índice de precios}} = \frac{\left(1+\frac{3,09}{100}\right)\cdot 100}{\left(1+\frac{1}{100}\right)\cdot 100} = 102,069 \Rightarrow$

$TVI(volumen) \approx 2,07\,\%$

b.3)

$$GCF_t^{\text{base t-1}} = \frac{GCF_t^{\text{base t}}}{IPC_{t-1}^t} \cdot 100 = \frac{200}{101} \cdot 100 = 198,01$$

, puesto que la tasa de variación interanual entre t-1 y t del deflactor fue del 1 %, por tanto $IPC_{t-1}^t = (TVI(IPC)_{t-1}^t + 1) \cdot 100 = (0,01+1) \cdot 100 = 101.$

c.1)

$$GCF_{\text{3er trim}} = GCF_{\text{2do trim}} \left[TVI(GCF)_{\text{2do trim}}^{\text{3er trim}}\right] = 52 \cdot (1,02) = 53,04$$

c.2)

$$GCF_t^{\text{base t}} = 200$$
$$GCF_t^{\text{base t}} = GCF_{\text{1er trim}} + GCF_{\text{2do trim}} + GCF_{\text{3er trim}} + GCF_{\text{4o trim}}$$
$$GCF_{\text{4o trim}} = 200 - 47 - 52 - 53,04 = 47,96$$

c.3)

$$TVI(GCF)_{t-1}^t \approx 3,09\,\%$$

Por tanto

$TVI(GCF)_{t-1}^t =$

$$= \frac{TVI(GCF)_{\text{1er tri}} + TVI(GCF)_{\text{2do tri}} + TVI(GCF)_{\text{3er tri}} + TVI(GCF)_{\text{4o tri}}}{4}$$

$$3,09 = \frac{2,5 + 3 + 3,5 + TVI(GCF)_{\text{4o tri}}}{4}$$

$$TVI(GCF)_{\text{4o tri}} = 3,37\,\%$$

Cuestión 3

Considere la siguiente ecuación salarial

$$ln(salario) = \beta_0+\beta_1 edu+\beta_2 edu^2+\beta_3 exper+\beta_4 exper^2+\beta_5(edu \text{ x } exper)+\beta_6 horasem+\epsilon$$

donde las variables explicativas son años de educación, años de experiencia y número de horas trabajadas a la semana. Los resultados de la estimación de esta ecuación y de algunos modelos derivados de la misma se muestran en la siguiente tabla. El número de observaciones con los que se ha realizado las estimaciones es de 1000.

Nota: Para la resolución de este ejercicio se entregan al opositor las tablas estadísticas de las distribuciones de probabilidad necesarias.

Variable					
	modelo I	**modelo II**	**modelo III**	**modelo IV**	**modelo V**
constante	1.055	1.252	1.573	1.917	0.904
	(0.266)	(0.190)	(0.188)	(0.080)	(0.096)
edu	0.0498	0.0289	0.0366		0.1006
	(0.0397)	(0.0344)	(0.0350)		(0.0063)
edu^2	0.00319	0.00352	0.00293		
	(0.00169)	(0.00166)	(0.00170)		
exper	0.0373	0.0303		0.0279	0.0295
	(0.0081)	(0.0048)		(0.0054)	(0.0048)
exper^2	0.000485	0.000456			
	(0.000090)	(0.000086)			
exper x edu	-0.000510			-0.00047	-0.00044
	(0.000482)			(0.000096)	(0.000086)
horasem	0.01145	0.01156	0.01345	0.01524	0.01188
	(0.00137)	(0.00137)	(0.00136)	(0.00151)	(0.00136)
SCR	222.4166	222.6674	233.8317	280.5061	223.6716
AIC	-1.489	-1.490	-1.445	X	-1.488
SC	X	-1.461	-1.426	-1.244	-1.463

a.- Usando un nivel del 5 %, indique qué coeficientes estimados no son significativamente diferentes de cero. Para ello elabore una tabla similar a la anterior en donde se indique con claridad lo pedido.

b.- Indique qué restricción en el modelo I arroja el modelo II. Utilice un contraste adecuado para contrastar la restricción señalada. Muestre si es posible obtener el mismo resultado usando un contraste tipo ratio de la t.

c.- Justifique qué modelo utilizaría si quisiera contrastar si la experiencia es relevante para la determinación del salario. Seguidamente realice el contraste. Haga lo mismo para el caso de que quisiera contrastar la relevancia de la educación en la determinación del salario.

d.- Explique a qué cuestión econométrica estamos tratando de responder cuando realizamos restricciones sobre el modelo I para obtener el V.

e.- A partir de los resultados de sus respuestas en los apartados anteriores y de algún otro contraste que considere necesario, indique qué modelo considera más adecuado.

f.- Las tres últimas líneas indican los acrónimos de "suma de cuadrados de los residuos", criterio de información de Akaike, y criterio de información de Schwarz. Observará que hay algunos que faltan. ¿Cómo lo completaría?¿Para qué sirven estos criterios?

Solución.

a) Dado que el tamaño muestral es T=1000, independientemente del número de regresores incluidos en el modelo, el valor teórico con el que comparar el estadístico $\left|\frac{\hat{\beta}_j}{\sigma(\hat{\beta}_j)}\right| \sim t_{T-k}$ es $1,96 \simeq t_{\infty;0,025}$ de forma que en caso de que $\left|\frac{\hat{\beta}_j}{\sigma(\hat{\beta}_j)}\right| > 1,96$ se rechazará la hipótesis nula $\beta_j = 0$, es decir, el parámetro será significativo.

	modelo I	modelo II	modelo III	modelo IV	modelo V
cte	3,967*	6,589*	8,367*	23,962*	9,416*
edu	1,254	0,840	1,045	—	15,968*
edu^2	1,887	2,120*	1,723	—	—
exper	4,604*	6,312*	—	5,166*	6,145*
exper^2	5,388*	5,302*	—	—	—
exper x edu	1,058	—	—	4,895*	5,116*
horasem	8,357*	8,437*	9,889*	10,092*	8,735*

Tabla 5.2: Recoge con los cocientes $\left|\frac{\hat{\beta}_j}{\sigma(\hat{\beta}_j)}\right|$ que comparándolos con el valor 1,96 proporciona la significatividad de cada uno de los coeficientes (se marcan con * los coeficientes significativos y en amarillo los coeficientes no significativos).

b) La restricción sería $\beta_5 = 0$, es decir, que el coeficiente de la variable "exper x edu"sea nulo. Utilizando un contraste de modelos restringidos, dado que solo hay una restricción bajo la hipótesis nula, sería $s = 1$.

$$F_{exp} = \frac{\frac{SRR_R - SRR_{SR}}{s}}{\frac{SRR_{SR}}{T-k-1}} \sim F_{s,T-k-1} \Leftrightarrow F_{exp} = \frac{\frac{222{,}6674 - 222{,}4166}{1}}{\frac{222{,}4166}{1000-6-1}} \sim F_{1,\infty}$$

De donde $F_{exp} = 1,119$ y $F_{\text{teórico}} = 3,88$. Al ser $F_{exp} < F_{\text{teórico}}$ no existe suficiente evidencia para rechazar la hipótesis nula y se concluye que β_5 no es significativo.

Utilizando un contraste tipo ratio de la t, la hipótesis nula también es $H_0 : \beta_5 = 0$ y se utiliza el estadístico de contraste (t-ratio) dado por $t_{\hat{\beta}_5} = \frac{\hat{\beta}_5}{\sigma(\hat{\beta}_5)} \sim t_{T-k-1}$. El valor de $t_{\hat{\beta}_5} = 1,058$. Dado que el contraste tiene una sola restricción, s=1 y $F_{s,T-k-1} \equiv t_{T-k-1}$, pero en este caso el valor proporcionado por la tabla de la t de Student es $t_\infty \simeq 1,96$, por lo que tampoco se rechazaría la hipótesis nula. Ambos contrastes proporcionan el mismo resultado ya que $\sqrt{1,119} = 1,058$ y $\sqrt{3,88} = 1,96$.

c) Para contrastar si la experiencia es relevante en la determinación del salario podríamos comparar los modelos I y III, donde el modelo III es un modelo anidado en I en el que se impone que $\beta_3 = 0$, $\beta_4 = 0$ y $\beta_5 = 0$, que son los coeficientes asociados de alguna forma a la experiencia. Se podría utilizar un estadístico F de contraste del modelo no restringido I al modelo restringido III. Así, la hipótesis nula sería $H_0 : \beta_3 = 0, \beta_4 = 0, \beta_5 = 0$ mientras que la hipótesis alternativa es que al menos uno de los coeficientes β_3, β_4 o β_5 sea distinto de cero. En este caso, el número de restricciones bajo H_0 son tres, es decir, s=3.

$$F_{exp} = \frac{\frac{SRR_R - SRR_{SR}}{s}}{\frac{SRR_{SR}}{T-k-1}} \sim F_{s,T-k-1} \Leftrightarrow F_{exp} = \frac{\frac{233{,}8317-222{,}4166}{3}}{\frac{222{,}4166}{1000-6-1}} \sim F_{3,\infty}$$

Donde $F_{exp} = 16,987$ y $F_{\text{teórico}} = 2,64$. Al ser $F_{exp} > F_{\text{teórico}}$, existe evidencia suficiente para rechazar la hipótesis nula y se concluye que al menos uno de los coeficientes será distinto de cero. Por tanto, la experiencia es un factor importante en la determinación del salario.

Para contrastar la relevancia de la educación en la determinación del salario tenemos el problema de que no hay ningún modelo de los propuestos en el que no aparezca la variable educación de alguna forma. Los modelos IV y V tienen el término de interacción de la experiencia con la educación y por tanto no se pueden comparar entre sí para tratar de afirmar si la variable educación es relevante, al igual que tampoco tendría sentido comparar los modelos I y IV, porque no desaparecería del modelo la interacción entre ambas. Aunque en los modelos IV y V individualmente la interacción no es significativa, en el modelo I con todas las variables sí que lo es, y por tanto no parece que tengamos las herramientas suficientes para poder decidir si esa variable es relevante o no.

d) Dado que las restricciones en el modelo I para obtener el V consisten en eliminar las variables explicativas del modelo no lineales, en concreto edu^2 y exper^2, se está intentando contrastar si la correcta especificación del modelo precisa la introducción de regresores no lineales.

Los errores de especificación en la forma funcional se analizan con el contraste Reset elaborado por Ramsey en 1969, el cual permite identificar si se está usando una forma lineal incorrecta y cualquier error de omisión o la presencia de correlaciones entre las variables explicativas y la perturbación. El test Reset probaría si esas variables al cuadrado pueden explicar la variable dependiente.

e) Podríamos utilizar el contraste de significación global para analizar en los modelos propuestos si todos los parámetros asociados a cada variable son iguales a cero, aunque su utilidad es limitada en cuanto a la valoración de la idoneidad de un modelo econométrico. Además, normalmente suele ser significativo cuando alguna de las variables explicativas lo es de forma individual, por lo que no ofrece una información completa. Las hipótesis serían:

$$\begin{cases} H_0 : \beta_1 = \beta_2 = \cdots = \beta_k = 0 \\ H_1 : \exists \beta_i \neq 0 \end{cases}$$

y el estadístico de contraste es $F = \frac{\frac{SSE}{k}}{\frac{SSR}{T-k-1}} = \frac{\frac{R^2}{k}}{\frac{1-R^2}{T-k-1}}$, que bajo la hipótesis nula sigue una distribución F-Snedecor(k,n-k-1).

En modelos con variables que no sean significativas de forma individual, pero sí globalmente puede aparecer el problema de la multicolinealidad, así que es necesario estar atento a esta problemática. El modelo IV tiene una suma residual más elevada que el resto, lo cual implica que su R^2 es más bajo y tiene menor poder explicativo, si bien es cierto que todas sus variables son significativas individualmente.

Además sería necesario efectuar un análisis de los residuos del modelo, de manera que hay que estudiar la posible presencia de heterocedasticidad y de autocorrelación, para lo que existen diversos contrastes como el de Durbin-Watson y el de Breusch-Pagan o el de White. Si los residuos no tienen un comportamiento adecuado el método de mínimos cuadrados no produciría unas estimaciones robustas y por lo tanto el modelo habría que corregirlo.

El principio de parsimonia también debe estar presente en cualquier análisis de un modelo econométrico, de forma que sea razonable y responda a la realidad del conocimiento general que se tiene sobre el área de estudio en la medida de lo posible.

A la vista de los modelos expuestos, y a falta de hacer las comprobaciones adicionales citadas para poder elegir con cierta seguridad uno de ellos, con los datos de los que disponemos parece que el modelo V es el más adecuado viendo en conjunto todas sus propiedades. Sus parámetros son todos significativos, es un modelo lineal, en el cual tanto la experiencia como la educación parecen influir en la determinación del salario, lo cual es lógico, y es más sencillo que el modelo I y II que tienen sumas residuales algo más bajas.

f) Dado que $AIC = \frac{2k}{n} + Ln\left(\frac{SCR}{N}\right)$ y $SBC = \frac{k}{n}Ln(n) + Ln\left(\frac{SCR}{N}\right)$, siendo k el número de parámetros del modelo, se tiene que:

$$AIC - \frac{2k}{n} = Ln\left(\frac{SCR}{N}\right)$$
$$SBC - \frac{k}{n}Ln(n) = Ln\left(\frac{SCR}{N}\right)$$

Por tanto, para el Modelo I se tiene:

$AIC - \frac{2k}{n} = SBC - \frac{k}{n}Ln(n) \Leftrightarrow -1,489 - \frac{2\cdot 7}{1000} + \frac{7}{1000} \cdot Ln(1000) = SBC \Leftrightarrow$ $SBC = -1,4546$

y para el valor faltante del modelo IV:

$AIC - \frac{2k}{n} = SBC - \frac{k}{n}Ln(n) \Leftrightarrow AIC = -1,244 - \frac{4}{1000}Ln(1000) + \frac{2\cdot 4}{1000} \Leftrightarrow$ $AIC = -1,2636$

Estos criterios de información son útiles para comparar modelos alternativos para la misma variable endógena. Hay que tener en cuenta que el criterio de decisión consiste en elegir el modelo que tenga un menor valor AIC, el cuál prima la capacidad predictiva del modelo y tiende a sobreparametrizarlo mientras que SBC prima la especificación correcta.

Cuestión 4

El número diario de ingresos en un servicio de urgencias sigue una distribución de Poisson de parámetro θ_1, con independencia entre los registros correspondientes a días diferentes. Se va a observar el número de ingresos $X_1, \ldots, X_r$ durante r días, a fin de estimar θ_1 para hacer las previsiones de mantenimiento de servicio. Se supone que los errores en la estimación de θ_1 producen pérdida cuadrática:

a) Demostrar que el estadístico $T(X_1, \ldots, X_r) = \sum_{i=1}^{r} X_i$ es suficiente.

b) Determinar k para que $T_k = kT$ sea un estimador insesgado para θ_1 y comprobar que es de mínima varianza para θ_1.

c) Suponiendo que dispusiéramos de una muestra aleatoria simple $Y_1, \cdots, Y_{120}$ de los ingresos mensuales de urgencias registrados a lo largo de los 10 años en dicho servicio, obtener en función de ella la expresión de un intervalo de confianza aproximado para θ_1 al nivel $1 - \alpha$ por el método de la cantidad pivotal.

d) Suponiendo que dispusiéramos de una muestra aleatoria simple correspondiente al número de ingresos medios registrados esos mismos meses en un segundo servicio de urgencias independiente del anterior, para el que los ingresos diarios siguen una distribución de Poisson de parámetro θ_2, obtener la expresión de un estadístico de contraste para resolver:

$$\begin{cases} H_0 : \theta_1 = \theta_2 \\ H_1 : \theta_1 \neq \theta_2 \end{cases}$$

Solución.

a) Dado que $X_i \sim P(\theta_1)$, es decir, $P_{\theta_1}(X_i = x_i) = e^{-\theta_1} \frac{\theta_1^{x_i}}{x_i!}$, la función de probabilidad de la muestra será:

$$P_{\theta_1}(X_1 = x_1, \ldots, X_n = x_n) = e^{-n\theta_1} \frac{\theta_1^{\sum x_i}}{\prod x_i!}$$

de forma que si se consideran $g_{\theta_1}(T) = e^{-n\theta_1}\theta_1^{\sum x_i}$, que solo depende de la muestra a través del estadístico $T(X_1,\ldots,X_r) = \sum_{i=1}^{r} X_i$, y $h(x_1,\ldots,x_n) = \frac{1}{\prod x_i!}$, que es una función que no depende del parámetro, el teorema de factorización proporciona que el estadístico es suficiente.

b)

$$E\left[T_k\right] = kE[T] = kE\left[\sum_{i=1}^{r} X_i\right] = k\sum_{i=1}^{r} E[X_i] = k\sum_{i=1}^{r}\theta_1 = kr\theta_1$$

Luego para que $E\left[T_k\right] = \theta_1$, tiene que ser $k = \frac{1}{r}$, es decir, la media muestral es el único estimador de la forma kT que es insesgado para θ_1. Veamos si es de mínima varianza.

Puesto que $X_i \sim P(\theta_1)$, por la propiedad reproductiva de la distribución de Poisson $T \sim P(r\theta_1)$ y por tanto:

$Var(T) = r\theta_1$ y $Var(T_k) = Var\left(\frac{1}{r}T\right) = \frac{1}{r^2}Var(T) = \frac{r\theta_1}{r^2} = \frac{\theta_1}{r}$.

Por otra parte, la Cota de Cramer Rao(CCR), al ser T_k un estimador insesgado será: $CCR = \frac{1}{I(\theta_1)} = \frac{1}{r\cdot I^*(\theta_1)}$

$$\begin{aligned} I^*(\theta_1) &= -E\left[\frac{\partial^2 \mathrm{Ln} f_{\theta_1}(x)}{\partial\theta_1^2}\right] = -E\left[\frac{\partial^2}{\partial\theta_1^2}\left(-\theta_1 + x\cdot log\theta_1 - x!\right)\right] = \\ &= -E\left[\frac{\partial}{\partial\theta_1}\left(-1 + \frac{x}{\theta_1}\right)\right] = -E\left[-\frac{x}{\theta_1^2}\right] = \frac{\theta_1}{\theta_1^2} = \frac{1}{\theta_1} \end{aligned}$$

Luego la $CCR = \frac{1}{r\cdot\frac{1}{\theta_1}} = \frac{\theta_1}{r}$, por lo tanto, puesto que $Var(T_k) = CCR$, es eficiente y por tanto de mínima varianza.

c) Dado que $X_i \sim P(\theta_1)$ son los ingresos diarios, los ingresos mensuales serán $Y = \sum_{r=1}^{30} X_r$ que, por la propiedad reproductiva de la Poisson, sigue una distribución $Y \sim P(30\cdot\theta_1)$ que se puede aproximar por $Y \sim N(30\cdot\theta_1, \sqrt{30\cdot\theta_1})$. Tomando una muestra aleatoria simple $Y_1, \cdots, Y_{120}$ de los ingresos mensuales, se tiene que

$$\overline{Y} = \frac{1}{120}\sum_{i=1}^{120} Y_i \sim N\left(30\cdot\theta_1, \sqrt{\frac{30\cdot\theta_1}{120}}\right) \equiv N\left(30\cdot\theta_1, \frac{\sqrt{\theta_1}}{2}\right)$$

de forma que

$$\frac{\overline{Y}-30\cdot\theta_1}{\frac{\sqrt{\theta_1}}{2}} \sim N(0,1)$$

Entonces, utilizando el método de la cantidad pivotal, siendo $\tilde{T}$ el pivote: $P(c_1 \leq \tilde{T}(x_1,\ldots,x_n,\theta_1) \leq c_2) = \alpha$, y tomando como cantidad pivotal a $\frac{\overline{Y}-30\cdot\theta_1}{\frac{\sqrt{\theta_1}}{2}}$ se tiene que:

$$P\left(-\lambda_{\frac{\alpha}{2}} \leq \frac{\overline{Y}-30\cdot\theta_1}{\frac{\sqrt{\theta_1}}{2}} \leq \lambda_{\frac{\alpha}{2}}\right) = \alpha$$

De donde:

$$\frac{\overline{Y}-30\cdot\theta_1}{\frac{\sqrt{\theta_1}}{2}} = \lambda_{\frac{\alpha}{2}} \Leftrightarrow \left(\frac{\overline{Y}-30\cdot\theta_1}{\frac{\sqrt{\theta_1}}{2}}\right)^2 = \lambda^2_{\frac{\alpha}{2}} \Leftrightarrow \overline{Y}^2 + 900\theta_1^2 - 60\overline{Y}\theta_1 = \lambda^2_{\frac{\alpha}{2}}\frac{\theta_1}{4}$$
$$\Leftrightarrow 900\theta_1^2 + \theta_1\left(-60\overline{Y} - \frac{\lambda^2_{\frac{\alpha}{2}}}{4}\right) + \overline{Y}^2 = 0$$

Y las raíces son:

$$\theta_1 = \frac{60\overline{Y} + \frac{\lambda^2_{\frac{\alpha}{2}}}{4} \pm \sqrt{3600\overline{Y}^2 + \frac{\lambda^4_{\frac{\alpha}{2}}}{16} + 30\overline{Y}\lambda^2_{\frac{\alpha}{2}} - 4\cdot 900\cdot\overline{Y}^2}}{1800} =$$
$$= \frac{60\overline{Y} + \frac{\lambda^2_{\frac{\alpha}{2}}}{4} \pm \sqrt{\lambda^2_{\frac{\alpha}{2}}\left(\frac{\lambda^2_{\frac{\alpha}{2}}}{16} + 30\overline{Y}\right)}}{1800} =$$
$$= \frac{\overline{Y}}{30} + \frac{\lambda^2_{\frac{\alpha}{2}}}{7200} \pm \sqrt{\frac{\lambda^2_{\frac{\alpha}{2}}}{3240000}\left(\frac{\lambda^2_{\frac{\alpha}{2}}}{16} + 30\overline{Y}\right)}$$

Luego

$$\theta_1 \in \left[\frac{\overline{Y}}{30} + \frac{\lambda^2_{\frac{\alpha}{2}}}{7200} - \sqrt{\frac{\lambda^2_{\frac{\alpha}{2}}}{3240000}\left(\frac{\lambda^2_{\frac{\alpha}{2}}}{16} + 30\overline{Y}\right)}, \frac{\overline{Y}}{30} + \frac{\lambda^2_{\frac{\alpha}{2}}}{7200} + \sqrt{\frac{\lambda^2_{\frac{\alpha}{2}}}{3240000}\left(\frac{\lambda^2_{\frac{\alpha}{2}}}{16} + 30\overline{Y}\right)}\right]$$

d) Además de la muestra $\overline{Y} \sim N\left(30 \cdot \theta_1, \frac{\sqrt{\theta_1}}{2}\right)$, se dispone de otra muestra aleatoria simple en otro servicio de urgencias $\overline{Z} \sim N\left(30 \cdot \theta_2, \frac{\sqrt{\theta_2}}{2}\right)$. Un estadístico para resolver el contraste propuesto, teniendo en cuenta que las varianzas poblacionales son desconocidas y distintas sería:

$$\frac{\overline{y} - \overline{z}}{\sqrt{\frac{S^2_{c\overline{y}}}{n} + \frac{S^2_{c\overline{z}}}{n}}} \sim t_f$$

Donde n=120 y $f = \frac{\left(\frac{S^2_{c\overline{y}}}{n} + \frac{S^2_{c\overline{z}}}{n}\right)^2}{\frac{1}{n+1}\left(\frac{S^2_{c\overline{y}}}{n}\right)^2 + \frac{1}{n+1}\left(\frac{S^2_{c\overline{z}}}{n}\right)^2} - 2 = \frac{\left(\frac{S^2_{c\overline{y}} + S^2_{c\overline{z}}}{120}\right)^2}{\frac{\left(\frac{S^2_{c\overline{y}}}{n}\right)^2 + \left(\frac{S^2_{c\overline{z}}}{n}\right)^2}{121}} - 2$ entendiendo como que hay que elegir el entero más próximo al resultado de esta expresión.

O bien, dado que el tamaño muestral es suficientemente grande, se podría aceptar la aproximación

$$\frac{\overline{y} - \overline{z}}{\sqrt{\frac{S^2_{c\overline{y}}}{n} + \frac{S^2_{c\overline{z}}}{n}}} \sim N(0,1)$$

de forma que no se rechazaría la igualdad de medias si $\left|\frac{\overline{y} - \overline{z}}{\sqrt{\frac{S^2_{c\overline{y}}}{n} + \frac{S^2_{c\overline{z}}}{n}}}\right| \leq z_{\frac{\alpha}{2}}$

Cuestión 5

Una oficina de estadística internacional nos facilita de la Región A los siguientes datos de población a 1 de enero de cada año (columnas [2] a [6]) y de defunciones durante los años de calendario señalados (columnas [7] a [11]). También se dispone de información detallada sobre salud autopercibida a partir de los datos de una encuesta de condiciones de vida para el país en que está incluída la mencionada Región A, que se supone bastante próximos a la realidad regional.

Grupode edad	Población a 1 de enero					Defunciones					% con mala salud
	2011	2012	2013	2014	2015	2010	2011	2012	2013	2014	2011-2012
[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]	[11]	[12]
0	51.239	49.222	47.619	44.308	44.453	180	140	127	104	125	0,7
1 - 4	218.970	215.759	209.305	199.674	192.383	45	34	31	29	25	0,9
5 - 9	255.800	261.281	265.724	269.985	269.398	17	24	22	30	19	0,5
10 - 14	234.010	237.601	239.937	244.617	248.975	29	25	23	15	17	0,3
15 - 19	238.863	235.736	231.256	228.371	229.842	51	53	47	34	37	0,5
20 - 24	272.590	265.689	258.557	250.598	244.678	83	63	52	64	52	0,6
25 - 29	340.803	321.447	301.253	285.500	273.215	108	103	110	88	66	0,8
30 - 34	437.132	419.965	394.196	365.084	341.900	212	200	199	143	136	1,0
35 - 39	435.477	440.117	439.598	435.451	425.701	320	309	290	265	236	1,8
40 - 44	408.628	413.213	412.158	411.964	415.151	532	501	482	454	443	2,9
45 - 49	374.921	382.673	389.565	392.446	392.895	785	814	767	746	719	3,8
50 - 54	333.903	343.692	348.318	353.085	359.996	1.081	1.102	1.131	1.103	1.165	5,1
55 - 59	283.511	289.539	295.568	304.140	315.019	1.439	1.427	1.376	1.476	1.477	7,7
60 - 64	269.593	271.651	274.308	271.056	270.957	1.854	1.942	1.982	1.878	1.901	11,4
65 - 69	234.820	243.864	252.303	258.610	260.472	2.469	2.510	2.413	2.666	2.721	13,3
70 - 74	202.194	200.317	199.933	206.853	221.483	3.545	3.522	3.419	3.368	3.574	16,9
75 - 79	177.808	179.428	181.192	177.782	168.125	5.636	5.553	5.669	5.285	5.052	21,3
80 - 84	128.901	131.505	135.258	139.654	142.328	7.624	7.936	8.031	7.506	7.608	29,4
85 y +	100.029	106.951	110.963	117.249	122.703	14.119	14.530	16.132	15.265	16.177	35,8

Se solicitan los siguientes resultados:

1. Elaborar una tabla de mortalidad correspondiente al intervalo de años para los que se dispone de datos de salud autopercibida que al menos incluya las siguientes series:

a) Probabilidades de muerte (${}_nq_x$) y supervivientes (l_x) entre dos aniversarios de edades cumplidas.

b) Población estacionaria de la tabla en cada intervalo de edades (${}_nL_x$) y esperanza de vida al inicio de cada intervalo de edades (e_x).

2. A partir de los datos anteriores se precisa calcular la esperanza de vida en buena salud (EVBS) al nacimiento y esperanza de vida en buena salud a los 65 años, igualmente para el intervalo de años para los que se dispone de datos de salud autopercibida. Se entiende como buena la salud informada como diferente a la reportada por los entrevistados como mala salud (columna [12]).

3. Comente brevemente, qué conclusiones se pueden sacar de los resultados de las diferencias obtenidas entre esperanza de vida y esperanza de vida en buena salud de la Región A en el contexto de la evolución de éstas magnitudes a nivel internacional en las últimas décadas.

Solución.

1.a) Dado que se dispone de datos de salud autopercibida para los años 2011 y 2012, se elaborará la tabla de mortalidad correspondiente a este intervalo. Para ello, en primer lugar, habría que calcular la tasa de mortalidad específica para cada grupo de edad en el periodo de dos años para el que disponemos de dichos datos de salud:

$$_nm_x = \frac{\text{defunciones}_x^{2011} + \text{defunciones}_x^{2012}}{2 \cdot \frac{\text{población media}_x^{2011} + \text{población media}_x^{2012}}{2}} \cdot 1000$$

donde: $\text{población media}_x^{2011} = \frac{P_x^{01-01-11} + P_x^{01-01-12}}{2}$ y $\text{población media}_x^{2012} = \frac{P_x^{01-01-12} + P_x^{01-01-13}}{2}$. Una vez obtenidas todas las tasas específicas de mortalidad, se calcularían a partir de estas las probabilidades de muerte o el riesgo de muerte:

$$_nq_x = \frac{n \cdot {}_nm_x}{1 + (1 - a_x)_n m_x \cdot n}$$

donde n es la amplitud del intervalo de edad. A continuación, utilizando como raíz de la serie supervivientes $l_0 = 100000$ se calcularían de forma recursiva las defunciones teóricas $_nd_x = {}_nq_x \cdot l_x$, siendo $l_{x+n} = l_x - {}_nd_x$.

1.b) Para el cálculo de la población estacionaria, se utiliza la fórmula

${}_nL_x = n \cdot (l_{x+n} + a_x \cdot {}_nd_x)$ y finalmente, la esperanza de vida se calcula con la fórmula $e_x = \frac{T_x}{l_x} = \frac{T_{x+n}+{}_nL_x}{l_x}$ teniendo en cuenta que para el intervalo final de la tabla $T_w = L_w$

2) El cálculo de la esperanza en buena salud puede llevarse a cabo con el método de Sullivan, que consiste en modificar la tabla de vida multiplicando la tasa específica de prevalencia por edad t_x por ${}_nL_x$ (el número de años vividos en el intervalo de edad) para obtener una serie biométrica modificada del número de años vividos sin mala salud en el intervalo de edad correspondiente. Para cada intervalo de edad las series que se precisan para el cálculo de la EVBS son:

l_x: Supervivientes a la edad x

${}_nL_x$: Número de años vividos en el intervalo de edad x, x+n

T_x: Años futuros de vida a partir de la edad x

t_x: Tasa de mala salud en el grupo de edad x, x+n

$(1–t_x) \cdot L_x$: Número de años vividos sin mala salud en el intervalo de edad x, x+n

$T_x^{'}$: Años futuros de vida en buena salud a partir de la edad x

e_x: Esperanza de vida a la edad x

$EVBS_x$: Esperanza de Vida en Buena Salud a la edad x

Por tanto tomando la columna [12], dividiéndola entre 100 se obtendría la serie t_x. Habría que calcular $(1–t_x)_nL_x$ para obtener el número de años vividos con Buena Salud en el intervalo de edad x, x+n y sumar estos valores para tener los años futuros de vida en buena salud a partir de la edad x ($T_x^{'}$):

$$T_x^{'} = \frac{\sum_x^w (1 - t_x)_n L_x}{l_x}$$

Es decir, se calcula igual que la esperanza de vida a la edad x pero en lugar de sumar el número de años vividos en el intervalo de edad x,x+n se suma el número de años vividos en ausencia de mala salud.

Para los casos pedidos en este ejercicio sería:

$$T_0^{'} = \frac{\sum_{0}^{85y+} (1 - t_x)_n L_x}{l_0}$$

$$T_{65}^{'} = \frac{\sum_{65}^{85y+} (1 - t_x)_n L_x}{l_{65}}$$

	lx	dx	qx	mx	Lx	Tx	ex	tx	1-tx	(1-tx)Lx	T'x	EVBSx
0	100.000	270,285	2,70285	2,70651	99864,8573	7932760,75	79,3276075	0,007	0,993	99165,8033	7441499,62	74,4149962
1-4	99.730	60,298	0,60461	0,15120	398798,263	7832895,89	78,5412444	0,009	0,991	395209,078	7342333,81	73,6223285
5-9	99.669	43,902	0,44048	0,08812	498237,328	7434097,63	74,5875503	0,005	0,995	495746,141	6947124,74	69,7016694
10-14	99.626	50,369	0,50559	0,10114	498001,648	6935860,3	69,6193174	0,003	0,997	496507,643	6451378,59	64,7562889
15-19	99.575	105,696	1,06147	0,21241	497611,485	6437858,65	64,6532693	0,005	0,995	495123,427	5954870,95	59,8027847
20-24	99.469	107,600	1,08174	0,21647	497078,244	5940247,17	59,7193131	0,006	0,994	494095,775	5459747,52	54,8886877
25-29	99.362	164,571	1,65628	0,33153	496397,816	5443168,92	54,7812767	0,008	0,992	492426,633	4965651,75	49,9754364
30-34	99.197	236,544	2,38458	0,47748	495395,03	4946771,11	49,8680127	0,01	0,99	490441,079	4473225,12	45,0942326
35-39	98.961	337,129	3,40669	0,68250	493960,849	4451376,08	44,9812354	0,018	0,982	485069,554	3982784,04	40,2461044
40-44	98.624	586,801	5,94991	1,19353	491651,024	3957415,23	40,1264506	0,029	0,971	477393,145	3497714,48	35,4652871
45-49	98.037	1.007,950	10,28134	2,06689	487664,147	3465764,2	35,3516644	0,038	0,962	469132,909	3020321,34	30,8080354
50-54	97.029	1.569,164	16,17214	3,26079	481221,362	2978100,06	30,6929323	0,051	0,949	456679,073	2551188,43	26,2930903
55-59	95.460	2.282,715	23,91287	4,84045	471591,664	2496878,69	26,1563669	0,077	0,923	435279,106	2094509,36	21,9412963
60-64	93.177	3.303,387	35,45282	7,21852	457626,409	2025287,03	21,7359174	0,114	0,886	405456,998	1659230,25	17,8072989
65-69	89.874	4.426,841	49,25630	10,10001	438300,84	1567660,62	17,4429513	0,133	0,867	380006,828	1253773,25	13,9504083
70-74	85.447	7.081,910	82,88098	17,29282	409528,963	1129359,78	13,2171184	0,169	0,831	340318,569	873766,423	10,2258593
75-79	78.365	11.362,388	144,99345	31,26532	363418,218	719830,819	9,18563528	0,213	0,787	286010,138	533447,854	6,80723485
80-84	67.002	17.624,729	263,04604	60,57640	290950,425	356412,6	5,31939656	0,294	0,706	205411	247437,717	3,69296523
85y+	49.378	26.184,870	530,29727	144,32776	65462,1752	65462,1752	1,32574316	0,358	0,642	42026,7165	42026,7165	0,85112711

Figura 5.1: Resumen de cálculo de la tabla de mortalidad con la matización del método de Sullivan.

Los resultados de realizar este proceso pueden verse en la figura 5.1, en la que EVBS(0)=74,14 años y la EVBS(65)=13,95 años. En esta tabla se ha puesto variable q_x en tanto por mil, puesto que en las tablas que publica el INE suele aparecer así.

3) En esta región la esperanza de vida al nacer en la población era 79,23 años, mientras que la esperanza de vida libre de incapacidad al nacer era 74,14 años respectivamente. Esto supone que más del 93,6 % de los años de

esperanza de vida eran vividos con buena salud, lo cual denota una región con unas buenas condiciones de salud.

Normalmente los estudios relativos a la esperanza de vida en buena salud distinguen el colectivo de hombres del colectivo de mujeres, analizando lo que se conoce como brecha de género (mujeres-hombres), que es la diferencia en años de la esperanza de vida en buena salud a una determinada edad de la mujer menos la esperanza de vida en buena salud del hombre. Se basa en la prevalencia (proporción de personas que sufren una enfermedad con respecto al total de la población en estudio) de la población a una edad específica con buena y mala salud, y en la información sobre mortalidad específica por edades.

En España, durante los años 2008, 2009, 2010 y 2013, el número de años de vida en buena salud al nacimiento fue superior en los hombres. En los años 2011 y 2012 esta situación cambió y fue superior el número esperado de años de vida en buena salud de las mujeres al nacer. En el año 2014 el número de años de vida en buena salud al nacimiento fue el mismo para los hombres y para las mujeres. En los años 2015, 2016 y 2017 el número de años de vida en buena salud al nacimiento fue superior en las mujeres. Y en el año 2018 el número de años de vida en buena salud al nacimiento fue el mismo para los hombres y para las mujeres (68,0 años). Si se consideran distintas edades, a los 50 años en el año 2018 en España, el horizonte de años de vida en buena salud era de (21,9 años) para los hombres y de (22,1 años) para las mujeres. A los 65 años en el año 2018, el número de años de esperanza de vida en buena salud es de (11,3 años) para las mujeres y de (11,5 años) para los hombres.

A los 65 años, en todo el periodo 2008-2015 es superior el número esperado de años en buena salud de los hombres. En el año 2016 el número de años de vida en buena salud a los 65 años fue el mismo para los hombres y para las mujeres. En el año 2017, la esperanza de vida en buena salud de las mujeres fue ligeramente superior a la de los hombres, 12,4 y 12,3 respectivamente. Y en el año 2018 esta situación cambió y fue superior el número esperado de años de vida en buena salud de los hombres a los 65 años (11,3 para las mujeres y 11,5 para los hombres).

En España, con información correspondiente al año 2018, los hombres al nacer viven el 84,5 % de sus años de esperanza de vida en condiciones de buena salud frente al 79,2 % que suponen los años de esperanza de vida en buena salud de las mujeres respecto a su horizonte total de vida. A los 65 años, los hombres viven el 59,8 % de sus años de horizonte de vida en buena salud frente al 49,0 % del horizonte de años de las mujeres. El mayor número de años de esperanza de vida a todas las edades de las mujeres va asociado a peores condiciones de salud que los hombres.

En los países europeos existe variedad respecto a si la brecha de género en los años de esperanza de vida en buena salud es a favor de las mujeres o de los hombres, tanto al nacer como a los 65 años.

En el año 2018 en la UE-28, el horizonte de años de esperanza de vida en buena salud al nacer de las mujeres alcanzaba un valor de (63,8 años) y en los hombres un valor de (63,4 años), es decir, el valor para la brecha de género en la esperanza de vida en buena salud al nacer es a favor de las mujeres (0,4 años). Países como Francia (1,1 años), Alemania (1,2 años) e Irlanda (2,0 años) tienen una brecha positiva a favor de la mujer en el año 2018. En otros países la brecha de género es a favor de los hombres, como en Países Bajos (3,9 años), Dinamarca (3,4 años), y Portugal (2,3 años).

A los 65 años en el año 2018, en el conjunto de la UE-28 la brecha de género en la esperanza de vida en buena salud es a favor de las mujeres (0,1 años). En España el valor de la brecha de género en el año 2018 es de 0,2 años a favor de los hombres.

5.2. Soluciones del tercer examen del Cuerpo de Diplomados en Estadística del Estado de 2016

Cuestión 1

De una población que sigue una distribución Poisson de parámetro λ desconocido, se obtiene una muestra aleatoria simple de tamaño 3, (x_1, x_2, x_3), y se construyen los siguientes estimadores de la varianza poblacional:

$\hat{V}_1 = \frac{x_1+x_3}{2}$	$\hat{V}_2 = x_1 + x_2 - 2x_3$	$\hat{V}_3 = \frac{\sum_{i=1}^{3}(x_i-\overline{x})^2}{2}$

Se pide:

a) ¿Son insesgados?

b) Calcular el error cuadrático medio de $\hat{V}_2$

c) ¿Es la media muestral, $\overline{x} = \frac{\sum_{i=1}^{3} x_i}{3}$, un estimador insesgado para la varianza poblacional?. Compararla con el estimador $\hat{V}_1$.

d) ¿Alguno de los estimadores propuestos es eficiente?

Solución.

a) Siendo $X_i \sim P(\lambda)$ se sabe que $E[X_i] = \lambda$ y que $Var(X_i) = \lambda$, por tanto:

$$E[\hat{V}_1] = \frac{1}{2}(E[x_1] + E[x_3]) = \frac{1}{2}(\lambda + \lambda) = \lambda$$

$$E[\hat{V}_2] = E[x_1] + E[x_2] - 2E[x_3] = \lambda + \lambda - 2\lambda = 0$$

$\hat{V}_3$ es la cuasivarianza muestral, estimador insesgado de la varianza poblacional, por tanto, estimador insesgado de λ pues:

$$\hat{V}_3 = \frac{\sum_{i=1}^{3}(x_i - \overline{x})^2}{2} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1} = \frac{n}{n-1}\frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n} = \frac{n}{n-1}\hat{\sigma^2}$$

Como $E[\hat{\sigma^2}] = \frac{n-1}{n}\sigma^2$, entonces $E[\hat{V}_3] = \frac{3}{2} \cdot E[\hat{\sigma^2}] = \frac{3}{2}\frac{2}{3}\lambda = \lambda$

De forma que tanto $\hat{V}_1$ como $\hat{V}_3$ son estimadores insesgados de la varianza poblacional.

b)

$$ECM(\hat{V}_2) = Var(\hat{V}_2) + b^2(\hat{V}_2), \text{ siendo } b(\hat{V}_2) = \text{“sesgo de } \hat{V}_2\text{“}$$

Calculando cada parte tenemos:

$$Var(\hat{V}_2) = Var(x_1+x_2-2x_3) = Var(x_1)+Var(x_2)+4Var(x_3) = \lambda+\lambda+4\lambda = 6\lambda$$

$$b(\hat{V}_2) = E[\hat{V}_2] - \lambda = -\lambda$$

$$ECM(\hat{V}_2) = 6\lambda + (-\lambda)^2 = \lambda(6+\lambda)$$

c) Dado que $E[\overline{x}] = \frac{1}{3} \cdot \sum_{i=1}^{3} E[x_i] = \lambda$, sí es un estimador insesgado para la varianza poblacional. El estimador $\hat{V}_1$ es también insesgado. Sin embargo,

$$Var(\hat{V}_1) = \frac{Var(x_1) + Var(x_3)}{4} = \frac{\lambda+\lambda}{4} = \frac{\lambda}{2}$$

$$Var(\overline{x}) = \frac{Var(x_1) + Var(x_2) + Var(x_3)}{9} = \frac{\lambda+\lambda+\lambda}{9} = \frac{\lambda}{3}$$

Luego la media muestral es preferible como estimador pues tiene menor varianza.

d) Vamos a calcular la información de Fisher y la cota de Cramer-Rao en cada caso:

$$P(X=x) = e^{-\lambda}\frac{\lambda^x}{x!} \Rightarrow logP = -\lambda + xlog\lambda - logx! \Rightarrow \frac{\partial logP}{\partial \lambda} = -1 + \frac{x}{\lambda}$$

$$\frac{\partial^2 logP}{\partial \lambda^2} = -\frac{x}{\lambda^2} \Rightarrow I^*(\lambda) = -E\left[\frac{\partial^2 logP}{\partial \lambda^2}\right] = -E\left[-\frac{x}{\lambda^2}\right] = \frac{\lambda}{\lambda^2} = \frac{1}{\lambda}$$

La cota de Cramer-Rao para los estimadores de la clase C(g), es decir, para aquellos tales que $E[T] = g(\theta) = \theta + b(\theta)$ tiene la expresión CCR=$\frac{(1+b'(\theta))^2}{I_n(\theta)}$, es decir, que :

$$Var(T) \geq \frac{(1+b'(\theta))^2}{I_n(\theta)}$$

En el caso de $\hat{V}_1$ y $\hat{V}_3$, al ser insesgados y tener una muestra de tamaño 3, la cota de Cramer-Rao vale CCR=$\frac{1}{I_3(\lambda)} = \frac{1}{3 \cdot I^*(\lambda)} = \frac{1}{3 \cdot \frac{1}{\lambda}} = \frac{\lambda}{3}$, por lo tanto viendo las varianzas de ambos se concluye que solo la media muestral es eficiente dentro de la clase de estimadores insesgados.

Para $\hat{V}_2$ tenemos que su sesgo vale $-\lambda$ por lo tanto, entre la clase de estimadores con este sesgo, la cota de Cramer-Rao vale CCR=$\frac{(1+(-1))^2}{3 \cdot \frac{1}{\lambda}} = 0$, pero hemos visto que la varianza de $\hat{V}_2$ no es nula, por lo que no es eficiente.

Cuestión 2

La duración de la vida de los elementos de una población es una variable aleatoria con función de densidad:

$$f(x) = \begin{cases} \frac{1}{a} e^{-\left(\frac{x}{a}\right)} \ , x > 0 \\ 0 \text{ en el resto} \end{cases}$$

siendo $a > 0$ un parámetro desconocido. Para una muestra de tamaño n calcular:

a) Estimador máximo verosímil de a.

b) Estimador de a por el método de los momentos.

c) ¿Son eficientes?

Solución.

a) La expresión de la función de densidad nos desvela que se trata de una variable aleatoria exponencial, es decir, $X \sim \text{Exp}(\frac{1}{a})$. Calculamos la función de verosimilitud:

$$f(x) = \frac{1}{a} e^{-\left(\frac{x}{a}\right)} \Leftrightarrow f(x_1, \ldots, x_n) = \frac{1}{a^n} e^{-\frac{1}{a} \sum\limits_{i=1}^{n} x_i} \Leftrightarrow log f(x_1, \ldots, x_n) = -n log a - \frac{1}{a} \sum_{i=1}^{n} x_i$$

$$\frac{\partial log f(x_1, \ldots, x_n)}{\partial a} = 0 \Leftrightarrow -\frac{n}{a} + \frac{\sum\limits_{i=1}^{n} x_i}{a^2} = 0 \Leftrightarrow \frac{n}{a} = \frac{\sum\limits_{i=1}^{n} x_i}{a^2} \Leftrightarrow \hat{a} = \frac{\sum\limits_{i=1}^{n} x_i}{n} = \overline{x}$$

b) Siendo $\lambda = \frac{1}{a}$, la condición del método de los momentos es igualar los momentos muestrales a los poblaciones, por lo que tenemos que:
$\frac{\sum_{i=1}^{n} x_i}{n} = E[X] \Leftrightarrow \overline{x} = \frac{1}{\frac{1}{a}} \Leftrightarrow \hat{a} = \overline{x}$, puesto que en una exponencial de parámetro $\frac{1}{a}$, integrando por partes es sencillo comprobar que $E(X) = \frac{1}{\frac{1}{a}} = a$.

c) Se trata de ver si la media muestral es un estimador eficiente de a. Primero calculamos la esperanza de la media muestral, es decir, $E(\overline{x}) = E\left(\frac{\sum_{i=1}^{n} x_i}{n}\right) = \frac{\sum_{i=1}^{n} E(x_i)}{n} = \frac{nE(x_i)}{n} = E(x_i) = a$, así que $\overline{x}$ es un estimador insesgado de a.

Calculamos la información de Fisher y la cota de Cramer-Rao:

$$I_n(a) = -E\left(\frac{\partial^2 logL(x_1,\ldots,x_n)}{\partial a^2}\right)$$

$$\frac{\partial^2 logL(x_1,\ldots,x_n)}{\partial a^2} = \frac{\partial}{\partial a}\left(\frac{\partial logL(x_1,\ldots,x_n)}{\partial a}\right) = \frac{\partial}{\partial a}\left(-\frac{n}{a} + \frac{\sum_{i=1}^{n} x_i}{a^2}\right) =$$

$$= \frac{n}{a^2} - \frac{2a\sum_{i=1}^{n} x_i}{a^4} = \frac{n}{a^2} - \frac{2\sum_{i=1}^{n} x_i}{a^3}$$

$$-E\left(\frac{\partial^2 logL(x_1,\ldots,x_n)}{\partial a^2}\right) = -E\left[\frac{n}{a^2} - \frac{2\sum_{i=1}^{n} x_i}{a^3}\right] = -\frac{n}{a^2} + \frac{2\sum_{i=1}^{n} E[x_i]}{a^3} =$$

$$= -\frac{n}{a^2} + \frac{2na}{a^3} = \frac{n}{a^2}$$

La Cota de Cramer-Rao para estimadores insesgados de a vale
CCR=$\frac{1}{I_n(a)} = \frac{1}{\frac{n}{a^2}} = \frac{a^2}{n}$, y como la varianza de la media muestral coincide con este valor, puesto que $Var(\overline{x}) = \frac{1}{n^2}\sum_{i=1}^{n} Var(x_i) = \frac{1}{n^2} \cdot n \cdot \frac{1}{(\frac{1}{a})^2} = \frac{a^2}{n}$, entonces $\overline{x}$ es eficiente.

Cuestión 3

La longitud (en mm.) de los tornillos fabricados diariamente por una máquina se distribuye según una Normal con media μ desconocida y desviación típica $\sigma = 2$. Un día determinado se toma una muestra de 10 tornillos cuya longitud media resultó ser de 96 mm. Se pide:
a) Calcular un intervalo de confianza al 99 % para la longitud media μ.
b) Determinar el tamaño de la muestra que sería necesario para que un intervalo al 99 % tuviera una longitud igual a 2mm.

Solución.
a) Siendo $X \equiv$ "longitud en mm de los tornillos fabricados en un día", $X \sim N(\mu, 2)$. Se tiene que $\overline{x} = 96$. Un intervalo de confianza para la longitud media con $\frac{\alpha}{2} = \frac{0{,}01}{2} = 0,005$, sería:

$$I_{\mu}^{(1-\alpha)} = \left[\overline{x} - z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}, \overline{x} + z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right], \text{ con } z_{\frac{\alpha}{2}} \text{ tal que } P(N(0,1) \geq z_{\frac{\alpha}{2}}) = \frac{\alpha}{2}$$

teniendo en cuenta que $\frac{\overline{x}-\mu}{\frac{\sigma}{\sqrt{n}}}$ es un pivote con distribución normal $N(0,1)$. Luego con los datos que tenemos el intervalo queda:

$$I_{\mu}^{(1-\alpha)} = \left[96 - 2,575 \cdot \frac{2}{\sqrt{10}}, 96 + 2,575 \cdot \frac{2}{\sqrt{10}}\right] = [94,371, 97,628]$$

b) El tamaño de muestra que sería necesario para que la longitud fuera igual a 2mm sería n tal que

$$\overline{x} + z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} - \left(\overline{x} - z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right) = 2 \Leftrightarrow 2 \cdot z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} = 2 \Leftrightarrow z_{\frac{\alpha}{2}} \cdot \sigma = \sqrt{n} \Leftrightarrow$$

$$\Leftrightarrow n = (2,575 \cdot 2)^2 = (5,15)^2 = 26,52 \approx 27$$

Cuestión 4

Dada la tabla adjunta de medidas del pH y de la concentración de un ácido en sangre en 15 individuos, se pide:

1. Calcular las medias de la concentración del ácido condicionadas por el pH.

2. Calcular las varianzas de la concentración del ácido condicionadas por el pH.

3. Calcular la razón de correlación $\eta^2_{\text{Conc:pH}}$ de la concentración respecto al pH.

4. Si ρ denota el coeficiente de correlación lineal de Pearson entre ambas variables, ¿es mayor ρ^2 o $\eta^2_{\text{Conc:pH}}$? Coméntese muy brevemente.

pH	4	5	7	5	6	5	5	5	5	6	6	5	7	5	5
Concentración	443	296	101	224	91	252	195	550	170	382	152	448	64	147	161

Nota: La varianza marginal de la concentración es 20396,33.

Solución.

1,2) Se calcula en primer lugar la distribución del ácido condicionada por el pH para los distintos valores de éste último. Denotamos por Y="pH" y X="Concentración", de forma que:

$x_i \vert y = 4$	$n_{i\vert j}$	$f_{i\vert j}$
443	1	$\frac{1}{15}$
	$n_{\bullet 4} = 1$	$f_{\bullet 4} = \frac{1}{15}$

$$\overline{x}_{i|4} = 443 \text{ y } S^2_{i|4}(x) = 0$$

$x_i \mid y=5$	$n_{i\mid j}$	$f_{i\mid j}$
147	1	$\frac{1}{9}$
161	1	$\frac{1}{9}$
170	1	$\frac{1}{9}$
195	1	$\frac{1}{9}$
224	1	$\frac{1}{9}$
252	1	$\frac{1}{9}$
296	1	$\frac{1}{9}$
448	1	$\frac{1}{9}$
550	1	$\frac{1}{9}$
	$n_{\bullet 5}=9$	$f_{\bullet 5}=\frac{9}{15}$

$$\overline{x}_{i|5} = \frac{147+\cdots+550}{9} = 271,44$$

$$S^2_{i|5}(x) = a_2 - a_1^2 = \frac{147^2+\cdots+550^2}{9} - 271,44^2 = 17312,914$$

$x_i \mid y=6$	$n_{i\mid j}$	$f_{i\mid j}$
91	1	$\frac{1}{3}$
152	1	$\frac{1}{3}$
382	1	$\frac{1}{3}$
	$n_{\bullet 6}=3$	$f_{\bullet 6}=\frac{3}{15}$

$$\overline{x}_{i|6} = \frac{91+152+382}{3} = 208,33$$

$$S^2_{i|6}(x) = a_2 - a_1^2 = \frac{91^2+152^2+382^2}{3} - 208,33^2 = 15700,22$$

$x_i \mid y=7$	$n_{i\mid j}$	$f_{i\mid j}$
64	1	$\frac{1}{2}$
101	1	$\frac{1}{2}$
	$n_{\bullet 7}=2$	$f_{\bullet 7}=\frac{2}{15}$

$$\overline{x}_{i|7} = \frac{64+101}{2} = 82,5$$

$$S^2_{i|7}(x) = a_2 - a_1^2 = \frac{64^2+101^2}{2} - 82,5^2 = 342,25$$

3) La razón de correlación de la concentración respecto al pH viene dada por

$$\eta^2_{\text{Conc:pH}} = 1 - \sum_{j=1}^{4} \frac{Var\,(Conc|ph=y)\cdot f_{\bullet j}}{Var(Conc)}$$

por lo tanto, sustituyendo convenientemente y teniendo en cuenta que el ejercicio nos proporciona la varianza de la concentración, que tiene un valor de 20396,33 obtenemos:

$$\eta^2_{\text{Conc:pH}} = 1 - \frac{\left(0 + 17312,914\cdot\frac{9}{15} + 15700,22\cdot\frac{3}{15} + 342,25\cdot\frac{2}{15}\right)}{20396,33} = 0,3345$$

4) Sean $\eta^2_{X|Y}$ la razón de correlación y $\rho^2_{X,Y}$ el coeficiente de correlación lineal. Se tiene que:

$$0 \leq \rho^2_{X,Y} \leq \eta^2_{Y|X} \leq 1$$

$$0 \leq \rho^2_{X,Y} \leq \eta^2_{X|Y} \leq 1$$

En caso de que $\rho^2_{X,Y} = \eta^2_{Y|X}$, la curva de regresión de Y sobre X coincide con la recta de Y sobre X. Y si $\rho^2_{X,Y} = \eta^2_{X|Y}$ ocurre lo análogo.

Cuestión 5

El número de viajeros y las pernoctaciones realizadas por los turistas alojados en establecimientos hoteleros de Castilla-La Mancha durante el mes de enero de 2017 vienen reflejados en la siguiente tabla:

	Viajeros	Pernoctaciones
Albacete	18,1	27,5
Ciudad Real	22,2	35,1
Cuenca	15,2	26,8
Guadalajara	16,4	27,1
Toledo	50,0	79,2

Tabla 5.3: Unidad: miles

Fuente: Encuesta de Ocupación Hotelera. INE (www.ine.es)

a) Obtenga los índices simples del número de viajeros para cada provincia tomando Toledo como referencia.

b) Elabore la serie de los índices simples del número de pernoctaciones para cada provincia tomando Toledo como referencia.

c) Para el numero de viajeros, dé el valor del índice complejo no ponderado de la media aritmética de la comunidad autónoma apoyándose en los índices calculados anteriormente.

d) Calcule el índice complejo ponderado de la media aritmética del número de viajeros de la Comunidad Autónoma, usando como ponderación el número de pernoctaciones de cada provincia.

e) Comente brevemente cuál de los dos índices complejos calculados anteriormente sería de mayor utilidad para analizar la evolución del turismo en Castilla-La Mancha.

Solución.

a) Los índices simples del número de viajeros se calcula dividiendo cada valor por el valor de Toledo, que es 50.

Albacete	$\frac{18{,}1}{50} = 0,362 \Rightarrow IViaj_{Toledo}^{Albacete} = 36,2$
Ciudad Real	$\frac{22{,}2}{50} = 0,444 \Rightarrow IViaj_{Toledo}^{CiudadReal} = 44,4$
Cuenca	$\frac{15{,}2}{50} = 0,304 \Rightarrow IViaj_{Toledo}^{Cuenca} = 30,4$
Guadalajara	$\frac{16{,}4}{50} = 0,328 \Rightarrow IViaj_{Toledo}^{Guadalajara} = 32,8$
Toledo	$\frac{50}{50} = 1 \Rightarrow IViaj_{Toledo}^{Toledo} = 100$

b) Los índices simples del número de pernoctaciones se obtiene de forma análoga a la anterior.

Albacete	$\frac{27{,}5}{79{,}2} = 0,3472 \Rightarrow IPernoc_{Toledo}^{Albacete} = 34,72$
Ciudad Real	$\frac{35{,}1}{79{,}2} = 0,4431 \Rightarrow IPernoc_{Toledo}^{CiudadReal} = 44,31$
Cuenca	$\frac{26{,}8}{79{,}2} = 0,3383 \Rightarrow IPernoc_{Toledo}^{Cuenca} = 33,83$
Guadalajara	$\frac{27{,}1}{79{,}2} = 0,3421 \Rightarrow IPernoc_{Toledo}^{Guadalajara} = 34,21$
Toledo	$\frac{79{,}2}{79{,}2} = 1 \Rightarrow IPernoc_{Toledo}^{Toledo} = 100$

c) El índice complejo no ponderado utilizando los datos del apartado b) se puede calcular así:

$$I_{cnp} = \frac{IViaj_{Toledo}^{Albacete} + IViaj_{Toledo}^{CiudadReal} + IViaj_{Toledo}^{Cuenca} + IViaj_{Toledo}^{Guadalajara} + IViaj_{Toledo}^{Toledo}}{5} =$$
$$= \frac{36,2 + 44,4 + 30,4 + 32,8 + 100}{5} = 48,76$$

d) Para el cálculo de este índice se tienen en cuenta cuántas pernoctaciones ha habido en cada provincia, siendo este el peso que afecta a cada índice individual Por tanto calculamos la estructura de las ponderaciones, teniendo en cuenta que la suma de la columna de pernoctaciones es 195,7, de la siguiente forma:

$$w_{Albacete}^{pernoc} = \frac{27,5}{195,7} = 0,1405 \ , \ w_{CiudadReal}^{pernoc} = \frac{35,1}{195,7} = 0,1794 \ ,$$
$$w_{Cuenca}^{pernoc} = \frac{26,8}{195,7} = 0,1369 \ , \ w_{Guadalajara}^{pernoc} = \frac{27,1}{195,7} = 0,1385 \ ,$$
$$w_{Toledo}^{pernoc} = \frac{79,2}{195,7} = 0,4047$$

$$I_{cp} = IViaj_{Toledo}^{Albacete} \cdot w_{Albacete}^{pernoc} + IViaj_{Toledo}^{CiudadReal} \cdot w_{CiudadReal}^{pernoc} +$$
$$+ IViaj_{Toledo}^{Cuenca} \cdot w_{Cuenca}^{pernoc} + IViaj_{Toledo}^{Guadalajara} \cdot w_{Guadalajara}^{pernoc} +$$
$$+ IViaj_{Toledo}^{Toledo} \cdot w_{Toledo}^{pernoc} = 36,2 \cdot 0,1405 + 44,4 \cdot 0,1794 + 30,4 \cdot 0,1369 +$$
$$+ 32,8 \cdot 0,1385 + 100 \cdot 0,4047 = 62,226$$

e) El índice ponderado en este caso utiliza la información auxiliar del número de pernoctaciones para su construcción, de forma que cada provincia tiene un peso diferente y parece más adecuado que el índice sin ponderar que trata a todas las provincias por igual en cuanto a su importancia, cuando no parece que el comportamiento de las mismas sea similar si uno se fija en las pernoctaciones, variable que es más elevada en provincias como Toledo que en el resto.

Cuestión 6

Utilizando los datos siguientes relativos a una economía imaginaria con ausencia de impuestos, calcule en términos del SEC2010:

a) La cuenta completa de bienes y servicios.

b) El Producto Interior Bruto de la economía a precios básicos y a precios de mercado.

c) La producción de I+D realizada en la economía es de 1.000 millones de euros, de la cual se vende el 25 % a otros países. Señale en qué rúbricas de la cuenta de bienes y servicios anterior estaría contabilizada esta operación.

Producción de bienes y servicios	9000
Exportaciones	400
Consumo público	1000
Consumo intermedio	5000
Importaciones	1600
Consumo privado	2500
Subvenciones a los productos	500
Variación de existencias	200

Tabla 5.4: Unidad:millones de euros

Solución.

a)

R	Cuenta de Bienes y Servicios		E
9000	(Producción) P.1	P.2 (Consumos Intermedios)	5000
1600	(Importaciones) P.7	P.3 (GCFind + GCFcol)	(1000+2500)
-500	(Subv's productos) D.31	P.5 (var existencias + FBCF)	200+1000
		P.6 (Exportaciones)	400
10100			10100

b) PIB(pm) = P.3(GCF) + P.5(FBC) + P.6(X) – P.7(M) = (1000+2500) + (200 +1000) + 400 – 1600 = 3500

PIB(pm) = PIB(pb) + D.21 – D.31 $\Rightarrow$

PIB(pb) = PIB(pm) – D.21 + D.31 = 3500 + 500 = 4000=Producción - CI = 9000-5000

c) La producción de I+D forma parte del agregado producción y por tanto su valor afectaría a la producción de bienes y servicios, pero a su vez este concepto de investigación y desarrollo se engloba dentro de la FBCF y como se exporta el 25 %, también se verían afectadas las exportaciones.

R	Cuenta de Bienes y Servicios		E
9000	(Producción) P.1	P.2 (Consumos Intermedios)	5000
1600	(Importaciones) P.7	P.3 (GCFind + GCFcol)	(1000+2500)
-500	(Subv's productos) D.31	P.5 (var existencias + FBCF)	200+750
		P.6 (Exportaciones)	650
10100			10100

Cuestión 7

A partir de la siguiente información y deduciendo aquellas magnitudes que necesite, calcule los siguientes saldos para el total de la economía:

a) El Excedente de Explotación / Renta Mixta Bruto.

b) La Renta Nacional Bruta a precios de mercado.

c) La Renta Nacional Bruta disponible.

d) El Ahorro Nacional Bruto.

e) La capacidad o necesidad de financiación del país. Interprete el resultado.

Datos	
Gasto en consumo final	19.200
Producto interior bruto a precios de adquisición	30.000
Formación bruta de capital fijo	9.375
Remuneración de asalariados (interior)	8.000
Remuneración de asalariados pagada al resto del mundo	52
Remuneración de asalariados recibida del resto del mundo	34
Rentas de la propiedad pagadas al resto del mundo	1.200
Rentas de la propiedad recibidas del resto del mundo	2.000
Impuestos sobre la producción y las importaciones	1.200
Subvenciones a la producción	360
Transferencias corrientes pagadas al resto del mundo	52
Transferencias corrientes recibidas del resto del mundo	272
Consumo de capital fijo	1.480
Exportaciones netas de bienes y servicios	1.425
Transferencias de capital recibidas del resto del mundo	170

Tabla 5.5: Unidad: millones de euros

Solución.

Podemos resolver los apartados con las siguientes igualdades:

a) EBE/RM = PIB(pm) – RA(interior) – (D.2(Imp sb prod e imp's) – D.3(Subv sob prod)) = 30.000 – 8.000 – (1.200 – 360) = 21.160 millones de euros

b) RNB(pm) = PIB(pm) – rentas primarias a pagar RM + rentas primarias a cobrar RM = 30.000 – (52+1.200) + (34+2.000) = 30.782 millones de euros

c) RNBD = RNB – transferencias corrientes a pagar RM + transferencias corrientes cobradas RM = 30.782 – 52 + 272 = 31.002 millones de euros

d) ANB = RNBD – GCF = 31.002 – 19.200 =11.802 millones de euros

e) CNF= Capacidad/Necesidad de financiación = (ANB – CCF) – transferencias de capital a pagar RM + transferencias capital recibidas RM – [FBC – CCF – adquisición de activos no fin no prod] =11.802 – 1.480 + 170 – (9.375 – 1.480) = 11.802 + 170 – 9.375 = 2597 millones de euros

, donde la FBC se puede calcular de la siguiente forma: PIBpm = P.3 + P.5 + P.6 – P.7

30.000 = 19.200 + P.5 + 1.425

P.5 = 9375 millones de euros

Puesto que la capacidad/necesidad de financiación es positiva esto significa que la economía tiene superavit y puede financiar al resto del mundo.

Cuestión 8

En una economía con dos sectores de actividad, se dispone del siguiente extracto de la tabla input-output simétrica:

	SECTOR 1	SECTOR 2	GCF	FBC	X
SECTOR 1	12	34	100	35	7
SECTOR 2	32	35	40	90	10

Tabla 5.6: Unidad:Millones de euros

Además, se proporcionan los siguientes datos para completar la tabla:

- Las importaciones han sido 15 millones de euros de productos del Sector1 y 25 millones de euros de productos del Sector2.
- La remuneración de asalariados ha sido de 75 millones de euros en el Sector1 y 62 millones de euros en el Sector2.
- Los impuestos netos sobre la producción que han sido pagados por el Sector1 ascienden a 2 millones de euros, mientras que en el caso del Sector2 ascienden a 6 millones.
- No existen impuestos sobre los productos en esta economía.

Se pide:

a) Calcular el total de la oferta a precios básicos.

b) Calcular el total de la oferta a precios de adquisición.

c) Calcular el Excedente Bruto de Explotación de cada sector.

d) Calcular el Valor Añadido Bruto de la economía a partir de la tabla por las tres vías (demanda, oferta, rentas).

Solución.

a) Total Oferta a precios básicos $\equiv$ Total empleos = 395

b) Total Oferta a precios de adquisición = Total Oferta a precios básicos + Impuestos netos sobre los productos = Total Oferta a precios básicos , puesto que no tenemos el dato de los impuestos netos.

c) EBE(SECTOR 1) = VAB(pb)-RA - Otros Imp netos sb Producción=129-75-2 = 52, y EBE(SECTOR 2) = 113-62-6 = 45, puesto que para el SECTOR 1 tenemos que:

VAB(pb)=Producción(pb)-C.I(pa)=173-44=129,

y para el SECTOR 2 es:

VAB(pb)=Producción(pb)-C.I(pa)=182-69=113

d) VAB_{rentas}=RA + Otros Imp.Netos sobre Productos+EBE/RMB=137+8+97=242

VAB$_{oferta}$=Producción - Consumos Intermedios=355-113=242

VAB$_{demanda}$=GCF+FBC+X-M=140+125+17-40=242.

Tabla 5.7: Tabla input-output simétrica completa.
T.D.I.(Total Demanda Intermedia),T.D.F.(Total Demanda Final)
Las celdas en color amarillo son los datos que nos proporciona el enunciado dentro de la tabla

	S1	S2	T.D.I.	GCF	FBC	X	T.D.F	Total Empleos
SECTOR 1	12	34	46	100	35	7	142	188 (46+142)
SECTOR 2	32	35	67	40	90	10	140	207 (67+140)
Total(pb)	44	69	113	140	125	17	282	395
Imp.Netos sb Prod's	-	-	-	-	-	-	-	-
Total(pa)	44	69	113	140	125	17	282	395
CI (pa)	44	69	113					
RA	75	62	137					
Otros Imp.Netos sb producción	2	6	8					
EBE/RM	52	45	97					
VAB (pb)	129	113	242					
Producción(pb)	173	182	355					
Importaciones	15	25	40					
Total Oferta(pb)	188	207	395					

Cuestión 9

A partir de estos datos construir la primera fila de estas series de la tabla de mortalidad.

Años	$a(x)$	$m(x)$	$q(x)$	$l(x)$	$d(x)$	$L(x)$
0 años	0,1155			100000		
1 año				99738,88		
2años						

Solución.

Los cálculos que hay que hacer teniendo en cuenta las propiedades de las series biométricas son los siguientes:

$d(0) = l(0) - l(1) = 100000 - 99738,88 = 261,12$

$L(0) = l(1) + a(0)d(0) = 99738,88 + 0,1155 \cdot 261,12 = 99769,03$

$m(0) = \frac{d(0)}{L(0)} \cdot 1000 = \frac{261{,}12}{99769{,}03} \cdot 1000 = 2,61725‰$

$q(0) = \frac{d(0)}{l(0)} = \frac{261{,}12}{100000} = 0,0026112$

de forma que la tabla con la primera fila queda así:

Años	$a(x)$	$m(x)$	$q(x)$	$l(x)$	$d(x)$	$L(x)$
0 años	0,1155	2,6172	2,6112	100000	261,12	99769,03
1 año				99738,88		
2años						

Escribimos en la tabla las series $m(x)$ y $q(x)$ en tanto por mil, que es como suele aparecer en las tablas de mortalidad que publica el INE.

Cuestión 10

Con los siguientes datos calcular:

1 de enero de 2016		
	Hombres	Mujeres
Total	22.809.420	23.636.408
0-15 años	3.852.848	3.624.085
16-64 años	15.203.861	15.065.141
65 y más años	3.752.711	4.947.182

a) El ratio de masculinidad.

b) El índice de envejecimiento.

c) La tasa de dependencia.

d) La tasa de dependencia de jóvenes.

e) La tasa de dependencia de mayores.

Solución.

Construimos la columna de totales en la tabla principal como primer paso:

1 de enero de 2016			
	Hombres	Mujeres	Total
Total	22.809.420	23.636.408	46.445.828
0-15 años	3.852.848	3.624.085	7.476.933
16-64 años	15.203.861	15.065.141	30.269.002
65 y más años	3.752.711	4.947.182	8.699.893

Tenemos las siguientes definiciones de cada uno de los apartados:

$$\text{a) Ratio masculinidad} = \frac{P_{TOTAL}^{hombres}}{P_{TOTAL}^{mujeres}} \cdot 100 = \frac{22.809.420}{23.636.408} \cdot 100 = 96,5\,\%$$

$$\text{b) Índice de envejecimiento} = \frac{P_{>64}}{P_{<16}} \cdot 100 = \frac{8.699.893}{7.476.933} \cdot 100 = 116,35\,\%$$

c) $\text{Tasa de dependencia} = \dfrac{P_{<16} + P_{>64}}{P_{16-64}} \cdot 100 = \dfrac{7.476.933 + 8.699.893}{30.269.002} \cdot 100 = 53,44\,\%$

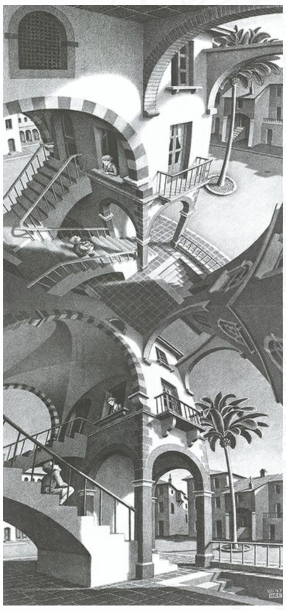
d) $\text{Tasa de dependencia}_{<16} = \dfrac{P_{<16}}{P_{16-64}} \cdot 100 = \dfrac{7.476.933}{30.269.002} \cdot 100 = 24,7\,\%$

e) $\text{Tasa de dependencia}_{>64} = \dfrac{P_{>64}}{P_{16-64}} \cdot 100 = \dfrac{8.699.893}{30.269.002} \cdot 100 = 28,74\,\%$

Capítulo 6

Año 2015

"If my mind can conceive it, if my heart can believe it,
then I can achieve it."
—Muhammad Ali



Maurits Cornelis Escher nació el 17 de junio de 1898 en Leenwarden (Países Bajos), hijo de un ingeniero hidráulico. Era un pésimo estudiante que tuvo que repetir curso dos veces. Para él la escuela era una pesadilla, excepto las clases de dibujo. Como tantos otros grandes artistas, era zurdo. Le gustaban el clima y los paisajes italianos, y a menudo los recorría a pie en larguísimas excursiones.
Escher también viajó a España, donde descubriría la Alhambra de Granada, el Generalife y la Mezquita de Córdoba, cuyas maravillas estudiaría con detalle. Lo que aprendió allí tendría fuertes influencias en muchos de sus trabajos, especialmente en los relacionados con la partición regular del plano y el uso de patrones que rellenan el espacio sin dejar ningún hueco.
Hasta 1951 vivió dependiendo económicamente de sus padres. A partir de entonces comenzó a vender sus grabados y obtener un buen dinero por ellos. Esto le permitió vivir sus últimos años con una economía personal excelente. Generalmente hacía copias de las litografías, grabados por encargo, diseños de sellos, portadas de libros, y algunas esculturas en marfil y madera.
Hasta 1962 su producción de trabajos fue muy constante. Entonces cayó enfermo y eso supuso un pequeño parón transitorio. En 1969 realizó su último trabajo original, Serpientes, que demostraba que su habilidad seguía intacta. Hacia 1970 ingresó en una residencia para artistas en Holanda, donde pudo mantener su propio taller.
Falleció el 27 de marzo de 1972.

$2015 = 5 \cdot 13 \cdot 31$

11111011111

6.1. Soluciones del segundo examen del Cuerpo Superior de Estadísticos del Estado de 2015

Cuestión 1

Sean $(X_1, \ldots, X_n)$ e $(Y_1, \ldots, Y_m)$ dos muestras aleatorias simples de dos distribuciones independientes exponenciales de medias θ_1 y θ_2, respectivamente con θ_1 y θ_2 mayores que 0.

a) Demostrar que la expresión

$$T_X = \frac{2}{\theta_1} \sum_{i=1}^{n} x_i$$

tiene una distribución independiente de θ_1. Identificar esta distribución.

b) Basándose en el resultado anterior, definir una variable pivotal para construir un intervalo de confianza para el cociente de medias $\frac{\theta_1}{\theta_2}$ con nivel de confianza $1 - \alpha$.

c) Calcular la expresión del intervalo de confianza para $\frac{\theta_1}{\theta_2}$ con nivel de confianza $1 - \alpha$.

d) Construir la región crítica para contrastar la hipótesis $H_0 : \theta_1 = \theta_2$ frente a $H_1 : \theta_1 \neq \theta_2$.

Solución.

a) En primer lugar identificamos la distribución de $T_X = \frac{2}{\theta_1} \sum_{i=1}^{n} x_i$,

$$F_{T_X}(x) = P(T_x \leq x) = P\left(\frac{2}{\theta_1} \sum_{i=1}^{n} x_i \leq x\right) = P\left(\sum_{i=1}^{n} x_i \leq \frac{\theta_1}{2} x\right) =$$

$$= F_T\left(\frac{\theta_1}{2}x\right) \text{ con } T = \sum_{i=1}^{n} x_i$$

Consecuentemente,

$$f_{T_X}(x) = \frac{\theta_1}{2} f_T\left(\frac{\theta_1}{2}x\right)$$

Del enunciado se sabe que $X_i \sim exp\left(\frac{1}{\theta_1}\right)$ y por tanto

$$T = \sum_{i=1}^{n} x_i \sim \gamma\left(n, \frac{1}{\theta_1}\right).$$

Una variable aleatoria con distribución $\gamma(p, a)$ tiene como función de densidad

$$f(x) = \frac{a^p x^{p-1} e^{-ax}}{\Gamma(p)}, \; x > 0$$

y por tanto, teniendo en cuenta que $\Gamma(n) = (n-1)!$,

$$f_T(x) = \begin{cases} \frac{\left(\frac{1}{\theta_1}\right)^n x^{n-1} e^{-\frac{x}{\theta_1}}}{(n-1)!} & \text{si } x > 0 \\ \\ 0 & \text{en otro caso} \end{cases}$$

De forma que la función de densidad de $f_{T_X}(x)$ será

$$f_{T_X}(x) = \frac{\theta_1}{2} \frac{\left(\frac{1}{\theta_1}\right)^n \left(\frac{\theta_1}{2}x\right)^{n-1} e^{-\frac{\frac{\theta_1}{2}x}{\theta_1}}}{(n-1)!} = \frac{\frac{\theta_1}{2}}{(n-1)!} \frac{1}{\theta_1^n} e^{-\frac{\theta_1 x}{\theta_1 2}} \left(\frac{\theta_1}{2}x\right)^{n-1} =$$

$$= \frac{\frac{1}{2}}{(n-1)!} e^{\frac{-1}{2}x} \left(\frac{1}{2}x\right)^{n-1},$$

, es decir, tiene una distribución $\gamma\left(n, \frac{1}{2}\right) = \gamma\left(\frac{2n}{2}, \frac{1}{2}\right) \equiv \chi^2_{2n}$, que no depende de θ_1.

b) Siguiendo un razonamiento análogo al anterior, se tiene que $T_Y = \frac{2}{\theta_2}\sum_{i=1}^{m} y_i \sim \chi^2_{2m}$. Por tanto

$$\frac{T_X}{T_Y} = \frac{\frac{2}{\theta_1}\sum_{i=1}^{n} x_i}{\frac{2}{\theta_2}\sum_{i=1}^{m} y_i} \sim \frac{\chi^2_{2n}}{\chi^2_{2m}} \Rightarrow \frac{\frac{2\sum_{i=1}^{n} x_i}{\theta_1 2n}}{\frac{2\sum_{i=1}^{m} y_i}{\theta_2 2m}} = \frac{\frac{1}{\theta_1}\overline{x}}{\frac{1}{\theta_2}\overline{y}} \sim \frac{\frac{\chi^2_{2n}}{2n}}{\frac{\chi^2_{2m}}{2m}} \equiv F_{2n,2m}$$

Y se puede utilizar como variable pivotal, cuya distribución no depende de los parámetros θ_1 y θ_2:

$$\frac{\frac{T_X}{2n}}{\frac{T_Y}{2m}} \equiv F_{2n,2m}.$$

c) Basándose en la variable pivotal obtenida, se tiene que

$$1-\alpha = P\left(F_{2n,2m;1-\frac{\alpha}{2}} \leq \frac{\frac{1}{\theta_1}\overline{x}}{\frac{1}{\theta_2}\overline{y}} \leq F_{2n,2m;\frac{\alpha}{2}}\right)$$

siendo $F_{2n,2m;\frac{\alpha}{2}}$ el punto tal que $P\left(F_{2n,2m} \geq F_{2n,2m;\frac{\alpha}{2}}\right) = \frac{\alpha}{2}$. Así,

$$\begin{aligned} 1-\alpha =& P\left(F_{2n,2m;1-\frac{\alpha}{2}} \leq \frac{\theta_2\overline{x}}{\theta_1\overline{y}} \leq F_{2n,2m;\frac{\alpha}{2}}\right) = \\ =& P\left(\frac{\overline{x}}{\overline{y}}\frac{1}{F_{2n,2m;\frac{\alpha}{2}}} \leq \frac{\theta_1}{\theta_2} \leq \frac{\overline{x}}{\overline{y}}\frac{1}{F_{2n,2m;1-\frac{\alpha}{2}}}\right) \end{aligned}$$

y resulta como intervalo de confianza para $\frac{\theta_1}{\theta_2}$ de nivel de confianza $1-\alpha$ el intervalo

$$I^{1-\alpha}_{\frac{\theta_1}{\theta_2}} = \left[\frac{\overline{x}}{\overline{y}}\frac{1}{F_{2n,2m;\frac{\alpha}{2}}}, \frac{\overline{x}}{\overline{y}}\frac{1}{F_{2n,2m;1-\frac{\alpha}{2}}}\right].$$

Dadas las propiedades siguientes,

$$X \equiv F_{n,m} \Rightarrow \frac{1}{X} \equiv F_{m,n} \qquad \text{y} \qquad F_{n,m;1-\alpha} = \frac{1}{F_{n,m;\alpha}},$$

el intervalo de confianza se puede reescribir como

$$I^{1-\alpha}_{\frac{\theta_1}{\theta_2}} = \left[\frac{\overline{x}}{\overline{y}}F_{2m,2n;1-\frac{\alpha}{2}}, \frac{\overline{x}}{\overline{y}}F_{2m,2n;\frac{\alpha}{2}}\right].$$

d) Contrastar la hipótesis

$$\left\{\begin{array}{ll} H_0: & \theta_1 = \theta_2 \\ H_1: & \theta_1 \neq \theta_2 \end{array}\right. \Leftrightarrow \left\{\begin{array}{ll} H_0: & \frac{\theta_1}{\theta_2} = 1 \\ H_1: & \frac{\theta_1}{\theta_2} \neq 1 \end{array}\right.$$

con nivel de significación α equivale a construir un intervalo de confianza para $\frac{\theta_1}{\theta_2}$ con nivel de confianza $1-\alpha$ y rechazar la hipótesis nula si 1 no pertenece a dicho intervalo.

$$1 \notin I^{1-\alpha}_{\frac{\theta_1}{\theta_2}} = \left[\frac{\overline{x}}{\overline{y}} F_{2m,2n;1-\frac{\alpha}{2}}, \frac{\overline{x}}{\overline{y}} F_{2m,2n;\frac{\alpha}{2}}\right]$$

equivale a ver si

$$\frac{\overline{y}}{\overline{x}} \notin \left[F_{2m,2n;1-\frac{\alpha}{2}}, F_{2m,2n;\frac{\alpha}{2}}\right]$$

por lo que la región crítica será

$$RC = \left\{\frac{\overline{y}}{\overline{x}} \in \left(-\infty, F_{2m,2n;1-\frac{\alpha}{2}}\right) \cup \left(F_{2m,2n;\frac{\alpha}{2}}, \infty\right)\right\}$$

Otra opción para contrastar la hipótesis de igualdad de dos poblaciones sería utilizar un contraste de Kolmogorov - Smirnov para dos muestras extraídas repectivamente de las poblaciones en consideración. Representando por $F_n^*(x)$ y $G_m^*(y)$ las correspondientes funciones de distribución empíricas, donde

$$F_n^*(x) = \frac{1}{n}\sum_{i=1}^{n} I_{(-\infty,x]}(x_i) \qquad \text{y} \qquad G_m^*(y) = \frac{1}{m}\sum_{i=1}^{m} I_{(-\infty,y]}(y_i),$$

y teniendo en cuenta que $F_n^*(x)$ y $G_m^*(y)$ son buenos estimadores de las correspondientes funciones de distribución F y G se utiliza el estadístico de contraste

$$D_{n,m} = \sup_x |F_n^*(x) - G_m^*(y)|.$$

Se rechazaría la hipótesis nula de igualdad de las poblaciones cuando el estadístico tome valores significativamente altos. Por tanto, fijado un nivel de significación α, la región crítica del contraste sería

$$D_{n,m} \geq d_{m,n;\alpha}$$

donde $d_{m,n;\alpha}$ es el valor de $D_{n,m}$ tal que $P(D_{n,m} \geq d_{m,n;\alpha}) = \alpha$.

Cuestión 2

En una población de 20.000 empresas, se quiere estimar la facturación media $\overline{Y}$. Para ello, se obtiene una muestra aleatoria simple (sin reemplazamiento) de tamaño n=5000.

a) Si se dispone de una información auxiliar, que es el número de asalariados para cada empresa, denotada por X, la media de asalariados en la población es $\overline{X} = 35$ y los datos muestrales disponibles son:

$$\widehat{\overline{Y}} = 5,5 \cdot 10^6 \text{ (Media muestral de } Y\text{)}$$

$$\widehat{\overline{X}} = 30 \text{ (Media muestral de } X\text{)}$$

$$S_Y^2 = \frac{\sum_{i=1}^n (y_i - \widehat{\overline{Y}})^2)}{n-1} = 25 \cdot 10^{10}$$

$$S_X^2 = \frac{\sum_{i=1}^n (x_i - \widehat{\overline{X}})^2)}{n-1} = 15$$

$$\widehat{\rho} = 0,7 \text{ (Coeficiente de correlación muestral entre } X, Y\text{)}$$

Se pide:

- Calcular el estimador de razón para $\overline{Y}$. ¿Es insesgado?
- Calcular un estimador de la varianza del estimador de razón.

b) Si no disponemos de esa información auxiliar del apartado anterior pero después de seleccionar la muestra, obtenemos información adicional sobre la actividad económica principal de las empresas que permite clasificarlas en 3 grandes grupos: Industria (post - estrato 1 con N_1=3000), Comercio (post - estrato 2 con N_2=6000) y Servicios (post - estrato 3 con N_3=11000). Si en este caso la información muestral disponible en cada uno de los post - estratos (indicados por el subíndice i = 1,2,3) es:

$$\widehat{\overline{Y}}_1 = 8 \cdot 10^6; \qquad S_{Y_1}^2 = 15 \cdot 10^{10}$$

$$\widehat{\overline{Y}}_2 = 4 \cdot 10^6; \qquad S^2_{Y_2} = 20 \cdot 10^{10}$$

$$\widehat{\overline{Y}}_3 = 5 \cdot 10^6; \qquad S^2_{Y_3} = 22 \cdot 10^{10}$$

Se pide:

- Calcular el estimador post - estratificado para $\overline{Y}$. ¿Es insesgado?
- ¿Cuál es la distribución de n_1 (tamaño muestral del post estrato 1)? Calcular la media y la varianza de n_1.
- Calcular un estimador de la varianza del estimador post - estratificado.
- Calcular un intervalo de confianza al 95 % para $\overline{Y}$. ¿Qué significa este intervalo de confianza?
- ¿Cuál de los dos estimadores, el de razón o el post - estratificado, es más eficiente? Razone la respuesta.

Solución.

a) El estimador de razón para $\overline{Y}$ se calcula como:

$$\widehat{\overline{Y}}_R = \widehat{R}\overline{X} = \frac{\widehat{\overline{Y}}}{\widehat{\overline{X}}}\overline{X} = \frac{5,5 \cdot 10^6}{30} 35 = 6,4166666 \cdot 10^6.$$

El estimador es sesgado pues en general la esperanza de un cociente de variables aleatorias no es igual al cociente de las esperanzas.

$$E\left[\widehat{\overline{Y}}_R\right] = E\left[\widehat{R}\overline{X}\right] = E\left[\frac{\widehat{\overline{Y}}}{\widehat{\overline{X}}}\right]\overline{X} \neq \overline{X}.$$

Un estimador de la varianza del estimador de razón será

$$\widehat{Var}\left(\widehat{\overline{Y}}_R\right) = \overline{X}^2 \cdot \widehat{Var}\left(\widehat{R}\right) = \overline{X}^2 \frac{1}{\overline{X}^2} \frac{1-f}{n}\left[\widehat{S}^2_Y + \widehat{R}^2 \widehat{S}^2_X - 2\widehat{R}\widehat{S}_{XY}\right]$$

donde

$$\widehat{S}_{XY} = \widehat{\rho}\widehat{S}_Y\widehat{S}_X = 0,7\sqrt{25 \cdot 10^{10}}\sqrt{15} = 1355544,171$$

y por tanto:

$$
\begin{aligned}
\widehat{Var}\left(\widehat{\overline{Y}}_R\right) &= \frac{1-\frac{5000}{20000}}{5000}\left[25\cdot 10^{10}+\left(\frac{5,5\cdot 10^6}{30}\right)^2 15-2\frac{5,5\cdot 10^6}{30}1355544,17\right] = \\
&= \frac{3}{20000}\left[25\cdot 10^{10}+\frac{5,5^2\cdot 10^{12}}{30^2}\cdot 15-\frac{5,5\cdot 10^5}{3}2711088,34\right] = \\
&= \frac{3}{20000}\left[\left(25+\frac{5,5^2}{3^2}\cdot 15\right)\cdot 10^{10}-5,5\cdot 10^5\cdot 903696,11\right] \simeq \\
&= \frac{3}{20000}\left[7,54\cdot 10^{11}-4,97\cdot 10^{11}\right] \simeq \\
&\simeq 38550000
\end{aligned}
$$

b) Para calcular el estimador post - estratificado, sean

$$
W_1=\frac{N_1}{N}=\frac{3000}{20000}=\frac{3}{20},\qquad W_2=\frac{N_2}{N}=\frac{3}{10},\qquad W_3=\frac{N_3}{N}=\frac{11}{20},
$$

entonces

$$
\widehat{\overline{Y}}_{POST}=\sum_{h=1}^{L}W_h\overline{y}_h=\frac{3}{20}8\cdot 10^6+\frac{3}{10}4\cdot 10^6+\frac{11}{20}5\cdot 10^6=5150000.
$$

Al ser los pesos de cada estrato W_h conocidos, se trata de un estimador insesgado. En caso de que estos pesos fuesen desconocidos, denotémosles por W'_h, entonces el estimador $\widehat{\overline{Y}}'' = \sum_{h=1}^{L} W'_h\overline{y}_h$ será sesgado y

$$
b\left(\widehat{\overline{Y}}''\right)=E\left(\widehat{\overline{Y}}''\right)-\overline{Y}=\sum_{h=1}^{L}W'_h\overline{y}_h-\sum_{h=1}^{L}W_h\overline{y}_h=\sum_{h=1}^{L}\left(W'_h-W_h\right)\overline{y}_h.
$$

El tamaño muestral del post estrato 1, n_1, sigue una distribución hipergeométrica de parámetros $H(N,N_1,n)\equiv H(20000,3000,5000)$ de forma que su esperanza será $E[n_1]=n\cdot\frac{N_1}{N}=750$ y su varianza será

$$
\begin{aligned}
Var(n_1) &= n\frac{N_1}{N}\left(1-\frac{N_1}{N}\right)\frac{N-n}{N-1} = \\
&= 5000\frac{3}{20}\left(1-\frac{3}{20}\right)\frac{20000-5000}{20000-1}=478,14
\end{aligned}
$$

Un estimador de la varianza del estimador post - estratificado será

$$\begin{aligned}
\widehat{Var}\left(\widehat{\overline{Y}}_{POST}\right) =& \frac{1-f}{n}\sum_{h=1}^{L} W_h \widehat{S}_h^2 + \frac{1-f}{n^2}\sum_{h=1}^{L}(1-W_h)\,\widehat{S}_h^2 = \\
&= \frac{1-\frac{5000}{20000}}{5000}\left(\frac{3}{20}\cdot 15\cdot 10^{10} + \frac{3}{10}\cdot 20\cdot 10^{10} + \frac{11}{20}\cdot 22\cdot 10^{10}\right) + \\
&+ \frac{1-\frac{5000}{20000}}{5000^2}\left[\left(1-\frac{3}{20}\right)\cdot 15\cdot 10^{10} + \left(1-\frac{3}{10}\right)\cdot 20\cdot 10^{10} + \right. \\
&\left. + \left(1-\frac{11}{20}\right)\cdot 22\cdot 10^{10}\right] = \\
&= \frac{3}{20000}2,035\cdot 10^{11} + 3\cdot 10^{-8}\left(3,665\cdot 10^{11}\right) = \\
&= 30535995
\end{aligned}$$

El intervalo de confianza será

$$\begin{aligned}
IC_{\overline{Y}}^{1-\alpha} &= \left[\widehat{\overline{Y}}_{POST} - \lambda_{\frac{\alpha}{2}}\sqrt{\widehat{Var}\left(\widehat{\overline{Y}}_{POST}\right)}, \widehat{\overline{Y}}_{POST} + \lambda_{\frac{\alpha}{2}}\sqrt{\widehat{Var}\left(\widehat{\overline{Y}}_{POST}\right)}\right] = \\
&= \left[5150000 - 1,96\sqrt{30535995}, 5150000 + 1,96\sqrt{30535995}\right] \\
&= [5139169,161; 5160830,839]
\end{aligned}$$

En general, el significado de cualquier intervalo de confianza que se pueda construir es que hay una probabilidad igual a 0,95 (por ejemplo, si usamos este nivel de confianza) de obtener una muestra de tal manera que el intervalo de extremos aleatorios que estamos calculando incluya al parámetro con el que se han realizado las observaciones muestrales, sea cual sea dicho valor. Ahora bien, cuando hemos extraído una muestra concreta, como es nuestro caso y tenemos un intervalo concreto, ahora no tiene sentido pensar en esa probabilidad del 0,95, puesto que ahora los extremos del mismo no son aleatorios. La idea es que ese 0,95 denota el margen de confianza con el que se puede decir que el valor del parámetro está entre los extremos del intervalo. Si obtuviésemos un gran número de muestras y para cada una construyéramos el intervalo de confianza resultando por tanto diferentes intervalos, podríamos decir que aproximadamente el 95 % de ellos contendrían el valor del parámetro.

A la vista de que la varianza del estimador post estratificado es menor se puede decir que este es más eficiente puesto que además es un estimador insesgado, con lo que su error cuadrático medio será menor que el del estimador de la razón.

Cuestión 3

La información contable de una economía imaginaria correspondiente al año t se recoge en la tabla 6.1. Las unidades monetarias que aparecen en la tabla son precios corrientes del año t. Sabiendo que el PIB en t-1 a precios corrientes asciende a 1030 unidades monetarias, realice los siguientes cálculos para el año t:

a) El PIB a precios corrientes y la tasa de variación anual en volumen.

b) El Excedente Bruto de Explotación /Renta mixta.

c) El Consumo final individual efectivo de los hogares.

d) El Ahorro bruto de la economía.

Tabla 6.1: Información contable año t (precios corrientes del año t).

VAB ramas primarias	25
VAB industria	170
VAB construcción	60
VAB servicios	700
Remuneración de los asalariados residentes	500
Remuneración de los asalariados no residentes por empleadores residentes	1
Remuneración de los asalariados residentes por empleadores no residentes	3
Consumo final colectivo efectivo de las AAPP	90
Impuestos netos de subvenciones sobre los productos	80
Otros Impuestos netos de subvenciones sobre la producción	10
Formación bruta de capital	205
Exportaciones de bienes y servicios	310
Importaciones de bienes y servicios	340
Rentas primarias netas recibidas del Resto del Mundo	-10
Transferencias corrientes netas recibidas del Resto del Mundo	-12
Tasa de variación anual en "t" del deflactor implícito del PIB	-0,5 %

Solución.

a) El PIB en el año t a precios corrientes es

$$PIB_t = \Sigma VAB + impuestos\ netos\ sobre\ los\ productos$$
$$= 25 + 170 + 60 + 700 + 80 \Rightarrow PIB_t^{base\ t} = 1035.$$

Sabemos que la Tasa de variación anual en "t"del deflactor implícito del PIB es -0,5 %, es decir,

$$TVI\,(IPC)_{t-1}^{t} = -0,5\,\% = -\frac{0,5}{100} = -0,005,$$

luego el índice de precios será

$$IPC_{t-1}^{t} = \left(TVI\,(IPC)_{t-1}^{t} + 1\right) \cdot 100 = 99,5$$

Por otra parte,

$$TVI\,(PIB\ valor)_{t-1}^{t} = \frac{PIB_t}{PIB_{t-1}} - 1 = \frac{1035}{1030} - 1 = 0,00485 = 0,485\,\%$$

por lo que el índice de valor será $100,485$. Teniendo en cuenta que

$$\acute{I}ndice\ de\ valor = \acute{I}ndice\ de\ precios \cdot \acute{I}ndice\ de\ volumen$$
$$\acute{I}ndice\ de\ volumen = \frac{100,485}{99,5} \cdot 100 = 100,99$$

la tasa de variación en volumen entre t - 1 y t es 0,99 %.

b) El Excedente Bruto de Explotación/Renta mixta se calcula a partir de la siguiente igualdad:

$$PIB = RA_{interior} + EBE/RM +$$
$$+\ impuestos\ netos\ sobre\ la\ producción\ e\ importaciones,$$

donde

$RA_{nacional} = RA_{interior} +$

$$+ RA \text{ recibidas del } RM(\text{pagadas por no residentes a residentes})$$
$$- RA \text{ pagadas al } RM(\text{pagada por residentes a no residentes})$$

Entendemos que el dato de la tabla: 'Remuneración de los asalariados residentes' es la remuneración de asalariados nacional, puesto que no se especifica quien la paga, si es un empleador residente o no, pero en definitiva eso no es relevante, sino saber que es la remuneración de trabajadores residentes.

Luego $RA_{interior} = 500-3+1 = 498$ y por tanto despejando en la ecuación del PIB anterior:

$$EBE/RM = 1035 - 498 - 80 - 10 = 447.$$

c) Dado que el único sector que puede efectuar gasto colectivo es el sector de las AAPP, el C_{col} de la economía coincide con el consumo final colectivo efectivo de las AAPP, que en este caso es 90. Así:

$$PIB_{gasto} = C_{ind} + C_{col} + FBK + X - M$$
$$\Rightarrow C_{ind} = 1035 - (90 + 205 + 310 - 340) = 770$$

y este coincide con el Consumo Final Individual Efectivo de los hogares que comprende el gasto en Consumo individual de las AAPP en forma de transferencias sociales en especie y el de las ISFLSH que también tienen consumo individual.

d) Finalmente se utilizan las cuentas de distribución secundaria de la renta y de utilización de la renta disponible para el cálculo del ahorro bruto de la economía.

E	Cuenta de asignación de la renta primaria		R
		PIB_{pm}	1035
		Rentas primarias netas recibidas del RM	-10
1025	Renta Nacional Bruta		

E	Cuenta de distribución secundaria de la renta		R
		Renta Nacional Bruta	1025
		Transferencias corrientes netas recibidas del resto del mundo	–12
1013	Renta Nacional Bruta Disponible		

E	Cuenta de utilización de la renta disponible		R
770 + 90	GCF	Renta Nacional Bruta Disponible	1013
153	Ahorro bruto		

Cuestión 4

Una investigadora con datos del año 2000 procedente de una encuesta longitudinal nacional realizada por la correspondiente institución oficial norteamericana dispone de información sobre las siguientes variables: salario por hora (SAL), en euros; años de escolarización (S), años de experiencia laboral (EXP); género; etnicidad (raza negra, raza hispana y raza blanca, estos últimos son aquellos que no son ni de raza negra ni hispana).

La investigadora elimina de la muestra a los hispanos, dejando 2135 blancos y 273 de raza negra, y define las variables binarias HOM y NEG. HOM toma el valor 1 para hombres, y 0 para mujeres. NEG es 1 para miembros de raza negra y 0 para blancos. Igualmente define LSAL como el logaritmo natural de la variable SAL.

La investigadora estima por MCO los modelos siguientes cuya variable dependiente siempre es LSAL.

Modelo 1.- Variables explicativas: S, EXP, HOM, para toda la muestra

Modelo 2.- Variables explicativas: S, EXP, HOM, y NEG para toda la muestra

Modelo 3.- Variables explicativas: S, EXP, HOM, para solo blancos

Modelo 4.- Variables explicativas: S, EXP, HOM, para solo negros.

Posteriormente, la investigadora define los siguientes términos de interacción (donde * se refiere a la operación producto):

SN = S*NEG

EN = EXP*NEG

HN = HOM*NEG

y ejecuta una quinta regresión, con la misma variable dependiente:

Modelo 5.- Variables explicativas: S, EXP, HOM, NEG, SN, EN, HN, y usa toda la muestra.

Los resultados se muestran en la siguiente Tabla, en la que no aparecen los datos de la regresión del modelo 4 y donde SCR es la suma cuadrática de los residuos y los valores entre paréntesis son los errores estándar.

	Modelo 1 toda la muestra	Modelo 2 Toda la muestra	Modelo 3 Solo blancos	Modelo4 Solo raza NEG	Modelo 5 Toda la muestra
S	0.124 (0.004)	0.121 (0.004)	0.122 (0.004)	V	0.122 (0.004)
EXP	0.033 (0.002)	0.032 (0.002)	0.033 (0.003)	W	0.033 (0.003)
HOM	0.278 (0.020)	0.277 (0.020)	0.306 (0.021)	X	0.306 (0.021)
NEG	—	–0.144 (0.032)	—	—	0.205 (0.225)
SN	—	—	—	—	–0.009 (0.016)
EN	—	—	—	—	–0.006 (0.007)
HN	—	—	—	—	–0.280 (0.065)
constante	0.390 (0.075)	0.459 (0.076)	0.411 (0.084)	Y	0.411 (0.082)
R-cuadrado	0.335	0.341	0.332	0.321	0.347
SCR	610.0	605.1	555.7	Z	600.0
n	2,408	2,408	2,135	273	2,408

Se pide lo siguiente (sea lo más completo y explícito que pueda en su respuesta, y utilice en su caso las tablas que acompañan este ejercicio).

1. ¿Qué significado tiene la variable constante en el modelo 5?

2. Calcular los coeficientes V, W, X e Y de la regresión del Modelo 4 (no calcular los errores estándar), y Z, que es la SCR, explicando tus cálculos.

3. Dar una interpretación del coeficiente NEG de la regresión del Modelo 2.

4. Realizar un test tipo F del poder explicativo de NEG, SN, EN, y HN en la regresión del Modelo 5. Utilice la tabla adecuada de entre las que se le han proporcionado.

5. Dar una interpretación a los coeficientes de NEG, y HN en la regresión

del Modelo 5.

6. Explicar si un test tipo t sobre el coeficiente NEG en la regresión del Modelo 2 es suficiente para mostrar que las ecuaciones de salarios son diferentes para trabajadores de raza blanca y negra.

7. La investigadora realiza dos regresiones adicionales, una en la que la variable dependiente está formada por los residuos de la regresión del Modelo 5 (RESID), y otra por su cuadrado (RESID2). Las variables explicativas son las de la regresión del Modelo 5. Los R-cuadrado de dichas respectivas regresiones son los siguientes:

   R-cuadrado (RESID) = 0.25

   R-cuadrado (RESID2) =0.26

   Teniendo en cuenta esta información, indique

   (i) para qué sería necesario llevar a cabo un análisis de los residuos;

   (ii) qué tipo de contraste podría plantear;

   (iii) calcúlelo, si es posible, con la información disponible; en caso contrario indique por qué no se podría llevar a cabo.

Solución.

1. Dado que el modelo 5 es

$$ln(SAL) = \beta_0 + \beta_1 S + \beta_2 EXP + \beta_3 HOM + \beta_4 NEG + \beta_5 SN+$$
$$+ \beta_6 EN + \beta_7 HN + u_{m5}$$

la constante hace referencia al logaritmo neperiano del salario por hora de una mujer blanca sin años de escolarización y sin experiencia laboral, pues se consideraría que en ese caso todas las variables explicativas toman el valor cero, al ser ese grupo el grupo base.

2. Comenzamos por el cálculo de la Suma de Cuadrados Residual (SCR) faltante, Z. Para ello, hay que tener en cuenta que se define la SCR como

$$SCR = \sum_{i=1}^{n} \widehat{u}_{mi}^{2} = \sum_{i=1}^{n} \left(In(SAL)_{mi} - \widehat{ln(SAL)}_{mi} \right)^{2}$$

y que se denomina modelo no restringido al modelo 5, cuya ecuación es

$$\begin{aligned} In(SAL) = \beta_0 + \beta_1 S + \beta_2 EXP + \beta_3 HOM + \beta_4 NEG + \beta_5 SN + \\ + \beta_6 EN + \beta_7 HN + u_{m5}, \end{aligned}$$

siendo el modelo estimado:

$$\begin{aligned} \widehat{ln(SAL)} =& 0,411 + 0,122S + 0,033EXP + 0,306HOM + \\ &+ 0,205NEG - 0,009SN - 0,006EN - 0,280HN \text{ (modelo 5)} \end{aligned}$$

El modelo 5 incluye la variable NEG y sus interacciones con el resto de variables del modelo 1, puesto que el modelo 5 trata de comprobar para toda la muestra si la constante y todas las pendientes son las mismas para los dos grupos de raza, negros y blancos. A este modelo 5 se le llama modelo no restringido y su suma residual se puede descomponer en la suma de las SCR de las dos regresiones resultantes de estimar el modelo 1 para los blancos y para los negros, que son los modelos 3 (solo para blancos) y 4 (solo para negros) respectivamente:

$$\widehat{ln(SAL)} = 0,411 + 0,122S + 0,033EXP + 0,306HOM \text{ (modelo 3)}$$
$$\widehat{ln(SAL)} = Y + V \cdot S + W \cdot EXP + X \cdot HOM \text{ (modelo 4)}$$

Así,

$$SCR_{\text{modelo 5;toda la muestra}} = SCR_{\text{modelo 3;raza blanca}} + SCR_{\text{modelo 4;raza negra}}.$$

$Z = SCR_{\text{modelo 4;raza negra}} = 600 - 555,7 = 44,3.$

Se comprueba de forma inmediata que el modelo 5 evaluado para el valor de la variable $NEG = 0$, que sería para los blancos, se reduce al modelo 3 y sus coeficientes son idénticos. Esto mismo debe ocurrir cuando $NEG = 1$ respecto del modelo 4 estimado. Veamos la situación:

Cuando $NEG = 1$ el modelo 5 queda:

$$\begin{aligned}
\widehat{ln(SAL)} &= 0,411 + 0,122S + 0,033EXP + 0,306HOM+ \\
&+ 0,205 - 0,009S - 0,006E - 0,280H = \\
&= 0,616 + (0,122 - 0,009)S + (0,033 - 0,006)EXP+ \\
&+ (0,306 - 0,280)HOM = \\
&= 0,616 + 0,113S + 0,027EXP + 0,026HOM \\
&\text{(modelo 5 para NEG=1)}
\end{aligned}$$

Los coeficientes pedidos en el ejercicio, que son Y, V, W y X por tanto tienen los valores 0,616, 0,113, 0,027 y 0,026 respectivamente. Es decir, la constante es la misma en ambos modelos, y por ejemplo, para el coeficiente V tenemos que:

$$V = \underset{(\text{modelo 4})}{\widehat{\beta}_1} = \underset{(\text{modelo 5})}{\widehat{\beta}_1} + \underset{(\text{modelo 5})}{\widehat{\beta}_5} = 0,122 + (-0,009) = 0,113.$$

En definitiva, ese modelo 4 es el modelo que resulta de considerar en el modelo 5 la variable $NEG = 1$.

$$\widehat{ln(SAL)} = \underset{(Y)}{0,616} + \underset{(V)}{0,113}S + \underset{(W)}{0,027}EXP + \underset{(X)}{0,026}HOM \text{ (modelo 4)}$$

3. La interpretación del coeficiente NEG de la regresión del modelo 2

$$ln(SAL) = \beta_0 + \beta_1 S + \beta_2 EXP + \beta_3 HOM + \beta_4 NEG + u_{m2},$$

que ha sido estimado obteniendo:

$$\widehat{ln(SAL)} = 0,459 + 0,121S + 0,032EXP + 0,277HOM + \\ - 0,144NEG \text{ (modelo 2).}$$

se basa en que es un modelo con factores cualitativos, y que en este caso la categoría base son las mujeres blancas, puesto que para este grupo el modelo sería:

$$E(ln(SAL_i)|i \text{ es mujer blanca}) = \beta_0 + \beta_1 S + \beta_2 EXP$$

Pero también tenemos:

$$E(ln(SAL_i)|i \text{ es hombre blanco}) = \beta_0 + \beta_3 + \beta_1 S + \beta_2 EXP$$

$$E(ln(SAL_i)|i \text{ es mujer negra}) = \beta_0 + \beta_4 + \beta_1 S + \beta_2 EXP$$

$$E(ln(SAL_i)|i \text{ es hombre negro}) = \beta_0 + \beta_3 + \beta_4 + \beta_1 S + \beta_2 EXP$$

Por tanto el coeficiente NEG de la regresión, que es β_4, se interpreta de la siguiente forma:

$$\beta_4 = E(ln(SAL_i)|i \text{ es mujer negra}) - E(ln(SAL_i)|i \text{ es mujer blanca}) = \\ = E(ln(SAL_i)|i \text{ es hombre negro}) - E(ln(SAL_i)|i \text{ es hombre blanco})$$

de manera que en este modelo β_4 mide las diferencias en el salario medio entre negros y blancos con la misma experiencia laboral, escolarización y que sean del mismo sexo. Este coeficiente es significativo de forma individual, por lo que su interpretación en la práctica es la que acabamos de comentar.

De la misma forma:

$$\beta_3 = E(ln(SAL_i)|i \text{ es hombre negro}) - E(ln(SAL_i)|i \text{ es mujer negra}) =$$
$$= E(ln(SAL_i)|i \text{ es hombre blanco}) - E(ln(SAL_i)|i \text{ es mujer blanca})$$

, es decir, β_3 mide las diferencias en el salario medio entre hombres y mujeres con la misma experiencia laboral, educación y que sean de la misma raza. Igual que antes, este coeficiente es significativo individualmente.

Por tanto, y con el modelo estimado que tenemos, manteniendo constantes los niveles de escolarización, los años de experiencia laboral y el sexo, una persona de raza negra vería reducido su salario un 14,4 % respecto al salario de una persona de raza blanca. La justificación es que si se comparan los valores de $ln\,(SAL)$ antes y después de que se haya producido una variación discreta, ΔNEG, en NEG, manteniendo el resto de variables constantes tenemos:

$$\begin{aligned} &ln\,(SAL + \Delta SAL) - ln\,(SAL) = \\ &\qquad = [\beta_0 + \beta_4\,(NEG + \Delta NEG)] - [\beta_0 + \beta_4 NEG] = \\ &\qquad = \beta_4\,(\Delta NEG)\,. \end{aligned}$$

De este modo, al considerar

$$ln\,(SAL + \Delta SAL) - ln\,(SAL) \simeq \frac{\Delta SAL}{SAL}$$

se obtiene la aproximación y por tanto

$$\frac{\Delta SAL}{SAL} \simeq \beta_4\,(\Delta NEG) = -0,144\,(\Delta NEG)\,,$$

es decir, un cambio unitario en la variable NEG genera un cambio en $\frac{\Delta SAL}{SAL}$ de $-0,144$, que implica una variación porcentual en SAL de

$-14,4\,\%$. Sin embargo, esto no es más que una aproximación, pues se trata de un modelo log - nivel en el que se toma el logaritmo de la variable dependiente pero las variables explicativas están en niveles. El cálculo empírico de la variación se lleva a cabo de la siguiente forma, teniendo en cuenta el modelo inicial, que es el modelo 2:

$$ln\,(SAL)_{raza\ negra} - ln\,(SAL)_{raza\ blanca} = ln\left(\frac{(SAL)_{raza\ negra}}{(SAL)_{raza\ blanca}}\right) = \beta_4$$

y tomando exponenciales a ambos lados,

$$\frac{(SAL)_{raza\ negra}}{(SAL)_{raza\ blanca}} = e^{\beta_4},$$

donde restando una unidad a izquierda y derecha del igual

$$\frac{(SAL)_{raza\ negra}}{(SAL)_{raza\ blanca}} - 1 = \frac{(SAL)_{raza\ negra} - (SAL)_{raza\ blanca}}{(SAL)_{raza\ blanca}} = e^{\beta_4} - 1.$$

De modo que la diferencia porcentual exacta entre los salarios de las personas de raza blanca y negra que sean del mismo sexo y con iguales niveles de escolarización y experiencia es

$$100 \cdot \left(e^{\beta_4} - 1\right) = 100 \cdot \left(e^{-0{,}144} - 1\right) = -13,41\,\%,$$

es decir, las personas de raza negra tienen en media un salario un 13,41 % menor que las personas de raza blanca en este país, manteniendo constantes el resto de variables.

4. El modelo 5 es:

$$\begin{aligned} ln(SAL) = \beta_0 + \beta_1 S + \beta_2 EXP + \beta_3 HOM + \beta_4 NEG + \beta_5 SN + \\ + \beta_6 EN + \beta_7 HN + u_{m5} \end{aligned}$$

La hipótesis nula es que las variables NEG, SN, EN y HN, de forma conjunta no tienen poder explicativo, es decir, ($H_0 : \beta_4 = \beta_5 = \beta_6 = \beta_7 = 0$)

mientras que la hipótesis alternativa sería que alguna de ellas lo tuviera. El estadístico de contraste sería

$$F = \frac{\frac{SCR_R - SCR_{NR}}{s}}{\frac{SCR_{NR}}{T-k-1}} \sim F_{s,T-k-1}.$$

Utilizando como modelo restringido el Modelo 1 y siendo $s = 4$,

$$F = \frac{\frac{SCR_R - SCR_{NR}}{s}}{\frac{SCR_{NR}}{T-k-1}} = \frac{\frac{610-600}{4}}{\frac{600}{2408-7-1}} = \frac{\frac{10}{4}}{\frac{1}{4}} = 10$$

De modo que $F_{exp} = 10$ y $F_{teórico} = F_{(s,T-k-1;\alpha)} = F_{(4,\infty;0{,}05)} = 2{,}41$ con el nivel de significación habitual del 5 %. Al ser $F_{exp} > F_{teórico}$, se rechaza la hipótisis nula de no significación y se concluye que las variables NEG, SN, EN y HN tienen poder explicativo.

5. Antes de analizar el caso que nos ocupa vamos a ilustrar un ejemplo más sencillo para comprender el significado de la interacción con claridad. En el apartado (3), el modelo 2 no permitía que las diferencias por sexo para un nivel de experiencia y raza dados fueran diferentes en función de la raza de una persona.

   Para permitir que las diferencias entre hombres y mujeres sean diferentes para blancos y negros, y que las diferencias salariales medias entre blancos y negros sean distintas entre hombres y mujeres, se consigue añadiendo a ese modelo 2 una variable que sea una interación entre la variable ficticia de Hombre y la variable ficticia Negro. Es decir, el modelo tendría que ser:

$$ln(SAL) = \beta_0 + \beta_1 S + \beta_2 EXP + \beta_3 HOM + \beta_4 NEG + \beta_5 HN + u_{modInter},$$

   Si calculamos en este modelo la media del salario para hombres y mujeres, negros y blancos, y que tengan los mismos años de experiencia laboral y de escolarización, tenemos las relaciones:

$$E(Ln(SAL_i)|\text{i es mujer blanca}) = \beta_0 + \beta_1 S + \beta_2 EXP$$
$$E(Ln(SAL_i)|\text{i es hombre blanco}) = \beta_0 + \beta_3 + \beta_1 S + \beta_2 EXP$$
$$E(Ln(SAL_i)|\text{i es mujer negra}) = \beta_0 + \beta_4 + \beta_1 S + \beta_2 EXP$$
$$E(Ln(SAL_i)|\text{i es hombre negro}) = \beta_0 + \beta_3 + \beta_4 + \beta_5 + \beta_1 S + \beta_2 EXP$$

Por lo tanto:

$$\beta_3 = E(Ln(SAL_i)|\text{hombre blanco}) - E(Ln(SAL_i)|\text{mujer blanca})$$
$$\beta_3 + \beta_5 = E(Ln(SAL_i)|\text{hombre negro}) - E(Ln(SAL_i)|\text{mujer negra})$$
$$\beta_4 = E(Ln(SAL_i)|\text{mujer negra}) - E(Ln(SAL_i)|\text{mujer blanca})$$
$$\beta_4 + \beta_5 = E(Ln(SAL_i)|\text{hombre negro}) - E(Ln(SAL_i)|\text{hombre blanco})$$

Introducir en el modelo la interación entre las dos variables ficticias nos permite que las diferencias entre hombres y mujeres sean distintas entre negros y blancos (la diferencia por sexos para individuos negros es β_5 unidades mayor que para los blancos). Asimismo, las diferencias entre blancos y negros es distinta entre hombres y mujeres (la diferencia entre negros y blancos es β_5 unidades mayor para los hombres que para las mujeres). De esta forma se puede entender el coeficiente de HN en ese modelo 2 modificado que comentábamos.

El modelo 5 al que se refiere este apartado es más complejo, puesto que añade más variables explicativas y además interacciones con las variables dicótomas.

En el modelo 5,

$$\widehat{Ln(SAL)} = \underset{(0{,}082)}{0,411} + \underset{(0{,}004)}{0,122S} + \underset{(0{,}0{,}003)}{0,033\,EXP} + \underset{(0{,}021)}{0,306HOM} +$$

$$+\underset{(0{,}225)}{0,205}NEG-\underset{(0{,}016)}{0,009}SN-\underset{(0{,}007)}{0,006}EN-\underset{(0{,}065)}{0,28}HN$$

el grupo base son las mujeres blancas, puesto que en este caso todas las variables binarias son nulas y las interacciones también, quedando el mismo simplificado de la siguiente forma:

$$\widehat{Ln(SAL)}_{MujBlan}=0,411+0,122S+0,033EXP$$

que podemos escribir equivalentemente:

$$E(Ln(SAL_i)|\text{i es mujer blanca})=0,411+0,122S+0,033EXP$$

La expresión del modelo 5 de forma genérica es:

$$ln(SAL)=\beta_0+\beta_1S+\beta_2EXP+\beta_3HOM+\beta_4NEG+\beta_5SN+$$
$$+\beta_6EN+\beta_7HN+u_{m5}$$

y tenemos las siguientes igualdades:

$$E(Ln(SAL_i)|\text{i es mujer blanca})=\beta_0+\beta_1S+\beta_2EXP$$
$$E(Ln(SAL_i)|\text{i es hombre blanco})=\beta_0+\beta_3+\beta_1S+\beta_2EXP$$
$$E(Ln(SAL_i)|\text{i es mujer negra})=\beta_0+\beta_4+(\beta_1+\beta_5)S+(\beta_2+\beta_6)EXP$$
$$E(Ln(SAL_i)|\text{i es hombre negro})=\beta_0+\beta_3+\beta_4+\beta_7+(\beta_1+\beta_5)S+$$
$$+(\beta_2+\beta_6)EXP$$

y consecuentemente, si los niveles de experiencia y escolaridad son fijos y se mantienen constantes:

$$\beta_3=E(Ln(SAL_i)|\text{hombre blanco})-E(Ln(SAL_i)|\text{mujer blanca})$$
$$\beta_3+\beta_7=E(Ln(SAL_i)|\text{hombre negro})-E(Ln(SAL_i)|\text{mujer negra})$$

y la interpretación teórica del coeficiente de HN: $\beta_7=-0,28$ sería que las diferencias entre hombres y mujeres sean distintas entre blancos y negros (la diferencia por sexos para individuos negros es un 28 % menor que para los blancos).

El coeficiente de $NEG(\beta_4 = 0,205)$ indicaría de forma teórica que una mujer negra, en una situación en la que no tiene experiencia ni años de escolaridad tendría un salario que en media sería un 20,5 % más alto que el de una mujer blanca en esas mismas condiciones. Decimos de forma teórica porque en el modelo estimado este coeficiente NEG, además de los coeficientes de SN y EN no son significativos individualmente y por tanto interpretarlos desvirtuaría la lectura real de estos colectivos. De hecho se observa, por ejemplo, que el error estandar de la variable NEG en este modelo 5 es mayor que el error estandar de la variable NEG en el modelo 2 de modo que el coeficiente de NEG en el modelo 5 se estima de forma menos precisa que en el modelo 2.

Por tanto, realmente, atendiendo a los coeficientes que son significativos en este modelo 5, es decir, a 0,411(constante), 0,122(S), 0,033(EXP), 0,306(HOM) y -0,280(HN), el grupo base ahora son las mujeres, no las mujeres blancas, porque el coeficiente de NEG no es adecuado interpretarlo. Como el coeficiente de HOM es 0,306 el modelo arroja que los hombres blancos ganan un 30 % más que las mujeres. La lectura del coeficiente de HN se hace en términos de que un hombre negro solo gana un 2,6 % más que las mujeres, puesto que los coeficientes de HOM y de HN interactúan obteniéndose 0,306-0,280=0,026.

En este modelo 5 lo que ocurre es que al introducir más variables en el modelo respecto al modelo 2, en el que la variable NEG relativa a la raza era significativa y podíamos interpretar la comparación entre razas, ahora esa variable deja de ser significativa porque la raza interacciona con el sexo a través de la variable HN.

6. El modelo 2 era:

   $ln(SAL) = \beta_0 + \beta_1 S + \beta_2 EXP + \beta_3 HOM + \beta_4 NEG + u_{m2}$

   $ln(\widehat{SAL}) = 0,459 + 0,121S + 0,032EXP + 0,277HOM - 0,144NEG.$

   La hipótesis nula $H_0 : \beta_4 = 0$ equivale a decir que una vez que la escolarización, la experiencia y el sexo han sido tomadas en cuenta, la raza no tiene efecto sobre el salario por hora. El estadístico de contraste es

   $$T = \frac{\widehat{\beta}_4}{\widehat{\sigma}(\widehat{\beta}_4)} \sim t_{T-k-1}.$$

   Calculamos:

   $$T = \frac{\widehat{\beta}_4}{\widehat{\sigma}(\widehat{\beta}_4)} = -4,5 \qquad t_{(T-k-1;\frac{\alpha}{2})} = t_{(\infty;0,975)} = 1,96$$

   y dado que el estadístico muestral en valor absoluto es mayor que el obtenido de las tablas, se rechaza la hipótesis nula y se concluye que la raza sí tiene efecto sobre el salario por hora. Sin embargo, no puede decirse que un test tipo t sobre el coeficiente NEG en la regresión del Modelo 2 sea suficiente para mostrar que las ecuaciones de salarios son diferentes para trabajadores de raza blanca y negra pues, a pesar de la significación individual de la variable, pueden presentarse distintas situaciones que influyan en la representatividad de esta variable en el modelo global. La inclusión de nuevas variables pueden suponer un cambio en la estimación del coeficiente de la variable NEG variando su magnitud y por tanto el valor del estadístico de contraste pudiendo dejar de ser significativa individualmente.

7. (I) Se lleva a cabo un análisis de los residuos para ver si la especificación del modelo es correcta. En caso de que los residuos no sean ruido blanco y presenten una estructura determinada se

tendrán indicios de mala especificación del modelo (presencia de autocorrelación o heteroscedasticidad).

(II) Tenemos el modelo 5

$$ln(SAL) = \beta_0 + \beta_1 S + \beta_2 EXP + \beta_3 HOM + \beta_4 NEG + \beta_5 SN + \\ + \beta_6 EN + \beta_7 HN + u_{m5}$$

sobre el que se construyen sendas regresiones adicionales:

$$\hat{u}_{m5} = \beta_0^u + \beta_1^u S + \beta_2^u EXP + \beta_3^u HOM + \beta_4^u NEG + \beta_5^u SN + \\ + \beta_6^u EN + \beta_7^u HN + v_{u;m5} \ , \ R^2_{\hat{u}_{m5}} = 0,25$$

$$\hat{u}^2_{m5} = \beta_0^{u^2} + \beta_1^{u^2} S + \beta_2^{u^2} EXP + \beta_3^{u^2} HOM + \beta_4^{u^2} NEG + \beta_5^{u^2} SN + \\ + \beta_6^{u^2} EN + \beta_7^{u^2} HN + v_{u^2;m5} \ , \ R^2_{\hat{u}^2_{m5}} = 0,26$$

Al disponer de los datos anteriores se podría plantear un contraste de heterocedasticidad utilizando los estadísticos F o ML. A este contraste se le conoce como contraste de heteroscedasticidad de Breusch - Pagan. La hipótesis nula sería

$$H_0 : Var\left(u_t|X_t\right) = E\left(u_t^2|X_t\right) = \sigma^2$$

de modo que si la hipótesis nula no es cierta $E\left(u_t^2|X_t\right)$ no será constante, sino que será función de al menos una de las variables explicativas del modelo. Así la hipótesis alternativa sería

$$H_1 : Var\left(u_t|X_t\right) = E\left(u_t^2|X_t\right) = g\left(Z_t\right) = \sigma_t^2$$

donde $Z_t = (Z_{1t}, \ldots, Z_{pt})'$ son algunas de las variables explicativas del vector de las X_t y/o funciones conocidas de ellas. Por tanto, una forma de comprobar si la esperanza condicionada de u_t^2 depende de ese subconjunto Z_t es mediante la regresión

$$u_t^2 = \delta_0 + \delta_1 Z_{1t} + \cdots + \delta_p Z_{pt} + v_t,$$

donde v_t se supone ruido blanco. De este modo, la hipótesis nula de homoscedasticidad se puede reescribir como

$$H_0 : \delta_1 = \cdots = \delta_p = 0,$$

El contraste de significatividad de estos coeficientes se puede llevar acabo utilizando el coeficiente de determinación R^2 de la regresión que tiene como variable dependiente los residuos al cuadrado. Podríamos pues utilizar, como se adelantaba al principio, los estadísticos F o LM que tienen las siguientes expresiones:

$$F = \frac{\frac{R^2_{\widehat{u}^2}}{p}}{\frac{1-R^2_{\widehat{u}^2}}{T-p-1}} \sim F_{k,T-p-1} \qquad LM = T \cdot R^2_{\widehat{u}^2} \sim \chi^2_p$$

(III) Utilizando el estadístico $LM = T \cdot R^2_{\widehat{u}^2} = 2408 \cdot 0,26 = 626,08$. Por otra parte, considerando el nivel de significación individual habitual, ($\alpha = 0,05$), el valor proporcionado por la tabla de Chi Cuadrado será $\chi^2_{7;0,05} = 14,07$ de forma que $LM > \chi^2_{7;0,025}$ y por tanto se rechaza la hipótesis nula de homoscedasticidad.

Cuestión 5

1) Calcular los siguientes parámetros a partir de los datos (en miles) de un país, referidos a un año T cualquiera, que figuran en la figura 6.1.

a) Tasas de migración exterior con destino al extranjero

b) Probabilidades de migrar al exterior

c) Sedentarios de la tabla en relación con la migración exterior

d) Añada otro parámetro básico que pueda establecerse en una tabla de migraciones

2) Calcule dos medidas agregadas relativas a la migración exterior a partir de los datos de la tabla. Comente sus resultados.

3) Qué medida de la intensidad inmigratoria exterior por grupos de edades puede elaborarse con los datos que suministran. Explique algunos elementos de esta medida: a) sus diferencias con una tabla de migración y b) su sentido probabilístico, así como otras consideraciones que resulten pertinentes.

4) Comente las principales magnitudes de la evolución demográfica para el total de la población mencionada durante el periodo.

	Población al incio del año	Población al final del año	Defunciones	Migraciones entre provincias	Emigraciones hacia el extranjero	Inmigraciones desde el extranjero
0-4 años	2.422,8	2.320,4	1,5	24,6	25,0	18,0
5-9 años	2.440,5	2.478,1	0,2	21,1	21,6	13,8
10-14 años	2.226,7	2.267,6	0,2	16,1	20,4	16,0
15-19 años	2.165,6	2.140,7	0,4	18,2	23,7	23,4
20-29 años	5.343,3	5.121,9	1,5	108,9	125,7	79,4
30-39 años	7.761,9	7.484,1	3,8	137,9	163,9	57,1
40-49 años	7.522,8	7.547,3	10,8	71,5	90,0	34,3
50-59 años	6.079,9	6.212,0	23,7	37,3	43,2	22,0
60-64 años	2.502,3	2.492,8	17,6	13,5	11,7	9,8
65 y más año	8.262,1	8.442,9	330,9	40,1	22,6	17,3

Figura 6.1: Datos de migraciones de un país en el año T

Solución.

1) a) Utilizando la fórmula

$$m_{x,x+n} = \frac{E_x}{\frac{P_{01,01.x}+P_{31,12.x}}{2}} \cdot 1000$$

se calculan las tasas de migración exterior con destino al extranjero. Puesto que tenemos las defunciones en cada grupo de edad, si los datos fueran longitudinales se podría corregir el denominador restando la mitad de las muertes, pero en este caso los datos del problema son transversales y no es necesaria dicha matización. En el denominador, para el cálculo de estas tasas, teóricamente hay que reflejar la población residente media en el ámbito de estudio, en el año T.

$$m_{0,4} = \frac{25,0}{\frac{2422,8+2320,4}{2}} \cdot 1000 = 10,5‰$$

$$m_{5,9} = \frac{21,6}{\frac{2440,5+2478,1}{2}} \cdot 1000 = 8,8‰$$

$$m_{10,14} = 9,1‰$$

$$m_{15,19} = 11‰$$

$$m_{20,29} = 24‰$$

$$m_{30,39} = 21,5‰$$

$$m_{40,49} = 11,9‰$$

$$m_{50,59} = 7‰$$

$$m_{60,64} = 4,7‰$$

$$m_{>65} = 2,7‰$$

b) La probabilidad de migrar al exterior se calcula usando

$$e_{x,n} = \frac{2n \cdot m_{x,x+n}}{2 + (n \cdot m_{x,x+n})}$$

$$e_{0,4} = 0,0514$$

$$e_{5,9} = 0,043$$

$$e_{10,14} = 0,0444$$

$$e_{15,19} = 0,0536$$

$$e_{20,29} = 0,2145$$

$$e_{30,39} = 0,1941$$

$$e_{40,49} = 0,1127$$

$$e_{50,59} = 0,0679$$

$$e_{60,64} = 0,0232$$

Para el último grupo de edad, se considera la amplitud del último intervalo como $n = 90 - 65 = 25$ años.

$$e_{>65} = \frac{2 \cdot 25 \cdot 0,0027}{2 + (25 \cdot 0,0027)} = 0,0654$$

c) Los sedentarios de la tabla en relación con la migración exterior son para el primer intervalo $S_{0,4} = 100000$. Para el resto se utiliza:

$$S_{x+n} = S_x - (S_x \cdot e_{x,n})$$

$$S_{5,9} = S_{0,4} - (S_{0,4} e_{0,4}) = 100000 - (100000 \cdot 0,0514) = 94864,6$$

$$S_{10,14} = 90788,2$$

$$S_{15,19} = 86758,7$$

$$S_{20,29} = 82111,7$$

$$S_{30,39} = 64501,6$$

$$S_{40,49} = 51979,5$$

$$S_{50,59} = 46120,9$$

$$S_{60,64} = 42989,1$$

$$S_{>65} = 41993,8$$

d) Otro parámetro básico que puede establecerse en una tabla de migraciones son los sucesos o migrantes teóricos. Recoge el número de sucesos objeto de estudio que se produce entre dos edades exactas utilizando:

$$M_{x,x+n} = S_x - S_{x+n}.$$

Por tanto:

$$M_{0,4} = S_0 - S_5 = 100000 - 94864,6 = 5135,4$$

$$M_{5,9} = 4076,5$$

$$M_{10,14} = 4029,5$$

$$M_{15,19} = 4646,9$$

$$M_{20,29} = 17610,1$$

$$M_{30,39} = 12522,1$$

$$M_{40,49} = 5858,6$$

$$M_{50,59} = 3131,8$$

$$M_{60,64} = 995,3$$

2) Dos medidas agregadas relativas a la migración exterior a partir de los datos de la tabla podrían ser la intensidad migratoria y la edad media a la migración. La intensidad migratoria se calcula como:

$$IM = \frac{\sum_{x=0}^{w-n} M_{x,x+n}}{100000} \cdot 100 = \frac{5135,4 + 4076,5 + \cdots + 3131,8 + 995,3}{1000} =$$
$$= \frac{58006,2}{1000} \simeq 58\,\%.$$

Es decir, un 58 % de una cohorte ficticia que mostrase los mismos patrones que la población en estudio emigraría al exterior alguna vez en su vida. Por otra parte, la edad media a la migración se calcula como

$$\overline{x}_m = \frac{\sum_{x=0}^{w-n} \left(x + \frac{n}{2}\right) M_{x,x+n}}{\sum_{x=0}^{w-1} M_{x,x+n}}$$

donde x hace referencia al inicio del intervalo y n a la amplitud del mismo. Así,

$$\overline{x}_m = \frac{\left(0 + \frac{5}{2}\right) 5135,4 + \left(5 + \frac{5}{2}\right) 4076,5 + \left(10 + \frac{5}{2}\right) 4029,5 + \cdots}{5135,4 + 4076,5 + 4029,5 + \cdots}$$

$$= \frac{1765539,51}{58006,2} = 26,75$$

Así, la edad media a la que se emigraría al exterior en una generación ficticia que mostrase los patrones de emigración al exterior del país en estudio sería a los 26,75 años.

3) Una medida de la intensidad inmigratoria exterior por grupos de edades que puede elaborarse con los datos que suministran es el índice sintético de movilidad:

$$ISM = \sum_{x=0}^{w-n} n \cdot m_{x,x+n} =$$

$$= 5 \cdot 0,0105 + 5 \cdot 0,0088 + 5 \cdot 0,0091 + 5 \cdot 0,011 + 10 \cdot 0,024+$$

$$+ 10 \cdot 0,0215 + 10 \cdot 0,0119 + 10 \cdot 0,007 + 5 \cdot 0,0047 + 25 \cdot 0,0027 =$$

$$= 4,6654.$$

Así, un individuo residente en este ámbito geográfico emigraría con destino al extranjero un número medio de 4,6654 veces a lo largo de su vida en caso de mantener la misma intensidad a la emigración por edad que la observada en el año t para dicho colectivo poblacional. El índice Sintético de Movilidad se presenta como el mejor indicador de intensidad pues al eliminar los efectos no deseados de la estructura por edad refleja, exclusivamente, la intensidad migratoria. Además no solo resume el patrón migratorio del ámbito seleccionado sino que es una buena medida comparativa. La única limitación asociada a él es que al ser un indicador transversal puede verse afectado por la coyuntura del momento analizado. Indica el número de desplazamientos, migraciones, que un individuo perteneciente a una generación ficticia realizaría como media si a lo largo de su vida pasara por las sucesivas tasas específicas de migración observadas en la tabla. Operativamente se obtiene a partir del sumatorio de las respectivas tasas específicas de migración por edad. Nótese que si se trabaja con grupos de edad, la suma debe multiplicarse por su amplitud pues se asume la hipótesis que los individuos permanecen dentro de cada grupo de edad el mismo número de años que la amplitud del mismo.

4) El método más frecuentemente utilizado para el estudio de la evolución demográfica de las poblaciones es el método de las componentes. El núcleo formal del método de los componentes parte de la llamada "ecuación compensadora" o "ecuación general de población". En ella, la población para un periodo determinado puede expresarse en función de una población anterior y los flujos de entrada y salida que se han producido entre esas dos fechas, es decir, una población crece o decrece en función de los efectos de cuatro determinantes básicos: nacimientos, defunciones, inmigraciones y emigraciones de la siguiente manera

$$P_{t+n} = P_t + N_{t,t+n} - D_{t,t+n} + I_{t,t+n} - E_{t,t+n}$$

donde P_t y P_{t+n} son la población en el año t y en el año $t+n$ respectivamente, $N_{t,t+n}$ son los nacimientos que se han producido entre los años t y $t+n$, $D_{t,t+n}$ son las defunciones que se han producido entre los años t y $t+n$, $I_{t,t+n}$ son los inmigrantes que han llegado entre los años t y $t+n$ y $E_{t,t+n}$ son los emigrantes que se han marchado entre los años t y $t+n$. Una forma equivalente de escribir la ecuación compensadora sería

$$P_{t+n} = P_t + SN_{t,t+n} + SM_{t,t+n}$$

donde $SN_{t,t+n}$ y $SM_{t,t+n}$ son el saldo natural o vegetativo y el saldo migratorio respectivamente. En el caso que nos ocupa, el periodo de estudio es un año por lo que $n = 1$. Podemos calcular el valor del saldo migratorio:

$$SM_{t,t+1} = I_{t,t+1} - E_{t,t+1} = 291,1 - 547,8 = -256,7.$$

El número de inmigrantes y emigrantes se calcula como la suma de las inmigraciones desde el extranjero y emigraciones hacia el extranjero para el total de la población respectivamente:

$$I_{t,t+1} = 18,0 + 13,8 + 16,0 + \cdots + 9,8 + 17,3 = 291,1$$
$$E_{t,t+1} = 25,0 + 21,6 + 20,4 + \cdots + 11,7 + 22,6 = 547,8$$

El saldo migratorio con el exterior es negativo lo cual quiere decir que el número de personas que se marchan del país supera al número de personas que llegan. A continuación, sustituyendo en la ecuación compensadora se obtiene el valor del saldo natural.

$$SN_{t,t+1} = P_{t+1} - P_T - SM_{t,t+1}$$

De forma análoga, la población total en el año $t+1$ y en el año t se calcula respectivamente como la suma de la población en cada uno de los grupos de edad.

$$P_{t+1} = 2320,4 + 2478,1 + 2267,6 + \cdots + 2492,8 + 8442,9 = 46507,8$$
$$P_t = 2422,8 + 2440,5 + 2226,7 + \cdots + 2502,3 + 8262,1 = 46727,9$$

Por lo que el saldo natural en este caso es positivo, es decir, durante el año estudiado nacen más personas de las que mueren. Sin tener en cuenta el saldo migratorio, este país aumentaría su población durante el periodo estudiado pues

$$SN_{t,t+1} = 46507,8 - 46727,9 - (-256,7) = 36,6.$$

Sin embargo, se observa que el saldo migratorio es mucho mayor que el saldo natural por lo que, teniendo en cuenta los dos saldos y la ecuación compensadora, se observa que el país está perdiendo población. En concreto, la diferencia entre la población al inicio y al final del año es

$$P_{t+1} - P_t = SN_{t,t+1} + SM_{t,t+1} = 36,6 + (-256,7) = -220,1$$

que equivalentemente, al disponer de los valores de la población total al inicio y al final de año, se podía haber calculado como

$$P_{t+1} - P_t = 46507,8 - 46727,9 = -220,1.$$

Los nacimientos en esta población son 427,2 mil personas (se obtiene este dato despejando del saldo natural), que supera a las defunciones, que son 390,6 mil personas fallecidas, por tanto que el saldo natural sea positivo es un buen síntoma de sostenibilidad demográfica, aunque se vea condicionado por la alta migración.

Por otra parte, cabe señalar la elevada intensidad migratoria que presenta el país, un 55,65 % de una cohorte ficticia que mostrase los mismos patrones que la población en estudio emigraría al exterior alguna vez en su vida. Por otra parte, el país presenta una edad media a la migración baja (26,75 años).

6.2. Soluciones del tercer examen del Cuerpo de Diplomados en Estadística del Estado de 2015

Cuestión 1

El tiempo de retraso, medido en minutos, del AVE Madrid-Sevilla sigue una variable aleatoria continua con función de distribución:

$$F(x) = \begin{cases} 0 & \text{si } x \leq -1 \\ \\ k(x+1) + \frac{x^2-1}{2} & \text{si } -1 < x \leq 0 \\ \\ k(x+1) - \frac{x^2+1}{2} & \text{si } 0 < x \leq 1 \\ \\ 1 & \text{si } x > 1 \end{cases}$$

1. Calcula el valor de k

2. Calcular la probabilidad de que el tren llegue antes de la hora prevista.

3. Calcular el tiempo esperado de retraso.

4. Calcular la probabilidad de que el tren llegue entre medio minuto de adelanto y un minuto de retraso.

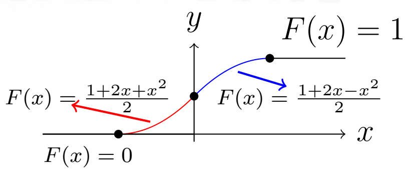
5. Sabiendo que el tren ha llegado con retraso, calcular la probabilidad de que lo haya hecho menos de 15 segundos después de lo previsto.

Solución.

Tiene que ocurrir que la función de distribución sea continua por la derecha, por lo que tenemos que analizar los posibles puntos de discontinuidad, es decir, el -1, 0 y el 1:

$$F(-1) = \lim_{\varepsilon \to 0^+} F(-1+\varepsilon) \Leftrightarrow 0 = k(-1+1) + \frac{1-1}{2} \Leftrightarrow 0 = 0$$
$$F(0) = \lim_{\varepsilon \to 0^+} F(0+\varepsilon) \Leftrightarrow k - \frac{1}{2} = k - \frac{1}{2} \Leftrightarrow 0 = 0$$
$$F(1) = \lim_{\varepsilon \to 0^+} F(1+\varepsilon) \Leftrightarrow 2k - \frac{1+1}{2} = 1 \Leftrightarrow k = 1.$$

Así $k = 1$. La función de distribución tendría esta forma:



La probabilidad de que el tren llegue antes de la hora prevista es

$$P(X \leq 0) = F(0) = 1 + \frac{0-1}{2} = \frac{1}{2}$$

Ya que la función de distribución es simétrica respecto al origen, tenemos que $E[X] = 0$. Podría verse empíricamente; se calcula en primer lugar la función de densidad:

$$f(x) = F'(x) = \begin{cases} 1+x & \text{si } -1 < x \leq 0 \\ 1-x & \text{si } 0 < x \leq 1 \\ 0 \text{ en el resto} & \end{cases}$$

De forma que:

$$E(X) = \int_{-1}^{0} x(1+x)dx + \int_{0}^{1} x(1-x)dx = \left[x^2 + \frac{x^3}{3}\right]_{-1}^{0} + \left[x^2 - \frac{x^3}{3}\right]_{0}^{1}$$
$$= -1 + \frac{1}{3} + 1 - \frac{1}{3} = 0$$

La probabilidad de que el tren llegue entre medio minuto de adelanto y un minuto de retraso será

$$P(-0,5 \leq X \leq 1) = F(1) - F(-0,5) = 1 - \left[0,5 + \frac{0,5^2 - 1}{2}\right] = 0{,}875$$

Sabiendo que el tren ha llegado con retraso, la probabilidad de que lo haya hecho menos de 15 segundos después de lo previsto. Teniendo en cuenta que 15 segundos equivalen a 0,25 minutos tenemos:

$$P(X < 0,25|X > 0) = \frac{P(0 < X < 0,25)}{P(X > 0)} = \frac{P(0 < X < 0,25)}{1 - P(X \leq 0)}$$
$$= \frac{F(0,25) - F(0)}{1 - F(0)} = 0,4375$$

Cuestión 2

El tiempo de vida de un tipo de bombillas se sabe que es una variable aleatoria que sigue una distribución exponencial de parámetro $\frac{1}{\Theta}$ con $\Theta > 0$.

Para estimar el tiempo medio de vida se extrae una muestra aleatoria simple de n bombillas y se consideran dos estimadores. Uno es la media muestral, $T_1(X_1,\ldots,X_n)=\frac{1}{n}\sum_{i=1}^{n}x_i$, y el otro se define como $T_2(X_1,\ldots,X_n)=$ $\text{mín}(X_1,\ldots,X_n)$. Se pide:

1. Calcular la media, el sesgo y la varianza del estimador T_1.

2. Calcular la media, el sesgo y la varianza del estimador T_2.

3. ¿Es T_2 insesgado? Si no lo es, define un estimador insesgado T_3 que sea función de él.

4. Entre los dos estimadores insesgados, T_1 y T_3 ¿cuál es más eficiente?

Solución.

Sea $X_i \sim Exp\left(\frac{1}{\Theta}\right)$, se tiene $E\left[X_i\right]=\Theta$ y $Var\left(X_i\right)=\Theta^2$ por lo que $E\left[(x_i)^2\right]=Var\left(x_i\right)+E\left[x_i\right]^2=2\Theta^2$.

$$E\left[T_1\right]=\frac{1}{n}\sum_{i=1}^{n}E\left[x_i\right]=\frac{1}{n}n\Theta=\Theta.$$

Así, se trata de un estimador insesgado y $b\left[T_1\right]=\text{sesgo}(T_1)=0$.

$$Var\left(T_1\right)=\frac{1}{n^2}\sum_{i=1}^{n}Var(x_i)=\frac{1}{n^2}n\Theta^2=\frac{\Theta^2}{n}.$$

Para el cálculo de la media, el sesgo y la varianza del estimador T_2 se obtiene en primer lugar su función de densidad.

$$X \sim exp\left(\frac{1}{\Theta}\right) \Rightarrow F\left(x\right)=1-e^{\frac{-1}{\Theta}x},\ \text{x>0} \Rightarrow f\left(x\right)=\frac{1}{\Theta}e^{-\frac{1}{\Theta}x},\ \text{x>0}$$

Por tanto, según las propiedades de los estadísticos ordenados, si denotamos por $X_{(1)}=min(X_1,\ldots,X_n)=T_2(X_1,\ldots,X_n)$, sabemos que por ser la variable aleatoria continua:

$$\begin{aligned}F_{X_{(1)}}\left(x\right)&=P(X_{(1)}\leq x)=1-P(X_{(1)}\geq x)=1-P(X_1\geq x,\ldots,X_n\geq x)=\\&=1-\prod_{i=1}^{n}P(X_i\geq x)=1-\prod_{i=1}^{n}(1-P(X_i<x))=1-(1-F_x(x))^n=\\&=1-(1-F(x))^n\end{aligned}$$

luego derivando obtenemos la función de densidad de T_2:

$$f_{T_2}(x) = n\,(1-F(x))^{n-1}\,f(x).$$

De forma que, operando con la información de la distribución exponencial queda,

$$f_{T_2}(x) = \begin{cases} n \cdot \frac{1}{\Theta} \cdot e^{-\frac{x}{\Theta}} \cdot \left(e^{-\frac{x}{\Theta}}\right)^{n-1} = \frac{n}{\Theta} e^{-\frac{n}{\Theta}x} & \text{si } x > 0 \\ 0 & \text{resto} \end{cases}$$

que coincide con una exponencial de parámetro $\frac{n}{\Theta}$.

Es decir $T_2(X_1, \ldots, X_n) \equiv Exp(\frac{n}{\Theta})$, luego

$$E\,[T_2] = \frac{\Theta}{n}$$

por lo que el estimador es sesgado y

$$b\,[T_2] = \frac{\Theta}{n} - \ \Theta = \Theta\left(\frac{1-n}{n}\right).$$

Además

$$Var\,(T_2) = \frac{\Theta^2}{n^2}.$$

Por tanto T_2 no es insesgado pero considerando $T_3 = nT_2 = n \cdot X_{(1)}$, se tiene un estimador insesgado ya que

$$E\,[T_3] = nE\,[T_2] = n\frac{\Theta}{n} = \Theta.$$

Para comparar la eficiencia entre T_1 y T_3 se comparan las varianzas. Siendo

$$Var\,(T_3) = n^2 Var\,(T_2) = n^2\frac{\Theta^2}{n^2} = \Theta^2$$

tenemos que $Var\,(T_1) = \frac{\Theta^2}{n} < Var\,(T_3) = \Theta^2$ y T_1 es más eficiente que T_3, siempre que $n > 1$, puesto que si es 1 son igual de eficientes.

Cuestión 3

Un fabricante de piezas de plomo tiene un stock de cien piezas clasificadas según su destino en 50 para Europa y 50 para el resto del mundo. Se desea estimar el peso total de las piezas y la única información disponible es relativa al peso (en toneladas) del envío en el año anterior tal y como viene en la tabla 6.2. Ignorando el factor de corrección para poblaciones finitas, se pide:

1. El fabricante decide seleccionar 10 piezas con probabilidades iguales y sin reemplazamiento ¿cuál es la varianza del estimador insesgado del peso total?
2. Un empleado de la fabrica le recomienda obtener estimaciones separadas según el destino de las piezas y deciden realizar un muestreo estratificado por destino y afijación proporcional ¿cual es la varianza del nuevo estimador insesgado para el peso total?
3. Si el muestreo estratificado se llevase a cabo con afijación de mínima varianza ¿cuantas piezas se seleccionan según su destino?

Tabla 6.2: Información relativa al peso (en toneladas) del envío en el año anterior

Destino	Media	Cuasivarianza
Europa	6	4
Resto	4	9

Solución.

El estimador insesgado del total en muestreo sin reposición y con probabilidades iguales es:

$$\hat{X} = \sum_{i=1}^{n} \frac{X_i}{\pi_i} = \sum_{i=1}^{n} \frac{X_i}{\frac{n}{N}} = \sum_{i=1}^{n} \frac{N}{n} X_i = N \cdot \overline{x}$$

La varianza del dicho estimador, ignorando el factor de corrección para poblaciones finitas, viene dada por

$$Var(\widehat{X}) = Var(N\widehat{\overline{X}}) = N^2 Var(\overline{x}) = N^2 \frac{S^2}{n}.$$

Puesto que tenemos que $N = 100$ y $n = 10$ tenemos que calcular el valor de S^2 para resolver la primera parte del problema. Pero sabemos el total de piezas con destino Europa y resto del mundo, por lo que $N_1 = 50$ y $N_2 = 50$.

Podemos calcular la media poblacional como una media ponderada de Europa y el resto del mundo:

$$\overline{X} = \frac{N_1}{N}\overline{X}_1 + \frac{N_2}{N}\overline{X}_2 = \frac{1}{2}6 + \frac{1}{2}4 = 5$$

y por tanto

$$S^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left(X_i - \overline{X}\right)^2 = \frac{1}{99}\sum_{i=1}^{100}\left[X_i^2 - 100\overline{X}^2\right] = \frac{1}{99}\left[\sum_{i=1}^{100}X_i^2\right] - \frac{100}{99}\overline{X}^2$$

así que hay que calcular esa suma de cuadrados para obtener S^2, pero como tenemos las cuasivarianzas de cada grupo, en cada uno, aunque es una propiedad general, se cumple que:

$$(N_i - 1)S_i^2 = \sum_{i=1}^{N_i} X_i^2 - N_i\overline{X}_i^2$$

siendo entonces

$$\sum_{i=1}^{100} X_i^2 = \sum_{i=1}^{50} X_i^2 + \sum_{i=51}^{100} X_i^2 = \left[(N_1 - 1)S_1^2 + N_1\overline{X}_1^2\right] + \left[(N_2 - 1)S_2^2 + N_2\overline{X}_2^2\right] =$$

$$= \left[49 \cdot 4 + 50 \cdot 6^2\right] + \left[49 \cdot 9 + 50 \cdot 4^2\right] = 196 + 1800 + 441 + 800 = 3237$$

y finalmente $S^2 = \frac{3237}{99} - \frac{100}{99} \cdot 25 = 32,69696 - 25,252525 = 7,44443$

Luego $Var(\widehat{X}) = N^2\frac{S^2}{n} = 100^2 \cdot \frac{7,44443}{10} = 7444,43$

En el caso de muestreo estratificado, utilizando la afijación proporcional $n_h = kN_h \Leftrightarrow k = \frac{n_h}{N_h} = \frac{n}{N} = \frac{1}{10} \Leftrightarrow n_h = 5$. Por tanto, si suponemos que $f_1 = f_2 = 0$,

$$Var\left(\widehat{X}_{st}\right) = \sum_{i=1}^{2} N_h^2\left(1 - f_h\right)\frac{S_h^2}{n_h} \simeq \sum_{i=1}^{2} N_h^2\frac{S_h^2}{n_h} = 50^2\frac{4}{5} + 50^2\frac{9}{5} = 6500.$$

En la afijación de mínima varianza,

$$n_h = n\frac{N_hS_h}{\sum_{i=1}^{n} N_hS_h} \text{ para h=1,2}$$

Por tanto,

$$n_1 = 10\frac{50\sqrt{4}}{50\sqrt{4}+50\sqrt{9}} = 4 \qquad n_2 = 10\frac{50\sqrt{9}}{50\sqrt{4}+50\sqrt{9}} = 6.$$

De modo que

$$Var\left(\widehat{X}_{st}\right) = 50^2\frac{4}{4} + 50^2\frac{9}{6} = 6250.$$

Cuestión 4

Dada la tabla 6.3 de medidas de la cintura (en pulgadas) y del peso (en libras) de 10 individuos, se pide:

1. Calcular todos los momentos respecto al origen y a la media de ordenes 1 y 2 de la distribución bidimensional.

2. Calcular el coeficiente de correlación lineal de Pearson de ambas variables y comentar su significado.

3. Calcular los cuantiles c_p^{Peso} y $c_p^{Cintura}$ de órdenes $p_i = 0, \frac{1}{3}, \frac{2}{3}, 1$ (i = 0, ..., 3) de sendas distribuciones unidimensionales.

4. Sean los intervalos $I_i^{Peso} = \left[c_{p_i-1}^{Peso}, c_p^{Peso}\right)$ e $I_j^{Cintura} = \left[c_{p_j-1}^{Cintura}, c_{p_j}^{Cintura}\right)$ con $i,j = 1,2,3$. Construir la tabla de contingencia del numero de individuos con medidas de peso y cintura en cada par de intervalos $\left(I_i^{Peso}, I_j^{Cintura}\right)$.
NB: $I_3^{Peso} = \left[c_2^{Peso}, c_3^{Peso}\right]$ e $I_3^{Cintura} = \left[c_2^{Cintura}, c_3^{Cintura}\right]$.

5. Calcular la distribución de frecuencias relativas del peso condicionada respecto a la cintura a partir de la tabla de contingencia anterior.

Tabla 6.3: Medidas de la cintura (en pulgadas) y del peso (en libras)

Cintura	32	36	38	33	39	40	41	35	38	38
Peso	175	181	200	159	196	192	205	173	187	188

Solución.

A partir de la tabla 6.4 de frecuencias absolutas, se calculan los momentos respecto al origen y a la media de órdenes 1 y 2 de la distribución.

$$a_{10} = \frac{\sum_{i=1}^{8} x_i n_{i\bullet}}{N} = \frac{32 \cdot 1 + \cdots + 38 \cdot 3 + \cdots + 41 \cdot 1}{10} = 37$$

$$a_{01} = \frac{\sum_{j=1}^{10} y_j n_{\bullet j}}{N} = \frac{175 \cdot 1 + \cdots + 188 \cdot 1}{10} = 185,6$$

$$a_{20} = \frac{\sum_{i=1}^{8} x_i^2 n_{i\bullet}}{N} = 1376,8$$

$$a_{02} = \frac{\sum_{j=1}^{10} y_j^2 n_{\bullet j}}{N} = 34621,4$$

$$a_{11} = \sum_{i=1}^{10} \sum_{j=1}^{10} x_i y_j f_{ij} = \frac{1}{10} (32 \cdot 175 + 36 \cdot 181 + \cdots + 41 \cdot 205) = 6899,7$$

$$m_{10} = 0$$

$$m_{01} = 0$$

$$m_{20} = a_{20} - a_{10}^2 = 7,8$$

$$m_{02} = a_{02} - a_{01}^2 = 174,04$$

$$m_{11} = a_{11} - a_{10} a_{01} = 32,5$$

El coeficiente de correlación lineal de Pearson, que mide el grado de asociación lineal, será:

$$\rho = \frac{S_{XY}}{S_X S_Y} = \frac{m_{11}}{\sqrt{m_{20}}\sqrt{m_{02}}} = 0,88.$$

Su rango de variación es [0, 1] siendo 0 la ausencia de relación lineal y 1 la relación lineal perfecta. Cintura y Peso están relacionadas linealmente de forma directa, pues $\rho = 0,88$.

Tabla 6.4: Frecuencias absolutas

	175	181	200	159	196	192	205	173	187	188	$n_{i\bullet}$
32	1										1
33				1							1
35								1			1
36		1									1
38			1						1	1	3
39					1						1
40						1					1
41							1				1
$n_{\bullet j}$	1	1	1	1	1	1	1	1	1	1	10

Comenzando por el calculo de los cuantiles c_p^{Peso}, ordenando de menor a mayor los datos correspondientes a las observaciones de pesos de los 10 individuos, es decir,

$$\underset{x_{(1)}}{159}, \underset{x_{(2)}}{173}, \underset{x_{(3)}}{175}, \underset{x_{(4)}}{181}, \underset{x_{(5)}}{187}, \underset{x_{(6)}}{188}, \underset{x_{(7)}}{192}, \underset{x_{(8)}}{196}, \underset{x_{(9)}}{200}, \underset{x_{(10)}}{205}$$

teniendo en cuenta que la posición del cuantil de orden p es:
$x_p^X = 1 + p(N-1)$, se tiene:

$$x_0^{Peso} = 1 \Rightarrow c_0^{Peso} = 159$$
$$x_1^{Peso} = 1 + \frac{1}{3}(N-1) = 1 + 9/3 = 4 \Rightarrow c_1^{Peso} = 181$$
$$x_2^{Peso} = 1 + \frac{2}{3}(N-1) = 7 \Rightarrow c_2^{Peso} = 192$$
$$x_3^{Peso} = 1 + \frac{3}{3}(N-1) = 1 + 9 = 10 \Rightarrow c_3^{Peso} = 205$$

Siguiendo el mismo razonamiento para la variable Cintura sera:

$$x_0^{Cintura} = 1 \Rightarrow c_0^{Cintura} = 32$$

$$x_1^{Cintura} = 1 + \frac{1}{3}(N-1) = 1 + 9/3 = 4 \Rightarrow c_1^{Cintura} = 36$$
$$x_2^{Cintura} = 1 + \frac{2}{3}(N-1) = 7 \Rightarrow c_2^{Cintura} = 38$$
$$x_3^{Cintura} = 1 + \frac{3}{3}(N-1) = 1 + 9 = 10 \Rightarrow c_3^{Cintura} = 41$$

Por tanto, los intervalos pedidos serán

$$I_1^{Peso} = [159, 181)$$
$$I_2^{Peso} = [181, 192)$$
$$I_3^{Peso} = [192, 205]$$
$$I_1^{Cintura} = [32, 36)$$
$$I_2^{Cintura} = [36, 38)$$
$$I_3^{Cintura} = [38, 41]$$

Así, la tabla de contingencia será:

Tabla 6.5: Tabla de contingencia con los intervalos de Peso y Cintura

	$I_1^{Peso} = [159, 181)$	$I_2^{Peso} = [181, 192)$	$I_3^{Peso} = [192, 205]$
$I_1^{Cintura} = [32, 36)$	3	0	0
$I_2^{Cintura} = [36, 38)$	0	1	0
$I_3^{Cintura} = [38, 41]$	0	2	4

Tabla 6.6: Distribución de frecuencias relativas del Peso condicionada a $I_1^{Cintura}$

	$n_{i\|j}$	$f_{i\|j}$
$I_1^{Peso}\|I_1^{Cintura}$	3	1
$I_2^{Peso}\|I_1^{Cintura}$	0	0
$I_3^{Peso}\|I_1^{Cintura}$	0	0

Tabla 6.7: Distribución de frecuencias relativas del Peso condicionada a $I_2^{Cintura}$

	$n_{i\|j}$	$f_{i\|j}$
$I_1^{Peso}\|I_2^{Cintura}$	0	0
$I_2^{Peso}\|I_2^{Cintura}$	1	1
$I_3^{Peso}\|I_2^{Cintura}$	0	0

Tabla 6.8: Distribución de frecuencias relativas del Peso condicionada a $I_3^{Cintura}$

	$n_{i\|j}$	$f_{i\|j}$
$I_1^{Peso}\|I_3^{Cintura}$	0	0
$I_2^{Peso}\|I_3^{Cintura}$	2	$\frac{2}{6}$
$I_3^{Peso}\|I_3^{Cintura}$	4	$\frac{4}{6}$

Cuestión 5

A partir de la información relativa al número total de asuntos judiciales resueltos en España (en miles), así como de las cantidades reconocidas a los trabajadores (en millones de euros) durante los años 2003, 2004 y 2005, que se muestra en la tabla 6.9, se pide calcular:

1. Los índices simples (base 2003 y en porcentaje) de las cantidades reconocidas a los trabajadores para las distintas materias objeto de la demanda.
2. La serie de números índice (base 2003 y en porcentaje) media aritmética ponderada de la cuantía de cantidades reconocidas a los trabajadores, utilizando como ponderaciones fijas el numero de asuntos judiciales en 2003.
3. La repercusión de los conflictos colectivos en la variación del Índice de las cantidades reconocidas a los trabajadores afectados por asuntos judiciales resueltos en España, entre 2004 y 2005.
4. La participación porcentual de los conflictos colectivos en la variación del

índice de las cantidades reconocidas a los trabajadores afectados por asuntos judiciales resueltos en España entre 2004 y 2005.

5. La suma de las participaciones en la variación del índice general.

Tabla 6.9: Número total de asuntos judiciales y cantidades reconocidas a los trabajadores en España.

Materia objeto de la demanda	Año 2003		Año 2004		Año 2005	
	Cuantía	Número	Cuantía	Número	Cuantía	Número
Conflictos colectivos	43	1,6	94	2,1	1347	2
Despido	263	64	223	62	230	62
Reclamaciones por contrato de trabajo	249	147	200	139,4	213	127

Solución.

Los índices simples de cantidad vienen dados por la expresión:

$$I_0^t = \frac{q_{it}}{q_{i0}} \cdot 100.$$

Así, se tienen los índices recogidos en la tabla 6.10.

Tabla 6.10: Índices simples (base 2003) de las cantidades

	Año 2003	Año 2004	Año 2005
Conflictos colectivos	$I_{03}^{03} = \frac{43}{43}100 = 100$	$I_{03}^{04} = \frac{94}{43}100 = 218,6$	$I_{03}^{05} = \frac{1347}{43}100 = 3132,55$
Despido	$I_{03}^{03} = \frac{263}{263}100 = 100$	$I_{03}^{04} = \frac{223}{263}100 = 84,79$	$I_{03}^{05} = \frac{230}{263}100 = 87,45$
Reclamaciones por contrato de trabajo	$I_{03}^{03} = \frac{249}{249}100 = 100$	$I_{03}^{04} = \frac{200}{249}100 = 80,32$	$I_{03}^{05} = \frac{213}{249}100 = 85,54$

La serie de números índice (base 2003 y en porcentaje) media aritmética ponderada (map) de la cuantía de cantidades reconocidas a los trabajadores, utilizando como ponderaciones fijas el numero de asuntos judiciales en 2003, será:

Año 2003:

$$I(map)_{03}^{03} = \frac{\sum_{i=1}^{n} w_{it} I_0^t}{\sum_{i=1}^{n} w_{it}} = \frac{1,6 \cdot 100 + 64 \cdot 100 + 147 \cdot 100}{1,6 + 64 + 147} = 100$$

Año 2004:

$$I(map)_{03}^{04} = \frac{\sum_{i=1}^{n} w_{it} I_0^t}{\sum_{i=1}^{n} w_{it}} = \frac{1,6 \cdot 218,6 + 64 \cdot 84,79 + 147 \cdot 80,32}{212,6} = 82,71$$

Año 2005:

$$I(map)_{03}^{05} = \frac{\sum_{i=1}^{n} w_{it} I_0^t}{\sum_{i=1}^{n} w_{it}} = \frac{1,6 \cdot 3132,55 + 64 \cdot 87,45 + 147 \cdot 85,54}{212,6} = 109,046$$

La repercusión de los conflictos colectivos en la variación

$$R_{conflictos} = \frac{\Delta I_{conflictos} w_{conflcitos}}{\sum_{i=1}^{n} w_{it}} = \frac{(3132,55 - 218,6)1,6}{212,6} = 21,93$$

Es decir, el hecho de que la cuantía de las cantidades reconocidas por conflictos colectivos haya aumentado de 2004 a 2005 supone que si el resto de los precios se mantuvieran constantes, la repercusión que tendría este efecto es que se incrementaría el índice de las cantidades reconocidas a los trabajadores por asuntos judiciales en 21,93 puntos. El índice entre 2004 y 2005 aumenta de 82,71 a 109,046, es decir, tiene un incremento de 26,336 puntos, de los cuales 21,93 se deben a la repercusión de los conflictos colectivos.

La participación porcentual de los conflictos colectivos en la variación del índice de las cantidades reconocidas a los trabajadores afectados por asuntos judiciales resueltos en España entre 2004 y 2005 es:

$$P_{conflictos} = \frac{R_{conflictos}}{\sum_{i=1}^{n} R_i} 100 = \frac{21,93}{21,93 + 0,8 + 3,61} 100 = 83,25\,\%$$

puesto que:

$$R_{despido} = \frac{\Delta I_{conflictos} w_{conflcitos}}{\sum_{i=1}^{n} w_{it}} = \frac{(87,45 - 84,79)64}{212,6} = 0,8$$

$$R_{reclam} = \frac{\Delta I_{conflictos} w_{conflcitos}}{\sum_{i=1}^{n} w_{it}} = \frac{(85,54 - 80,32)147}{212,6} = 3,61$$

La suma de las participaciones en la variación del índice general.

$$\sum_{i=1}^{n} P_i = P_{conflictos} + P_{despido} + P_{reclamaciones} =$$

$$= \left(\frac{R_{despido}}{\sum_{i=1}^{3} R_i} + \frac{R_{conflictos}}{\sum_{i=1}^{3} R_i} + \frac{R_{reclamaciones}}{\sum_{i=1}^{3} R_i} \right) 100 = \left(\frac{\sum_{i=1}^{3} R_i}{\sum_{i=1}^{3} R_i} \right) 100 = 100\,\%$$

Cuestión 6

Utilizando los datos de la tabla 6.11 relativos a una economía imaginaria, calcule en términos del SEC 2010:

1. El gasto en consumo final.
2. La formación bruta de capital fijo.
3. El saldo neto exterior de bienes y servicios.
4. El deflactor implícito del PIB para 2014 en base 2010 = 100.
5. El valor añadido bruto a precios de mercado (o precios de adquisición).
6. Los impuestos netos sobre productos e importaciones.
7. Las subvenciones a los productos.
8. La remuneración de los factores de producción.
9. La remuneración de los asalariados.

Tabla 6.11: Datos relativos a una economía imaginaria.

Consumo individual efectivo	1.200
Formación neta de capital	155
Exportaciones de bienes y servicios	400
Rentas de la propiedad y de la empresa	700
VAB a precios básicos	2.100
Consumo colectivo efectivo	600
PIBpm a precios constantes de 2010	2.000
Consumo de capital fijo	20
Otros impuestos netos sobre la producción [D.29 - D.39]	220
Variación de existencias	5
Impuestos sobre productos e importaciones [D.21]	175
Adquisiciones menos cesiones de objetos valiosos	5
Gasto del Estado en armamento	25
Importaciones de bienes y servicios	150

Datos a precios corrientes de 2014 (salvo indicación contraria).

Solución.

1. GCF = CFE individual + CFE colectivo = 1.800

2. FBC = FBCF + "Var. Existencias" + "Adq menos ces objs valiosos"

   $\Downarrow$

   FBCF = FNC + CCF - "Var. Existencias" - "Adq menos ces objs valiosos"

   = 155 + 20 - 5 - 5 =165

3. Saldo neto exterior = Exportaciones de bienes y servicios - Importaciones de bienes y servicios= 400 -150= 250

4. $\text{PIB}^{2014}_{precCorrientes}$ = GCF + FBC + X - I = 1.800 + 175 + 250 = 2.225
   $\text{PIB}^{2010}_{precCorrientes}$ = 2.000 $\Rightarrow$
   $\text{DPIB}^{2014}_{2010}=\frac{PIB^{2014}_{precCorrientes}}{PIB^{2010}_{precConstantes}}100=\frac{2225}{2000}100=1{,}1125\cdot 100=111{,}25$

5. VAB(pm) = PIB(pm) para el total de la economía, por tanto
   VAB(pm) = 2.225

6. PIB(pm) = PIB (pb) + impuestos sobre los productos e importaciones - subvenciones a los productos e importaciones $\Rightarrow$
   impuestos sobre los productos e importaciones - subvenciones a los productos e importaciones = 2.225 - 2.100 = 125, puesto que para el total de la economía $PIB(pb) = VAB(pb) = 2100$.

7. D.21 - D.31 = 125, por tanto, siendo D.21 = 175 se tiene D.31 = 50

8. Se tiene que: Remuneración de factores de producción = RA + EBE/RM; por otra parte: PIBpm - impuestos netos sobre la producción e importaciones = RA + EBE/RM; luego:
   Remuneración de factores de producción = 2225 - 125 (que es D.21-D.31) - 220 (que esD.29-D.39)= 1880

9. Rentas de la propiedad y la empresa = EBE + RM, luego EBE + RM = 700 (puesto que es un dato del problema), por tanto: RA = 1880 - 700 = 1180

Cuestión 7

Con los datos de la tabla 6.12 en millones de euros de la CNE-Base 2010 para la economía española en 2010, calcule:

1. El Valor Añadido Bruto a coste de los factores (VAB cf).
2. El Valor Añadido Bruto a precios de mercado (VAB pm).

3. El Valor Añadido Bruto a precios básicos (VAB pb).

4. El Producto Interior Bruto a precios de mercado (PIB pm).

Tabla 6.12: Datos, en millones de euros, de la CNE-Base 2010 para la economía española en 2010

Producción Total de bienes y servicios a precios básicos	2.038.315
Consumos intermedios	1.048.042
IVA que grava los productos	58.458
Otros impuestos ligados a la producción	14.654
Impuestos sobre los productos, excluidos el IVA y los impuestos sobre las importaciones	37.199
Otras subvenciones a la producción	12.095
Subvenciones a los productos	6.297
Impuestos y derechos sobre las importaciones (excluido el IVA)	1.640

Solución.

1. VAB(pb) = producción (pb) - CI (padq) =
   =2.038.315 - 1.048.042 = 990.273 millones de €

2. PIB(pm) = VAB (pb) + impuestos sobre los productos - subvenciones a los productos = VAB (pb) + IVA + impuestos sobre las importaciones (excl. IVA) + otros impuestos sobre los productos (excl. IVA e impuestos sobre importaciones) - subvenciones a los productos = 990.273 + 58.458 + 1.640 + 37.199 - 6.297 = 1.081.273 millones de €
   Para el total de la economía, el VAB(pm) = PIB(pm)

3. VAB(cf) = VAB(pb) - otros impuestos (menos las subvenciones) sobre la producción = 990.273 - (14.654 - 12.095) = 987.714 millones de €

Cuestión 8

En una economía con tres sectores de actividad y en ausencia de impuestos, subvenciones e importaciones, construya la tabla input-output simétrica con los siguientes datos:

1. La utilización de los consumos intermedios por parte de los sectores productivos se realiza de la siguiente manera:

- El sector1 consume 10€ procedente del sector2; 4€ del sector3 y 12€ del sector1.
- El sector2 consume 25€ procedente del sector1; 12€ del sector2 y 8€ del sector3.
- El sector3 consume 17€ procedente del sector 1; 20€ del sector2 y 15€ del sector3.

2. El gasto en consumo final ha sido de 53€ en productos del sector1; 53€ del sector2 y 41€ del sector3.

3. La formación bruta de capital ha sido 28€ de productos del sector1; 30€ del sector2 y 18€ del sector3.

4. Las exportaciones han sido 38€ de productos de sector1; 15€ del sector2 y 5€ del sector3.

5. El factor capital utilizado por cada sector ha sido: 60, 38 y 23, respectivamente.

Interprete el significado económico de cada una de las cifras que encontramos en la columna correspondiente al sector2 y en la fila relativa al sector3.

Solución.

Tabla 6.13: Tabla input-output simétrica completa

	Sector 1	Sector 2	Sector 3	Total demanda intermedia	GCF	FBC	Exp	Total empleos
Sector 1	12	25	17	54	53	28	38	173
Sector 2	10	12	20	42	53	30	15	140
Sector 3	4	8	15	27	41	18	5	91
Total (pb)	26	45	52	123	147	76	58	404
CI (p.adq)	26	45	52	123				
VAB (pb)	147	95	39	281				
RA	87	57	16	160				
EBE/RM	60	38	23	121				
Producción (pb)	173	140	91	404				
Oferta (pb)	173	140	91	404				

Significado económico de las cifras en la columna del sector2:

En las columnas lo que se registra es la utilización, consumo, o uso productivo, de productos intermedios o de inputs primarios para obtener la producción de la rama de que se trate.

- 25: el sector 2, en su proceso de producción, se ha gastado o ha consumido 25€ de la producción del sector 1.
- 12: el sector 2, en su proceso de producción, se ha gastado 12€ de productos de su propia producción.

- 8: el sector 2 ha utilizado productos del sector 3 por valor de 8€ para obtener su producción

- 45: cantidad total de los consumos intermedios realizados por este sector 2 en su proceso productivo. Al no haber impuestos este total a precios básicos coincide con el valor a precios de adquisición.

- 95: el valor añadido bruto a precios básicos del sector 2 es 95€

- 57: en el sector 2, se pagan salarios por valor de 57€

- 38: el excedente bruto de explotación es de 38€

- 140: el valor de la producción del sector 2 a precios básicos es la suma de los productos intermedios y los inputs utilizados (CI + VAB), es decir $45 + 95 = 140$

- 140: oferta total coincide con producción porque no hay importaciones

Significado económico de las cifras en la fila del sector3:
De igual modo, por filas no se computan las ventas, sino los destinos intermedios o finales de los productos de una rama.

- 4: el sector 3 ha destinado productos por valor de 4€ para consumos intermedios del sector 1. Es decir, de la producción de este sector 3, el sector 1 demanda productos por valor de 4€ para consumirlos en su proceso de producción.

- 8: de la producción del sector 3, el sector 2 demanda productos por valor de 8€ para consumirlos en su proceso de producción.

- 15: el sector 3 ha destinado productos por valor de 15€ para consumos intermedios del propio sector 3

- 27: en total, la demanda intermedia de productos producidos por el sector 3 que realizan los tres sectores son 27€.
- 41: el sector 3 ha tenido 41€ de gasto en consumo final
- 18: el sector 3 ha registrado un valor 18€ de formación bruta de capital
- 5: el sector 3 ha destinado productos por valor de 5€ para exportaciones
- 91: en total, los empleos totales del sector 3 son de 91€

Cuestión 9

Se tienen los datos extraídos de las tablas de nacimientos y de la población española recogidos en las tablas 6.14 y 6.15. Se pide:

1. Representar en el esquema de Lexis los datos de la tabla ”Nacimientos. Año 2013”.
2. Calcular la tasa de fecundidad específica para la generación de 1989.
3. Calcular la tasa de fecundidad específica para las mujeres de 24 años.

Tabla 6.14: Nacimientos. Año 2013.

Edad y generación de la madre	Total nacimientos
22 años: 1991	3.066
22 años: 1990	3.219
23 años: 1990	3.656
23 años: 1989	3.799
24 años: 1989	4.259
24 años: 1988	4.463
25 años: 1988	5.184

Tabla 6.15: Población residente en España

	1 de enero de 2013			1 de enero de 2014		
Edad	Ambos sexos	Hombres	Mujeres	Ambos sexos	Hombres	Mujeres
22 años	483.688 2	45.742	237.947	471.980	240.541	231.439
23 años	500.067	253.050	247.016	482.902	244.467	238.435
24 años	516.666	260.090	256.576	497.608	250.963	246.645
25 años	532.641	267.563	265.078	512.336	257.292	255.043
26 años	552.538	276.701	275.837	526.425	263.779	262.647

Solución.

Las tasas se calculan fácilmente desde la representación del diagrama de Lexis.

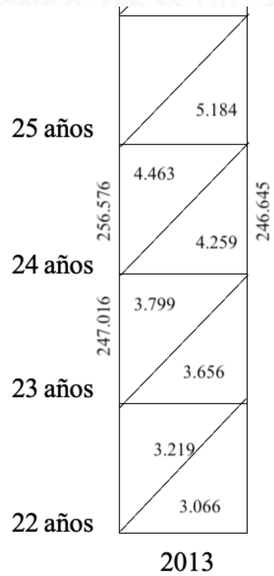


Figura 6.2: Representación del esquema de Lexis de los nacimientos del año 2013.

$$tef_{gen89}^{2013} = \frac{nacimientos_{gen89}^{2013}}{población\ media\ mujeres_{gen89}^{2013}} \cdot 1000 = \frac{3799 + 4259}{\frac{247{,}016 + 246{,}645}{2}} \cdot 1000 = 32,64‰$$

$$tef_{24}^{2013} = \frac{nacimientos_{24años}^{2013}}{población\ media\ mujeres_{24años}^{2013}} \cdot 1000 = \frac{4259 + 4463}{\frac{256{,}576 + 246{,}645}{2}} \cdot 1000 = 34,66‰$$

Cuestión 10

A partir de la información incluida en la figura 6.3 referida al total de España y la Comunidad Autónoma de Andalucía para 2013, calcular:

1. La tasa de mortalidad infantil.
2. La tasa de mortalidad infantil neonatal.
3. La tasa de mortalidad infantil neonatal temprana.
4. La tasa de mortalidad infantil neonatal tardía.
5. La tasa de mortalidad infantil post-neonatal.

	Defunciones en 2013												
	Todos los menores de 1 año	Menos de 24 horas	De 1 día	De 2 días	De 3 días	De 4 días	De 5 días	De 6 días	De 7 a 13 días	De 14 a 20 días	De 21 a 27 días	Nacimientos 2013	Población media 0 años en 2013
España	1.164	293	65	62	42	29	43	23	140	68	41	425.715	437.015
Andalucía	256	65	13	11	5	13	13	6	31	11	11	81.470	82.012

Figura 6.3: Información referida al total de España y a Andalucía.

Solución.

ESPAÑA

1. $TMI^{2013} = \frac{D_0^{2013}}{NV^{2013}} \cdot 1000 = \frac{defunciones_{<1 año}}{nacimientos} \cdot 1000 = \frac{1164}{425715} \cdot 1000 = 2,734‰$

2. $TMIN^{2013} = \frac{D_{<28\ días}^{2013}}{NV^{2013}} \cdot 1000 = \frac{defunciones_{<28 días}}{nacimientos} \cdot 1000 =$
$= \frac{293+65+62+\cdots+68+41}{425715} \cdot 1000 = 1,893‰$

3. $TMINtemprana^{2013} = \frac{D^{2013}_{<7\text{ días}}}{NV^{2013}} \cdot 1000 = \frac{defunciones_{<7días}}{nacimientos} \cdot 1000 =$
$= \frac{293+65+62+42+29+43+23}{425715} \cdot 1000 = 1,308‰$

4. $TMINtardía^{2013} = \frac{D^{2013}_{7\text{ días-27 días}}}{NV^{2013}} \cdot 1000 = \frac{defunciones_{7a28días}}{nacimientos} \cdot 1000 =$
$= \frac{140+68+41}{425715} \cdot 1000 = 0,585‰$

5. $TMIPN^{2013} = \frac{D^{2013}_{28\text{ días-364 días}}}{NV^{2013}} \cdot 1000 = \frac{defunciones_{28a364días}}{nacimientos} \cdot 1000 =$
$= \frac{1164-(293+65+62+\cdots+68+41)}{425715} \cdot 1000 = 0,841‰$

ANDALUCÍA

1. $TMI^{2013} = \frac{D^{2013}_{0}}{NV^{2013}} \cdot 1000 = \frac{defunciones_{<1año}}{nacimientos} \cdot 1000 = \frac{256}{81470} \cdot 1000 = 3,142‰$

2. $TMIN^{2013} = \frac{D^{2013}_{<28\text{ días}}}{NV^{2013}} \cdot 1000 = \frac{defunciones_{<28días}}{nacimientos} \cdot 1000 =$
$= \frac{65+13+11+\cdots+11+11}{81470} \cdot 1000 = 2,197‰$

3. $TMINtemprana^{2013} = \frac{D^{2013}_{<7\text{ días}}}{NV^{2013}} \cdot 1000 = \frac{defunciones_{<7días}}{nacimientos} \cdot 1000 =$
$= \frac{65+13+11+5+13+13+6}{81470} \cdot 1000 = 1,546‰$

4. $TMINtardía^{2013} = \frac{D^{2013}_{7\text{ días-27 días}}}{NV^{2013}} \cdot 1000 = \frac{defunciones_{7a28días}}{nacimientos} \cdot 1000 =$
$= \frac{31+11+11}{81470} \cdot 1000 = 0,651‰$

5. $TMIPN^{2013} = \frac{D^{2013}_{28\text{ días-364 días}}}{NV^{2013}} \cdot 1000 = \frac{defunciones_{28a364días}}{nacimientos} \cdot 1000 =$
$= \frac{256-(65+13+11+5+\cdots+11+11)}{81470} \cdot 1000 = 0,945‰$

Capítulo 7

Año 2014

"You cannot train alone and expect to run a fast time. There is a formula: 100 % of me is nothing compared to 1 % of the whole team."

—Eliud Kipchoge

Eliud Kipchoge nació en Kapsisiywa, Kenia, el 5 de noviembre de 1984 y es el reciente ganador de la medalla de oro en la maratón de los Juegos Olímpicos de Rio 2016. Eliud es fiel a sí mismo y por eso tiene una nota en su pared del autor brasileño Paulo Coelho, lo cual lo identifica: "Si usted quiere tener éxito, debe respetar una regla: Nunca se mienta a sí mismo".

$2014 = 12 + \cdots + 64$

2014 is an interprime number because it is at equal distance from previous prime (2011) and next prime (2017)

7.1. Soluciones del segundo examen del Cuerpo Superior de Estadísticos del Estado de 2014

Cuestión 1

Sea $X_1, \ldots, X_n$ una muestra aleatoria simple de la distribución

$$f_\theta(x) = \theta e^{-\theta x}\ ,\ x > 0$$

a) Calcular el estimador de máxima verosimilitud para $P(X > t)$.

b) Supongamos que se miden 10 componentes cuya duración sigue la distribución anterior y que si se suma el tiempo de funcionamiento de todas ellas se contabiliza un total de 1200 horas. A partir de esta muestra, estimar la probabilidad de que una componente vaya a durar menos de 130 horas.

c) Probar si el estadístico $T(X_1, \ldots, X_n) = \frac{n-1}{\sum_{i=1}^n X_i}$ es centrado para θ.

d) Probar si $\sum_{i=1}^n X_i$ es estadístico suficiente y completo. ¿Es $T(X_1, \ldots, X_n)$ estimador centrado uniformemente de mínima varianza?

Solución.

a)

$$P\left(X > t\right) = 1 - P\left(X \leq t\right) = 1 - F(t) = 1 - \left(1 - e^{-\theta t}\right) = e^{-\theta t}$$

Es decir, se busca el estimador de máxima verosimilitud de $g\left(\theta\right) = e^{-\theta t}$. La función de verosimilitud de la muestra es

$$f_\theta(x_1, \ldots, x_n) = \theta^n e^{-\theta \sum_{i=1}^n x_i} \qquad x_1, \ldots, x_n > 0$$

con lo cual

$$log f_\theta(x_1, \ldots, x_n) = n log\theta - \theta \sum_{i=1}^{n} x_i$$

$$\frac{\partial log f_\theta(x_1, \ldots, x_n)}{\partial\theta} = \frac{n}{\theta} - \sum_{i=1}^{n} x_i,$$

donde igualando a cero esta derivada y despejando θ queda:

$$\widehat{\theta} = \frac{n}{\sum_{i=1}^{n} x_i} = \frac{1}{\overline{x}}$$

Así, el estimador de máxima verosimilitud de θ es el inverso de la media muestral y, consecuentemente, la estimación de máxima verosimilitud de $e^{-\theta t}$ será:

$$\widehat{g(\theta)} = e^{-\frac{t}{\overline{x}}}$$

pues el estimador de máxima verosimilitud es invariante ante transformaciones biunívocas y la función $e^{-\theta t}$ es decreciente en θ.

b) Suponiendo $n = 10$ y $\sum_{i=1}^{n} x_i = 1200$

$$P(X < 130) = 1 - P(X \geq 130) = 1 - P(X > 130) =$$
$$= 1 - e^{-\frac{130}{\overline{x}}} = 1 - e^{-\frac{130}{120}} = 0,6615$$

c) Para ver si el estadístico T es centrado se calcula

$$E[T] = E\left[\frac{n-1}{\sum_{i=1}^{n} X_i}\right] = (n-1)E\left[\frac{1}{Y}\right]$$

donde $Y = \sum_{i=1}^{n} X_i$ y $X_i \sim Exp\,(\theta) \equiv \gamma\,(1, \theta)$. Por la propiedad reproductiva de la distribución Gamma, $Y \sim \gamma\,(n, \theta)$ de forma que:

$$E\left[\frac{1}{Y}\right] = \int_0^\infty \frac{1}{y} \frac{\theta^n}{(n-1)!} e^{-\theta y} y^{\theta-1} dy = \frac{\theta}{n-1} \int_0^\infty \frac{\theta^{n-1}}{(n-2)!} e^{-\theta y} y^{\theta-2} dy$$

Considerando los parámetros $p = n - 1$ y $a = \theta$, la integral anterior es la integral de la función de densidad de una distribución $\gamma(n-1, \theta)$ por lo que el valor de dicha integral es 1. Así,

$$E\left[\frac{1}{Y}\right] = \frac{\theta}{n-1}$$

y consecuentemente, la esperanza del estadístico T es

$$E[T] = (n-1)E\left[\frac{1}{\sum_{i=1}^{n} X_i}\right] = (n-1)\frac{\theta}{n-1} = \theta$$

y queda demostrado que el estadístico T es centrado para θ.

d) Para el análisis de la suficiencia de $\sum_{i=1}^{n} X_i$ se utiliza el teorema de factorización. La función de verosimilitud de la muestra es $f_\theta(x_1, \ldots, x_n) = \theta^n e^{-\theta \sum_{i=1}^{n} x_i}$. Considerando la siguiente descomposición:

$$g(S) = \theta^n e^{-\theta \sum_{i=1}^{n} x_i} = \theta^n e^{-\theta S}, \text{ donde } S = \sum_{i=1}^{n} X_i$$

$$h(x_1, \ldots, x_n) = 1$$

se observa que $g(S)$ depende de la muestra solo a través del estadístico S y que $h(x_1, \ldots, x_n)$ no depende del parámetro. Por el teorema de factorización queda demostrada la suficiencia del estadístico.

En cuanto a la completitud, $S = \sum_{i=1}^{n} X_i$ será completo si $E_\theta[f(S)] = 0$ implica que $f(S) = 0$. Como hemos visto que $\sum_{i=1}^{n} X_i \sim \gamma(n, \theta)$,

$$E_\theta[f(S)] = \int_0^\infty f(y)\frac{\theta^n}{(n-1)!}e^{-\theta y}y^{n-1}dy.$$

Por tanto, $E_\theta[f(S)] = 0$ para cada $\theta > 0$, significa que la transformada de Laplace $\int_0^\infty f(y)e^{-\theta y}y^{n-1}dy$, de la función $f(y)y^{n-1}$ se anula. Según el teorema de unicidad de la transformada de Laplace, ello solo es posible si $f(y)y^{n-1} = 0$, o bien $f(y) = 0$ para todo $y > 0$. Es decir, S es un estadístico suficiente y completo.

Finalmente, por el teorema de Lehmann - Scheffé, al ser $S = \sum_{i=1}^{n} X_i$ suficiente y completo y T función de dicho estadístico suficiente y completo, que es a su vez centrado, entonces es ECUMV.

Cuestión 2

Sea un diseño estratificado con muestreo aleatorio simple sin reemplazamiento en cada estrato. Se dispone de H estratos de tamaño N_h, $h = 1, \ldots, H$. El objetivo es estimar la media poblacional $\overline{Y}$ de una cierta característica Y. Sean $\overline{X}_h$, $h = 1, \ldots, H$ las medias poblacionales de los estratos de una característica auxiliar X, que suponemos conocidos.

a) Para estimar la media poblacional $\overline{Y}$ se propone el siguiente estimador

$$\widehat{\overline{Y}}_D = \widehat{\overline{Y}}_{HT} + \overline{X} - \widehat{\overline{X}}_{HT}$$

donde $\overline{X}$ es la media poblacional de X y $\widehat{\overline{Y}}_{HT}$, $\widehat{\overline{X}}_{HT}$ son los estimadores de Horvitz-Thompson de las medias de Y y X, respectivamente. Se pide:

a.1. Demostrar que $\widehat{\overline{Y}}_D$ es un estimador insesgado para $\overline{Y}$.

a.2. Calcular la varianza de $\widehat{\overline{Y}}_D$.

a.3. ¿Cuál es la asignación óptima del tamaño muestral en h, n_h, cuando se minimiza la varianza de $\widehat{\overline{Y}}_D$? (Se supone que el coste por unidad muestral no depende del estrato.)

a.4. ¿En qué condiciones $\widehat{\overline{Y}}_D$ sería preferible a $\widehat{\overline{Y}}_{HT}$?

b) Bajo las mismas condiciones anteriores consideramos ahora el siguiente estimador:

$$\widehat{\overline{Y}}_b = \widehat{\overline{Y}}_{HT} + b\left(\overline{X} - \widehat{\overline{X}}_{HT}\right)$$

donde b es un número real fijo. ¿Cuál es el valor óptimo para b?

c) Ahora se supone que no conocemos los valores poblacionales de X ni se dispone de la población estratificada. Entonces, en primer lugar se obtiene una muestra aleatoria simple sin reemplazamiento, S^*, de tamaño n^* entre las N unidades de la población y calculamos el estimador de Horvitz-Thompson $\widehat{\overline{X}}^*$ de la media poblacional $\overline{X}$. Después, se obtiene una submuestra aleatoria simple sin reemplazamiento de n unidades a partir de S^*. Se propone el siguiente estimador:

$$\widehat{\overline{Y}}_c = \widehat{\overline{Y}} + c\left(\widehat{\overline{X}}^* - \widehat{\overline{X}}\right)$$

donde

$$\widehat{\overline{Y}} = \frac{1}{n}\sum_{i=1}^{n} y_i \qquad \widehat{\overline{X}} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad \widehat{\overline{X}}^* = \frac{1}{n^*}\sum_{i=1}^{n^*} x_i$$

Demostrar que:

$$Var\left(\widehat{\overline{Y}}_c\right) = \frac{N-n^*}{Nn^*}S_y^2 + \frac{n^*-n}{nn^*}S_u^2$$

donde

$$u_i = y_i - cx_i, \qquad S_y^2 = \frac{\sum_{i=1}^{N}(y_i-\overline{y})^2}{N-1}, \qquad S_u^2 = \frac{\sum_{i=1}^{N}(u_i-\overline{u})^2}{N-1}.$$

Solución.

a) a.1.

$$E\left[\widehat{\overline{Y}}_D\right] = E\left[\widehat{\overline{Y}}_{HT} + \overline{X} - \widehat{\overline{X}}_{HT}\right] = \overline{Y} + \overline{X} - \overline{X} = \overline{Y}$$

por ser insesgados los estimadores de Horvitz y Thompson, $\widehat{\overline{Y}}_D$ es un estimador insesgado para $\overline{Y}$.

a.2.

$$Var\left(\widehat{\overline{Y}}_D\right) = Var\left(\widehat{\overline{Y}}_{HT} + \overline{X} - \widehat{\overline{X}}_{HT}\right) =$$

$$= Var\left(\widehat{\overline{Y}}_{HT}\right) + Var\left(\overline{X} - \widehat{\overline{X}}_{HT}\right) + 2Cov\left(\widehat{\overline{Y}}_{HT}, \overline{X} - \widehat{\overline{X}}_{HT}\right) =$$

$$= Var\left(\widehat{\overline{Y}}_{HT}\right) + Var\left(\widehat{\overline{X}}_{HT}\right) - 2Cov\left(\widehat{\overline{Y}}_{HT}, \widehat{\overline{X}}_{HT}\right) =$$

$$= \sum_{h=1}^{H} W_h^2 \frac{1-f_h}{n_h}\left[S_{Y_h}^2 + S_{X_h}^2 - 2S_{X_hY_h}\right].$$

a.3. La asignación óptima del tamaño muestral en h, n_h, cuando se minimiza la varianza de $\widehat{\overline{Y}}_D$ vendrá dada por la solución del siguiente problema de optimización con restricciones:

$$\begin{cases} \text{mín}\, Var\left(\widehat{\overline{Y}}_D\right) \\ s.a.\ C = \sum c \cdot n_h \end{cases} \Leftrightarrow \begin{cases} \text{mín} \sum_{h=1}^{H} W_h^2 \frac{1-f_h}{n_h}\left[S_{Y_h}^2 + S_{X_h}^2 - 2S_{X_hY_h}\right] \\ s.a.\ C = \sum c \cdot n_h \end{cases}$$

Este problema se resuelve aplicando el método de los multiplicadores de Lagrange, considerando como función objetivo:

$$\Phi\left(n_h, \lambda\right) = \sum_{h=1}^{H} W_h^2 \frac{1-f_h}{n_h}\left[S_{Y_h}^2 + S_{X_h}^2 - 2S_{X_hY_h}\right] + \lambda\left(\sum c \cdot n_h - C\right).$$

Operando:

$$\Phi = \sum_{h=1}^{H} W_h^2 \left(1 - \frac{n_h}{N_h}\right) \frac{S_{Y_h}^2 + S_{X_h}^2 - 2S_{X_hY_h}}{n_h} + \lambda\left(\sum c \cdot n_h - C\right) =$$

$$= \sum_{h=1}^{H} W_h^2 \frac{S_{Y_h}^2 + S_{X_h}^2 - 2S_{X_hY_h}}{n_h} - \sum_{h=1}^{H} W_h^2 \frac{S_{Y_h}^2 + S_{X_h}^2 - 2S_{X_hY_h}}{N_h} +$$

$$+ \lambda\left(\sum c \cdot n_h - C\right) =$$

$$= \sum_{h=1}^{H} W_h^2 \frac{S_{Y_h}^2 + S_{X_h}^2 - 2S_{X_hY_h}}{n_h} - \sum_{h=1}^{H} W_h \frac{S_{Y_h}^2 + S_{X_h}^2 - 2S_{X_hY_h}}{N} +$$

$$+ \lambda\left(\sum c \cdot n_h - C\right).$$

El paso siguiente es derivar la función objetivo respecto de sus variables, igualar a cero y eliminar el parámetro λ. Se tiene:

$$\left.\begin{array}{l} \frac{\partial \Phi(n_h,\lambda)}{\partial n_h} = 0 \ (h = 1, 2, \ldots, H) \\ \frac{\partial \Phi(n_h,\lambda)}{\partial \lambda} = 0 \end{array}\right\}$$

Derivando se tiene:

$$\left.\begin{array}{l}\frac{\partial\Phi}{\partial n_h}=-W_h^2\frac{S_{Y_h}^2+S_{X_h}^2-2S_{X_hY_h}}{n_h^2}+\lambda\cdot c=0\ (h=1,2,\ldots,H)\\ \frac{\partial\Phi}{\partial\lambda}=\sum c\cdot n_h-C\end{array}\right\}$$

Por tanto,

$$\lambda=\frac{W_h^2}{c}\cdot\frac{S_{Y_h}^2+S_{X_h}^2-2S_{X_hY_h}}{n_h^2}=\frac{N_h^2}{N^2}\cdot\frac{S_{Y_h}^2+S_{X_h}^2-2S_{X_hY_h}}{c\cdot n_h^2}$$

$$\Rightarrow\sqrt{\lambda}=\frac{N_h}{N}\cdot\frac{\sqrt{S_{Y_h}^2+S_{X_h}^2-2S_{X_hY_h}}}{\sqrt{c}\cdot n_h}.$$

De donde:

$$N\sqrt{\lambda}=\frac{N_h\sqrt{\frac{S_{Y_h}^2+S_{X_h}^2-2S_{X_hY_h}}{c}}}{n_h}$$

$$\Rightarrow\frac{n_h}{N_h\sqrt{\frac{S_{Y_h}^2+S_{X_h}^2-2S_{X_hY_h}}{c}}}=\frac{1}{N\sqrt{\lambda}}=Cte.$$

para $h=1,2,\ldots,H$. Desarrollando la igualdad para todo h y aplicando las propiedades de las proporciones se tiene:

$$\frac{n_1}{N_h\sqrt{\frac{S_{Y_1}^2+S_{X_1}^2-2S_{X_1Y_1}}{c}}}=\cdots=\frac{n_H}{N_H\sqrt{\frac{S_{Y_H}^2+S_{X_H}^2-2S_{X_HY_H}}{c}}}=Cte.$$

Esto es,

$$\begin{aligned}&\frac{n_h}{N_h\sqrt{\frac{S_{Y_h}^2+S_{X_h}^2-2S_{X_hY_h}}{c}}}=\\&=\frac{n_1+n_2+\cdots n_H}{N_1\sqrt{\frac{S_{Y_1}^2+S_{X_1}^2-2S_{X_1Y_1}}{c}}+\cdots+N_H\sqrt{\frac{S_{Y_H}^2+S_{X_H}^2-2S_{X_HY_H}}{c}}}=\\&=\frac{n}{\sum_{h=1}^H N_h\sqrt{\frac{S_{Y_h}^2+S_{X_h}^2-2S_{X_hY_h}}{c}}}.\end{aligned}$$

Despejando n_h se obtiene el valor de la afijación de mínima varianza:

$$n_h=n\frac{N_h\sqrt{\frac{S_{Y_h}^2+S_{X_h}^2-2S_{X_hY_h}}{c}}}{\sum_{h=1}^H N_h\sqrt{\frac{S_{Y_h}^2+S_{X_h}^2-2S_{X_hY_h}}{c}}}$$

y dividiendo numerador y denominador por N y simplificando se puede escribir equivalentemente como

$$n_h = n \cdot \frac{W_h \sqrt{S^2_{Y_h} + S^2_{X_h} - 2S_{X_hY_h}}}{\sum_{h=1}^{H} W_h \sqrt{S^2_{Y_h} + S^2_{X_h} - 2S_{X_hY_h}}}$$

a.4. Teniendo en cuenta que $Var\left(\widehat{\overline{Y}}_D\right) = \sum_{h=1}^{H} W_h^2 \frac{1-f_h}{n_h}\left[S^2_{Y_h} + S^2_{X_h} - 2S_{X_hY_h}\right]$ y $Var\left(\widehat{\overline{Y}}_{HT}\right) = \sum_{h=1}^{H} W_h^2 \frac{1-f_h}{n_h} S^2_{Y_h}$, $\widehat{\overline{Y}}_D$ será preferible a $\widehat{\overline{Y}}_{HT}$ cuando $Var\left(\widehat{\overline{Y}}_D\right)$ sea menor que $Var\left(\widehat{\overline{Y}}_{HT}\right)$. Esto es,

$$\begin{aligned}
Var\left(\widehat{\overline{Y}}_D\right) < Var\left(\widehat{\overline{Y}}_{HT}\right) &\Leftrightarrow \sum_{h=1}^{H} W_h^2 \frac{1-f_h}{n_h}\left[S^2_{X_h} - 2S_{X_hY_h}\right] < 0 \\
&\Leftrightarrow \sum_{h=1}^{H} W_h^2 \frac{1-f_h}{n_h} S^2_{X_h} < \sum_{h=1}^{H} W_h^2 \frac{1-f_h}{n_h} 2S_{X_hY_h} \\
& Var\left(\widehat{\overline{X}}_{HT}\right) < 2 \cdot Cov\left(\widehat{\overline{Y}}_{HT}, \widehat{\overline{X}}_{HT}\right)
\end{aligned}$$

b) El valor óptimo para b en el estimador

$$\widehat{\overline{Y}}_b = \widehat{\overline{Y}}_{HT} + b\left(\overline{X} - \widehat{\overline{X}}_{HT}\right)$$

será aquel que minimice la varianza. Siendo

$$\begin{aligned}
Var\left(\widehat{\overline{Y}}_b\right) &= Var\left(\widehat{\overline{Y}}_{HT} + b\left(\overline{X} - \widehat{\overline{X}}_{HT}\right)\right) = \\
&= Var\left(\widehat{\overline{Y}}_{HT}\right) + b^2 Var\left(\overline{X} - \widehat{\overline{X}}_{HT}\right) + 2b\, Cov\left(\widehat{\overline{Y}}_{HT}, \overline{X} - \widehat{\overline{X}}_{HT}\right) = \\
&= Var\left(\widehat{\overline{Y}}_{HT}\right) + b^2 Var\left(\widehat{\overline{X}}_{HT}\right) - 2b\, Cov\left(\widehat{\overline{Y}}_{HT}, \widehat{\overline{X}}_{HT}\right) = \\
&= \sum_{h=1}^{H} W_h^2 \frac{1-f_h}{n_h}\left[S^2_{Y_h} + b^2 S^2_{X_h} - 2bS_{X_hY_h}\right].
\end{aligned}$$

Para hallar el valor de b que minimiza esta expresión, se considera la varianza función de b y se iguala a cero su derivada respecto de b.

$$\Phi(b) = \sum_{h=1}^{H} W_h^2 \frac{1-f_h}{n_h}\left[S^2_{Y_h} + b^2 S^2_{X_h} - 2bS_{X_hY_h}\right]$$

$$\Phi^{'}(b) = 2b\sum_{h=1}^{H} W_h^2 \frac{1-f_h}{n_h} S_{X_h}^2 - 2\sum_{h=1}^{H} W_h^2 \frac{1-f_h}{n_h} S_{X_hY_h} = 0.$$

Despejando b se tiene que el valor que minimiza la varianza del estimador $\widehat{\overline{Y}}_b$ es:

$$b = \frac{\sum_{h=1}^{H} W_h^2 \frac{1-f_h}{n_h} S_{X_hY_h}}{\sum_{h=1}^{H} W_h^2 \frac{1-f_h}{n_h} S_{X_h}^2} = \frac{Cov\left(\widehat{\overline{Y}}_{HT}, \widehat{\overline{X}}_{HT}\right)}{Var\left(\widehat{\overline{X}}_{HT}\right)}$$

c) En un muestreo como este por fases tenemos dos conjuntos de unidades de muestreo, cuya selección da lugar a tener dos tipos de variación, que son la correspondiente al muestreo de unidades primarias, que lo denotaremos con el subíndice n^*, y la que es debida al submuestreo dentro de un conjunto fijo de unidades primarias, que lo denotamos con el subíndice n.

En primer lugar, se demuestra que el estimador $\widehat{\overline{Y}}_c$ es insesgado:

$$\begin{aligned} E\left[\widehat{\overline{Y}}_c\right] &= E_{n^*}\left[E_n\left(\widehat{\overline{Y}}_c\right)\right] = E_{n^*}\left[E_n\left(\widehat{\overline{Y}} + c\left(\widehat{\overline{X}}^* - \widehat{\overline{X}}\right)\right)\right] = \\ &= E_{n^*}\left[\overline{Y} + cE_n\left(\widehat{\overline{X}}^* - \widehat{\overline{X}}\right)\right] = E_{n^*}\left[\overline{Y} + c\left(\widehat{\overline{X}}^* - \overline{X}\right)\right] = \\ &= \overline{Y} + cE_{n^*}\left[\widehat{\overline{X}}^* - \overline{X}\right] = \overline{Y} + c\left(\overline{X} - \overline{X}\right) = \overline{Y}. \end{aligned}$$

Para el cálculo de la varianza, aplicando el teorema de Madow tenemos:

$$\begin{aligned} Var\left(\widehat{\overline{Y}}_c\right) &= Var_{n^*}\left(E_n\left(\widehat{\overline{Y}}_c\right)\right) + E_{n^*}\left(Var_n\left(\widehat{\overline{Y}}_c\right)\right) = \\ &= Var_{n^*}\left(E_n\left(\widehat{\overline{Y}}\right) + cE_n\left(\widehat{\overline{X}}^* - \widehat{\overline{X}}\right)\right) + E_{n^*}\left(Var_n\left(\widehat{\overline{Y}}_c\right)\right) = \\ &= Var_{n^*}\left(\widehat{\overline{Y}}^* + c\left(\widehat{\overline{X}}^* - E_n\left(\widehat{\overline{X}}\right)\right)\right) + E_{n^*}\left(Var_n\left(\widehat{\overline{Y}}_c\right)\right) = \\ &= Var_{n^*}\left(\widehat{\overline{Y}}^* + c\left(\widehat{\overline{X}}^* - \widehat{\overline{X}}^*\right)\right) + E_{n^*}\left(Var_n\left(\widehat{\overline{Y}}_c\right)\right) = \\ &= Var_{n^*}\left(\widehat{\overline{Y}}^*\right) + E_{n^*}\left(Var_n\left(\widehat{\overline{Y}}_c\right)\right) \end{aligned}$$

donde $\widehat{\overline{Y}}^*$ es el estimador de la media poblacional con la muestra de tamaño n^*, por tanto:

$$Var\left(\widehat{\overline{Y}}_c\right) = \left(1 - \frac{n^*}{N}\right)\frac{S_Y^2}{n^*} + E_{n^*}\left(Var_n\left(\widehat{\overline{Y}}_c\right)\right) =$$

$$= \left(\frac{N-n^*}{Nn^*}\right) S_Y^2 + E_{n^*}\left(Var_n\left(\widehat{\overline{Y}}_c\right)\right).$$

Por otra parte se calcula $Var_n\left(\widehat{\overline{Y}}_c\right)$:

$$Var_n\left(\widehat{\overline{Y}}_c\right) = Var_n\left(\widehat{\overline{Y}} + c\left(\widehat{\overline{X}}^* - \widehat{\overline{X}}\right)\right) = Var_n\left(\widehat{\overline{Y}} + c\widehat{\overline{X}}^* - c\widehat{\overline{X}}\right)$$

y como $Var_n\left(c\widehat{\overline{X}}^*\right) = 0$ porque $c\widehat{\overline{X}}^*$ no depende de los valores observados en la submuestra de n elementos extraída de la primera de n^* y por tanto es constante, entonces se tiene:

$$Var_n\left(\widehat{\overline{Y}}_c\right) = Var_n\left(\widehat{\overline{Y}} - c\widehat{\overline{X}}\right) = Var_n\left(\widehat{\overline{U}}\right)$$

siendo $u_i = y_i - cx_i$. Así,

$$Var_n\left(\widehat{\overline{U}}\right) = Var_n\left(\overline{u}\right) = \left(1 - \frac{n}{n^*}\right)\frac{\widehat{S}_U^2}{n}$$

donde

$$\widehat{S}_U^2 = \frac{\sum_{i=1}^{n^*}\left(u_i - \overline{u}\right)^2}{n^* - 1}.$$

Continuando con el desarrollo del teorema de Madow:

$$\begin{aligned} Var\left(\widehat{\overline{Y}}_c\right) &= \left(\frac{N-n^*}{Nn^*}\right) S_Y^2 + E_{n^*}\left(Var_n\left(\overline{u}\right)\right) = \\ &= \left(\frac{N-n^*}{Nn^*}\right) S_Y^2 + \left(1 - \frac{n}{n^*}\right)\frac{E_{n^*}\left(\widehat{S}_U^2\right)}{n} = \\ &= \left(\frac{N-n^*}{Nn^*}\right) S_Y^2 + \left(1 - \frac{n}{n^*}\right)\frac{S_U^2}{n} \end{aligned}$$

siendo ahora

$$S_U^2 = \frac{\sum_{i=1}^{N}\left(u_i - \overline{u}\right)^2}{N - 1}.$$

Por tanto, finalmente se tiene lo que se quería demostrar:

$$Var\left(\widehat{\overline{Y}}_c\right) = \left(\frac{N-n^*}{Nn^*}\right) S_Y^2 + \left(\frac{n^*-n}{n\ n^*}\right) S_U^2.$$

Cuestión 3

La estimación del PIB a precios de mercado de España para el año 2013 en la nueva base 2010 de la Contabilidad Nacional ha sido de 1049,2 miles de millones de euros (mm€). Se conocen también los datos por ramas de actividad relativos al Valor Añadido Bruto a precios básicos (VAB) y a la Remuneración de Asalariados (RA) para el citado año (mm€) que se recogen en la tabla 7.1. Para el conjunto de la economía se tiene además la siguiente información (mm€):

Otros impuestos menos subvenciones sobre la producción____________________9,6
Gasto en consumo final____________________814,5
Formación bruta de capital ____________________198,9
Remuneración de los asalariados residentes por empleadores no residentes__2,3
Remuneración de los asalariados no residentes por empleadores residentes__0,2
Impuestos menos subvenciones sobre la producción y las importaciones pagados al resto del mundo____________________-4,6
Rentas netas de la propiedad pagadas al resto del mundo____________________13,8

Tabla 7.1: Datos del VAB y la RA para el año 2013

	VAB	RA
Agricultura, ganadería, silvicultura y pesca	26,6	4,3
Industria	168,6	82,3
Construcción	55,1	27,6
Servicios	708,2	376,1

a. Se pide hallar los siguientes agregados:

a.1. EBE/RM.

a.2. Saldo de intercambios exteriores de bienes y servicios.

a.3. Renta nacional bruta.

b. Sabiendo que la estimación del PIB a precios corrientes para el año 2012 fue de 1055,2 y que la variación interanual en volumen del PIB entre 2012 y 2013 ha sido de -1,2 %:

b.1. Calcular la tasa de variación interanual del deflactor implícito de la economía.

b.2. Si se sabe que la variación interanual en volumen de los tres primeros trimestres del años 2013 fue respectivamente de -2,2 %, -1,7 % y -1,0 %, ¿se podría conocer la variación interanual aproximada para el cuarto trimestre de dicho año?

Solución.

a) a.1. Para cada rama de actividad se tiene que

$$\begin{aligned} VAB =& Producción - Consumos\ intermedios \\ VAB =& \sum rentas\ primarias\ generadas\ en\ el\ proceso\ productivo = \\ =& RA + EBE/RMB + otros\ impuestos\ menos\ subvenciones \\ & a\ la\ producción \end{aligned}$$

Por una parte,

$$VAB_{total} = 26,6 + 168,6 + 55,1 + 708,2 = 958,5.$$

Por otra,

$$\begin{aligned} VAB_{total} =& RA + EBE/RMB + otros\ impuestos\ menos\ subvenciones \\ & a\ la\ producción \end{aligned}$$

$$\begin{aligned} VAB_{total} =& 4,3 + 82,3 + 27,6 + 376,1 + EBE/RM + 9,6 \\ VAB_{total} =& 499,9 + EBE/RM. \end{aligned}$$

Luego

$$EBE/RM = 958,5 - 499,9 = 458,6 \text{ mm€}$$

a.2. El saldo de intercambios exteriores es la diferencia entre las importaciones hechas por España y las exportaciones hechas por España, es decir: $M - X$. De la ecuación del cálculo del PIB por el enfoque del gasto se sabe que:

$$\begin{aligned} PIB =& GCF + FBC + X - M \\ &\Rightarrow M - X = GCF + FBC - PIB \\ &\Rightarrow M - X = 814,5 + 198,9 - 1049,2 = -35,8 \text{ mm€} \end{aligned}$$

a.3. Finalmente, el valor de la RNB es:

$$\begin{aligned} RNB = PIB_{pm}^{2013} + & \\ & + RA \textit{ de residentes pagada por no residentes} \\ & - RA \textit{ de no residentes pagada por residentes} \\ & - \textit{rentas netas de la propiedad pagadas al resto del mundo} \\ & - \textit{impuestos netos sobre la producción y las importaciones} \\ & \textit{pagados al resto del mundo} = \\ & = 1049,2 + 2,3 - (0,2 + 13,8 - 4,6) = 1042,1 \text{ mm€} \end{aligned}$$

b) b.1. Sabiendo que la estimación del PIB a precios corrientes para el año 2012 fue de 1055,2 y en el año 2013 de 1049,2, el índice de valor es:

$$IV(PIB)_{2012}^{2013} = \frac{1049,2}{1055,2} \cdot 100 = 99,43.$$

Siendo la variación interanual en volumen del PIB entre 2012 y 2013 de -1,2 %, se tiene que:

$$IQ(PIB)_{2012}^{2013} = IVol(PIB)_{2012}^{2013} = (TVI\,(Vol,\ PIB)_{2012}^{2013} + 1) \cdot 100 =$$
$$= \left(\frac{-1,2}{100} + 1\right) \cdot 100 = 98,8.$$

Luego, al ser

$$\textit{Índice de valor} = \textit{Índice de precios} \cdot \textit{Índice de volumen},$$

$$\textit{Índice de precios} \equiv IPC_{2012}^{2013} = \frac{99,43}{98,8} \cdot 100 = 100,639$$

y por tanto la tasa de variación interanual del deflactor implícito de la economía es de un 0,64 %.

b.2. Por el apartado b.1., se sabe que $TVI(Vol,\ PIB)_{2012}^{2013} = -1,2\,\%$ luego la variación interanual aproximada para el cuarto trimestre de dicho año, x, será:

$$-1,2 = \frac{-2,2 - 1,7 - 1 + x}{4}$$
$$-4,8 = x - 4,9$$
$$x = 0,1$$

Es decir, la tasa interanual del cuarto trimestre es 0,1 %.

Cuestión 4

Considere la siguiente expresión de la curva de Phillips a corto plazo aumentada con expectativas:

$$y_t - E_{t-1}\,(y_t) = \beta\,(u_t - u_t^*) + e_t \tag{1}$$

donde y_t es la inflación en el periodo t y $E_{t-1}\,(y_t)$ es el valor esperado en $t-1$ para la inflación en t, u_t es la tasa de paro observada en t y u^* se refiere a la tasa natural de paro. Haga el supuesto de que la esperanza de la inflación en $t-1$ es precisamente la inflación observada en $t-1$ de manera que

$$y_t - y_{t-1} = \beta_1 + \beta_2 u_t + e_t \tag{2}$$

a) Justifique si el modelo (2) está anidado en (1) o si el modelo (1) está anidado en (2), o si, por el contrario, no son modelos anidados.

b) Con los datos de una economía europea para el periodo abarcado entre el primer trimestre de 1975 y el último de 1990 se obtiene la siguiente estimación del modelo (2) para los parámetros β_1, β_2:

Variables	*Coeficiente*	*Desv. típica*	*Estadístico t*	*valor p*
Constante	1,4126	2,32829	0,6067	0,54629
"paro"	-0,205876	0,323353	-0,6367	0,52671

Media de la var. dependiente = -0,0432317
Desviación típica de la var. dependiente. = 3,46601
Suma de cuadrados de los residuos = 739,902
Desviación típica de los residuos = 3,48275
R^2 = 0,0066016
R^2 corregido = -0,00968362
Grados de libertad = 61
Estadístico de Durbin-Watson = 2,61265
Coef. de autocorr. de primer orden. = -0,339296
Log-verosimilitud = -166,99
Criterio de información de Akaike = 337,979
Criterio de información Bayesiano de Schwarz = 342,266
Criterio de Hannan-Quinn = 339,665

A partir del modelo estimado, estime la tasa natural de paro.

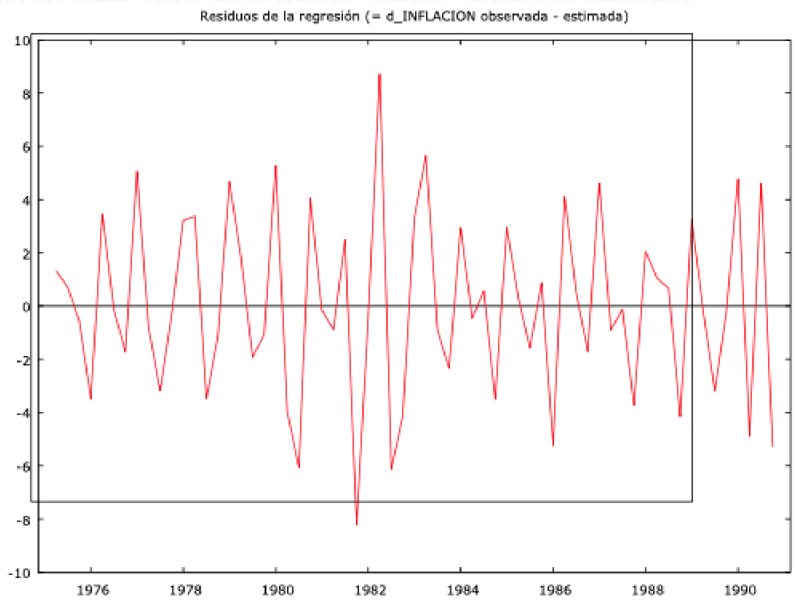
c) La tasa de paro estimada es una función no lineal de los parámetros estimados. Utilice el método delta para estimar su varianza. Para ello considere la siguiente matriz de covarianzas de los coeficientes estimados (tabla 7.2). Posteriormente, construya un intervalo de confianza del 95 % para dicha estimación. Utilice el valor crítico de 2.

Tabla 7.2: Matriz de covarianzas de los coeficientes

constante	"paro"	
5.42093	-0,739369	constante
	0,104557	"paro"

d) Un investigador independiente, con datos para las mismas variables y la misma economía, pero para el periodo 1950-2000, obtienen una estimación de la tasa natural de 5,46 %. Contraste si la obtenida en este ejemplo es significativamente diferente y, en su caso, trate de explicar la diferencia o semejanza.

e) Los residuos del modelo estimado (modelo (2)) son los siguientes:



Sobre estos residuos se realiza un contraste estadístico de Breusch y Godfrey. El resultado del mismo es:

Contraste de Breusch-Godfrey para autocorrelación hasta el orden 4estimaciones MCO
utilizando las 59 observaciones 1976:2-1990:4
Variable dependiente: ehat

VARIABLE	COEFICIENTE	DESV.TÍP.	ESTAD T	VALOR P	
const	1,52799	1,97027	0,776	0,44148	
PARO	-0,206071	0,276687	-0,745	0,45970	
ehat_1	-0,695497	0,139759	-4,976	<0,00001	***
ehat_2	-0,763386	0,165642	-4,609	0,00003	***
ehat_3	-0,340544	0,163138	-2,087	0,04167	**
ehat_4	-0,119786	0,148330	-0,808	0,42295	

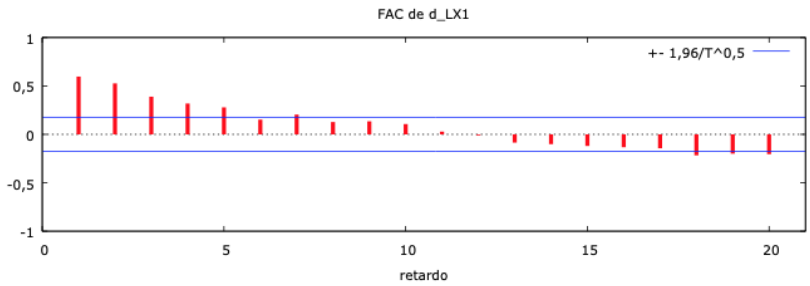
R-cuadrado = 0,426673
Estadístico de contraste: LMF = 10,249645, con valor p = P(F(4,53) > 10,2496) = 3,2e-006

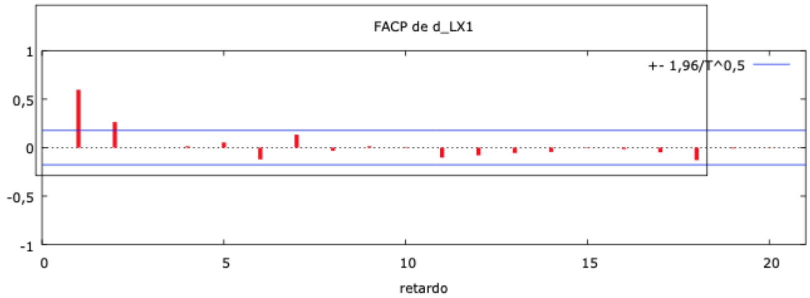
Estadístico alternativo: TR^2 = 25,173689, con valor p = P(Chi-cuadrado(4) > 25,1737) = 4,64e-005

Ljung-Box Q' = 20,3203 con valor p = P(Chi-cuadrado(4) > 20,3203) = 0,000432

Explique (i) en qué consiste el contraste, (ii) qué conclusión obtiene sobre el modelo y (iii) conteste justificadamente a la siguiente pregunta: ¿los resultados obtenidos eran esperados a la luz del gráfico de residuos del modelo?

f) Considere ahora que es pertinente modelizar los residuos. Para ello realiza un análisis de las funciones de autocorrelación que proporcionamos a continuación:





Y responda a las siguientes cuestiones: ¿Cómo modelizaría los residuos? ¿Qué modelo plantearía ahora para estimar el modelo 2?

g) Un colega le proporciona la siguiente estimación del modelo (2)

VARIABLE	COEFICIENTE	DESV.TÍP.	ESTAD T	VALOR P
const	1,34076	0,683277	1,962	0,05462
PARO	-0,191344	0,0955573	-2,002	0,05001

Estimaciones de los coeficientes AR:

e_1	-0,678453	0,136016	-4,988	<0,00001
e_2	-0,740201	0,160427	-4,614	0,00002
e_3	-0,314694	0,157269	-2,001	0,05034
e_4	-0,100201	0,144001	-0,696	0,48946

Suma de los coeficientes AR = -1,83355
Estadísticos basados en los datos rho-diferenciados:
Suma de cuadrados de los residuos = 419,857
Desviación típica de los residuos = 2,71402
R-cuadrado = 0,424775
R-cuadrado corregido = 0,414683
Grados de libertad = 57
Estadístico de Durbin-Watson = 1,85502
Coef. de autocorr. de primer orden. = 0,0245641
Criterio de información de Akaike (AIC) = 287,215
Criterio de información Bayesiano de Schwarz (BIC) = 291,37
Criterio de Hannan-Quinn (HQC) = 288,837

Responda justificadamente a las siguientes cuestiones: ¿Considera que su colega está utilizando la información del correlograma proporcionada en el apartado anterior? ¿Diría que el modelo ha mejorado respecto del primer modelo estimado en el apartado b)? ¿De qué manera se le ocurre mejorar el modelo? ¿Puede decir que se trata de una estimación de MCG?

Solución.

a) Se dice que un modelo B está anidado en otro modelo A si es posible obtener el modelo B únicamente imponiendo restricciones en A. De esta forma, el modelo (2) está anidado en (1) porque es un caso particular de este al imponerse las restricciones

$$\beta \cdot u_t^* = \beta_1 \qquad \text{y} \qquad E_{t-1}(y_t) = y_{t-1}.$$

b) La estimación del modelo (2) proporcionada por los datos de una economía europea para el periodo abarcado es, llamando $z_t = y_t - y_{t-1}$:

$$\begin{aligned} \widehat{z}_t =& \widehat{\beta}_1 + \widehat{\beta}_2 u_t \\ \Rightarrow& \widehat{z}_t = 1,4126 - 0,205876 u_t. \end{aligned}$$

A partir de las ecuaciones de los modelos (1) y (2) proporcionadas en el enunciado y sabiendo que $E_{t-1}(y_t) = y_{t-1}$, se tiene

$$\beta\left(u_t - u_t^*\right) + e_t = \beta_1 + \beta_2 u_t + e_t \Rightarrow \beta u_t - \beta u_t^* = \beta_1 + \beta_2 u_t,$$

y por tanto:

$$\widehat{\beta}_2 = \widehat{\beta} = -0,205876.$$

$$\widehat{\beta}_1 = -\widehat{\beta} u_t^* \Rightarrow u_t^* = \frac{\widehat{\beta}_1}{-\beta} = -\frac{\widehat{\beta}_1}{\widehat{\beta}_2} = -\frac{1,4126}{-0,205876} = 6,8615$$

y la tasa natural de paro estimada es $u_t^* = 6,8615$.

c) Dado que la tasa de paro estimada es una función no lineal de los parámetros estimados, se utiliza el método delta para estimar su varianza siendo la matriz de covarianzas de los coeficientes estimados

$$\Sigma_{\widehat{\beta}_1,\widehat{\beta}_2} = \begin{pmatrix} 5,42093 & -0,739369 \\ 0 & 0,104557 \end{pmatrix}.$$

Para ello se define $h(\beta)$ como:

$$h(\beta) \equiv u_t^* = \frac{\beta_1}{-\beta_2}$$

que es una función no lineal de los parámetros del modelo y se calcula el gradiente:

$$\nabla h(\beta)^T = \left(\frac{\partial h(\beta)}{\partial \beta_1} \; , \quad \frac{\partial h(\beta)}{\partial \beta_2} \right) = \left(\frac{-1}{\beta_2} \; , \quad \frac{\beta_1}{\beta_2^2} \right).$$

A continuación se aplica el método delta, que implica que:

$$\sqrt{n}\,(h(B) - h(\beta)) \xrightarrow{n\to\infty} N\left(0, [\nabla h(\beta)]^T \Sigma_{\beta_1,\beta_2} \nabla h(\beta)\right)$$

$$Var(\sqrt{n}\cdot h(\beta)) = Var(\sqrt{n}\cdot u_t^*) \simeq [\nabla h(\beta)]^T \Sigma_{\beta_1,\beta_2} \nabla h(\beta) =$$
$$= \left(\frac{-1}{\beta_2} \; , \quad \frac{\beta_1}{\beta_2^2} \right) \Sigma_{\beta_1,\beta_2} \begin{pmatrix} \frac{-1}{\beta_2} \\ \frac{\beta_1}{\beta_2^2} \end{pmatrix}.$$

De modo que la estimación será:

$$\widehat{Var}(\sqrt{n}\cdot \widehat{u_t^*}) = \left[\nabla h\left(\widehat{\beta}\right)\right]^T \Sigma_{\widehat{\beta}_1,\widehat{\beta}_2} \nabla h\left(\widehat{\beta}\right) =$$
$$= \left(\frac{-1}{\beta_2} \; , \quad \frac{\beta_1}{\beta_2^2} \right) \Sigma_{\widehat{\beta}_1,\widehat{\beta}_2} \begin{pmatrix} \frac{-1}{\beta_2} \\ \frac{\beta_1}{\beta_2^2} \end{pmatrix} =$$
$$= \left(\frac{-1}{-0{,}205876} \; , \quad \frac{1{,}4126}{(-0{,}205876)^2} \right) \begin{pmatrix} 5,42093 & -0,739369 \\ 0 & 0,104557 \end{pmatrix} \begin{pmatrix} \frac{-1}{-0{,}205876} \\ \frac{1{,}4126}{(-0{,}205876)^2} \end{pmatrix} =$$
$$= \left(4,8573 \; , \quad 33,3279 \right) \begin{pmatrix} 5,42093 & -0,739369 \\ 0 & 0,104557 \end{pmatrix} \begin{pmatrix} 4,8573 \\ 33,3279 \end{pmatrix} =$$

$$= \begin{pmatrix} 26,3310 & -0,106634 \end{pmatrix} \begin{pmatrix} 4,8573 \\ 33,3279 \end{pmatrix} = 124,3436$$

y por tanto $\widehat{Var}\left(\widehat{u_t^*}\right) = \widehat{Var}\left(h\left(\widehat{\beta}\right)\right) = \frac{124,3436}{64} = 1,9428.$

El intervalo de confianza del 95 % para la estimación de la tasa de paro será:

$$IC_{\widehat{u_t^*}}^{1-\alpha} = \left[\widehat{u_t^*} - z_{\alpha/2}\sqrt{\widehat{Var}\left(\widehat{u_t^*}\right)}, \widehat{u_t^*} + z_{\alpha/2}\sqrt{\widehat{Var}\left(\widehat{u_t^*}\right)}\right]$$

$$IC_{\widehat{u_t^*}}^{1-\alpha} = \left[6,8615 - 2\sqrt{1,9428}; 6,8615 + 2\sqrt{1,9428}\right]$$

$$IC_{\widehat{u_t^*}}^{1-\alpha} = [6,8615 - 2 \cdot 1,3938; 6,8615 + 2 \cdot 1,3938]$$

$$IC_{\widehat{u_t^*}}^{1-\alpha} = [4,0738; 9,6491]$$

d) En el periodo 1975 - 1990 la estimación obtenida es $\widehat{u_1^*} = 6,8615$ mientras que para el periodo 1950 - 2000 esta es $\widehat{u_2^*} = 5,46$. Para contrastar si ambas tasas son significativamente diferentes basta con ver si $\widehat{u_2^*}$ pertenece al intervalo de confianza calculado en c). Dado que

$$\widehat{u_2^*} = 5,46 \in [4,0738; 9,6491],$$

se puede concluir que las tasas no son significativamente distintas.

e) El contraste de Breusch y Godfrey es un contraste de autocorrelación en el que se regresan los residuos del modelo sobre un modelo autorregresivo. Es decir, el modelo de regresión $z_t = \beta_1 + \beta_2 u_t + e_t$, en el que los residuos pueden seguir un esquema autoregresivo AR(4), es decir,

$$e_t = \Phi_1 e_{t-1} + \Phi_2 e_{t-2} + \Phi_3 e_{t-3} + \Phi_4 e_{t-4} + \varepsilon_t$$

se estima por MCO para obtener los residuos muestrales $\widehat{e}_t$, y después se ajusta la regresión auxiliar:

$$\widehat{e}_t = \beta_1 + \beta_2 u_t + \Phi_1 \widehat{e}_{t-1} + \Phi_2 \widehat{e}_{t-2} + \Phi_3 \widehat{e}_{t-3} + \Phi_4 \widehat{e}_{t-4} + \varepsilon_t$$

En este ajuste obtenemos el valor de R^2 y se puede utilizar la distribución asintótica del estadístico $T \cdot R^2 \sim \chi^2_4$ para tomar una decisión sobre la hipótesis nula del contraste, que sería $H_0 : \{\Phi_1 = \Phi_2 = \Phi_3 = \Phi_4 = 0\}$, que significa ausencia de autocorrelación.

En el caso que nos ocupa, el modelo autorregresivo que se propone es de orden 4 y por eso tiene esa expresión la regresión auxiliar que se lleva a cabo para el contraste.

Sustituyendo con las estimaciones obtenidas se tiene:

$$\begin{aligned}\widehat{e}_t =& 1,52799 - 0,206071 u_t - 0,695497 \widehat{e}_{t-1} - 0,763386 \widehat{e}_{t-2} + \\ & - 0,340544 \widehat{e}_{t-3} - 0,119786 \widehat{e}_{t-4}\end{aligned}$$

Para tomar la decisión de rechazar o no la hipótesis nula del contraste hay comparar como decíamos el valor del estadístico $T \cdot R^2$ y el valor proporcionado por la distribución de la $\chi^2_{4;\alpha}$. Se observa que $T \cdot R^2 > \chi^2_{4;\alpha}$, puesto que el p-valor es muy pequeño (0,0000464) y por tanto se rechaza la hipótesis nula de ausencia de autocorrelación y se puede concluir que existe autocorrelación. Dado que el cuarto retardo no es significativo, podría decirse que la autocorrelación es de orden 3. En el gráfico se observa que los residuos presentan un comportamiento cíclico por lo que los resultados obtenidos van en la línea de que es necesario considerar un modelo con autocorrelación.

f) En las funciones de autocorrelación proporcionadas se observa lo siguiente:

- FAC: no se anula, su memoria es infinita. El correlograma tiene un comportamiento sinusoidal pero tendente a cero.
- FACP: su representación gráfica presenta únicamente dos valores distintos de cero que están fuera de los márgenes de confianza, los corres-

pondientes a los dos primeros retardos. Consecuentemente, el orden del modelo es dos.

A la luz de estas caracetrísticas observadas en las funciones de autocorrelación proporcionadas se plantea un modelo AR(2) para modelizar los residuos de forma que el modelo propuesto es:

$$e_t = \Phi_1 e_{t-1} - \Phi_2 e_{t-2} + \varepsilon_t.$$

g) El modelo propuesto es:

$$z_t = \beta_0 + \beta_1 u_t + \Phi_1 e_{t-1} + \Phi_2 e_{t-2} + \Phi_3 e_{t-3} + \Phi_4 e_{t-4} + \varepsilon_t$$

y la estimación del modelo propuesto por el compañero es la siguiente:

$$\begin{aligned}\widehat{z}_t &= \widehat{\beta}_0 + \widehat{\beta}_1 u_t + \widehat{\Phi}_1 e_{t-1} + \widehat{\Phi}_2 e_{t-2} + \widehat{\Phi}_3 e_{t-3} + \widehat{\Phi}_4 e_{t-4} = \\ &= 1,34076 - 0,191344 u_t - 0,678453 e_{t-1} - 0,740201 e_{t-2} \\ &- 0,314694 e_{t-3} - 0,100201 e_{t-4}\end{aligned}$$

Sin embargo, no parece que el compañero esté utilizando la información del correlograma proporcionada en el apartado anterior. En los mencionados gráficos de las funciones de autocorrelación se observa que la FACP se anula para retardos superiores al segundo. Es decir, el compañero ha utilizado un modelo con más autocorrelación que la que realmente se observa. Sin embargo la estimación sí que mejora respecto al primer modelo estimado. El R^2 aumenta considerablemente pues con el modelo del compañero se explica mayor proporción de la variable dependiente. Por otra parte disminuyen los criterios de información de Akaike (AIC) y Bayesiano de Schwarz (BIC) y esta disminución es reflejo de la mejora en el modelo. El estadístico de Durbin Watson también disminuye y su nuevo valor indica que ya no existe autocorrelación en el modelo ya que el valor de las cotas del contraste de Durbin Watson para este modelo, con $n = 60$ y $\alpha = 0,05$ son:

$$d_U = 1,62; \quad d_L = 1,55; \quad 4 - d_U = 2,38; \quad 4 - d_L = 2,45.$$

Consecuentemente,

$$DW_{(\text{modelo apartado b})} = 2,61265 \in (4 - d_L, 4) \Rightarrow \text{Autocorrelación negativa}$$

$$DW_{(\text{modelo apartado g})} = 1,85502 \in (d_U, 4 - d_U) \Rightarrow \text{No existe autocorrelación}$$

Para mejorar este modelo, se podrían eliminar el tercer y cuarto retardos tratando de simplificar al máximo el mismo. El tercer retardo tiene un p-valor cercano a 0,05 y el cuarto no es significativo. En cuanto a la pregunta de si se puede decir que se trata de una estimación de MCG hay que tener en cuenta lo siguiente:

En el modelo lineal general, la matriz de varianzas y covarianzas del vector de perturbaciones ε es un escalar: $Var\left(\varepsilon\right) = \sigma_\varepsilon^2 I_T$, de forma que las perturbaciones son homocedásticas y no están correlacionadas. Sin embargo, en un modelo que presente autocorrelación, la forma general de la matriz de varianzas y covarianzas de las perturbaciones es:

$$E\left(\varepsilon\varepsilon^T\right) = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1T} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2T} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{T1} & \sigma_{T2} & \cdots & \sigma_T^2 \end{pmatrix} = \Omega.$$

Se trabaja por tanto dentro del marco más general del modelo de regresión lineal con matrices de varianzas y covarianzas no escalares. Al flexibilizar el supuesto de perturbaciones esféricas, hay que buscar un método de estimación alternativo al de Mínimos Cuadrados Ordinarios pues las estimaciones obtenidas utilizando MCO dejarán de ser insesgadas y óptimas al incumplirse los supuestos del teoremas de Gauss - Markov. Este método alternativo tiene en cuenta la información que recoge la matriz de covarianzas Ω y se conoce con el nombre de Mínimos Cuadrados Generalizados (MCG). El método de estimación de MCG se basa en el criterio de estimación mínimo cuadrática pero la función de distancia a minimizar es distinta a la de este criterio pues incorpora, como ya se adelantaba, la información adicional de las pertubaciones en la matriz de varianzas y covarianzas Ω. La función objetivo a minimizar viene dada por

$$\min \left(Y - X\widehat{\beta}\right)^T \Omega^{-1} \left(Y - X\widehat{\beta}\right)$$

o equivalentemente se ecribe $\Omega = \sigma^2\Sigma$ donde Σ es conocida y σ^2 es un factor de escala:

$$\min \left(Y - X\widehat{\beta}\right)^T \sigma^2\Sigma^{-1} \left(Y - X\widehat{\beta}\right).$$

De las condiciones de primer orden del problema de minimización se obtiene un sistema de k ecuaciones normales

$$\left(X\Sigma^{-1}X\right)\widehat{\beta}_{MCG} = X^T\Sigma^{-1}Y$$

cuya solución es el estimador de mínimos cuadrados generalizados

$$\widehat{\beta}_{MCG} = \left(X\Sigma^{-1}X\right)^{-1} X^T\Sigma^{-1}Y.$$

Se puede demostrar que el estimador MCG de β es lineal, insesgado y óptimo dentro del marco del modelo de regresión lineal generalizado.

Por otra parte, cuando se realiza una modelización de los residuos, se está estimando el modelo de regresión por MCO incorporando la estructura de autocorrelación sin más que añadir a los regresores del modelo la especificación concreta de la perturbación, es decir, el tipo de proceso y el orden del mismo. En este caso concreto, en el que proponemos como mejora de la estimación del modelo propuesto por el compañero un modelo AR(2), la información recogida en la matriz Ω que se utiliza en la estimación de MCG recien explicada coincide con la información recogida en la matriz de varianzas y covarianzas del modelo AR(2). Así, puede decirse que se trata de una estimación de MCG.

Cuestión 5

La oficina estadística de un país facilita el siguiente repertorio de informaciones para un año T todo él referido a la población de mujeres (cifras en miles):

Todos los datos referidos a la población femenina en el año T							
Cantidades en miles	Población a 1 de Enero	Nacimientos por edad de la madre	Defunciones por edad del difunto	Migraciones exteriores		Migraciones interiores	Población a 31 de Diciembre
				Flujo emigración con destino al extranjero	Flujo inmigración con origen en el extranjero	Flujo de migraciones entre provincias	
TOTAL	23.710,2	206,6	190,6	239,2	146,1	232,1	23.633,7
0-4 años	1.172,5	-	0,7	12,0	8,6	11,8	1.123,8
5-9 años	1.184,1	-	0,1	10,5	6,7	10,3	1.200,4
10-14 años	1.080,6	0,1	0,1	9,9	7,5	7,9	1.102,2
15-19 años	1.051,1	4,3	0,1	11,0	11,4	9,3	1.038,5
20-29 años	2.645,4	53,7	0,4	58,7	43,5	56,0	2.538,2
30-39 años	3.795,6	134,4	1,3	64,8	26,7	61,2	3.681,0
40-49 años	3.705,0	14,1	3,7	34,6	16,1	29,5	3.719,6
50-59 años	3.067,9	-	7,6	19,5	11,6	16,8	3.134,2
60-64 años	1.286,7	-	5,2	6,0	5,1	6,3	1.281,0
65 y más años	4.721,3	-	171,4	12,2	8,9	23,0	4.814,8

Se consideran como migraciones interiores todos y cada uno de los cambios de residencia permanente que traspasan la frontera de una provincia y tienen como destino otra provincia del mismo país, sin distinción de orden de la migración u otra tipología cualquiera.

A la vista del anterior enunciado resuelva los siguientes asuntos, formulando y razonando cada uno de los extremos:

a) Calcule las tasas de migración interior para cada grupo de edades de mujeres.

b) Calcule las funciones de una tabla de migración interior femenina de acuerdo a los datos del mencionado año T, considerando que en el registro de las migraciones la incidencia de migraciones repetidas realizadas por un mismo individuo durante dicho periodo se considera empíricamente irrelevante:

   b.1. Las probabilidades de migrar dentro del país entre provincias diferentes.

   b.2. Sedentarios de la tabla de migración interprovincial

   b.3. Migraciones de la generación ficticia de la tabla

c) De acuerdo con la anterior tabla de migración: ¿cuál es la intensidad del fenómeno demográfico migración interior interprovincial?

d) ¿Cuál es la edad media a la migración?, ¿bajo que hipótesis se calcula?

e) Otros comentarios generales a la vista de los flujos demográficos que se presentan.

Solución.

a) Dado que se trata de datos transversales, las tasas de migración interior para cada grupo de edades de mujeres se calcula utilizando

$$m_x = \frac{I_x}{\frac{P_{01/01/T}+P_{31/12/T}}{2}} \cdot 1000$$

donde I_x es el flujo de migraciones entre provincias, $P_{01/01/T}$ es la población a 1 de enero, $P_{31/12/T}$ es la población a 31 de diciembre. Así, se tiene, expresado en tanto por mil:

$$m_{0,4} = \frac{11,8}{\frac{1172,5+1123,8}{2}} \cdot 1000 = 10,28‰$$

De forma análoga para el resto de grupos de edad:

$$m_{5,9} = 8,64‰$$

$$m_{10,14} = 7,24‰$$

$$m_{15,19} = 8,90‰$$

$$m_{20,29} = 21,61‰$$

$$m_{30,39} = 16,37‰$$

$$m_{40,49} = 7,95‰$$

$$m_{50,59} = 5,42‰$$

$$m_{60,64} = 4,91‰$$

$$m_{65+} = 4,82‰$$

b) b.1. Las probabilidades de migrar dentro del país entre provincias diferentes se calculará utilizando

$$q_x = \frac{n \cdot m_{x,x+n}}{1 + (1 - a_x)\, n \cdot m_{x,x+n}}$$

donde se hace el supuesto de que las personas que migran en el interior lo hacen de manera uniforme a lo largo de un año. Por tanto, el valor de las medias de estas fracciones de intervalo de tiempo de permanencia en un territorio de aquellos que migran durante el intervalo será $a_x = 0,5$. Se tienen las siguientes probabilidades de migrar:

$$q_{0,4} = \frac{n \cdot m_{0,4}}{1 + (1 - a_x)\, n \cdot m_{0,4}} = \frac{5 \cdot 0,01027}{1 + \left(1 - \frac{1}{2}\right) 5 \cdot 0,01027} = 0,0501$$

$$q_{5,9} = 0,0423$$

$$q_{10,14} = 0,0355$$

$$q_{15,19} = 0,0435$$

$$q_{20,29} = 0,1950$$

$$q_{30,39} = 0,1513$$

$$q_{40,49} = 0,0764$$

$$q_{50,59} = 0,0527$$

$$q_{60,64} = 0,0242$$

$$q_{+65} = 0,0471$$

Para el cálculo de la probabilidad de migrar del grupo de más de 65 años se establece como edad máxima 75 años y por tanto se considera la amplitud del intervalo $n = 10$.

b.2. Los sedentarios de la tabla de migración interprovincial representa el número de personas de la generación ficticia inicial que permanecen sin migrar a la edad exacta x y se calculan utilizando

$$l_{x+n} = l_x - M_x$$

Habitualmente, en la construcción de tablas se parte de una raíz de 100.000 individuos, así $l_{0,4} = 100000$. Una vez se tiene calculado el número de migraciones interiores de la cohorte teórica ocurridas entre las edades x y x+n, que se representa por M_{x+n}, se obtiene l_{x+n} de manera recurrente. El número de migraciones interiores de la cohorte teórica se calcula como

$$M_x = l_x \cdot q_x.$$

De esta forma, los sedentarios de la tabla de migración interprovincial en cada grupo de edad son:

$$l_{5,9} = 100000 - (100000 \cdot 0,0501) = 100000 - 5010,0 = 94990,0$$

$$l_{10,14} = 94990,0 - (94990,0 \cdot 0,0423) = 94990,0 - 4016,4 = 90973,6$$

$$l_{15,19} = 87739,6$$

$$l_{20,29} = 83919,7$$

$$l_{30,39} = 67555,4$$

$$l_{40,49} = 57332,6$$

$$l_{50,59} = 52950,7$$

$$l_{60,64} = 50157,8$$

$$l_{+65} = 48942,0$$

b.3. A continuación se recogen las migraciones de la generación ficticia de la tabla para cada grupo de edad, ya calculadas en el apartado anterior utilizando la fórmula $M_x = l_x \cdot q_x$.

$$M_{0,4} = 5010,0$$

$$M_{5,9} = 4016,4$$

$$M_{10,14} = 3234,0$$

$$M_{15,19} = 3819,9$$

$$M_{20,29} = 16364,3$$

$$M_{30,39} = 10222,8$$

$$M_{40,49} = 4381,9$$

$$M_{50,59} = 2793,0$$

$$M_{60,64} = 1215,7$$

$$M_{+65} = 2305,3$$

c) La intensidad del fenómeno demográfico migración interior interprovincial se mide utilizando

$$I = \sum_{x=0}^{w-1} \frac{M_{x,x+n}}{l_0} \cdot 100 = \sum_{x=0}^{w-1} \frac{M_{x,x+n}}{1000} =$$
$$= \frac{5010,0 + 4016,4 + \cdots + 1215,7 + 2305,3}{1000} = 53,3632.$$

Esta intensidad de migración hace referencia a que el 53,36 % de una generación que mostrase un comportamiento migratorio como el de las mujeres observadas realizaría una migración interior alguna vez en su vida.

d) Utilizando

$$\overline{e} = \frac{\sum_{x=0}^{w-1} \left(x + \frac{n}{2}\right) M_{x,x+n}}{\sum_{x=0}^{w-1} M_{x,x+n}}$$

se calcula la edad media a la migración donde x es la marca de clase que se obtiene para cada intervalo mediante el cálculo de la media de los extremos de dicho intervalo. Así,

$$\overline{e} = \frac{\sum_{x=0}^{w-1} \left(x + \frac{n}{2}\right) M_{x,x+n}}{\sum_{x=0}^{w-1} M_{x,x+n}} = \frac{1504974,2}{53363,2} = 28,2.$$

Esto es, la edad media a la que que migraría una generación ficticia que mostrase los patrones de migraciones entre provincias es 28,2 años. Las hipótesis bajo las que se calcula la edad media a la migración son que las personas que migran en el interior lo hacen de manera uniforme a lo largo de un año y que las personas migran como máximo hasta los 75 años de edad, de manera que la marca de clase del último intervalo se sitúa en 70.

e) Vistos los flujos demográficos que obtenidos, llama la atención la elevada intensidad migratoria que refleja que el 53,36 % de una generación que mostrase un comportamiento migratorio como el de las mujeres perteneciente a la población en estudio realizaría una migración interior alguna vez en su vida. Por otra parte, la edad media a la migración, 28,2 años, es baja. Estos dos indicadores hacen pensar que podría tratarse de un país en el que parte de su población, en este caso, la femenina, necesita emigrar para poder trabajar, puesto que esa migración se produce en edades tempranas, y por tanto, no parece un país muy desarrollado.

En la siguiente figura pueden verse todos los cálculos:

MUJERES AÑO T										
	$P_{01\ ENERO}$	N	D	E	I	MIGINT	$P_{31\ DICIEMBRE}$	P_{MEDIA}	$_5mx$	
total	23.710,20	206,6	190,6	239,2	146,1	232,1	23.633,70	23.671,95	0,00980485	
0-4	1.172,50	0	0,7	12	8,6	11,8	1.123,80	1.148,15	0,0102774	
5-9	1.184,10	0	0,1	10,5	6,7	10,3	1.200,40	1.192,25	0,00863913	
10-14	1.080,60	0,1	0,1	9,9	7,5	7,9	1.102,20	1.091,40	0,00723841	
15-19	1.051,10	4,3	0,1	11	11,4	9,3	1.038,50	1.044,80	0,00890123	
20-29	2.645,40	53,7	0,4	58,7	43,5	56	2.538,20	2.591,80	0,02160661	
30-39	3.795,60	134,4	1,3	64,8	26,7	61,2	3.681,00	3.738,30	0,01637108	
40-49	3.705,00	14,1	3,7	34,6	16,1	29,5	3.719,60	3.712,30	0,00794656	
50-59	3.067,90	0	7,6	19,5	11,6	16,8	3.134,20	3.101,05	0,00541752	
60-64	1.286,70	0	5,2	6	5,1	6,3	1.281,00	1.283,85	0,00490712	
65+	4.721,30	0	171,4	12,2	8,9	23	4.814,80	4.768,05	0,00482377	
	$_nmx$	e_{inicio}	n	a_x	$_nq_x$	l_x	$_nM_x$		e_{inicio} +n/2	$(e_{inicio}$ +n/2)$\cdot{}_nM_x$
0-4	0,0102774	0	5	0,5	0,05009977	100.000,0	5.010,0		2,5	12.524,9
5-9	0,00863913	5	5	0,5	0,04228243	94.990,0	4.016,4		7,5	30.123,1
10-14	0,00723841	10	5	0,5	0,03554876	90.973,6	3.234,0		12,5	40.425,0
15-19	0,00890123	15	5	0,5	0,04353729	87.739,6	3.819,9		17,5	66.849,0
20-29	0,02160661	20	10	0,5	0,19499965	83.919,7	16.364,3		25	409.107,7
30-39	0,01637108	30	10	0,5	0,15132409	67.555,4	10.222,8		35	357.796,4
40-49	0,00794656	40	10	0,5	0,07642883	57.332,6	4.381,9		45	197.183,9
50-59	0,00541752	50	10	0,5	0,05274642	52.950,7	2.793,0		55	153.612,9
60-64	0,00490712	60	5	0,5	0,02423823	50.157,8	1.215,7		62,5	75.983,5
65+	0,00482377	65	10	0,5	0,04710171	48.942,0	2.305,3		70	161.367,8
							53.363,2			1.504.974,2
										28,2

7.2. Soluciones del tercer examen del Cuerpo de Diplomados en Estadística del Estado de 2014

Cuestión 1

Sea X una variable aleatoria que representa la producción semanal de ácido nítrico de una fábrica medido en miles de m^3. La función de densidad de esta variable viene dada por: $f(x) = k(1-x)^3$ si $x \in (0,1)$. Calcular:
a) El valor de k para que $f(x)$ sea función de densidad.
b) La capacidad de los depósitos para que la probabilidad de que los depósitos se desborden sea 0,001.
c) Si el tiempo de vida, medido en meses, de los peces está relacionado con la producción de acido nítrico según la siguiente función: $V = 12e^{-x}$, calcular la probabilidad de que un pez viva más de tres meses y la vida media de los peces.

Solución. Para que sea función de densidad tiene que ocurrir que

$$\int_0^1 f(x)dx = 1.$$

Por tanto:

$$\int_0^1 k(1-x)^3 dx = k\left[-\frac{(1-x)^4}{4}\right]_0^1 = \frac{k}{4}$$

Luego $k = 4$, con lo que la función de densidad será:

$$f(x) = \begin{cases} 4(1-x)^3 & \text{si} \quad 1 > x > 0 \\ 0 & \text{resto} \end{cases}$$

La probabilidad de que los depósitos desborden es la probabilidad de que la producción semanal de ácido supere la capacidad del depósito c, es decir, se debe cumplir que $P(X > c) = 0,001$. Luego:

$$1-P(X \leq c) = 1-\int_0^c 4(1-x)^3 dx = 1-[-(1-x)^4]_0^c = 1-[-(1-c)^4+1] = 0,001$$

Por tanto:

$$(1-c)^4 = 0,001 \Leftrightarrow 1-c = 0,001^{\frac{1}{4}} \Leftrightarrow c = 1-0,001^{\frac{1}{4}} = 0,822172 \text{ miles de } m^3$$

Es decir, la capacidad del depósito debe ser 822,172 m^3.

Para resolver el apartado c) primeramente calculamos la función de distribución de la variable X:

$$F(x) = \int_0^x 4(1-x)^3 dx = 4\left[-\frac{(1-x)^4}{4}\right]_0^x = 1-(1-x)^4 \ , \ x \in (0,1)$$

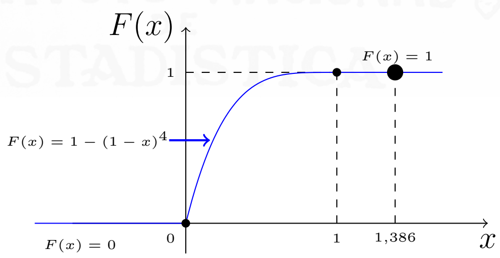
por tanto,

$$F(x) = \begin{cases} 0 & \text{si } x < 0 \\ 1-(1-x)^4 & \text{si } 0 \leq x < 1 \\ 1 & \text{si } x \geq 1 \end{cases}$$

Finalmente, sea $V = 12e^{-x}$,

$$P(V \geq 3) = P\left(12e^{-x} \geq 3\right) = P\left(x \leq -ln\left(\frac{3}{12}\right)\right) = F\left(-ln\left(\frac{3}{12}\right)\right) =$$
$$= F\left(-ln(0,25)\right) = F(1,386) = 1$$

por lo que la probabilidad de que un pez viva más de tres meses es 1, según puede comprobarse en el gráfico siguiente:



En cuanto a la esperanza, integrando por partes tres veces, se puede obtener:

$$E[V] = \int_0^1 12e^{-x} \cdot 4(1-x)^3 dx = 48\left(\frac{6}{e} - 2\right) \simeq 9,95 \text{ meses}$$

, que sería la vida media de los peces.

Cuestión 2

Sea (X_1, X_2) una muestra aleatoria simple de tamaño 2 de la variable aleatoria X que sigue una distribución normal con media cero y varianza $\frac{1}{\theta}$, siendo θ un parámetro desconocido. Consideramos la siguiente función de la muestra $T(X_1, X_2) = \frac{1}{2}\left(X_1^2 + X_2^2\right)$. Se pide:

a) Calcular la distribución de $2\theta T(X_1, X_2)$.

b) Hallar un intervalo de confianza basado en $2\theta T(X_1, X_2)$ para el parámetro θ y para un nivel $1-\alpha$ con $\alpha \in (0,1)$.

Solución.

Tenemos que $X_i \sim N\left(0, \sqrt{\frac{1}{\theta}}\right) \Rightarrow \sqrt{\theta} X_i \sim N(0,1) \Rightarrow \theta X_i^2 \sim \gamma\left(\frac{1}{2}, \frac{1}{2}\right) \equiv \chi_1^2$

luego por la propiedad reproductiva de la distribución gamma es:

$$\theta X_1^2 + \theta X_2^2 = \theta\left(X_1^2 + X_2^2\right) \sim \chi_2^2 \Rightarrow 2\theta \frac{X_1^2 + X_2^2}{2} = 2\theta T\left(X_1, X_2\right) \sim \chi_2^2$$

Al ser $2\theta T(X_1, X_2)$ un pivote se tiene que:

$$1 - \alpha = P\left(k_1 \leq 2\theta T(X_1, X_2) \leq k_2\right) \text{ , siendo } k_1 = \chi^2_{2,1-\frac{\alpha}{2}} \text{ y } k_2 = \chi_{2,\frac{\alpha}{2}}$$

y $P\left(\chi_2^2 > \chi^2_{2,1-\frac{\alpha}{2}}\right) = 1 - \frac{\alpha}{2}$.

De esta forma $1 - \alpha = P\left(\frac{\chi^2_{2,1-\frac{\alpha}{2}}}{2T(X_1,X_2)} \leq \theta \leq \frac{\chi^2_{2,\frac{\alpha}{2}}}{2T(X_1,X_2)}\right)$ y por tanto:

$$I_\theta^{1-\alpha} = \left[\frac{\chi^2_{2;1-\frac{\alpha}{2}}}{2T\left(X_1, X_2\right)}, \frac{\chi^2_{2;\frac{\alpha}{2}}}{2T\left(X_1, X_2\right)}\right]$$

Cuestión 3

Los taxis en servicio de una ciudad están numerados del 1 al N y se desea conocer cuántos taxis hay. Para ello se observa una muestra de n taxis y se apuntan sus números. Se pide:

a) Obtener un estimador por el método de los momentos. ¿Es insesgado?

b) Obtener un estimador por el método de máxima verosimilitud.

Solución.

a) Se entiende la distribución de la numeración de los N taxis como el resultado de una experiencia aleatoria de un conjunto finito de N posibles resultados todos ellos igualmente probables. Es decir, la variable X número de taxis puede tomar N valores ($k = 1, 2, ..., N$), todos con igual probabilidad. Su función de densidad será:

$$f(k) = P[X = k] = \frac{1}{N} \text{ , k=1,}\cdots\text{,N}$$

Su esperanza es por tanto:

$$E[X] = \sum_{i=1}^{N} x_i \cdot P(X = x_i) = \sum_{i=1}^{N} i \cdot \frac{1}{N} = \frac{1}{N} \cdot \frac{N(N+1)}{2} = \frac{N+1}{2}$$

Aplicando el método de los momentos,

$$E(X) = \frac{N+1}{2} = \overline{x} \Leftrightarrow \widehat{N}_{MM} = 2\overline{x} - 1$$

Veamos si es insesgado:

$$E\left[\widehat{N}_{MM}\right] = 2E\left[\overline{x}\right] - 1 = \frac{2}{n}E\left[\sum_{i=1}^{n} x_i\right] - 1 = \frac{2}{n}nE\left[x_i\right] - 1 = 2\left(\frac{N+1}{2}\right) - 1 = N$$

Luego es insesgado.

b) Para el cálculo del estimador por el método de máxima verosimilitud no se cumplen las condiciones de regularidad pues el soporte de la función depende del parámetro. Consecuentemente, no se puede aplicar el procedimiento habitual para el cálculo del estimador máximo verosímil.

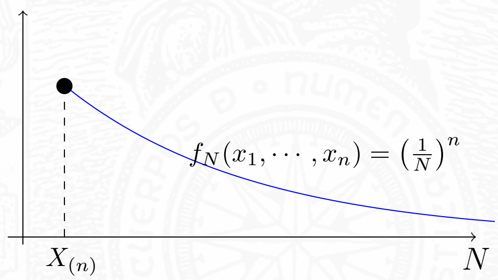
La función de verosimilitud sería:

$$f(x_1, \cdots, x_n) = \left(\frac{1}{N}\right)^n \cdot I_{\{X_{(n)}<N\}}$$

siendo $X_{(n)} = \max\limits_{1\leq i\leq n} X_i$, y

$$I_{\{X_{(n)}<N\}} = \begin{cases} 1 & \text{si } X_{(n)} < N \\ \\ 0 & \text{resto} \end{cases}$$

pero esta función, como función de N (ver figura siguiente), se hace máxima cuando N es lo menor posible, pero N siempre es mayor que $X_{(n)}$, luego toma el mayor valor para $\hat{N}_{MV} = X_{(n)}$, que sería el estimador máximo verosímil.



Cuestión 4

Sean los valores de altura (en centímetros) y de peso (en kilogramos) de un conjunto de 20 personas recogidos en la figura 7.1. Se pide:

a) El valor de la media de la distribución de alturas.

b) El valor de la desviación típica de la distribución de alturas.

c) El valor de la mediana de la distribución de alturas.

d) Agrúpense los datos en intervalos con una longitud razonadamente escogida y calcúlese la mediana de la distribución de datos agrupados.

e) El intervalo que contiene el 40 % central de la distribución de valores agrupados de alturas construida en el punto anterior.

f) Sean las variables binarias A y P que indican, respectivamente, si una persona tiene una altura superior a la mediana de las alturas y un peso superior a la mediana de los pesos. ¿Son A y P independientes? Demuéstrese.

Altura	164	164	166	167	169	169	169	169	170	170	171	171	171	173	173	173	174	174	175	177
Peso	68	70	69	69	69	70	69	67	71	69	70	71	76	71	70	68	70	74	71	70

Figura 7.1: Altura (en cm) y peso (en kg) de 20 personas

Solución.

Tabla 7.3: Distribución de frecuencias de la variable Altura

x_i	164	166	167	169	170	171	173	174	175	177
n_i	2	1	1	4	2	3	3	2	1	1
N_i	2	3	4	8	10	13	16	18	19	20

A la vista de la distribución de frecuencias de la tabla 7.3, se tiene que la media de las alturas es:

$$a_{10} = \frac{\sum_{i=1}^{n} x_i n_i}{N} = \frac{164 \cdot 2 + \cdots + 177 \cdot 1}{20} = 170,45 \text{ cm}.$$

El valor del momento de orden 2 respecto a la media muestral de la distribución de alturas es

$$m_{20} = a_{20} - a_{10}^2 = \frac{164^2 \cdot 2 + \cdots + 177^2 \cdot 1}{20} - 170,45^2 = 11,6475.$$

Por tanto, la desviación típica es $S = \sqrt{m_{20}^2} = 3,41$.

Al ser N par, habrá dos valores medianos de la distribución de alturas y la mediana será

$$Me = \frac{Me_1 + Me_2}{2}$$

Dado que $1 + \frac{N-1}{2} = 1 + \frac{19}{2} = 10,5$, será $Me_1 = 10$ y $Me_2 = 11$ de modo que

$$Me = \frac{170 + 171}{2} = 170,5 \text{ cm}.$$

Un criterio habitualmente utilizado para conocer el número de intervalos a formar para agrupar los datos en clases es formar $k = \sqrt{N}$ grupos, en este caso sería $\sqrt{N} = \sqrt{20} = 4,47 \simeq 4$. Por tanto, ya que el rango de la distribución es $177 - 164 = 13$, la amplitud de cada intervalo será $\frac{13}{4} = 3,25$ siendo los intervalos:

$$I_1 = [164; 167,25)$$

$$I_2 = [167,25; 170,5)$$

$$I_3 = [170,5; 173,75)$$

$$I_4 = [173,75; 177]$$

También se podría aplicar la fórmula de Stürges, que consiste en utilizar un número de intervalos igual a $k = ENT\left[1,5 + \frac{logN}{log2}\right] = ENT\left[1,5 + \frac{1,3010}{0,3010}\right] = ENT\,[5,822] = 5$

Tabla 7.4: Distribución de frecuencias de la variable Altura en intervalos

x_i	[164; 167,25)	[167,25;170,5)	[170,5;173,75)	[173,75; 177]
n_i	4	6	6	4
N_i	4	10	16	20

A partir de la distribución de frecuencias de la tabla 7.4, se calcula la mediana de la distribución agrupada. Siendo $\frac{N}{2} = 10$, el intervalo mediano será aquel con $N_i \geq \frac{N}{2}$, es decir, $[167,25; 170,5)$. Consecuentemente, la mediana de la distribución de datos agrupados será:

$$Me = L_{i-1} + \frac{\frac{N}{2} - N_{i-1}}{n_i} c_i = 167,25 + \frac{10-4}{6} 3,25 = 170,5$$

Para obtener el intervalo que contiene el 40 % central de la distribución de valores agrupados de alturas construida en el punto anterior, hay que calcular el percentil 30 y el percentil 70 utilizando la fórmula

$$Q_{r/q} = L_{i-1} + \frac{\frac{r}{q}N - N_{i-1}}{n_i} c_i$$

Siendo el intervalo I_2 el primer intervalo con una frecuencia acumulada $N_{i-1} \geq \frac{30}{100} 20 = 6$, entonces:

$$Q_{30/100} = 167,25 + \frac{6-4}{6} 3,25 = 168,33.$$

Para el caso del percentil 70, el primer intervalo con una frecuencia acumulada $N_{i-1} \geq \frac{70}{100}20 = 14$ es el intervalo I_3:

$$Q_{70/100} = 170,5 + \frac{14-10}{6}3,25 = 172,66.$$

De forma que el intervalo que contiene el 40 % central de la distribución es

$$[168,33; 172,66]$$

Para ver si A y P son independiente, calculamos sus distribuciones. Para ello hay que calcular la mediana de la variable Peso siguiendo un razonamiento análogo al utilizado para el cálculo de la mediana de la variable Altura.

Tabla 7.5: Distribución de frecuencias de la variable Peso

y_i	67	68	69	70	71	74	76
n_i	1	2	5	6	4	1	1
N_i	1	3	8	14	18	19	20

Ya que N es par, habrá dos valores medianos $Me_1 = 10$ y $Me_2 = 11$ de modo que

$$Me = \frac{70+70}{2} = 70.$$

La mediana de la variable peso es 70 kg y la de la altura es 170,5 cm, por lo que para cada observación podemos asignar el valor de las 2 variables binarias A y P comparando los valores (x_i, y_i) con estos dos valores de referencia (170,5,70), obteniendo la tabla 7.6.

Tabla 7.6: Distribución variables A y P

x_i	164	164	166	167	169	169	169	169	170	170	171	171	171	173	173	173	174	174	175	177
y_i	68	70	69	69	69	70	69	67	71	69	70	71	76	71	70	68	70	74	71	70
A	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1
P	0	0	0	0	0	0	0	0	1	0	0	1	1	1	0	0	0	1	1	0

Así, las variables A y P tendrán la distribución de frecuencias recogida en la tabla 7.7.

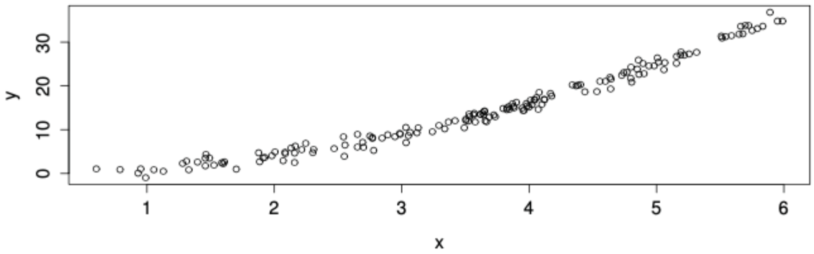
Tabla 7.7: Distribución de frecuencias A y P

	P=0	P=1	*Total*
A=0	9	1	10
A=1	5	5	10
Total vertical	14	6	20

Para comprobar si son independientes A y P, al ser una tabla de contingencia podemos tratar de comprobar si $n_{ij} = \frac{n_{i\bullet} \cdot n_{\bullet j}}{N}$, y es claro que no se cumple para cualquier valor de i y de j, puesto que por ejemplo $9 \neq \frac{14 \cdot 10}{20} = 7$, por lo tanto no son independientes.

Cuestión 5

Sean x e y variables cuantitativas con $n = 162$ pares de valores representados en la gráfica siguiente:



Se tiene el siguiente conjunto de cantidades:

$$\sum_{k=1}^{n} x_k = 592{,}4786 \qquad \sum_{k=1}^{n} y_k = 2469{,}627$$

$$\sum_{k=1}^{n} x_k^2 = 2469{,}245 \qquad \sum_{k=1}^{n} x_k y_k = 11111{,}01$$

$$\sum_{k=1}^{n} x_k^3 = 11093{,}92 \qquad \sum_{k=1}^{n} x_k^2 y_k = 52449{,}87$$

$$\sum_{k=1}^{n} x_k^4 = 52351{,}37 \qquad \sum_{k=1}^{n} x_k^3 y_k = 256396{,}9$$

$$R^2_{y\cdot x} = 0{,}9745407,\ R^2_{y\cdot xx^2} = 0{,}9950063 \qquad \sum_{k=1}^{n} y_k^2 = 51455{,}63$$

$$X^{(1)} = \begin{bmatrix} \mathbf{1} & \mathbf{x} \end{bmatrix},\ H^{(1)} = X^{(1)} \cdot (X^{(1)T} X^{(1)})^{-1} X^{(1)T} \quad ,\ y^T(\mathbb{I}_n - H^{(1)})y = 759{,}0528$$

$$X^{(2)} = \begin{bmatrix} \mathbf{1} & \mathbf{x} & \mathbf{x^2} \end{bmatrix},\ H^{(2)} = X^{(2)} \cdot (X^{(2)T} X^{(2)})^{-1} X^{(2)T} \quad ,\ y^T(\mathbb{I}_n - H^{(2)})y = 150{,}4276$$

Se pide:

a) Los coeficientes de la recta $y = \beta_0 + \beta_1 x$ ajustada por mínimos cuadrados.

b) La varianza residual del modelo anterior.

c) La varianza residual del modelo $y = \beta_0 + \beta_1 x + \beta_2 x^2$ ajustado por mínimos cuadrados.

d) Apartir de los valores de las varianzas residuales obtenidos anteriormente, arguméntese qué modelo es más apropiado.

e) En un diagrama con ejes de abscisas los valores predichos $\widehat{y}_k$ y de ordenadas los residuos e_k, indíquese cualitativamente cómo se espera que sea la relación entre $\widehat{y}_k$ y e_k.

Solución.

El ajuste de mínimos cuadrados se obtiene a través de

$$\widehat{\beta} = \left(X^t X\right)^{-1} X^t Y$$

donde

$$X^t X = \begin{pmatrix} n & \sum_i x_i \\ \sum_i x_i & \sum_i x_i^2 \end{pmatrix} = \begin{pmatrix} 162 & 592,4786 \\ 592,4786 & 2469,245 \end{pmatrix}$$

cuya inversa es

$$\left(X^t X\right)^{-1} = \begin{pmatrix} 0,05 & -0,012 \\ -0,012 & 0,0033 \end{pmatrix}.$$

Por otra parte,

$$X^t Y = \begin{pmatrix} \sum_i y_i \\ \sum_i x_i y_i \end{pmatrix} = \begin{pmatrix} 2469,627 \\ 11111,01 \end{pmatrix}$$

de esta forma:

$$\widehat{\beta} = \begin{pmatrix} 0,05 & -0,012 \\ -0,012 & 0,0033 \end{pmatrix} \begin{pmatrix} 2469,627 \\ 11111,01 \end{pmatrix} = \begin{pmatrix} -9,85 \\ 7,03 \end{pmatrix}.$$

De modo que la ecuación de la recta ajustada es

$$\widehat{y} = -9,95 + 7,03x.$$

La varianza residual está formada por los valores

$$e = Y - \widehat{Y} = Y - X\hat{\beta} = Y - X(X'X)^{-1}X'Y = Y - HY = (I - H)\,Y \Rightarrow$$

$$e^t e = \sum_{i=1}^{n} e_i^2 = Y^t\,(I-H)^t\,(I-H)\,Y = Y^t\,(I-H)\,Y$$

puesto que la matriz $(I-H)$ es idempotente, es decir, $(I-H)^2 = (I-H)$.

Así, $e^t e = 759,0528$ por lo que la varianza residual es

$$\widehat{S}_r^2 = \frac{e^t e}{n-k-1} = \frac{\sum_{i=1}^{n} e_i^2}{n-k-1} = \frac{759,0528}{162-1-1} = 4,74$$

Para estimar el modelo $y = \beta_0 + \beta_1 x + \beta_2 x^2$ se utiliza el cambio de variable $z = x^2$ obteniendo $y = \beta_0 + \beta_1 x + \beta_2 z$ por lo que la estimación mínimo cuadrática del modelo es

$$\widehat{\beta} = \left(\begin{pmatrix} 1 & 1 & \cdots & 1 \\ X_1 & X_2 & \cdots & X_n \\ \vdots & \vdots & \ddots & \vdots \\ Z_1 & Z_2 & \cdots & Z_n \end{pmatrix} \begin{pmatrix} 1 & X_1 & \cdots & Z_1 \\ 1 & X_2 & \cdots & Z_2 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_n & \cdots & Z_n \end{pmatrix} \right)^{-1} \begin{pmatrix} 1 & 1 & \cdots & 1 \\ X_1 & X_2 & \cdots & X_n \\ \vdots & \vdots & \ddots & \vdots \\ Z_1 & Z_2 & \cdots & Z_n \end{pmatrix} \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} =$$

$$= \begin{pmatrix} n & \sum_i x_i & \sum_i z_i \\ \sum_i x_i & \sum_i x_i^2 & \sum_i x_i z_i \\ \sum_i z_i & \sum_i z_i x_i & \sum_i z_i^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum_i y_i \\ \sum_i x_i y_i \\ \sum_i z_i y_i \end{pmatrix} =$$

$$= \begin{pmatrix} n & \sum_i x_i & \sum_i x_i^2 \\ \sum_i x_i & \sum_i x_i^2 & \sum_i x_i^3 \\ \sum_i x_i^2 & \sum_i x_i^3 & \sum_i x_i^4 \end{pmatrix}^{-1} \begin{pmatrix} \sum_i y_i \\ \sum_i x_i y_i \\ \sum_i x_i^2 y_i \end{pmatrix}.$$

La varianza del ajuste cuadrático será

$$\widehat{S}_r^2 = \frac{e^t e}{n-k-1} = \frac{\sum_{i=1}^n e_i^2}{n-k-1} = \frac{150,4276}{162-2-1} = 0,95$$

A partir de los valores de las varianzas residuales obtenidos, para argumentar qué modelo se ajusta mejor debe analizarse la varianza explicada. Siendo $R^2 = 1 - \frac{SR}{ST}$ denotamos por SR a la suma residual, ST es la suma total y SE la suma explicada. Así, para el primer modelo,

$$R_{y \cdot x}^2 = 0{,}9745407 = 1 - \frac{759,0528}{ST} \Rightarrow ST = 29814,56678 \Rightarrow$$

$SE = ST - SR = 29055,5$, al tener término independiente el modelo

$$R_{y \cdot xx^2}^2 = 0{,}9950063 = 1 - \frac{150,4276}{ST} \Rightarrow ST = 30123,47558 \Rightarrow SE = 29973,04798$$

nota:——————————————

Puede apreciarse que la suma total no es la misma en los dos casos, y esto no es correcto porque la variable dependiente es la misma, por lo que los datos en los que aparecen las matrices $H^{(i)}$ tienen alguna incongruencia y no son correctos.

Por este motivo calculamos las sumas residuales directamente utilizando los coeficientes de determinación de una forma sencilla. Primeramente se calcula la suma total de cuadrados:

$$ST = \sum_{i=1}^{n}(y_i - \overline{y})^2 = \sum_{i=1}^{n} y_i^2 - n \cdot \overline{y}^2 = 51455,63 - 162 \cdot \left(\frac{2469,262}{162}\right)^2 = 13818,25$$

y ahora se utiliza esta suma para cada caso en ambas regresiones:

$$SR_{y \cdot x} = (1 - R^2_{y \cdot x}) \cdot ST = (1 - 0,974507) \cdot 13818,25 = 352,268$$

$$SR_{y \cdot xx^2} = (1 - R^2_{y \cdot xx^2}) \cdot ST = (1 - 0,9950063) \cdot 13818,25 = 69,0042$$

De esta forma las varianzas residuales en ambos casos serían:

$$\widehat{S}^2_{r,y \cdot x} = \frac{SR_{y \cdot x}}{162 - 1 - 1} = \frac{352,268}{160} = 2,2017$$

$$\widehat{S}^2_{r,y \cdot xx^2} = \frac{SR_{y \cdot xx^2}}{162 - 2 - 1} = \frac{69,0042}{159} = 0,4399$$

Calculado en porcentaje respecto de la suma total (ST) se tiene que un 97,45 % de la varianza total se explica a través de este modelo de un parámetro y término independiente. Para el caso del segundo modelo, siguiendo un razomiento análogo se tiene que el porcentaje de varianza explicada es del 99,5 %, por tanto parece que el segundo modelo explica mejor la variable dependiente, aunque habría que valorar si ese 2 % de ganancia es lo suficientemente importante como para complicar el modelo, puesto que el principio de parsimonia es importante para plantear modelos coherentes y lo más sencillos posibles que faciliten la interpretación.

Para analizar si el ajuste de un modelo de regresión lineal con una variable explicativa es adecuado basta con representar un gráfico de dispersión de las dos variables en cuestión, pero si tenemos varias variables explicativas ya se complica la situación y lo que suele hacerse es graficar los residuos del modelo frente a las estimaciones que proporciona el modelo. Si en este gráfico los residuos no están situados en torno al cero, es decir, que se aprecia un comportamiento que no es aleatorio, sino sistemático, esto significa que la función de regresión no es lineal y que el ajuste no es correcto. Puede ocurrir que la correlación entre dos variables sea elevada y sin embargo la relación entre ambas sea fuertemente no lineal. En nuestro caso, al ser los coeficientes de determinación muy altos sugieren que el ajuste es muy bueno y sería esperable que los residuos e_k estuvieran situados en torno al cero con una distribución aleatoria sin que se apreciase ninguna pauta en ellos de crecimiento o decrecimiento.

Cuestión 6

Las empresas del sector informático de cierta región facturaron durante los años 2001, 2002 y 2003 las cantidades que se indican y a los precios que figuran en la tabla 7.8.

a) Construir, con base 2001, los índices de precios y cantidades de Laspeyres, Paasche y Fischer para el año 2002.

b) Calcular, con base 2001, el índice de valor para el año 2002 a partir de los índices anteriores.

c) Hallar la repercusión de los ordenadores portátiles en la variación del índice de precios de Laspeyres entre los años 2002 y 2003. Conocemos que la ponderación de los ordenadores portátiles es de 42,31 % y es constante en el tiempo.

Tabla 7.8: Facturación empresas sector informático

		Ordenadores de sobremesa		Ordenadores portátiles				
t	Año	p_{1t}	q_{1t}	p_{2t}	q_{2t}	$\sum_{i=1}^{2} p_{it}q_{i0}$	$\sum_{i=1}^{2} p_{i0}q_{it}$	$\sum_{i=1}^{2} p_{it}q_{it}$
0	2001	750	30	1100	15	39000	39000	39000
1	2002	805	31	1150	20	41400	45250	47955
3	2003	820	40	1175	25	42225	57500	62175

N.B.: p_{it} y q_{it} denotan precio (en euros) por unidad y cantidad vendida, respectivamente, del producto i en el período de tiempo t.

Solución.

$$L_{P2001}^{2002} = \frac{\sum_{i=1}^{2} p_{i02}q_{i01}}{\sum_{i=1}^{2} p_{i01}q_{i01}} \cdot 100 = \frac{41400}{39000} \cdot 100 = 106,15$$

$$L_{Q2001}^{2002} = \frac{\sum_{i=1}^{2} q_{i02}p_{i01}}{\sum_{i=1}^{2} q_{i01}p_{i01}} \cdot 100 = \frac{45250}{39000} \cdot 100 = 116,02$$

$$P_{P2001}^{2002} = \frac{\sum_{i=1}^{2} p_{i02}q_{i02}}{\sum_{i=1}^{2} p_{i01}q_{i02}} \cdot 100 = \frac{47955}{45250} \cdot 100 = 105,977$$

$$P_{Q2001}^{2002} = \frac{\sum_{i=1}^{2} q_{i02}p_{i02}}{\sum_{i=1}^{2} q_{i01}p_{i02}} \cdot 100 = \frac{47955}{41400} \cdot 100 = 115,83$$

$$F_{P2001}^{2002} = \sqrt{L_{P2001}^{2002} P_{P2001}^{2002}} = \sqrt{106,15 \cdot 105,97} \Rightarrow F_{P2001}^{2002} = 106,05$$

$$F_{Q2001}^{2002} = \sqrt{L_{Q2001}^{2002} P_{Q2001}^{2002}} = \sqrt{116,02 \cdot 115,83} \Rightarrow F_{Q2001}^{2002} = 115,92$$

A partir de los índices anteriores, el índice de valor para el año 2002 se puede calcular utilizando cualquiera de los siguientes productos:

$$IV_{2001}^{2002} = L_{P2001}^{2002} P_{Q2001}^{2002} = L_{Q2001}^{2002} P_{P2001}^{2002} = F_{P2001}^{2002} F_{Q2001}^{2002} = 122,9$$

Finalmente, se calcula la repercusión de los ordenadores portátiles en la variación del índice de precios de Laspeyres. Se tiene:

$$I_{portatiles\ base\ 2001}^{2002} = \frac{1150}{1100} \cdot 100 = 104,545$$

$$I^{2003}_{portatiles\ base\ 2001} = \frac{1175}{1100} \cdot 100 = 106,82$$

Por tanto, entre los años 2002 y 2003,$\Delta I^{2002-2003}_{portatiles} = 106,82 - 104,55 \simeq 2,27$ y así, la repercusión de este grupo será:

$$R_{portatiles} = \Delta I^{2002-2003}_{portatiles} w_{portatiles} = 2,27 \cdot 42,31 = 0,9615$$

Cuestión 7

Conociendo los siguientes saldos de una Balanza de Pagos (MBP6), en miles de millones de euros, recogidos en la tabla 7.9 calcular:

a) El saldo de la Cuenta de Bienes y Servicios.

b) El saldo de la Cuenta Corriente y explicar el significado económico de ese saldo.

c) El saldo de la Cuenta de Capital.

d) La Capacidad (+) Necesidad (-) de Financiación y su significado económico.

e) El saldo de la Cuenta Financiera.

Tabla 7.9: Saldos Balanza de Pagos (MBP6), en m.m. de euros.

Exportaciones de bienes FOB	166,0
Exportaciones de servicios	25,7
Importaciones de bienes FOB	162,9
Importaciones de servicios	6,4
Transferencias de capital	3,7
Adquisición / Enajenación de activos no financieros no producidos	0,3
Rentas primarias y secundarias	-16,7
Errores y omisiones	1,9

Solución.

El saldo de la Cuenta de Bienes y Servicios es el resultado de las transacciones de bienes y servicios, es decir:

$$\text{Saldo Cuenta de Bienes y Servicios} = \textit{Exportaciones de bienes FOB}+$$
$$+\textit{Exportaciones de servicios} - \textit{Importaciones de bienes FOB}-$$
$$-\textit{Importaciones de servicios} = 166,0 + 25,7 - 162,9 - 6,4 = 22,4 \text{ mm€}$$

El saldo de la Cuenta Corriente es el resultado de las transacciones de bienes, servicios, ingreso primario e ingreso secundario entre residentes y no residentes. El saldo de la cuenta corriente equivale a la brecha de ahorro-inversión de la economía.

$$\begin{aligned} \textit{Saldo Cuenta Corriente} &= \textit{saldo Cuenta de Bienes y Servicios} \\ &+ \textit{saldo Cuenta del Ingreso Primario} \\ &+ \textit{saldo Cuenta Ingreso Secundario} \\ &= 22,4 - 16,7 = 5,7 \text{ mm€} \end{aligned}$$

El saldo de la Cuenta de Capital es el resultado de las adquisiciones y disposiciones de activos no financieros no producidos y las transferencias de capital entre residentes y no residentes. Así:

$$\textit{Saldo Cuenta de Capital} = 0,3 + 3,7 = 4 \text{ mm€}$$

La Capacidad (+) Necesidad (-) de Financiación es la suma de los saldos de las cuentas corriente y de capital. Representa el préstamo (superávit) o endeudamiento (déficit) neto de la economía frente al resto del mundo. En nuestro caso veremos que la economía muestra una capacidad de financiación al ser este agregado positivo.

$$\text{CNF}=\textit{Saldo Cuenta Corriente}+\textit{Saldo Cuenta de Capital} = 5,7+4 = 9,7 \text{ mm€}$$

El saldo de la Cuenta Financiera muestra las adquisiciones y disposiciones netas de activos y pasivos financieros. Por definición, la balanza de pagos (BP) siempre está en equilibrio, siendo su saldo 0 pero en la práctica surgen desequilibrio por imperfecciones en los datos. Así:

$$\begin{aligned}\textit{Errores y omisiones} &= \textit{Saldo Cuenta Financiera}\\ &\quad - (\textit{Saldo Cuenta Corriente} + \textit{Saldo Cuenta de Capital}).\end{aligned}$$

Por tanto,

$$\textit{Saldo Cuenta Financiera} = 1,9 + 9,7 = 11,6 \text{ mm€}$$

Desde 2014, en las estadísticas que realiza el Banco de España sobre la balanza de pagos, los saldos de las subcuentas de la Balanza Financiera se calculan restando la entrada de capitales a la salida de capitales, es decir, VNA-VNP, lo que implica que la estructura de la balanza de pagos (BP) es: $BP = BCC + BK - BF +$ Errores y Omisiones $= 0$, o lo que es lo mismo, $BCC + BK = BF -$ Errores y Omisiones

Cuestión 8

A partir de la información recogida en la tabla 7.10 sobre una economía, calcule: a) El PIB a precios de mercado.

b) El Excedente de explotación / Renta mixta bruto.

c) El Producto interior neto a precios de mercado.

d) La Renta Nacional Bruta a precios de mercado.

e) La Renta Nacional Disponible Bruta a precios de mercado.

f) La capacidad o necesidad de financiación de esta economía.

Tabla 7.10: información sobre una economía

Importación de bienes y servicios(M)	3600
Exportación de bienes y servicios (X)	2900
Impuestos sobre productos	250
Formación bruta de capital (FBC)	2000
Consumo intermedio	11000
Gasto en consumo individual (CF individual)	6900
Gasto en consumo colectivo (CF colectivo)	3580
Consumo de capital fijo	150
Rentas de la propiedad recibidas del exterior	220
Rentas de la propiedad pagadas al exterior	230
Rentas de los asalariados recibidas del exterior	160
Rentas de los asalariados pagadas al exterior	180
Impuestos netos sobre producción e importaciones	900
Remuneraciones de asalariados	8500
Transferencias corrientes netas del exterior	35
Transferencias de capital netas del exterior	185

Solución.

$$\begin{aligned}
PIB(pm) =& GCF + FBC + X - M = GCF_{ind} + GCF_{col} + FBC + X - M = \\
=& 6900 + 3580 + 2000 + 2900 - 3600 = 11780 \\
PIB(pm) =& RA_{interior} + EBE/RM + \\
& + impuestos\ netos\ sobre\ producción\ e\ importaciones \Rightarrow \\
EBE/RM =& PIB(pm) - RA_{interior} - impuestos\ netos\ sobre\ prod\ e\ impor = \\
=& 11780 - 8500 - 900 = 2380 \\
PIN =& PIB - CCF = 11780 - 150 = 11630 \\
RNB =& PIB + rentas\ primarias\ a\ cobrar\ RM - \\
& rentas\ primarias\ a\ pagar\ RM = \\
=& 11780 + (220 + 160) - (230 + 180) = 11750 \\
RNBD =& RNB + transferencias\ corrientes\ a\ cobrar \\
& - transferencias\ corrientes\ a\ pagar = \\
=& 11750 + 35 = 11785 \\
Ahorro\ Bruto =& RNBD - GCF = 11785 - (6900 + 3580) = 1305 \\
VPNDHYTC =& Ahorro\ Bruto - CCF \\
& + trasferencias\ de\ capital\ a\ cobrar \\
& - trasferencias\ de\ capital\ a\ pagar = \\
=& 1305 - 150 + 185 = 1340 \\
CNF =& \ Capacidad\ (+)o\ necesidad\ (\text{-})de\ financiación = \\
& = VPNDHYTC - \\
& - (FBC - CCF + Adq\ de\ activos\ no\ financieros) = \\
& = 1340 - (2000 - 150 + 0) = -510
\end{aligned}$$

Cuestión 9

Se dispone de la tabla 7.11 de nacimientos abreviada y la tabla 7.12 de las cifras de población. Calcular, siempre que sea posible:

a) Edad media a la maternidad.

b) Tasa global de fecundidad.

c) Tasa global de natalidad.

d) Tasa de fecundidad específica para el grupo 15-19 años.

Tabla 7.11: Nacimientos. Año 2013.

Total	425.494
De 15 a 19 años	8.816
De 20 a 24 años	32.251
De 25 a 29 años	78.929
De 30 a 34 años	155.810
De 35 a 39 años	120.720
De 40 a 44 años	27.313
De 45 a 49 años	1.655

Tabla 7.12: Población residente. Mujeres.

Edad	1 de enero de 2013	1 de enero de 2014
De 15 a 19 años	1.051.130	1.038.329
De 20 a 24 años	1.201.940	1.166.856
De 25 a 29 años	1.443.413	1.372.353
De 30 a 34 años	1.811.757	1.711.917
De 35 a 39 años	1.983.801	1.970.685
De 40 a 44 años	1.886.012	1.890.390
De 45 a 49 años	1.818.983	1.829.884

Solución.

La edad media a la maternidad se calcularía utilizando la fórmula:

$$EMM^t = \frac{\sum_x \left(x + \frac{n}{2}\right) tef_x}{ICF}$$

donde tef_x es la tasa específica de fecundidad por edad expresada en tanto por uno. Se define la tasa específica como $tef_x = \frac{\textit{nacimientos de madre de edad } x}{\textit{población media de edad } x} \cdot 1000$, siendo x es el extremo inferior del intervalo de edad y n es la amplitud del intervalo. Por otra parte, en el denominador, ICF es el índice sintético de fecundidad y se calcula como la suma de las tasas específicas de fecundidad en tanto por uno, es decir, $ICF = \sum_x tef_x$. En primer lugar se calcula la población media para cada grupo de edad como la media de la población a 1 de enero de 2013 y la población a 1 de enero de 2014. A continución, se calculan las tasas específicas. Los cálculos para el primer grupo de edad seríam

$$\textit{Población media} = \frac{1051130 + 1038329}{2} = 1044729,5$$
$$tef_x = \frac{8816}{1044729,5} \cdot 1000 = 8,44‰.$$

Procediendo de igual forma para cada grupo de edad se tiene:

$$\sum_x \left(x+\frac{n}{2}\right) tef_x = \left(15+\frac{5}{2}\right) 0,00844 + \cdots \left(45+\frac{5}{2}\right) 0,00091 = 8,1236$$

Por otra parte, el índice sintético de fecundidad sería:

$$ICF = 0,00844 + 0,02723 + \cdots + 0,00091 = 0,2566$$

Así, la edad media a la maternidad es

$$EMM^t = \frac{8,1236}{0,2566} = 31,6595 \text{ años.}$$

Conviene considerar un número significativo de decimales en este tipo de operaciones para ser más exactos en el cálculo de la edad media, puesto que sino se subestima la misma.

La tasa global de fecundidad se calcula como cociente entre el total de nacimientos y la población media total de mujeres en edad fecunda.

$$TGF^{2013} = \frac{N^{2013}}{M^{2013}_{15-49}} 1000$$

siendo N^{2013} : 425494 y $M^{2013}_{15-49}(i)$ la suma de las poblaciones de cada uno de los grupos de edades en ambos momentos y dividida por dos, de donde se obtiene que la población media del periodo es 11088725. Por tanto:

$$TGF^{2013} = \frac{425494}{11088725} 1000 \simeq 38,37‰$$

Para el calculo de la tasa global de natalidad no hay datos suficientes pues haría falta conocer la población media total (tanto masculina como femenina).

$$TBN^t = \frac{N^t}{P^t} 1000$$

La tasa de fecundidad específica para el grupo 15-19 años es

$$tef_{15-19} = \frac{8816}{\frac{1051130+1038329}{2}} 1000 \simeq 8,439‰$$

Cuestión 10

El número de nacimientos de varones en una determinada región en 2013 fue de 42.104. Además, se tienen los datos de defunciones y población de varones recogidos en la tabla 7.13. Rellene todos los datos del extracto 7.14 de tabla de mortalidad abreviada. Suponer que las defunciones se reparten de forma homogénea, es decir, $a_x = 0,5$ para todas las edades.

Tabla 7.13: Defunciones y población de varones.

	Población		Defunciones
	Varones		Varones
Edad	01-ene-2013	01-ene-2014	Año 2013
0 años	42.634	41.539	137
1-4 años	191.260	184.086	27
5-9 años	249.865	252.464	28
10-14 años	230.342	233.090	17
15-19 años	226.043	224.299	51
20-24 años	253.177	248.243	99

Tabla 7.14: Extracto de tabla de mortalidad abreviada

Edad	Supervivientes a edad exacta x	Defunciones teóricas con edad (x,x+n)	Riesgo o probabilidad de muerte con edad (x,x+n)	Tasas específicas de mortalidad m(x,x+n)	Población estacionaria ${}_n\mathbf{L}_x$
0					
1					
5					
10					
15					
20					

Solución.

En primer lugar se calculan las tasas de mortalidad específicas:

$$m_x = \frac{defunciones_x}{\frac{P^{01-01-13}+P^{01-01-14}}{2}} \cdot 1000$$

que se suelen especificar en tanto por mil.

El dato del problema relativo a los nacimientos de varones se podría utilizar para calcular la tasa de mortalidad infantil (TMI), es decir, dividir las defunciones de 0 años entre los nacidos en ese año. Esto daría:

TMI=$\frac{137}{42104} \cdot 1000 = 3,25385‰$, que es ligeramente menor que utilizar el criterio de la población media. Esta tasa se podría utilizar en los sucesivos cálculos de la tabla, aunque las variaciones serían muy pequeñas entre ambos métodos. Algunos autores utilizan esta TMI como la tasa en el año 0 que hay que utilizar, y en el INE se corrigen las tasas de mortalidad con la información demográfica disponible, por lo que vamos a utilizar este dato en esta primera

Tabla 7.15: Resultados de la tabla de mortalidad.

	n	l(x)	$_nd_x$	$_nq_x(‰)$	$_nm_x(‰)$	$_nL_x$
0	1	100000	324,8562452	3,24856245	3,253847615	99837,57188
1	4	99675,14375	57,34345899	0,57530350	0,143867258	398585,8881
5	5	99617,8003	55,51185065	0,55724831	0,111480723	497950,2219
10	5	99562,28845	36,5155734	0,36676109	0,073365672	497720,1533
15	5	99525,77287	112,6464317	1,13183177	0,226494531	497347,2483
20	5	99413,12644	196,0869772	1,97244553	0,394878545	495594,9750

tasa, y en las demás ya operamos con la población media.

A continuación, utilizando

$$_nq_x = \frac{n \cdot m_x}{1 + n(1 - a_x)m_x} = \frac{2 \cdot n \cdot m_x}{2 + n \cdot m_x}, \text{ considerando } a_x = 0,5$$

se puede obtener la probabilidad de muerte, utilizando las tasas en tanto por uno en el cálculo. Sabiendo que $l_x - l_{x+1} = d_x$ y siendo $q_x = \frac{d_x}{l_x}$ se tiene que $l_{x+1} = l_x - d_x = l_x - q_x l_x = l_x (1 - q_x)$. Finalmente

$$_nL_x = n \left(l_{x+1} + a_x d_x \right).$$

Los resultados se muestran en la tabla 7.15, en la que hay que puntualizar que las tasas están en tanto por mil, y las probabilidades de morir también.

Escribimos en la tabla las series $m(x)$ y $q(x)$ en tanto por mil, que es como suele aparecer en las tablas de mortalidad que publica el INE.

Bibliografía

Agresti, A. (1990). Categorical Data Analysis. New York: John Wiley.

Agresti, A. (2015). Foundations of Linear and Generalized Linear Models. John Wiley.

Ash, R.B. (1970). Basic Probability Theory. John Wiley.

Aznar, G,A., García, A, y Martín, A.(1994). Ejercicios de econometría I y II. Pirámide

Casella, G. y Berger, R.L. (1992). Statistical Inference (2nd ed.). Duxbury Advanced Series.

Castro, J. (2009). Statistical disclosure control in tabular data (Report DR 2009-11), Universitat Politécnica de Catalunya.

Celestino Rey, F. (2010). Historia de los cuerpos especiales de estadística de la Administración General del Estado, 1860-2010. Visión Libros.

Cuadras, C.M. (2002). Problemas de Probabilidad y Estadística. PPU Barcelona.

Cuadras, C.M. (2012). Nuevos métodos de análisis multivariante.

García, A., Quesada, V. (1998). Lecciones de Cálculo de Probabilidades. Díaz de Santos.

Feller, W. (1966). An Introduction to Probability Theory and Its Applications. Vol. 1. John Wiley.

Feller, W. (1966). An Introduction to Probability Theory and Its Applications. Vol.2. John Wiley.

Fernández, C.C., y Fuentes, F. (1995). Curso de Estadística Descriptiva. Teoría y Práctica. Madrid: Ariel.

Fernández, F.R. y Mayor Gallego, J. A. (1995). Ejercicios y prácticas de muestreo en poblaciones finitas. EUB. Ediciones Universitarias de Barcelona.

Gibbons, J. and S. Chakraborti (2011). Nonparametric Statistical Inference. Chapman and Hall/CRC; 5ª ed.

Hundepool, A., Domingo-Ferrer, J., Franconi, L., Giessing, S., Lenz, R., Nordholt, E. S., Seri, G. and De Woldf, P. P. (2010). Handbook on statistical disclosure control. ESSNet SDC, Eurostat.

Ibarrola, P., Pardo, L., Quesada, V. (1997). Teoría de la Probabilidad. Síntesis.

Leguina, J. (1989). Fundamentos de Demografía.

Keppel, G. y Zedeck, S. (1989). Data Analysis for Research. Analysis of variance and multiple regression/ correlation approaches. New York: Freeman and Company.

Lequiller, F. y Derek Blades, (2014). Understanding National Accounts, OECD.

Livi Bacci, M. (1993). Introducción a la demografía, Ariel.

Martín-Guzmán, P. y Martín Pliego, F. (1985). Curso Básico de Estadística Económica. Madrid: AC.

Martín Pliego, F.J. (1990). Curso práctico de estadística económica.

Matilla, M; Pérez, P y Sanz, B. (2017). Econometría y Predicción. McGraw-Hill .

Merediz, A. (2004). Historia de la estadística oficial como institución pública en España.

Montero, J., Pardo, L, Morales, D., Quesada, V. (1988). Ejercicios y Problemas de Cálculo de Probabilidades. Díaz de Santos.

Peña, D. (1987). Estadística. Modelos y métodos 1: Fundamentos. Madrid: Alianza Universidad.

Peña, D. (1987). Estadística. Modelos y métodos 2. Modelos Lineales y Series Temporales. Madrid: Alianza Universidad.

Peña, D. (2002). Analisis de datos multivariantes. McGraw-Hill Interamericana de España/UNED.

Pérez L. C. (2005). Muestreo estadístico: conceptos y problemás resueltos. Alhambra.

Pérez L. C. (2014). Modelos Econometricos Con Datos de Panel: Conceptos y Ejercicios Resueltos. Createspace.

Pfeffermann, D. y Rao, C.R. (2009). Handbook od statisitics. Vol. 29A Sample surveys: design, methods and applications, North-Holland.

Pressat, Roland (1987). “Diccionario de la Demografía”. Oikos-Tau.

Pressat, Roland (1983). “El análisis demográfico”. Fondo de cultura económica.

Ruíz Maya, L. y Martín Pliego, F.J. (2004). Fundamentos de inferencia estadística Ed. Thompson-Paraninfo.

Särndal, C.-E, Swensson, B. y Wretman, J. (1992). Model assisted survey sampling. Springer.

Särndal, C.-E, y Lündstrom, S. (2005). Estimation in surveys with nonresponse. Wiley.

Schattschneider, D., Escher, M. C. (2004). M.C. Escher: Visions of symmetry. New York: H.N. Abrams.

Sistema Europeo de Cuentas, SEC 2010, (2013). EUROSTAT (Comisión Europea).

Statistics Canada (2010). Survey methods and practices. Statistics Canada.

Uriel, E. y Peiró, A. (2000. Introducción al análisis de series temporales. Madrid: AC.

Valliant, R., Dorfmann, A.H. y Royall, R.M. (2000). Finite population sampling and inference: a prediction approach. Wiley.

Vélez, R., Hernández, V. (1995). Cálculo de Probabilidades 1. UNED

Vélez, R., y Garcia, A. (2012). Principios de Inferencia Estadistica. UNED.

Vinuesa, J. (1994). Demografia, análisis y proyecciones (Espacios y sociedades) .

Vinuesa y Puga (2007). Técnicas y ejercicios de demografía.

Wooldridge, J.B. (2001). Econometric Analysis of Cross-Section and Panel Data. MIT Press.

Wooldridge, J.B. (2010). Introducción a la econometría. Un enfoque moderno. Cengage Learning.

Apéndices

Apéndice A

Exámenes de problemas de 2019

A.1. Segundo examen del Cuerpo Superior de Estadísticos del Estado

EJERCICIO 2

CUESTIÓN 1:

Sea X una variable aleatoria con ley de probabilidad definida como

$$P_\theta(x) = (\frac{\theta}{2})^{|x|}(1-\theta)^{1-|x|}, \quad x = -1, 0, 1; \quad 0 \leq \theta \leq 1$$

Y sea X_1, X_2 ,..., X_n una muestra aleatoria simple de ella. Resolver razonadamente las siguientes cuestiones:

a) Encontrar el estimador de máxima verosimilitud para θ y obtener la estimación que tendríamos con él si para 20 observaciones de X se obtienen frecuencias $n_1 = 7$, $n_2 = 5$, $n_3 = 8$ para los valores -1, 0 y 1, respectivamente.

b) Encontrar el estimador para θ por el método de los momentos y analizar si es insesgado. Obtener la estimación de θ que propone este estimador para la muestra descrita en el apartado a).

c) Obtener la varianza para ambos estimadores y discutir, para aquel que sea insesgado, si es de mínima varianza.

CUESTIÓN 2:

En un laboratorio se llevan a cabo ensayos clínicos para comprobar la eficacia de algunos fármacos sobre el control del peso y se sospecha que hay dos de ellos que son equivalentes en cuanto a los efectos que producen. Supongamos que se disponen de las siguientes muestras de pesos medidos en libras.

	Participantes								
	1	2	3	4	5	6	7	8	9
Muestra 1	132	139	126	114	122	132	142	119	126
Muestra 2	124	141	118	116	114	132	145	123	121

Resolver razonadamente si los datos nos sugieren rechazar dicha sospecha a un nivel $\alpha = 0,01$ en las siguientes situaciones:

a) Las muestras 1 y 2, corresponden al peso de las mismas 9 personas adultas, después de administrarle a cada una de ellas los 2 tratamientos. Para ello, se ha diseñado el ensayo de forma que el participante i-ésimo estuviera en idénticas condiciones clínicas (con un peso de partida p_i idéntico) cuando se le administró uno y otro tratamiento.

b) Suponiendo que las muestras 1 y 2 corresponden a 18 personas en una misma situación clínica de partida (con un mismo peso de partida p). Se les divide al azar en dos grupos de 9 de personas cada uno de ellos, y a cada grupo se les administra uno y solo uno de los dos fármaco.

c) Interpretar las discrepancias observadas entre las metodologías de resolución de los apartados a) y b), y de cómo se concretan con las muestras analizadas.

Si fuera necesario, admitir normalidad para la variable peso y homocedasticidad para la variabilidad que presenta por grupo.

1

EJERCICIO 2

CUESTIÓN 3:

A) En una población con 5.000 viviendas, determinar el tamaño muestral necesario para que, con un nivel de confianza del 95 % ($F^{-1}_{N(0,1)}(1-\frac{0{,}05}{2})=1,96$), la estimación de la proporción de viviendas en alquiler no difiera en más del $0,1$ del valor verdadero. El muestreo se realiza sin reposición. No tenemos información sobre la proporción poblacional P de viviendas en alquiler y por tanto se supone que $P=1/2$.

B) La población de las 5.000 viviendas se estratifica en 3 grupos. Repartir el tamaño muestral obtenido en el apartado A) entre los estratos aplicando una afijación de mínima varianza o de Neyman y aplicando una afijación proporcional. Para ello se conoce el número de viviendas por estrato (N_h) y la proporción de viviendas de alquiler por estrato (P_h). Comentar los resultados.

N_h	P_h
2500	0,8
2500	0,5
500	0,1

CUESTIÓN 4:

Sea una población de 500 empresas con un total de 6.000 de asalariados. Se toma una muestra aleatoria simple sin reemplazamiento de 25 empresas con el objetivo de estimar el total de la cifra de negocios. Se denota por y_i la cifra de negocios de la empresa i, en miles de euros y x_i el número de asalariados de la empresa i. Los datos disponibles son:

$\sum_{i=1}^{25} x_i$=265; $\sum_{i=1}^{25} y_i$=710; $\sum_{i=1}^{25} x_i y_i$=9.772; $\sum_{i=1}^{25} x_i^2$=4.147; $\sum_{i=1}^{25} y_i^2$=25.702

Se pide:

a) Estimar el total de la cifra de negocios y su error de muestreo usando el método de la razón.

b) Estimar el total de la cifra de negocios y su error de muestreo usando el método de regresión lineal.

c) Comentar cuál de los dos métodos es más eficiente.

2

EJERCICIO 2

CUESTIÓN 5:

Se dispone de la siguiente información respecto a los agregados de una economía (datos en miles de millones de euros):

Producción de bienes y servicios (a precios básicos)	P.1	2.038
Consumo intermedio (a precios de adquisición)	P.2	1.048
Variación de existencias	P. 52	5
Formación bruta de capital fijo	P.51 g	249
Gasto en consumo final de los hogares	P.3	608
Gasto en consumo final de las ISFLSH*	P.3	11
Remuneración de los asalariados	D.1	541
Adquisiciones menos cesiones de objetos valiosos	P. 53	1
Exportaciones de bienes y servicios	P.6	276
Importaciones de bienes y servicios	P.7	290
Excedente bruto de explotación y renta mixta	B2G+B3G	446
Impuestos sobre producción e importaciones	D.2	112
Subvenciones	D.3	18
Impuestos sobre los productos	D.21	97

*Instituciones sin Fines de Lucro al Servicio de los Hogares

Se pide:

1. Calcule el PIB a precios de mercado (B1*G).

2. Calcule el valor de las subvenciones a los productos (D.31).

3. Calcule la formación bruta de capital (P.5).

4. Calcule el valor del gasto en consumo final de las Administraciones Públicas (P.3).

3

EJERCICIO 2

CUESTIÓN 6:

Registre las transacciones descritas a continuación, correspondientes al hogar de los Martínez durante el año de referencia t, en la secuencia de cuentas del sector hogares. Deberá comenzar por la cuenta de producción e ir avanzando especificando la cuenta de explotación, la cuenta de asignación de la renta primaria, la cuenta de distribución secundaria de la renta hasta finalizar con la cuenta de utilización de la renta disponible.

El hogar de los Martínez está compuesto por la pareja formada por Juan y María y sus hijos, Carlota y Miguel.

Durante el año t:

María ha recibido un salario de 2.000 unidades monetarias (u.m). Su empleador ha pagado 20 u.m. en concepto de cotizaciones sociales. María ha pagado 25 u.m. de impuesto sobre la renta y 15 u.m. de cotización social. Ha gastado 100 u.m. en transporte, y comidas fuera del hogar. También se ha comprado una moto por 280 u.m. que ha financiado con un préstamo por el que paga de interés 5 u.m. durante todo el año. El resto de su sueldo se lo cede a Juan que ha sido el responsable este año de llevar las cuentas en el hogar.

Juan ha estado desempleado recibiendo durante el año una prestación por desempleo de 350 u.m. Ha gastado 1100 u.m. en alimentos, bebidas, artículos de droguería y vestido y calzado para toda la familia y 900 u.m. en alquiler de la vivienda.

Carlota ha recibido pagas de sus padres por valor de 30 u.m. que ahorra para poder comprarse una bicicleta en el futuro.

Miguel no tiene empleo fijo pero hace ocasionalmente trabajos de pintura de pequeña entidad para los vecinos, que no declara. Ha cobrado por su producción 800 u.m. pero ha gastado 300 u.m. en pintura y brochas. Ocasionalmente ha llamado a un amigo para que le ayude y le ha pagado 40 u.m. Por otra parte, ha gastado 200 u.m. en bebidas, tabaco y tickets para eventos deportivos.

CUESTIÓN 7:

Usando datos de distintas Comunidades Autónomas, un investigador desea examinar la relación entre la renta media pagada (RENT) por alquiler en función de los precios medianos de la vivienda (MDHOUSE en miles de euros). Usamos una variable de control adicional que es el porcentaje de la población de la comunidad autónoma que vive en una zona urbana (PCTURBAN). Las columnas de la primera fila de esta tabla indican cuáles son las variables "dependientes" en cada regresión; C hace referencia a la denominada constante de regresión. Errores estándar en paréntesis.

	(1) $RENT$	(2) $MDHOUSE$	(3) $MDHOUSE$	(4) $RENT$	(5) $RENT$	(6) $EHAT$
C	125,9 (14,19)	−18,67 (12,00)	7,225 (8,936)	120,7 (12,43)	120,7 (15,71)	−62,85 (26,95)
$PCTURBAN$	0,525 (0,249)	0,182 (0,115)	0,616 (0,131)	0,0815 (0,244)	0,0815 (0,305)	−0,283 (0,258)
$MDHOUSE$	1,521 (0,228)			2,240 (0,268)	2,240 (0,339)	
$FAMINC$		2,731 (0,682)				4,448 (1,532)
$REG2$		−5,095 (4,122)				−6,768 (9,262)
$REG3$		−1,778 (4,073)				4,847 (9,151)
$REG4$		13,41 (4,048)				−18,77 (9,096)
$VHAT$				−1,589 (0,398)		
N	50	50	50	50	50	50
R^2	0.669	0.691	0.317	0.754	0.599	0.226
SCR	20259.6	3767.6	8322.2	15054.0	24565.7	19019.9

a) Las estimaciones de mínimos cuadrados del modelo se encuentran en la columna (1). ¿Es factible que MDHOUSE, el precio mediano de las casas, sea endógeno en esta regresión? Justifique su respuesta

b) Un investigador considera cuatro instrumentos: ingreso mediano de la familia (FAMINC 1,000€) y región del país (REG1, REG2, REG3). A continuación considere las regresiones indicadas en los modelos de las columnas (2) y (3). A partir de aquí, indique cómo contrastar si los instrumentos son débiles.

c) En la columna (4) los residuos de mínimo cuadráticos (VHAT) de la regresión en la columna (2) se agregan como un regresor a la regresión básica. Las estimaciones estan obtenidas utilizando mínimos cuadrados. ¿Cuál es la utilidad de esta regresión? ¿Qué indica sobre los resultados en (1)?

d) En la columna (5) estan las estimaciones VI/ MC2E utilizando los instrumentos enumerados en la parte (b). ¿Qué diferencias observas entre estos resultados y los resultados de los mínimos cuadrados en la columna (1)? Observe que las estimaciones (aunque no los errores estándares) son las mismas en las columnas (4) y (5). ¿Es un error? Justifique sus respuestas.

e) En la columna (6) los residuos de la estimación en (5) son regresados sobre las variables mostradas. ¿Qué información está contenida en estos resultados?

CUESTIÓN 8:

La tabla siguiente incluye estimaciones del correlograma (fac) y del correlograma parcial (fap) de un conjunto de serie temporales. Las series A, B y C contienen 200 observaciones temporales, mientras que las restantes tienen 300 observaciones. En algunas series se ofrecen estimaciones de la "serie en Niveles" y de la "serie en Diferencias". Teniendo esto en cuenta conteste a las siguientes preguntas:

a) Identifique las series A, B, C, D y E basándose en las FAC y FAP estimadas (explicite los argumentos en los que se basa)

b) Estime, cuando sea posible, los parámetros de los modelos identificados (desarrolle explícitamente sus cálculos y en qué se basa) en el apartado a)

Tabla

Correlogramas (fac) y correlogramas parciales (fap) estimados

	Retardos												Error
	1	2	3	4	5	6	7	8	9	10	11	12	estandár
Serie A													
Correlación (fac)	0.13	-0.05	0.00	-0.01	0.08	0.04	-0.02	0.01	-0.05	-0.08	-0.02	0.06	0.07
Correlación parcial (fap)	0.13	-0.08	0.02	-0.02	0.09	0.01	-0.01	0.02	-0.05	-0.07	-0.01	0.07	0.07
Serie B													
Correlación	0.42	0.03	-0.03	-0.12	0.15	-0.11	-0.06	0.02	0.05	0.07	0.02	-0.10	0.07
Correlación parcial	0.42	-0.18	0.04	-0.14	-0.05	-0.05	-0.01	0.04	-0.00	0.04	-0.06	-0.09	0.07
Serie C													
Correlación	0.58	0.35	0.21	0.13	0.02	-0.03	0.05	0.04	0.08	0.05	-0.08	-0.06	0.07
Correlación parcial	0.58	0.03	-0.01	0.01	-0.09	-0.03	-0.03	0.08	0.04	0.01	-0.12	0.01	0.07
Serie D													
Niveles correlación	0.90	0.84	0.78	0.71	0.66	0.61	0.57	0.53	0.49	0.45	0.42	0.37	0.06
Diferencias	-0.23	0.03	-0.01	-0.01	-0.08	0.03	-0.04	-0.01	0.03	-0.06	0.08	-0.03	0.06
Correlación parcial (niveles)	0.90	0.19	-0.01	-0.03	-0.02	0.06	-0.02	0.01	-0.00	-0.04	0.03	-0.09	0.06
Diferencias	-0.23	-0.02	-0.00	-0.02	-0.10	-0.01	-0.04	-0.03	0.02	-0.06	0.06	-0.01	0.06
Serie E													
Niveles correlación	0.98	0.95	0.92	0.89	0.86	0.84	0.81	0.78	0.75	0.72	0.69	0.66	0.06
Diferencias	0.42	-0.01	0.04	0.07	0.02	-0.08	-0.08	0.05	0.13	0.02	-0.06	0.03	0.06
Correlación parcial (niveles)	0.98	-0.21	0.07	0.01	-0.05	-0.03	0.03	0.01	-0.10	-0.06	-0.01	0.01	0.06
Diferencias	0.42	-0.06	-0.04	-0.10	0.01	-0.03	0.05	0.03	0.02	-0.08	-0.03	-0.07	0.06

CUESTIÓN 9:

Dada la siguiente tabla de mortalidad:

Edad	Tasa de mortalidad*	Promedio de años vividos el último año de vida	Riesgo de muerte*	Supervivientes	Defunciones teóricas	Población estacionaria	Tiempo por vivir	Esperanza de vida
4 años	0,103	0,504	0,103	99697,657	10,262		7920017,651	79,440
5 años	0,096	0,413		99687,395	9,575	99681,776	7820325,081	
(...)								
8 años	0,060	0,518	0,060	99671,017	6,022	99668,117	7620968,585	76,461
9 años	0,080	0,480	0,080		7,970	99660,851	7521300,468	75,466
10 años	0,057	0,499	0,057	99657,025		99654,168	7421639,617	74,472

* Riesgos y tasas de mortalidad vienen referidos a 1.000 habitantes

Se pide calcular:

a) El número de supervivientes de 9 años.

b) Población estacionaria de 4 años.

c) Defunciones teóricas de 10 años.

d) Esperanza de vida a los 5 años.

e) Riesgo de muerte a los 5 años .

EJERCICIO 2

CUESTIÓN 10:

Considere la siguiente matriz de migración de las regiones A, B y C de un país en un año dado t.

	Región de residencia actual				
Región de nacimiento	A	B	C	Fuera de la región	Total
A	10.359	53	74	127	10.613
B	100	8.250	22	79	8.451
C	88	15	9.750	94	9.947
Fuera de la región	250	48	69	-	-
En el extranjero	303	99	108	-	-
Total	11.100	8.465	10.023	-	-

Calcular:

a) El índice de aloctonía de A.

b) La proporción de emigrantes de B.

c) La proporción de inmigrantes de C.

d) La tasa bruta de intercambio entre A y B.

e) La tasa neta de intercambio entre B y C.

f) El índice de compensación entre A y B.

A.2. Tercer examen del Cuerpo de Diplomados en Estadística del Estado

OPOSICIONES AL CDEE. CONVOCATORIA 2019. EJERCICIO 3

CUESTIÓN 1. Los jugadores A y B lanzan un dado cada uno: si el dado de A marca un valor mayor o igual que el de B, comienza a jugar A; en caso contrario, comienza B. El juego consiste en sacar una bola de una urna que contiene una bola roja, una verde y una negra.
Si comienza el jugador A y la bola es roja, gana el jugador A. Si la bola no es roja, se devuelve a la urna y saca una bola el jugador B, que gana si la bola es verde o negra; en caso contrario, gana A.
Si empieza el jugador B, y la bola es roja o verde, gana el jugador B; en caso contrario, gana A.

Se pide calcular las siguientes probabilidades, razonando la respuesta:

a) La probabilidad de que el jugador A inicie el juego. ¿Y de que lo inicie el jugador B?
b) La probabilidad de que el jugador A gane el juego. ¿Y de qué lo gane el jugador B?
c) Si el juego lo ha ganado el jugador B ¿Cuál es la probabilidad de que lo haya iniciado?

CUESTIÓN 2. Una muestra aleatoria simple con reemplazamiento de 900 familias es seleccionada en una ciudad formada por 24.000 familias obteniendo los siguientes resultados:

$$\sum_{i=1}^{900} x_i = 9.000 \quad y \quad \sum_{i=1}^{900} x_i^2 = 162.819$$

Donde x_i representa el dinero en la cuenta corriente bancaria de la familia *i* investigada (en miles de euros). Se desea estimar la cantidad media de dinero depositado en las cuentas corrientes bancarias:

a) Obtenga un intervalo de confianza al 95% para la cantidad que se desea estimar.
b) Se agrupan las familias de la ciudad en dos estratos de 4.000 familias de renta alta y 20.000 de renta baja. Además, se sabe que la cantidad media depositada por las familias de renta alta es nueve veces la de las rentas bajas y que las desviaciones típicas en cada estrato son el doble de las medias correspondientes. ¿cómo distribuiría la muestra de 900 familias entre los dos estratos de manera que el error de muestreo sea mínimo?

(Nota: Para un valor α=0,025, el valor de la distribución normal que deja esa probabilidad a su derecha es 1,96)

CUESTIÓN 3. Se realiza el experimento de lanzar una moneda tres veces y registrar el número de caras obtenidas. Al repetir 80 veces el experimento se ha obtenido 7 veces ninguna cara, 24 veces una cara, 35 veces dos caras y 14 veces tres caras. Si *p* es "la probabilidad de obtener una cara al lanzar la moneda", se pide lo siguiente:

a) Obtenga el estimador de *p* por el método de los momentos.
b) ¿Es insesgado el estimador obtenido en el apartado anterior?
c) Calcule la varianza del estimador obtenido en el apartado (a).
d) Plantee y resuelva el contraste adecuado para estudiar si la moneda está equilibrada.

(Nota: para un valor $\alpha = 0.05$, el valor crítico necesario de la distribución del estadístico es 7,81)

CUESTIÓN 4. En un proceso de fabricación se analizan dos variables cuantitativas X e Y, obteniéndose los siguientes resultados: (0, 2), (1, 6), (3, 14), (-1, -2) y (2, 10). Calcule:

a) Las distribuciones marginales.
b) La distribución de las frecuencias relativas de X para valores de Y>2,5
c) El coeficiente de correlación lineal de ambas variables.
d) Los valores de X e Y para las siguientes observaciones: (-3, y) y (x, 4)

OPOSICIONES AL CDEE EJERCICIO 3

CUESTIÓN 5. Un ayuntamiento proporcionó los siguientes datos relacionados con la edad de la población extranjera residente en su ciudad en el año 2017.

	Edad mínima L_{i-1}	Edad máxima L_i	Edad media X_i	nº extranjeros n_i	nº acum. N_i	amplitud interv. c_i	X_i*n_i	$X_i^2*n_i$
Menos de 16 años	0	16	8	930	930	16	7.440	59.520
De 16 a 24 años	16	25	20,5	985	1.915	9	20.192,5	413.946,25
De 25 a 44 años	25	45	35	3.840	5.755	20	134.400	4.704.000
De 45 a 64 años	45	65	55	5.345	11.100	20	293.975	16.168.625
De 65 años y más	65	90	77,5	720	11.820	25	55.800	4.324.500
Suma				11.820			511.808	25.670.591

Responda a los siguientes apartados, especificando en cada caso, la medida estadística propuesta, su cálculo y la interpretación del resultado:
a) ¿Cuál es la edad más frecuente entre los residentes extranjeros?
b) ¿Entre qué edades se encuentra la mitad central de los residentes extranjeros?
c) ¿Cuál es la edad máxima que tiene el 50% de los extranjeros más jóvenes?
d) ¿Cuál es la edad media de los residentes extranjeros? ¿Es representativa de todo el colectivo?
e) Si el coeficiente de asimetría de Fisher de la distribución es 0,367 ¿puede afirmar que la distribución de edades de los residentes extranjeros es simétrica?

CUESTIÓN 6. A partir de la información que aparece en la tabla adjunta, calcule, para el total de la economía:

1) La cuenta de bienes y servicios.
2) Las cuentas de producción y explotación.
3) La Renta Nacional Bruta.
4) El Ahorro Nacional Bruto.
5) La capacidad o necesidad de financiación del país. Interprete el resultado.

Gasto en consumo final	900
Producción	1.660
Formación bruta de capital	680
Consumos intermedios	240
Remuneración de asalariados (interior)	740
Remuneración de asalariados pagada al resto del mundo	40
Remuneración de asalariados recibida del resto del mundo	30
Rentas de la propiedad pagadas al resto del mundo	250
Rentas de la propiedad recibidas del resto del mundo	158
Impuestos netos sobre los productos	110
Otros impuestos sobre la producción	354
Otras subvenciones a la producción	116
Transferencias corrientes pagadas al resto del mundo	190
Transferencias corrientes recibidas del resto del mundo	70
Consumo de capital fijo	180
Transferencias de capital recibidas	28
Transferencias de capital pagadas	20
Transferencias sociales en especie	45
Importaciones de bienes y servicios	120
Exportaciones de bienes y servicios	70

2

CUESTIÓN 7. En una economía con tres ramas de actividad, en la que cada una elabora un solo tipo de producto, y en ausencia de impuestos, subvenciones y márgenes de distribución, se conocen los siguientes datos:

1) Datos de la utilización de la producción como consumo intermedio:

i) De la producción de la rama 1 se destinan 28 millones de euros (M. €) a consumo intermedio de la propia rama; 18 M. € se adquieren como consumo intermedio por la rama 2; y 23 M. € como consumo intermedio por la rama 3.

ii) De la producción de la rama 2 se destinan 15 M. € a consumo intermedio de la propia rama, 22 M. € a consumo intermedio de la rama 1 y 15 M. € a consumo intermedio de la rama 3.

iii) De la producción de la rama 3 se destinan 19 M. € de su producción a consumo intermedio de la propia rama, 20 M. € a consumo intermedio de la rama 1 y 28 M. € a consumo intermedio de la rama 2.

2) La utilización de factores productivos por cada rama es la siguiente: el factor trabajo utilizado ha sido remunerado con 70 M. € por la rama 1, 60 M. € por la rama 2 y 50 M. € por la rama 3; y el factor capital utilizado, ha sido remunerado con 12 M. € por la rama 1, 36 M. € por la rama 2 y 29 M. € por la rama 3.

3) Las importaciones, clasificadas de acuerdo con los tres grupos de productos que se elaboran en esa economía, han sido: 25 M. € del producto 1; 20 M. € del producto 2; y 15 M. € del producto 3. Las importaciones se destinan exclusivamente a demanda final.

Se pide:

a) Construir la tabla input-output simétrica.
b) Calcular la demanda final para cada producto.
c) Determinar el Valor Añadido Bruto de cada rama.
d) Calcular los coeficientes técnicos de la rama 1. Explique su significado económico.

CUESTIÓN 8. Se conocen las siguientes operaciones de la balanza de pagos de España, todas ellas realizadas en el mismo periodo (un año t):

1) Se importan textiles de China por 500 M. €. Se pagan 300 M. € al contado y por los 200 M. € restantes se recibe un crédito del vendedor.
2) Inversores no residentes adquieren en la bolsa acciones emitidas por empresas residentes por 370 M. €. Pagan 150 M. € al contado y el resto con un préstamo de un banco nacional a devolver durante el año siguiente.
3) Empresas residentes pagan dividendos a sus accionistas residentes en el extranjero por un importe de 275 M. €.
4) Hogares residentes reciben remesas procedentes del extranjero por un importe de 45 M. €.

Se pide lo siguiente:

a) Realice las correspondientes anotaciones contables en la Balanza de Pagos, indicando balanza y sub-balanza donde se registran.

b) Indique, como resultado del conjunto de todas las operaciones, cuál es la capacidad o necesidad de financiación para España y si se ha producido una pérdida o una ganancia de reservas.

CUESTIÓN 9. Se muestra en las siguientes tablas la población residente en España, junto con los nacimientos y defunciones producidas en las fechas que se especifican en las tablas:

Año	Población total a 1 de enero
2010	46.500.000
2015	46.000.000

Período	Nacimientos	Defunciones
1 de enero de 2010 a 1 de enero de 2015	2.200.000	2.000.000

A continuación, se proporciona la población y los migrantes por grupos de edad.

Grupo de edad	Población a 1 de enero de 2010	Población a 1 de enero de 2015	Inmigrantes procedentes del extranjero entre el 1 de enero de 2010 y de 2015	Emigrantes con destino al extranjero entre el 1 de enero de 2010 y de 2015
0-15	7.500.000	7.000.000	260.000	300.000
16-64	31.000.000	30.000.000	1.200.000	1.050.000
65-80	8.000.000	9.000.000	100.000	75.000

Calcule:

a) Saldo migratorio y tasa de migración neta para el quinquenio.
b) Índice de atracción para el grupo de edad 16-64.
c) Tasa de migración neta para el grupo de edad 0-15.
d) Edad media a la emigración.

CUESTIÓN 10. Complete todos los datos del siguiente extracto de tabla de mortalidad abreviada. (Es suficiente redondear a dos decimales, excepto en las tasas específicas de mortalidad).

Edad	Tasas específicas de mortalidad	Promedio de años vividos el último año de vida	Riesgo o probabilidad de muerte	Supervivientes		Población estacionaria	Tiempo por vivir	Esperanza de vida
0		0,123939	0,002645464	100.000				
1		0,481492	0,000223966					
5		0,500314	0,000155676				8.209.014,8	

Apéndice B

Exámenes de problemas de 2018

B.1. Segundo examen del Cuerpo Superior de Estadísticos del Estado

Ejercicio 2

CUESTIÓN 1: De una distribución $N(100, \sigma)$, se toman dos muestras aleatorias simples independientes entre sí de tamaño 4 y 5.

Muestra 1	Muestra 2
98	97,8
103,4	101,3
100,5	97,9
99,7	100,7
	100,3

En la muestra 1: $\bar{x} = 100,4$; $\sum_{i=1}^{4}(x_i - \bar{x})^2 = 15,26$

En la muestra 2: $\bar{y} = 99,6$; $\sum_{i=1}^{5}(y_i - \bar{y})^2 = 10,72$

Calcular la $P(\bar{X} - \bar{Y} \geq 2)$, donde $\bar{X}$ e $\bar{Y}$ son las medias muestrales, indicando la distribución muestral necesaria para su cálculo.

CUESTIÓN 2: Contrastar con un nivel de significación 2,5 % que la duración de un determinado tipo de bombillas eléctricas es una variable aleatoria con función de densidad:

$$f(x) = 1/\theta exp(-\frac{1}{\theta}x) \qquad x > 0$$

con $\theta = 200$ horas, teniendo en cuenta que en una muestra aleatoria de 75 bombillas probadas hasta fundirse, se han observado las siguientes duraciones:

Duración	Número de bombillas
Hasta 200 horas	40
De 200 a 300 horas	15
De 300 a 400 horas	8
De 400 a 500 horas	6
Más de 500 horas	6
	N=75

1

Ejercicio 2

CUESTIÓN 3: De una población de $N = 500$ hogares, se obtiene una muestra aleatoria simple (sin reposición) de tamaño $n = 50$. En cada hogar de la muestra se mide el gasto en alimentación (Y) y el gasto total (X). Los datos, en miles de euros, son los siguientes:

$$\sum_{i=1}^{50} Y_i = 213; \quad \sum_{i=1}^{50} X_i = 452; \quad \sum_{i=1}^{50} Y_i^2 = 964; \quad \sum_{i=1}^{50} X_i^2 = 4552; \quad \sum_{i=1}^{50} Y_i X_i = 2060$$

Se pide:
A) Estimar el porcentaje de gasto en alimentación y su error de muestreo.
B) Estimar el gasto total en alimentación usando el método de la razón y su error de muestreo. Para ello se conoce que el gasto total de la población es de 5000.

CUESTIÓN 4: Se quiere estimar la proporción de pinos que hay en una zona forestal. Para ello, se divide la zona en 20 conglomerados de tamaños diferentes $M_i, i = 1, ..., 20$ conociendo que el total de árboles es $M = \sum_{1=1}^{20} M_i = 1000$. Se utiliza un muestreo de conglomerados con submuestreo donde en ambas etapas el procedimiento de selección es con probabilidades iguales sin reposición. En la primera etapa se seleccionan 4 conglomerados y en la segunda etapa se aplica una fracción de muestreo de $f_{2i} = 10/M_i$. Los valores del tamaño de los conglomerados muestrales y el número de pinos obtenido en cada uno de ellos vienen en la siguiente tabla:

Número de árboles= M_i	Número de pinos
60	8
80	7
50	6
30	4

Se pide:
A) Una estimación insesgada de la proporción de pinos y su error de muestreo.
B) Una estimación de la proporción de pinos utilizando el estimador de la razón al tamaño y su error de muestreo.
C) Comentar las ventajas e inconvenientes del estimador B) respecto al A).

CUESTIÓN 5: Se dispone de la siguiente información respecto a los agregados de oferta, demanda y rentas de una economía en miles de millones de euros (entre guiones, el código de las operaciones y saldos en *SEC*2010):

Producción de bienes y servicios (a precios básicos) -P.1-	2.000
Formación bruta de capital -P.5-	250
Subvenciones sobre los productos -D.31-	8
Gasto en consumo final - P.3 -	900
Importaciones de bienes y servicios - P.7 -	360
Exportaciones de bienes y servicios -P. 6 -	400
Formación bruta de capital fijo -P.51.g-	230
Excedente de explotación bruto y Renta mixta bruta - B.2g + B.3.g -	500
Impuestos sobre la producción y las importaciones - D.2 -	130
Consumo intermedio (precio de adquisición) - P.2 -	1.000
Subvenciones -D.3 -	20

Se pide:
1. Calcule el valor añadido a precios básicos.
2. Calcule el PIB a precios de mercado (*PIBpm*).
3. Calcule el valor de la variación de existencias. Suponga para ello que las adquisiciones menos cesiones de objetos valiosos son nulas.
4. Calcule la remuneración de los asalariados

CUESTIÓN 6: El *PIBpm* de una determinada economía en el año t ha sido de 1,100 (miles de millones de euros). Se conoce también que la tasa de variación de este agregado a precios corrientes entre el año $t-1$ y t fue de $4,9\,\%$ mientras que la correspondiente tasa de variación anual en volumen fue del $3,4\,\%$.

Además de la cuenta del resto del mundo se tiene la siguiente información del año $t-1$ (en miles de millones de euros):

Remuneración de asalariados recibida del resto del mundo	20
Remuneración de asalariados pagada al resto del mundo	8
Impuestos sobre la producción y las importaciones pagados al resto del mundo	10
Subvenciones recibidas del resto del mundo	50
Rentas de la propiedad recibidas del resto del mundo	50
Rentas de la propiedad pagadas al resto del mundo	70

Se pide:
1. Calcule el *PIBpm* y la Renta Nacional Bruta (*RNB*) del año $t-1$.
2. Calcule la variación entre $t-1$ y t del deflactor implícito del *PIBpm* .
3. ¿Considera que según los principios recogidos en el *SEC 2010* tiene sentido el cálculo de la *RNB* en términos reales? Razone su respuesta.

CUESTIÓN 7: ¿Cuál es la relación entre crimen y castigo? Esta importante pregunta fue estudiada mediante un panel de datos de Carolina del Norte. Las secciones transversales son 90 condados, y los datos son anuales para los años 1981-1987. En estos modelos la tasa de delincuencia pretende ser explicada a partir de variables como el efecto disuasivo del sistema legal, los salarios en el sector privado (que representan el retorno a las actividades legales). Los autores comentan que puede haber heterogeneidad entre los condados (características no observadas de cada condado).

En este marco de analítico, considere un modelo en el cual la tasa de criminalidad (y) es una función de la probabilidad de detención (X_1), probabilidad de ser un convicto (preso) (X_2), la probabilidad de una pena de prisión (X_3, el promedio de las penas de prisión (X_4), y el salario semanal promedio en el sector manufacturero (X_5). Importante: en **todos los casos se utilizan los logaritmos de las variables**.

(1) Indique los signos esperados en un modelo regresión lineal múltiple

(2) Una analista de datos propone estimar el modelo indicado en (1) mediante mínimos cuadrados ordinarios. Los resultados son los siguientes (errores estándar en paréntesis)

$$\widehat{y} = \underset{(0{,}3654)}{-6{,}0861} - \underset{(0{,}0403)}{0{,}6566}X_1 - \underset{(0{,}0277)}{0{,}4466}X_2 + \underset{(0{,}0727}{0{,}2082}X_3 - \underset{(0{,}0606)}{0{,}0586}X_4 + \underset{(0{,}0619)}{0{,}2921}X_5$$

Analice los signos de los coeficientes estimados y su significación (al 95 % utilizando para ello la distribución normal). ¿Son como esperaba? Interprete el coeficiente de X_1.

(3) Otra económetra sin embargo considera oportuno estimar el modelo (1) usando un estimador de efectos fijos. El estimador de efectos fijos arroja es siguiente modelos estimado

$$\widehat{y} = \underset{(0{,}3236)}{-3{,}2288} - \underset{(0{,}0376)}{0{,}2313}X_1 - \underset{(0{,}0222)}{0{,}1378}X_2 - \underset{(0{,}0393)}{0{,}1431}X_3 + \underset{(0{,}0310)}{0{,}0183}X_4 - \underset{(0{,}0553)}{0{,}1666}X_5$$

¿Los coeficientes coinciden ahora con lo que esperaba? Interpreta el coeficiente en X_1 y comparalo con la estimación en (2). ¿Qué concluye sobre el efecto disuasivo de la probabilidad de arresto? Por último, interpreta el coeficiente en X_4. ¿Qué conclusión tienes sobre la gravedad del castigo como elemento disuasivo?

(4) Tras estimar los dos modelos, argumente cuáles han podido ser los motivos para que la económetra ha tenido para proponer un modelo de efectos fijos. Justifique cuál de los dos modelos le parece más adecuado para estimar la relación entre crimen y castigo.

(5) ¿Puede haber endogeneidad en el modelo? Justifique la respuesta en términos del modelo propuesto.

CUESTIÓN 8: Una variable macroeconómica Y_t se modelizada con

$$Y_t = 0,01 + 0,5\epsilon_{t-1} + 0,1\epsilon_{t-2} + \epsilon_t$$

donde $\{\epsilon_t\}$ es i.i.d. cono media cero y varianza σ^2.

(i) Calcula la esperanza y varianza incondicionada de Y_t

(ii) Calcula la autocovarianza de primer y segundo orden de Y_t

(iii) ¿Cómo es la autocovarianza para retardos superiores a 2?

(iv) A partir de la información anterior, ¿puede concluirse que el proceso es estacionario débil? Justifique la respuesta

(v) ¿Es un proceso invertible? Justifique la respuesta

(vi) ¿Cuál es la esperanza condicionada de Y_{t+1} dada toda la información disponible en el periodo t?

CUESTIÓN 9: Dada la siguiente tabla de mortalidad:

	Tasa de mortalidad	Promedio de años vividos el último año de vida	Riesgo de muerte	Supervivientes	Defunciones teóricas	Población estacionaria	Tiempo por vivir	Esperanza de vida
0 años		0,123939	2,645464					
1 año		0,481492	0,223966				8209014,8	

Se pide:
a) Calcular los supervivientes a 0 años.
b) Defunciones teóricas a 0 años.
c) Población estacionaria a 0 años.
d) Tasa de mortalidad a 0 años.
e) Calcular los supervivientes a 1 años.
f) Defunciones teóricas a 1 años.
g) Población estacionaria a 1 años.
h) Tasa de mortalidad a 1 años.
i) Esperanza de vida para las personas de 1 año.
j) Tiiempo por vivir para las personas de 0 años.
k) Esperanza de vida para las personas de 0 año.

CUESTIÓN 10: A partir de la información de la Tabla A, y sabiendo que la población total a 1 de enero de 2015 era de 46.450,0 miles de personas mientras que a 1 de enero de 2016 era de 46.440,0 miles de personas, se pide calcular:
a) La Tasa Bruta de Natalidad (TBN^{2015}).
b) La Tasa General o Global de Fecundidad (TGF^{2015}).
c) Las Tasas Específicas de Fecundidad por edad de la madre (TEF_x).
d) El Índice Sintético de Fecundidad (ISF^{2015}).
e) La edad media a la maternidad en 2015.

Grupos de edad	Mujeres residentes (miles)		Nacimientos por edad de las madres, año 2015 (miles)
	1.enero.2015	1.enero.2016	
15-19	1.045,3	1.060,2	8,2
20-24	1.137,8	1.117,0	29,8
25-29	1.318,5	1.279,8	75,2
30-34	1.627,1	1.549,0	148,8
35-39	1.938,5	1.897,6	125,5
40-44	1.907,2	1.926,2	30,6
45-49	1.829,7	1.838,4	2,0
Total 15-49	10.804,1	10.668,2	420,1

B.2. Tercer examen del Cuerpo de Diplomados en Estadística del Estado

OPOSICIONES AL CDEE. CONVOCATORIA 2018. EJERCICIO 3

CUESTIÓN 1. Sea (X_1, X_2,… X_n) una muestra aleatoria simple de una variable aleatoria X discreta que toma los valores {-1, 0, 1} con las siguientes probabilidades:

$$P(X=-1)=\frac{1-\theta}{2}, \quad P(X=0)=\frac{\theta+\lambda}{2}, \quad P(X=1)=\frac{1-\lambda}{2} \quad 0<\theta<1,\ 0<\lambda<1$$

Calcule, razonando la respuesta:

a) Los estimadores de θ y de λ por el método de los momentos.
b) ¿Es insesgado el estimador de θ obtenido en el apartado anterior?
c) Análogamente ¿es insesgado el estimador de λ ?

CUESTIÓN 2. Sea (X, Y) una variable aleatoria bidimensional con función de densidad:

$$f(x,y)=\begin{cases} c(x-y) & si\ 0<x<2\ ,\ -x<y<x \\ 0 & en\ el\ resto\ de\ los\ casos \end{cases}$$

Calcular:

a) El valor de *c* para que f(x,y) sea función de densidad.
b) Las funciones de densidad marginales de X e Y.
c) ¿Son X e Y variables aleatorias independientes? Razone la respuesta.

CUESTIÓN 3. En una población dividida en dos estratos, se desea investigar el valor medio de una determinada característica cuantitativa y para ello se propone un muestreo estratificado considerando una función del coste de los trabajos de campo del tipo:

$$\sum_{h=1}^{2} \boldsymbol{c_h n_h}$$

Donde: h indica el estrato; c_h el coste por unidad encuestada en el estrato *h;* n_h el número total de unidades encuestadas en el estrato *h.*

La información de la que se dispone de los estratos es la siguiente:

Estrato	$\boldsymbol{W_h}$	$\boldsymbol{S_h}$	$\boldsymbol{c_h}$
1	0,4	10	4 euros
2	0,6	20	9 euros

Donde W_h y S_h representan el peso y la raíz de la cuasivarianza poblacional de la característica estudiada, respectivamente, en el estrato *h*. Si se desea minimizar el coste de la encuesta para un valor dado de la varianza del estimador de la media, se pide:

a) Hallar la relación entre los tamaños muestrales en los estratos, n_1 *y* n_2.
b) Hallar el coste total de los trabajos de campo para un tamaño total de la muestra de *n=300*
c) Despreciando la fracción de muestreo en los estratos, calcular el error de muestreo para *n=300*

CUESTIÓN 4. La siguiente tabla presenta datos de edades sobre un conjunto de 24 individuos seleccionados de dos poblaciones distintas.

P1	38	10	10	60	20	38	5	40	40	90	40	40	40	50	50	50	50	10	60	70	80	80	90	40
P2	40	30	30	30	30	60	35	35	35	40	98	40	40	40	20	40	40	45	45	50	35	60	60	40

Se pide:
1) Representar gráficamente ambas distribuciones de datos.
2) Comparar los grados de dispersión y simetría de ambas distribuciones.

CUESTIÓN 5. En la siguiente tabla se recogen los datos de la evolución de los precios de un producto entre 2013 y 2017, así como el valor del IPC de esa economía. A partir de los mismos, se pide:
a) Obtener la serie del IPC para todo el periodo señalado, tomando como año base 2013=100
b) Calcular la serie de precios de venta del artículo, expresados en euros de 2013
c) Calcular las variaciones anuales experimentadas en el precio en términos corrientes del artículo y compare su evolución respecto a la variación de los precios en términos constantes

Año	Precio venta	IPC (media anual)
2013	114,23	100,859
2014	119,87	100,707
2015	125,03	100,203
2016	129,31	100,000
2017	135,12	101,956

CUESTIÓN 6. Con los siguientes datos relativos al sector Hogares, calcule la sucesión completa de cuentas del sector, además, razone las siguientes preguntas:
1) ¿Se puede calcular el Producto Interior Bruto de este sector?
2) En el sector Hogares ¿se puede distinguir, a nivel teórico, la renta mixta del excedente bruto de explotación?
3) ¿Qué significado económico tiene la operación D.8 – Ajuste por la variación de los derechos por pensiones?
4) ¿Cuál es el consumo final colectivo del sector?
5) ¿Qué significado económico tiene una capacidad de financiación negativa?

Datos del sector Hogares (S.14) *(Cifras en millones de euros)*

Producción	206
Formación bruta de capital fijo	80
Adquisiciones menos cesiones de activos no financieros no producidos	2
Consumos intermedios	80
Remuneración de asalariados pagada	38
Remuneración de asalariados recibida	530
Impuestos netos sobre la producción y las importaciones	6
Impuestos corrientes	86
Gasto en consumo final	632
Cotizaciones sociales netas	150
Prestaciones sociales	190
Transferencias corrientes netas	-8
Transferencias de capital netas	-2
Transferencias sociales en especie	134
Variación de existencias	2
Adquisiciones menos cesiones de objetos valiosos	1
Ajuste por la variación de los derechos por pensiones	-14
Rentas de la propiedad netas	40

2

CUESTIÓN 7. Se conoce la siguiente información del sector S12 "Instituciones financieras" extraída de la Contabilidad Nacional, (cifras en millones de euros)

Valor añadido bruto	37.798
Impuestos corrientes sobre los beneficios	4.798
Formación bruta de capital	5.514
Remuneración de asalariados	19.944
Cotizaciones sociales netas recibidas	7.514
Prestaciones sociales pagadas	8.879
Rentas de la propiedad recibidas	76.499
Rentas de la propiedad pagadas	61.076
Impuestos sobre la producción y las importaciones	3.826
Subvenciones a la producción recibidas	95
Transferencias de capital a cobrar	998
Transferencias de capital a pagar	1.651
Consumo de capital fijo	4.415
Transferencias corrientes recibidas	27.530
Transferencias corrientes pagadas	26.768
Adquisiciones menos cesiones de activos no producidos	-25

Se pide:
1) Defina lo que son los servicios de intermediación financiera medidos indirectamente (SIFMI) según el SEC-2010 y en qué rúbrica/s estarían incluidos en la cuenta de producción.
2) Calcule la cuenta de explotación y su saldo.
3) Calcule la cuenta de asignación de la renta primaria y su saldo.
4) Calcule la renta disponible bruta.
5) Sabiendo que el ahorro bruto del sector es 25.510 millones de euros, calcule el gasto el consumo final y la capacidad o necesidad de financiación del sector.

CUESTIÓN 8. Indique cómo se registraría cada una de las siguientes operaciones en la Balanza de Pagos española, comprobando que la balanza queda equilibrada. ¿Cuál sería la capacidad o necesidad de financiación del país?

a) El Gobierno español envía una ayuda de 100.000€ para la construcción de un hospital en un país en vías de desarrollo.

b) Inmigrantes residentes en España envían remesas a sus países de origen por importe de 20.000€.

c) Una fábrica textil española importa materias primas por valor de 30.000€ (FOB), pagando la mitad al contado y la otra mitad mediante un crédito a 4 años concedido por los proveedores no residentes.

d) Una empresa española de instrumentos musicales exporta guitarras por valor de 60.000€, de los cuales 3.000€ corresponden a los fletes y seguros prestados por empresas residentes y pagados al contado por la empresa importadora no residente.

e) Un grupo inversor chino compra una empresa española de videojuegos por valor de 400.000€.

f) Una compañía española compra el 5% de una empresa de telefonía móvil noruega, por importe de 80.000€.

3

CUESTIÓN 9. La siguiente tabla se refiere a la población de un país contabilizada a 1 de julio.

Edad	Población residente a 1 de Julio	Nacimientos por edad de la madre	Defunciones
0	551	0	4
1-9	5.461	0	3
10-19	5.187	22	1
20-29	5.945	230	0
30-39	6.376	135	2
40-49	5.880	18	9
50 y más	10.600	0	230

Calcule:

1. La tasa bruta de natalidad, sabiendo que 5 nacimientos fueron de madres no residentes, todas ellas de más de 30 años.
2. La tasa específica de mortalidad de menores de 1 año
3. La tasa de mortalidad infantil
4. La tasa específica de fecundidad de las mujeres de 20 a 29 años, dada una razón de masculinidad del 103 para ese intervalo

---

CUESTIÓN 10. En un país con dos regiones, A y B, se ha construido la siguiente matriz de migraciones (datos en número de personas), a partir de la información de los censos de 2001 y 2011.

Territorio de residencia en 2001	Territorio de residencia en 2011		
	Región A	Región B	Total (A+B)
Región A	5.400.000	300.000	5.700.000
Región B	250.000	7.500.000	7.750.000
En el extranjero	80.000	100.000	180.000
No aplicable (*)	130.000	180.000	310.000
Total	5.860.000	8.080.000	13.940.000

(*): "No aplicable" recoge la población nacida en el periodo intercensal.

A partir de esa matriz, calcule:

a) El número total de emigrantes durante el periodo 2001-2011 y el número total de inmigrantes durante el periodo 2001-2011.
b) La proporción de emigración y la tasa de emigración de la región B durante el periodo 2001-2011.
c) El índice de atracción de la región B durante el periodo 2001-2011.
d) El saldo migratorio y las tasas de migración bruta y neta de la región B durante el periodo.

4

Apéndice C

Exámenes de problemas de 2017

C.1. Cuarto examen del Cuerpo Superior de Estadísticos del Estado

CUESTIÓN 1: Dos muestras independientes, ambas de tamaño 7, de sendas distribuciones poblacionales normales de varianza común σ^2 arrojan medias muestrales de 4.8 y 5.4, y varianzas muestrales de 8.38 y 7.62, respectivamente. Encontrar la expresión que permitiría obtener un intervalo de confianza para $\mu_1 - \mu_2$ (diferencia de sus medias poblacionales) al nivel 0.95.

CUESTIÓN 2: Se supone que el número de erratas por página de un libro sigue una distribución de Poisson. Elegidas al azar 95 páginas se obtuvo que había 0, 1, 2, 3, 4, 5 en 40, 30, 15, 7, 1 y 0 páginas respectivamente. ¿Contiene la muestra evidencias estadísticamente significativas para rechazar dicho supuesto? Justifica la respuesta.

CUESTIÓN 3: Sea una población de 3.245 empresas dedicadas al sector de comercio estratificadas en 3 grupos de tamaño, según el número de asalariados. Los datos disponibles son los siguientes:

Estrato	N_h	X_h	S_h
1 (entre 0 y 9 asalariados)	3.200	3.400	1,5
2 (entre 10 y 49 asalariados)	40	650	10,2
1 (de 50 o más asalariados)	5	580	150

Donde:
N_h= Número de empresas en el estrato h
X_h= Total de asalariados en el estrato h
S_h= Cuasivarianza poblacional de X en el estrato h

El tamaño muestral es de 50 empresas y el muestreo es aleatorio simple sin reposición. Se pide:

a) Calcular el tamaño muestral por estrato, usando la afijación de mínima varianza o de Neyman.
b) Calcular la varianza y el coeficiente de variación para el estimador insesgado del total de asalariados.

CUESTIÓN 4: Se quiere conocer la superficie dedicada a la plantación de pinos de una región. Dicha región, con un total de 50.000 Km^2, se divide en 100 áreas o conglomerados. Se extrae una muestra de 5 conglomerados con reemplazamiento y con probabilidades proporcionales a sus superficies. Las proporciones de superficie dedicadas a la plantación de pinos, en cada uno de los conglomerados de la muestra, son las siguientes:

Conglomerado	1	2	3	4	5
Proporción	0,05	0,2	0,1	0,15	0,25

Se pide calcular un estimador insesgado de la superficie total dedicada a la plantación de pinos y su error de muestreo relativo o coeficiente de variación.

CUESTIÓN 5: En una determinada economía, para la rama de actividad A, se tiene la siguiente información:

	año t	año t+1
Producción (p.b.) a precios corrientes (millones de euros)	450.000	482.000
Consumos intermedios (p.a.) a precios corrientes (millones de euros)	275.000	292.000
Índice de precios de la producción de la rama (en %)	100	105
Índice de precios de los consumos intermedios de la rama (en %)	100	103

Por otra parte, el PIB (p.m.) a precios corrientes de esa economía en el año t fue de 1.118.000 millones de euros; la tasa de variación anual de este agregado a precios corrientes entre t y t+1 fue del 4,0 % mientras que su tasa de variación anual en volumen fue del 3,1 %. Se pide:

a) Calcule la tasa de variación del valor añadido (p.b) a precios corrientes de la rama de actividad A entre t y t+1. Obtenga asimismo la tasa de variación en volumen del valor añadido (p.b) entre t y t+1.
b) Calcule el valor del PIB (p.m) en t+1 a precios corrientes y la variación anual entre t y t+1 del deflactor implícito del PIB (p.m)
Nota: p.b.(precio básico), p.a (precio de adquisición), p.m.(precios de mercado)

CUESTIÓN 6: Se dispone de la siguiente información de los principales agregados de oferta y rentas de una economía (en millones de euros):

	Cod. SEC 2010	
Producción de bienes y servicios (a precios básicos)	P1	2.000.000
Consumo intermedio (a precios de adquisición)	P2	1.020.000
Impuestos sobre la producción y las importaciones	D.2	112.000
Impuestos sobre los productos	D.21	90.000
Subvenciones	D.3	18.400
Subvenciones sobre los productos	D.31	6.000
Excedente de explotación de renta mixta	B.2.G+B.3.G	445.000

Por otra parte, de la cuenta del resto del mundo se obtienen los datos siguientes (en millones de euros):

	Cod. SEC 2010	
Remuneración de asalariados recibida del resto del mundo	D.1	1.100
Remuneración de asalariados pagada al resto del mundo	D.1	300
Impuestos sobre la producción e importaciones pagados al resto del mundo	D.2	1.500
Subvenciones recibidas del resto del mundo	D.3	6.000
Otras transferencias corrientes recibidas del resto del mundo	D.7	2.000
Rentas de la propiedad recibidas del resto del mundo	D.4	45.000
Rentas de la propiedad pagadas al resto del mundo	D.4	66.000
Otras transferencias corrientes pagadas al resto del mundo	D.7	1.500

Calcule, a partir de esta información, el valor añadido total a precios básicos, el PIB a precios de mercado, la remuneración de los asalariados y la renta nacional bruta de la economía.

2

CUESTIÓN 7: En un Estado (E) de nuestro interés acontece que los accidentes de tráfico son la causa principal de muerte para jóvenes entre 5 y 32 años. A través de varias políticas de gasto, el gobierno central hace un tiempo promulgó una de ley de uso obligatorio del cinturón de seguridad, y así reducir el número de muertes y de lesiones graves. Su departamento de análisis econométrico ubicado en su Instituto Nacional (INE) está interesado en examinar cómo de efectiva ha sido esta ley. Para ello, tiene disponible un conjunto de datos de panel de 50 unidades político-administrativas del país (piénsese en provincias, por ejemplo) para los años 1983-1997. Su conjunto de datos incluye las siguientes variables:

- $fatalidades$ es el número de muertes por miles de millas de tráfico
- CS_uso es el nivel de uso del cinturón de seguridad
- $velo90$ es una dummy $= 1$ si limite de velocidad es de 90 kms por hora, $= 0$ al contrario
- $velo120$ es una dummy $= 1$ si limite de velocidad es de 120 kms por hora, $= 0$ al contrario
- $ba08$ es una dummy $= 1$ si el límite de alcohol en la sangre es $\leq 0{,}08\,\%$, $= 0$ al contrario
- $beber - edad18$ es una dummy $= 1$ si tiene 18 años de edad para beber, $= 0$ al contrario
- ing es el ingreso per cápita
- $edad$ media de edad
- $provincia$ es un conjunto de dummies de provincias
- $año$ es un conjunto de dummies de años

La siguiente tabla contiene los resultados de varias regresiones (MCO agrupados, MCO con efectos fijos de estado, MC generalizados con efectos aleatorios de estado y MCO con efectos fijos de estado y año). Las siguientes preguntas se basan en estos resultados.

	1	2	3	4
CS_uso	4,07	−5,77	−4,50	−3,72
	(1,22)	(1,215)	(1,12)	(1,13)
$velo90$	0,148	−0,425	−0,341	−0,783
	(0,403)	(0,334)	(0,337)	(0,424)
$velo120$	2,40	1,23	1,34	0,804
	(0,511)	(0,329)	(0,328)	(0,340)
$ba08$	−1,92	−1,38	−1,36	−0,822
	(0,445)	(0,373)	(0,367)	(0,352)
$edad - beber18$	0,079	0,745	0,767	−1,13
	(0,876)	(0,507)	(0,510)	(0,535)
$ln(ing)$	−18,1	−13,5	−12,6	6,26
	(0,931)	(1,42)	(1,14)	(3,86)
$edad$	−0,007	0,979	0,232	1,32
	(0,109)	(0,382)	(0,239)	(0,383)
$constante$	196,5		137,9	
	(8,22)		(8,92)	
Efectos provincia	No	FE	RE	FE
Efectos años	No	No	No	FE
R^2	0.544	0.874	0.683	0.897

FE (Efectos fijos); RE(Efectos aleatorios)

1. Centrándonos en los resultados de la regresión MCO de datos agrupados (fusionados) en la columna 1,

a) ¿la regresión estimada sugiere que un mayor uso del cinturón de seguridad reduce significativamente las muertes?

b) ¿indique si este resultado tiene sentido? Si es así, explica por qué. Si no, explica que piensas que está pasando aquí.

2. Las columnas 2 y 3 contemplan las estimaciones por efectos fijos de provincia efectos aleatorios de provincia.

   a) ¿Qué conclusión puede sacarse con respecto al impacto de nivel de uso del cinturón de seguridad en las muertes cuando agregamos los efectos fijos?

3. Los resultados de la prueba de Hausman para los modelos de las columnas 3 y 4 son los siguientes

	Coeficientes			
	(b)	(B)	(b-B)	(diag(var_(b)-var_(B)))$^{-1}$
	fijos	.	diferencia	S.E.
CB_uso	-5.774782	-4.503969	-1.270813	0.2698476
velo90	-0.4250387	-0.3405939	-0.0844448	0.0644951
velo120	1.23329	1.335134	-0.1018444	0.0209829
ba08	-1.377456	-1.364296	-0.0131597	0.0648635
edad-beber18	0.7453195	0.766994	-0.0216745	.
ln(ing)	-13.5144	-12.61544	-0.8989608	0.8380231
edad	0.9786802	0.2318357	0.7468445	0.2976956

b=consistente bajo H_0 y H_a;
B=inconsistente bajo H_a, eficiente bajo H_0;
Test: H_0: discrepancias de coeficiente no sistemáticas

$$\begin{aligned} chi2(7) &= (b-B)^{'}[(var(b)-var(B)^{-1}](b-B) \\ &= 26{,}62 \\ Prob > chi2 &= 0{,}0004 \end{aligned}$$

Indica qué conclusión principal puedes sacar respecto de la información suministrada de cara a plantear el modelo más adecuado.

4. El modelo en la columna 4 agrega efectos fijos anuales al modelo de la columna 2. Una prueba F de la significación conjunta de estas dummies de 14 años nos da el estadístico F de 8.85. ¿Qué concluye acerca de la significación conjunta de las dummies del año? ¿Los resultados con respecto al impacto del uso del cinturón de seguridad cambian ahora que hemos agregado EF de los años?

4

CUESTIÓN 8: Está interesado en estimar una demanda de pescado vendido en el Mercado de Luarca, para ello usted va al mercado y recoge el precio diario y la cantidad vendida de 97 días consecutivos (dado que el mercado está cerrado los fines de semana, recopila datos de lunes a viernes). Específicamente, tiene datos de las siguientes variables:

- totqty - la cantidad total de pescado vendido ese día
- avgprc - el precio promedio del pescado vendido ese día
- lun - una dummy =1 por si el día es lunes
- mar - una dummy =1 por si el día es martes
- mie - una dummy =1 por si el día es miércoles
- jue - una dummy =1 por si el día es jueves
- wave2 - la altura máxima promedio de olas durante dos días anteriores a los datos de precio y cantidad
- wave3 - la altura media promedio de ola durante tres y cuatro días anteriores a los datos de precio y cantidad

Nota: aunque usemos subíndices de tiempo a lo largo de esta pregunta, no utilizaremos métodos de series temporales. También mantendremos la suposición de que todos los errores son homoscédasticos.

1. Suponga que la ecuación de demanda se puede escribir para cada período de tiempo como $ln(totqty_t) = \beta_0 + \beta_1 ln(avgprc_t) + \beta_2 lun_t + \beta_3 mar_t + \beta_4 mie_t + \beta_5 jue_t + u_t$ la demanda puede variar durante días de la semana. ¿Por qué no es apropiado usar MCO para estimar esta ecuación de demanda? ¿Qué información adicional necesitamos para tener estimadores consistentes de los parámetros de ecuación de demanda?

2. Las variables $wave2_t$ y $wave3_t$ son medidas de las alturas de las olas oceánicas en los últimos días. ¿Qué dos suposiciones necesitamos para usar $wave2_t$ y $wave3_t$ como instrumentos para $ln(avgprc_t)$ al estimar la ecuación de la demanda? Asegúrese de analizar cómo estas suposiciones están relacionadas con las ecuaciones de demanda y oferta.

3. La primera etapa de una regresión de mínimos cuadrados de dos etapas nos da los siguientes resultados:
$ln(\widehat{avgprc}) = \underset{(0,14)}{-1{,}02} - \underset{(0,114)}{0{,}012} lun_t - \underset{(0,1119)}{0{,}0090} mar_t + \underset{(0,112)}{0{,}051} mie_t + \underset{(0,111)}{0{,}124} jue_t + \underset{(0,021)}{0{,}094} wave2_t +$
$\underset{(0,020)}{0{,}053} wave3_t$
$\hat{R}^2 = 0{,}165$
y la prueba de la significación conjunta de $wave2_t$ y $wave3_t$ nos da un estadístico F: $F - stat = 19{,}1$, mientras la prueba de significación conjunta de las dummies de día de la semana da $F - stat = 0{,}53$. ¿Son $wave2_t$ y $wave3_t$ individualmente significativos al nivel de 1 %? ¿Qué revelan los resultados de la regresión de esta primera etapa sobre nuestros instrumentos?

4. Estimación de la ecuación de demanda por mínimos cuadrados en dos etapas nos da los siguientes resultados:
$ln(\widehat{totqty_t}) = \underset{(0,18)}{8{,}16} - \underset{(0,327)}{0{,}816} ln(avgprc_t) - \underset{(0,229)}{0{,}307} lun_t - \underset{(0,226)}{0{,}685} mar_t + \underset{(0,224)}{0{,}521} mie_t + \underset{(0,225)}{0{,}095} jue_t$
donde los errores estándar entre paréntesis son los correctos (es decir, toman en cuenta el procedimiento de dos etapas). ¿Cuál es la interpretación del coeficiente en $ln(avgprc_t)$?

5

¿Su magnitud parece razonable? Construya un intervalo de confianza del 95 % para este coeficiente.

5. Dado que tenemos dos instrumentos y una variable endógena, la ecuación de demanda esta sobre identificada. La prueba de sobre identificación de restricciones nos da un estadístico F de 0,013. ¿Qué concluye?

6. Dado que la ecuación de la oferta (no especificada) evidentemente depende de las variables de olas(wave), ¿qué dos suposiciones deberíamos hacer para estimar la elasticidad del precio de la oferta?

7. Aquí están los resultados de la estimación de una posible ecuación de oferta para esta industria por mínimos cuadrados en dos etapas:
$ln(\hat{totqty_t}) = \underset{(2,23)}{10{,}82} + \underset{(2,24)}{2{,}13} ln(avgprc_t) - \underset{(0,212)}{0{,}267} wave2_t - \underset{(0,139)}{0{,}169} wave3_t$
¿Cuál es la interpretación del coeficiente en $ln(avgprc_t)$? ¿Es estadísticamente significativo al nivel de 10 %?

CUESTIÓN 9: Dada la siguiente tabla de mortalidad:

	Tasa de mortalidad	Promedio de años vividos el último año de vida	Riesgo de muerte	Supervivientes	Defunciones teóricas	Población estacionaria	Tiempo por vivir	Esperanza de vida
98 años	296,19	0,4758	256,38	6.263,77	1.605,93	5.421,93	16.422,50	2,622
99 años		0,4688			1.223,66			
100 y más años		2,0362						

Se pide:
a) Calcular los supervivientes a las edades de 99 años y de 100 y más años.
b) Calcular el riesgo de muerte a las edades de 99 años y de 100 y más años.
c) Calcular la población estacionaria a las edades de 99 años y de 100 y más años.
d) Calcular el tiempo por vivir a las edades de 99 años y de 100 y más años.
e) Calcular la esperanza de vida a las edades de 99 años y de 100 y más años.
f) Calcular la tasa de mortalidad a las edades de 99 años y de 100 y más años.

CUESTIÓN 10: Sea un país con dos regiones, A y B, y considere la siguiente matriz migratoria por regiones en el periodo 2001-2011:

Región de residencia 2001	Región de residencia 2011		
	Región A	Región B	Total
Región A	1.244,4	14,4	1.258,8
Región B	18,2	889,4	907,6
En el extranjero	72,2	99,3	171,5
No aplicable	130,1	123,2	253,3
Total	1.464,9	1.126,3	2.591,2

Se pide:
a) El número total de emigrantes durante el periodo, la proporción de emigración y la tasa de emigración en las regiones A y B.
b) El número total de inmigrantes durante el periodo, la proporción de inmigración y el índice de atracción en las regiones A y B.
c) El saldo migratorio y las tasas de migración brutas y netas durante el periodo en las regiones A y B.
d) El índice de efectividad migratoria y el índice de migración diferencial en las regiones A y B.

Nota: 'No aplicable' recoge la población nacida en el período intercensal 2001 - 2011.

7

1. Enumerar las ventanas que existen en una sesión de SAS, indicando brevemente para qué sirve cada una de ellas.

2. Describir brevemente la estructura general de un fichero SAS.

3. ¿Qué es una librería SAS? Escríbase un ejemplo de sentencia de creación de una librería SAS.

4. Enumerar los componentes principales de un programa SAS, indicando brevemente qué función cumple cada componente.

5. Proporcionar tres ejemplos de operadores entre variables alfanuméricas en SAS.

6. Escríbase un paso **DATA** para leer un fichero de texto (sin formato) de nombre `file.txt` con variables `var1` y `var2` alfanuméricas en columnas separadas por una coma (`,`). El fichero `file.txt` se encuentra en la ruta `C:\Documents`.

7. Un fichero SAS de nombre `inputFile` contiene dos variables numéricas `var1` y `var2`. ¿Escríbase un breve código SAS para generar un nuevo fichero SAS de nombre `outputFile` que contenga la nueva variable `sumVar` que resulta de la suma de `var1` y `var2` para cada registro.

8. Explicar brevemente la diferencia entre estos dos pasos **DATA**:

```
DATA outputFile;
SET inputFile (DROP = var);
RUN;
```

```
DATA outputFile (DROP = var);
SET inputFile;
RUN;
```

9. Dados dos ficheros SAS `file1` y `file2` ordenados por la variable `ID`, escribir un paso **DATA** para obtener un fichero SAS de nombre `output` que solo contenga los registros del fichero `file1` que no estén en el fichero `file2`.

10. ¿Qué sentencias pueden especificarse en un procedimiento (**PROC**) **PRINT**?

C.2. Tercer examen del Cuerpo de Diplomados en Estadística del Estado

OPOSICIONES AL CDEE. CONVOCATORIA 2017. EJERCICIO 3

CUESTIÓN 1. La probabilidad de que una sucursal de un Banco reciba un cheque sin fondos es del 1%.

a) Si una sucursal en una hora recibe 20 cheques, ¿cuál es la probabilidad de que reciba algún cheque sin fondos en una hora?
b) Si la media del valor de los cheques sin fondos es de 580 € y la sucursal trabaja 6 horas diarias. Calcular la cantidad total esperada de euros en un día, correspondiente a los cheques sin fondo.
c) El banco dispone de 12 sucursales en una determinada ciudad, ¿cuál es la probabilidad de que al menos 4 de las sucursales reciban algún cheque sin fondos en 1 hora?
d) Por un error informático los sistemas de una sucursal del Banco detectan sólo la mitad de los cheques sin fondos que reciben ¿cuál es la probabilidad de que esta sucursal reciba un cheque sin fondos después de ser procesado por sus sistemas informáticos?

CUESTIÓN 2. Sea (X,Y) una variable aleatoria bidimensional con función de densidad conjunta:

$$f(x,y)=\begin{cases} e^{-(x+y)} & si\ x>0, y>0 \\ 0 & en\ el\ resto \end{cases}$$

a) Calcular las funciones de densidad marginales.
b) ¿Son independientes X e Y?
c) Calcular la función de distribución conjunta.
d) Calcular la matriz de varianzas-covarianzas de la variable (W,Z) con W=X+Y y Z=X-Y

CUESTIÓN 3. Una empresa láctea, que dispone de 100 plantas de producción distribuidas por toda la geografía, desea estimar la cantidad de leche envasada al mes. Para ello selecciona una muestra, con probabilidades iguales y con reposición, de 8 plantas en las que examina todas las máquinas para envasar. El resultado de la muestra es el siguiente:

Planta	Nº de máquinas	Cantidad total de leche envasada por las máquinas (miles de litros)
1	20	20
2	40	25
3	30	22
4	40	24
5	20	20
6	70	30
7	70	28
8	50	25

Se pide:

a) Proporcione una estimación insesgada del total de leche envasada y una estimación de su varianza.
b) Si se conoce que la empresa láctea tiene 4000 máquinas en total, obtenga una estimación del total de leche envasada por un método diferente al del apartado anterior.

1

CUESTIÓN 4. Dada la siguiente distribución bidimensional de frecuencias:

X \ Y	0	1	2
0	15	15	10
1	5	20	5
2	10	5	15

Se pide calcular: a) La media aritmética y la varianza de la variable X-Y
b) La recta de regresión de Y sobre X y la varianza residual de Y

CUESTIÓN 5. El importe medio y el número de hipotecas constituidas sobre fincas rústicas, viviendas y otras fincas, inscritas en los registros de la propiedad en el periodo 2015-2017, han sido los siguientes:

Hipotecas constituidas	Fincas rústicas		Viviendas		Otras fincas	
	Importe medio (miles €)	Número (miles)	Importe medio (miles €)	Número (miles)	Importe medio (miles €)	Número (miles)
2015	154,5	18,7	105,9	246,7	214,5	102,5
2016	165,7	16,7	109,8	282,7	195,3	101,4
2017	152,5	16,5	116,7	310,1	185,6	106,5

Fuente: Estadística de hipotecas, INE.

A partir de esos datos, se pide:

a) Calcular el Índice de Laspeyres del Importe medio de las hipotecas entre 2016 y 2017, tomando como año base 2015.

b) Determinar la variación interanual del Índice de Laspeyres entre 2016 y 2015.

c) ¿Cuál es la repercusión de la componente individual "Viviendas" en la variación del índice de Laspeyres calculado anteriormente, entre 2015 y 2016?

d) Calcular la participación de las "Viviendas" en el cambio total experimentado en el Índice general de 2016.

CUESTIÓN 6. A partir de la siguiente información, extraída de la Contabilidad Nacional Anual del año 2015, correspondiente al sector de Administraciones Públicas, se pide:

a) Calcular la producción del sector.
b) Calcular el excedente neto de explotación.
c) Calcular el consumo final efectivo.
d) Sabiendo que el ahorro bruto del sector es -28.617 millones de euros, calcule la capacidad o necesidad de financiación del sector.

Datos del sector de administraciones públicas (S.13)

Formación bruta de capital	26.965
Remuneración de asalariados	119.356
Impuestos netos sobre la producción y las importaciones pagados	476
Impuestos netos sobre la producción y las importaciones recibidos	127.499
Transferencias de capital a cobrar	10.330
Transferencias de capital a pagar	10.826
Consumo de capital fijo	27.628
Gasto en consumo final individual	119.628
Gasto en consumo final colectivo	89.295
Transferencias sociales en especie: producción adquirida en el mercado	28.202
Consumos intermedios	57.474
Adquisiciones menos cesiones de activos no producidos	926
Préstamos recibidos	1.250
Valores representativos de deuda emitidos	55
Participaciones en el capital y en fondos de inversión	180

(Cifras en millones de euros)

2

CUESTIÓN 7. En una economía con dos ramas de actividad, se dispone del siguiente extracto de la tabla input-output simétrica:

	Consumos intermedios		Gasto en consumo final (GCF)	Formación bruta de capital (FBC)	Exportaciones
	Rama 1	Rama 2			
Producto 1	30	10	50	10	20
Producto 2	10	40	20	20	10

(Cifras en millones de euros)

Además, se proporcionan los siguientes datos para completar la tabla:

- La producción de la rama 1 (producto 1) es de 100 millones de euros y la de la rama 2 (producto 2) es de 50 millones de euros.
- La remuneración de asalariados ha sido de 45 millones de euros en la rama 1 y 50 millones de euros en la rama 2.
- Los impuestos netos sobre la producción pagados por la rama 1 ascienden a 5 millones de euros, mientras que en el caso de la rama 2 ascienden a 10 millones.
- No existen impuestos sobre los productos ni márgenes de distribución (comercio y transporte) en esta economía.
- Todas las exportaciones se consideran de servicios de mercado. No existen exportaciones de bienes en esta economía.

Se pide:

a) Calcular las importaciones realizadas de cada producto (a excepción de las exportaciones e importaciones, suponga que todas las demás operaciones del Resto del Mundo se consideran nulas).

b) Calcular el Excedente Bruto de Explotación de la rama 2 y explique su significado económico.

c) Calcular el Producto Interior Bruto de la economía a partir de la tabla, por las tres vías (demanda, oferta, rentas).

d) Calcular la matriz de coeficientes técnicos. Interprete el resultado obtenido en la primera fila de esa matriz.

CUESTIÓN 8. Con los siguientes datos en millones de euros de la Contabilidad Nacional de España para el año 2010, calcule:

a) El Excedente de Explotación / Renta Mixta Bruto.

b) La Renta Nacional Bruta disponible a precios de mercado.

c) El Ahorro Nacional Neto.

d) La capacidad o necesidad de financiación del país. Interprete el resultado.

Producto interior bruto a precios de mercado	1.080.913
Consumo de capital fijo	182.025
Formación bruta de capital	254.549
Adquisiciones menos cesiones de activos no financieros no producidos	-119
Remuneración de asalariados (interior)	541.475
Rentas primarias netas del exterior	-15.155
Transferencias de capital netas con el exterior	5.732
Transferencias corrientes netas con el exterior	-12.718
Impuestos netos sobre la producción y las importaciones	93.559
Consumo final efectivo	840.470

CUESTIÓN 9. Los siguientes datos han sido tomados de un Censo de Población de referencia 31/12/1991

Año de nacimiento	Población
1989	59.984
1990	60.120
1991	60.850

La estadística de Movimiento Natural de Población proporciona la siguiente tabla sobre defunciones, relativa a la generación de 1990

Edad	Año	Defunciones
0	1990	104
0	1991	95
1	1991	70
1	1992	45
2	1992	27
2	1993	20

Suponiendo que el efecto migratorio es nulo, calcular el número de efectivos de la generación de 1990 y cuántos de éstos sobrevivieron a su segundo aniversario.

CUESTIÓN 10. La siguiente tabla presenta, por grupos de edad y sexo, datos correspondientes a la población de un país a 1 de Enero del año 2018 (en millones de habitantes):

Edad	Hombres	Mujeres	Ambos sexos
0-15	?	?	?
16-64	?	15,1	?
65 y más años	3,9	?	?
Total	?	?	46,7

Complete la tabla, conocidos los siguientes indicadores:
- Índice de envejecimiento de la población (ambos sexos) = 121,6%
- Ratio de masculinidad de la población mayor de 64 años = 76,5%
- Tasa de dependencia de la población masculina menor de 16 años = 25,0%
- Tasa de dependencia de la población mayor de 64 años (ambos sexos) = 29,7%

(Nota: es suficiente con que los resultados se aproximen con un solo decimal).

4

Apéndice D

Exámenes de problemas de 2016

D.1. Segundo examen del Cuerpo Superior de Estadísticos del Estado

Tribunal de la Oposición al Cuerpo Superior de Estadísticos del Estado

Pruebas selectivas para el ingreso en el Cuerpo Superior de Estadísticos del Estado. Orden ECC/1693/2015, de 5 de septiembre 2016 (BOE8/09/2016).

SEGUNDO EJERCICIO

1.- <u>Muestreo</u>

1. Sea una población de M=40.000 individuos distribuida en N=10.000 hogares. Notamos por:

   M_i= tamaño del hogar i (medido en número de individuos)

   A_i= número de mujeres en el hogar i

   A= total de mujeres en la población

   Se obtiene una muestra aleatoria simple (sin reemplazamiento) de n=20 hogares y se encuesta a todos los individuos de cada hogar seleccionado. Se pide:

   a) Una estimación insesgada de A y su error de muestreo.
   b) Una estimación de la eficiencia relativa del muestreo por conglomerados respecto a la del muestreo aleatorio simple.
   c) Una estimación de A y su error de muestreo, aplicando el estimador de razón al tamaño.
   d) Comentar las ventajas e inconvenientes del estimador de c) respecto al a).

Los datos necesarios para los cálculos son los siguientes:

$$\sum_{i=1}^{n} M_i = 69\,;\ \sum_{i=1}^{n} A_i = 35\,;\ \sum_{i=1}^{n} M_i^2 = 269\,;\ \sum_{i=1}^{n} A_i^2 = 81\,;\ \sum_{i=1}^{n} M_i A_i = 142$$

2. Los individuos de la población del apartado anterior se distribuyen en 200 secciones censales de 50 hogares cada una. Se quiere estimar el total de mujeres (A) en la población y para ello, se obtiene una muestra de 4 secciones censales, de las que se obtiene a su vez una submuestra de 5 hogares de cada sección censal seleccionada. El muestreo se realiza con reposición y probabilidades iguales en ambas etapas. Calcular una estimación insesgada de A y su error de muestreo.

   Los datos necesarios para los cálculos son los siguientes:

Secciones censales	1	2	3	4
Número de mujeres en cada sección censal	6	8	11	10

EX LIBRIS
CUERPOS D NUMEN
NUMERUS ESTADISTICA
INSTITUTO NACIONAL
DE
ESTADISTICA

2

2.- Economía

En la siguiente tabla se muestran datos correspondientes a operaciones y saldos de las cuentas del sector Administraciones Públicas (S.13) de una economía para un año t:

Formación bruta de capital (P.5)	25
Consumo de capital fijo (P.51c)	26
Consumo final colectivo efectivo (P.42)	80
Información sobre impuestos y subvenciones:	
Impuestos sobre la producción y las importaciones (recursos) (D.2)	125
Subvenciones (empleos) (D.3)	10
Impuestos corrientes netos sobre la renta, el patrimonio, etc. (D.5)	100
Información sobre las rentas de la propiedad (D.4):	
Intereses recibidos (D.41)	4
Intereses pagados (D.41)	40
Otras rentas de la propiedad netas recibidas (D.42-D.45)	3
Cotizaciones sociales netas (D.61)	130
Prestaciones sociales distintas de las transferencias sociales en especie (D.62)	180
Transferencias sociales en especie (D.63)	120
Información sobre otras transferencias:	
Otras transferencias corrientes netas pagadas (D.7)	10
Transferencias de capital netas a pagar (D.9)	1
Excedente de explotación bruto	30

Unidades: miles de millones de euros (mm €)

Nota: para esta economía se considera que las adquisiciones menos cesiones de activos no producidos son despreciables en el año t

Se pide:

a) Calcular para el sector S.13:

1. El saldo de rentas primarias, bruto y neto
2. La renta disponible ajustada bruta
3. El gasto en consumo final de las AAPP
4. Sabiendo que el concepto de déficit o superávit público es equivalente al concepto de capacidad o necesidad de financiación de las Administraciones Públicas y que el PIB a precios corrientes de esta economía en el año t ha sido de 1070 mm €, calcule el déficit público como porcentaje del PIB para el citado año

b) El gasto en consumo final de las AAPP a precios corrientes en el año t-1 fue de 194 mm € y la tasa de variación interanual (t/t-1) del deflactor de este agregado fue del 1%,

3

1. Calcule la tasa de variación interanual del gasto en consumo final de las AAPP a precios corrientes
2. Indique qué parte de esta variación sería atribuible a variaciones de volumen
3. Calcule el gasto en consumo final de las AAPP del año t valorado a precios del año anterior (t-1)

c) Sabiendo que en el año t el valor a precios corrientes del gasto en consumo final de las AAPP en los dos primeros trimestres fue de 47 y 52 mm €, respectivamente y que la tasa intertrimestral del 3º trimestre fue del 2%, calcule:

1. El valor del agregado a precios corrientes en el 3º trimestre
2. El valor del agregado a precios corrientes en el 4º trimestre
3. Indique a qué tasa interanual aproximada evolucionó en el 4º trimestre del año t, sabiendo que las tasas interanuales correspondientes a los tres primeros fueron 2,5%, 3% y 3,5%, respectivamente.

3. - Econometría

Considere la siguiente ecuación salarial

$$\ln(salario) = \beta_0 + \beta_1 edu + \beta_2 edu^2 + \beta_3 exper + \beta_4 exper^2$$
$$+\beta_5(edu \times exper) + \beta_6 horasem + \varepsilon$$

donde las variables explicativas son años de educación, años de experiencia y número de horas trabajadas a la semana. Los resultados de la estimación de esta ecuación y de algunos modelos derivados de la misma se muestran en la siguiente tabla. El número de observaciones con los que se han realizado las estimaciones es de 1000.

Variable					
	modelo I	**modelo II**	**modelo III**	**modelo IV**	**modelo V**
constante	1.055	1.252	1.573	1.917	0.904
	(0.266)	(0.190)	(0.188)	(0.080)	(0.096)
edu	0.0498	0.0289	0.0366		0.1006
	(0.0397)	(0.0344)	(0.0350)		(0.0063)
edu²	0.00319	0.00352	0.00293		
	(0.00169)	(0.00166)	(0.00170)		
exper	0.0373	0.0303		0.0279	0.0295
	(0.0081)	(0.0048)		(0.0054)	(0.0048)
exper²	0.000485	0.000456			
	(0.000090)	(0.000086)			
exper x **edu**	-0.000510			-0.00047	-0.00044
	(0.000482)			(0.000096)	(0.000086)
horasem	0.01145	0.01156	0.01345	0.01524	0.01188
	(0.00137)	(0.00137)	(0.00136)	(0.00151)	(0.00136)
SCR	222.4166	222.6674	233.8317	280.5061	223.6716
AIC	-1.489	-1.490	-1.445	X	-1.488
SC	X	-1.461	-1.426	-1.244	-1.463

a.- Usando un nivel del 5%, indique qué coeficientes estimados no son significativamente diferentes de cero. Para ello elabore una tabla similar a la anterior en donde se indique con claridad lo pedido.

b.- Indique qué restricción en el modelo I arroja el modelo II. Utilice un contraste adecuado para contrastar la restricción señalada. Muestre si es posible obtener el mismo resultado usando un contraste tipo ratio de la t.

c.- Justifique qué modelo utilizaría si quisiera contrastar si la experiencia es relevante para determinación del salario. Seguidamente realice el contraste. Haga lo mismo para

5

el caso de que quisiera contrastar la relevancia de la educación en la determinación del salario.

d.- Explique a qué cuestión econométrica estamos tratando de responder cuando realizamos restricciones sobre el modelo I para obtener el V.

e.- A partir de los resultados de sus respuestas en los apartados anteriores y de algún otro contraste que considere necesario, indique qué modelo considera más adecuado.

f.- Las tres última líneas indican los acrónimos de "suma cuadrada de los residuos", criterio de información de Akaike, y criterio de información de Schwarz. Observará que hay algunos que faltan. ¿Cómo lo completaría? ¿Para qué sirven estos criterios?

NOTA: Para la resolución de este ejercicio se entregan al opositor las tablas estadísticas de las distribuciones de probabilidad necesarias.

4.- <u>Estadística</u>

El número diario de ingresos en un servicio de urgencias sigue una distribución de Poisson de parámetro $\theta_{1,}$ con independencia entre los registros correspondientes a días diferentes. Se va a observar los números de ingresos X_1, X_2, ..., X_r durante r días, a fin de estimar θ_1 para hacer las previsiones de mantenimiento del servicio. Se supone que los errores en la estimación de θ_1 producen pérdida cuadrática:

a) Demostrar que el estadístico $T(X1, X2, \ldots, Xr) = \sum_{i=1}^{r} X_i$ es suficiente.

b) Determinar k para que $T_k = kT$ sea un estimador insesgado para θ_1 y comprobar que es de mínima varianza para θ_1.

c) Suponiendo que dispusiéramos de una muestra aleatoria simple Y_1, Y_2, ..., Y_{120} del número medio mensual de ingresos de urgencias registrados a lo largo de los últimos 10 años en dicho servicio, obtener en función de ella la expresión de un intervalo de confianza aproximado para θ_1 al nivel 1-α por el método de la cantidad pivotal.

d) Suponiendo que dispusiéramos de una muestra aleatoria simple correspondiente al número de ingresos medios registrados esos mismos meses en un segundo servicio de urgencias independiente del anterior, para el que los ingresos diarios siguen una distribución de Poisson de parámetro $\square_2$, obtener la expresión de un estadístico de contraste para resolver H_0: $\square_1=\square_2$ frente H_1: $\square_1\neq\square_2$.

5.- Demografía

Problema Demografía:

Una oficina estadística internacional nos facilita de la Región A los siguientes datos de población a 1 de enero de cada año (columnas [2] a [6]) y de defunciones durante los años de calendario señalados (columnas [7] a [11]).

También se dispone de información detallada sobre salud autopercibida a partir de los datos de una encuesta de condiciones de vida para el país en que está incluida la mencionada Región A, que se suponen bastante próximos a la realidad regional.

Grupode edad	Población a 1 de enero					Defunciones					% con mala salud
	2011	2012	2013	2014	2015	2010	2011	2012	2013	2014	2011-2012
[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]	[11]	[12]
0	51.239	49.222	47.619	44.308	44.453	180	140	127	104	125	0,7
1 - 4	218.970	215.759	209.305	199.674	192.383	45	34	31	29	25	0,9
5 - 9	255.800	261.281	265.724	269.985	269.398	17	24	22	30	19	0,5
10 - 14	234.010	237.601	239.937	244.617	248.975	29	25	23	15	17	0,3
15 - 19	238.863	235.736	231.256	228.371	229.842	51	53	47	34	37	0,5
20 - 24	272.590	265.689	258.557	250.598	244.678	83	63	52	64	52	0,6
25 - 29	340.803	321.447	301.253	285.500	273.215	108	103	110	88	66	0,8
30 - 34	437.132	419.965	394.196	365.084	341.900	212	200	199	143	136	1,0
35 - 39	435.477	440.117	439.598	435.451	425.701	320	309	290	265	236	1,8
40 - 44	408.628	413.213	412.158	411.964	415.151	532	501	482	454	443	2,9
45 - 49	374.921	382.673	389.565	392.446	392.895	785	814	767	746	719	3,8
50 - 54	333.903	343.692	348.318	353.085	359.996	1.081	1.102	1.131	1.103	1.165	5,1
55 - 59	283.511	289.539	295.568	304.140	315.019	1.439	1.427	1.376	1.476	1.477	7,7
60 - 64	269.593	271.651	274.308	271.056	270.957	1.854	1.942	1.982	1.878	1.901	11,4
65 - 69	234.820	243.864	252.303	258.610	260.472	2.469	2.510	2.413	2.666	2.721	13,3
70 - 74	202.194	200.317	199.933	206.853	221.483	3.545	3.522	3.419	3.368	3.574	16,9
75 - 79	177.808	179.428	181.192	177.782	168.125	5.636	5.553	5.669	5.285	5.052	21,3
80 - 84	128.901	131.505	135.258	139.654	142.328	7.624	7.936	8.031	7.506	7.608	29,4
85 y +	100.029	106.951	110.963	117.249	122.703	14.119	14.530	16.132	15.265	16.177	35,8

Se solicitan los siguientes resultados:

1. Elaborar una tabla de mortalidad correspondiente al intervalo de años para los que se dispone de datos de salud autopercibida, que al menos incluya las siguientes series:
   a. Probabilidades de muerte (${}_{n}q_{x}$) y supervivientes (l_x) entre dos aniversarios de edades cumplidas.
   b. Población estacionaria de la tabla en cada intervalo de edades (${}_{n}L_{x}$) y esperanza de vida al inicio de cada intervalo de edades (e_x).
2. A partir de los datos anteriores se precisa calcular la esperanza de vida en buena salud al nacimiento y esperanza de vida en buena salud a los 65 años para los años, igualmente para el intervalo de años para los que se dispone de datos de salud autopercibida. Se entiende como buena la salud la informada como diferente a la reportada por los entrevistados como mala salud (columna [12]).

8

3. Comente brevemente, qué conclusiones se pueden sacar de los resultados de las diferencias obtenidas entre esperanza de vida y esperanza de vida en buena salud de la Región A en el contexto de la evolución de éstas magnitudes a nivel internacional en las últimas décadas.

D.2. Tercer examen del Cuerpo de Diplomados en Estadística del Estado

CUESTIÓN 1: De una población que sigue una distribución Poisson de parámetro λ desconocido, se obtiene una muestra aleatoria simple de tamaño 3, (x_1, x_2, x_3), y se construyen los siguientes estimadores de la varianza poblacional:

$$\hat{V}_1 = \frac{x_1 + x_3}{2}; \qquad \hat{V}_2 = x_1 + x_2 - 2x_3; \qquad \hat{V}_3 = \frac{\sum_{i=1}^{3}(x_i - \bar{x})^2}{2}$$

Se pide:
a) ¿Son insesgados?
b) Calcular el error cuadrático medio de $\hat{V}_2$
c) ¿Es la media muestral, $\bar{x} = \dfrac{x_1 + x_2 + x_3}{3}$, un estimador insesgado para la varianza poblacional? Compararla con el estimador $\hat{V}_1$.
e) ¿Alguno de los estimadores propuestos es eficiente?

CUESTIÓN 2: La duración de la vida de los elementos de una población es una variable aleatoria con función de densidad:

$$f(x) = \begin{cases} \frac{1}{a} e^{\left(\frac{-x}{a}\right)} & si \quad x > 0 \\ \\ 0 & en\ el\ resto \end{cases}$$

siendo $a > 0$ un parámetro desconocido. Para una muestra de tamaño n calcular:
a) Estimador máximo verosímil de a.
b) Estimador de a por el método de los momentos.
c) ¿Son eficientes?

CUESTIÓN 3: La longitud (en mm.) de los tornillos fabricados diariamente por una máquina se distribuye según una Normal con media μ desconocida y desviación típica $\sigma = 2$. Un día determinado se toma una muestra de 10 tornillos cuya longitud media resultó ser de 96 mm. Se pide:
a) Calcular un intervalo de confianza al 99 % para la longitud media μ.
b) Determinar el tamaño de la muestra que sería necesario para que un intervalo al 99 % tuviera una longitud igual a 2mm.

CUESTIÓN 4: Dada la tabla adjunta de medidas del pH y de la concentración de un ácido en sangre en 15 individuos, se pide:

1. Calcular las medias de la concentración del ácido condicionadas por el pH.
2. Calcular las varianzas de la concentración del ácido condicionadas por el pH.
3. Calcular la razón de correlación $\eta^2_{\text{Conc;pH}}$ de la concentración respecto al pH.
4. Si ρ denota el coeficiente de correlación lineal de Pearson entre ambas variables, ¿es mayor ρ^2 o $\eta^2_{\text{Conc;pH}}$? Coméntese muy brevemente.

pH	4	5	7	5	6	5	5	5	5	6	6	5	7	5	5
Concentración	443	296	101	224	91	252	195	550	170	382	152	448	64	147	161

Nota: La varianza marginal de la concentración es 20396,33.

1

CUESTIÓN 5: El número de viajeros y las pernoctaciones realizadas por los turistas alojados en establecimientos hoteleros de Castilla-La Mancha durante el mes de enero de 2017 vienen reflejados en la siguiente tabla:

	Viajeros	**Pernoctaciones**
Albacete	18,1	27,5
Ciudad Real	22,2	35,1
Cuenca	15,2	26,8
Guadalajara	16,4	27,1
Toledo	50,0	79,2

Unidad: miles

Fuente: Encuesta de Ocupación Hotelera. INE (www.ine.es)

a) Obtenga los índices simples del número de viajeros para cada provincia tomando Toledo como referencia.
b) Elabore la serie de los índices simples del número de pernoctaciones para cada provincia tomando Toledo como referencia.
c) Para el número de viajeros, dé el valor del índice complejo no ponderado de la media aritmética de la comunidad autónoma apoyándose en los índices calculados anteriormente.
d) Calcule el índice complejo ponderado de la media aritmética del número de viajeros de la Comunidad Autónoma, usando como ponderación el número de pernoctaciones de cada provincia.
e) Comente brevemente cuál de los dos índices complejos calculados anteriormente sería de mayor utilidad para analizar la evolución del turismo en Castilla-La Mancha.

CUESTIÓN 6: Utilizando los datos siguientes relativos a una economía imaginaria con ausencia de impuestos, calcule en términos del SEC2010:
a) La cuenta completa de bienes y servicios.
b) El Producto Interior Bruto de la economía a precios básicos y a precios de mercado.
c) La producción de I+D realizada en la economía es de 1.000 millones de euros, de la cual se vende el 25 % a otros países. Señale en qué rúbricas de la cuenta de bienes y servicios anterior estaría contabilizada esta operación.

Datos	
Producción de bienes y servicios	9.000
Exportaciones	400
Consumo público	1.000
Consumo intermedio	5.000
Importaciones	1.600
Consumo privado	2.500
Subvenciones a los productos	500
Variación de existencias	200

Unidad: millones de euros

CUESTIÓN 7: A partir de la siguiente información y deduciendo aquellas magnitudes que necesite, calcule los siguientes saldos para el total de la economía:
a) El Excedente de Explotación / Renta Mixta Bruto.
b) La Renta Nacional Bruta a precios de mercado.
c) La Renta Nacional Bruta disponible.
d) El Ahorro Nacional Bruto.
e) La capacidad o necesidad de financiación del país. Interprete el resultado.

Datos	
Gasto en consumo final	19.200
Producto interior bruto a precios de adquisición	30.000
Formación bruta de capital fijo	9.375
Remuneración de asalariados (interior)	8.000
Remuneración de asalariados pagada al resto del mundo	52
Remuneración de asalariados recibida del resto del mundo	34
Rentas de la propiedad pagadas al resto del mundo	1.200
Rentas de la propiedad recibidas del resto del mundo	2.000
Impuestos sobre la producción y las importaciones	1.200
Subvenciones a la producción	360
Transferencias corrientes pagadas al resto del mundo	52
Transferencias corrientes recibidas del resto del mundo	272
Consumo de capital fijo	1.480
Exportaciones netas de bienes y servicios	1.425
Transferencias de capital recibidas del resto del mundo	170

Unidad: millones de euros

CUESTIÓN 8: En una economía con dos sectores de actividad, se dispone del siguiente extracto de la tabla input-output simétrica:

	Sector 1	Sector 2	GCF	FBC	Exportaciones
Sector 1	12	34	100	35	7
Sector 2	32	35	40	90	10

Unidad: millones de euros

Además, se proporcionan los siguientes datos para completar la tabla:
- Las importaciones han sido 15 millones de euros de productos del Sector1 y 25 millones de euros de productos del Sector2.
- La remuneración de asalariados ha sido de 75 millones de euros en el Sector1 y 62 millones de euros en el Sector2.
- Los impuestos netos sobre la producción que han sido pagados por el Sector1 ascienden a 2 millones de euros, mientras que en el caso del Sector2 ascienden a 6 millones.
- No existen impuestos sobre los productos en esta economía.

Se pide:
a) Calcular el total de la oferta a precios básicos.
b) Calcular el total de la oferta a precios de adquisición.
c) Calcular el Excedente Bruto de Explotación de cada sector.
d) Calcular el Valor Añadido Bruto de la economía a partir de la tabla por las tres vías (demanda, oferta, rentas).

3

CUESTIÓN 9: A partir de estos datos construir la primera fila de estas series de la tabla de mortalidad.

Edad	a(x)	m(x)	q(x)	l(x)	d(x)	L(x)
0	0,1155			100.000		
1				99.738,88		
2						

CUESTIÓN 10: Con los siguientes datos calcular:
a) El ratio de masculinidad.
b) El índice de envejecimiento.
c) La tasa de dependencia.
d) La tasa de dependencia de jóvenes.
e) La tasa de dependencia de mayores.

1 de enero de 2016		
	Hombres	Mujeres
Total	22.809.420	23.636.408
0-15 años	3.852.848	3.624.085
16-64 años	15.203.861	15.065.141
65 y más años	3.752.711	4.947.182

Apéndice E

Exámenes de problemas de 2015

E.1. Segundo examen del Cuerpo Superior de Estadísticos del Estado

Tribunal de la Oposición al Cuerpo Superior de Estadísticos del Estado

Pruebas selectivas para el ingreso en el Cuerpo Superior de Estadísticos del Estado. Orden ECC/1517/2015, de 16 de Julio (BOE 27/07/2015).

SEGUNDO EJERCICIO

1.- <u>Estadística.</u>

Sean $(X_1,\ldots,X_n)$e $(Y_1,\ldots,Y_m)$ dos muestras aleatorias simples de dos distribuciones independientes exponenciales de medias θ_1 y θ_2, respectivamente con θ_1 y θ_2 mayores que 0.

a) Demostrar que la expresión

$$T_X = \frac{2}{\theta_1}\sum_{i=1}^{n} X_i$$

tíene una distribución independiente de θ_1. Identificar esta distribución.

b) Basándose en el resultado anterior, definir una variable pivotal para construir un intervalo de confianza para el cociente de mediasθ_1/θ_2.

c) Calcular la expresión del intervalo de confianza para θ_1/θ_2con nivel de confianza 1-α.

d) Construir la región crítica para contrastar la hipótesis H_0: $\theta_1=\theta_2$frente H_0: $\theta_1\neq\theta_2$.

1

EX LIBRIS
CUERPOS D NUMEN
NUMERUS ESTADISTICA
INSTITUTO NACIONAL
DE
ESTADISTICA

2

Tribunal de la Oposición al Cuerpo Superior de Estadísticos del Estado

Pruebas selectivas para el ingreso en el Cuerpo Superior de Estadísticos del Estado. Orden ECC/1517/2015, de 16 de Julio (BOE 27/07/2015).

2.- <u>Muestreo.</u>

En una población de 20.000 empresas, se quiere estimar la facturación media $\bar{Y}$. Para ello, se obtiene una muestra aleatoria simple (sin reemplazamiento) de tamaño n=5000.

a) Si se dispone de una información auxiliar que es el número de asalariados para cada empresa, denotada por X, la media de asalariados en la población es $\bar{X} = 35$ y los datos muestrales disponibles son:

$\hat{\bar{Y}} = 5{,}5 \times 10^6$ (Media muestral de Y)

$\hat{\bar{X}} = 30$ (Media muestral de X)

$$s_y^2 = \frac{\sum_{i=1}^{n} (y_i - \hat{\bar{Y}})^2}{n-1} = 25 \times 10^{10}$$

$$s_x^2 = \frac{\sum_{i=1}^{n} (x_i - \hat{\bar{X}})^2}{n-1} = 15$$

$\hat{\rho} = 0{,}7$ (Coeficiente de correlación lineal muestral entre X, Y)

Se pide:

- Calcular el estimador de la razón para $\bar{Y}$. ¿Es insesgado?
- Calcular un estimador de la varianza del estimador de razón.

b) Si no disponemos de esa información auxiliar del apartado anterior, pero después de seleccionar la muestra, obtenemos información adicional sobre la actividad económica principal de las empresas, que permite clasificarlas en 3 grandes grupos: Industria (post-estrato 1 con N_1=3000), Comercio (post-estrato 2 con N_2=6000) y Servicios (post-estrato 3 con N_3=11.000).

Si en este caso la información muestral disponible en cada uno de los post-estratos (indicados por el subíndice i=1, 2, 3) es:

$$\hat{\bar{Y}}_1 = 8\times 10^6; \quad s^2_{y_1} = 15\times 10^{10}$$
$$\hat{\bar{Y}}_2 = 4\times 10^6; \quad s^2_{y_2} = 20\times 10^{10}$$
$$\hat{\bar{Y}}_3 = 5\times 10^6; \quad s^2_{y_3} = 22\times 10^{10}$$

Se pide:

- Calcular el estimador post-estratificado para $\bar{Y}$. ¿Es insesgado?
- ¿Cuál es la distribución de n_1 (tamaño muestral del post-estrato 1)? Calcular la media y la varianza de n_1.
- Calcular un estimador de la varianza del estimador post-estratificado.
- Calcular un intervalo de confianza al 95% para $\bar{Y}$. ¿Qué significa este intervalo de confianza?
- ¿Cuál de los dos estimadores, el de razón o el post-estratificado, es más eficiente? Razone la respuesta.

4

Tribunal de la Oposición al Cuerpo Superior de Estadísticos del Estado

Pruebas selectivas para el ingreso en el Cuerpo Superior de Estadísticos del Estado. Orden ECC/1517/2015, de 16 de Julio (BOE 27/07/2015).

3. –Economía.

La información contable de una economía imaginaria correspondiente al año t es la siguiente:

VAB ramas primarias	25
VAB industria	170
VAB construcción	60
VAB servicios	700
Remuneración de los asalariados residentes	500
Remuneración de los asalariados no residentes por empleadores residentes	1
Remuneración de los asalariados residentes por empleadores no residentes	3
Consumo final colectivo efectivo de las AAPP	90
Impuestos netos de subvenciones sobre los productos	80
Otros Impuestos netos de subvenciones sobre la producción	10
Formación bruta de capital	205
Exportaciones de bienes y servicios	310
Importaciones de bienes y servicios	340
Rentas primarias netas recibidas del Resto del Mundo	-10
Transferencias corrientes netas recibidas del Resto del Mundo	-12
Tasa de variación anual en "t" del deflactor implícito del PIB	-0,5%

Las unidades monetarias que aparecen en la tabla son precios corrientes del año t.

Sabiendo que el PIB en t-1 a precios corrientes asciende a 1030 unidades monetarias, realice los siguientes cálculos para el año t:

a) El PIB a precios corrientes y la tasa de variación anual en volumen

b) El Excedente Bruto de Explotación /Renta mixta

5

c) El Consumo final individual efectivo de los hogares

d) El Ahorro bruto de la economía

Tribunal de la Oposición al Cuerpo Superior de Estadísticos del Estado

Pruebas selectivas para el ingreso en el Cuerpo Superior de Estadísticos del Estado. Orden ECC/1517/2015, de 16 de Julio (BOE 27/07/2015).

4. - Econometría.

Una investigadora con datos del año 2000 procedente de una encuesta longitudinal nacional realizada por la correspondiente institución oficial norteamericana dispone de información sobre las siguientes variables: salario por hora (SAL), en euros; años de escolarización (S), años de experiencia laboral (EXP); género; etnicidad (raza negra, raza hispana y raza blanca, estos últimos son aquellos que no son ni de raza negra ni hispana).

La investigadora elimina de la muestra a los hispanos, dejando 2135 blancos y 273 de raza negra, y define las variables binarias HOM y NEG. HOM toma el valor 1 para hombres, y 0 para mujeres. NEG es 1 para miembros de raza negra y 0 para blancos. Igualmente define LSAL como el logaritmo natural de la variable SAL.

La investigadora estima por MCO los modelos siguientes cuya variable dependiente siempre es LSAL

Modelo 1.- Variables explicativas: S, EXP, HOM, para toda la muestra

Modelo 2.- Variables explicativas: S, EXP, HOM, y NEG para toda la muestra

Modelo 3.- Variables explicativas: S, EXP, HOM, para solo blancos

Modelo 4.- Variables explicativas: S, EXP, HOM, para solo negros

Posteriormente, la investigadora define los siguientes términos de interacción (donde * se refiere a la operación producto):

SN = S*NEG

EN = EXP*NEG

HN = HOM*NEG

y ejecuta una quinta regresión, con la misma variable dependiente:

Modelo 5.- Variables explicativas: S, EXP, HOM, NEG, SN, EN, HN, y usa toda la muestra.

Los resultados se muestran en la siguiente Tabla, en la que no aparecen los datos de la regresión del modelo 4 y donde SCR es la suma cuadrática de los residuos y los valores entre paréntesis son los errores estándar.

7

	Modelo 1 toda la muestra	Modelo 2 Toda la muestra	Modelo 3 Solo blancos	Modelo4 Solo raza NEG	Modelo 5 Toda la muestra
S	0.124 (0.004)	0.121 (0.004)	0.122 (0.004)	V	0.122 (0.004)
EXP	0.033 (0.002)	0.032 (0.002)	0.033 (0.003)	W	0.033 (0.003)
HOM	0.278 (0.020)	0.277 (0.020)	0.306 (0.021)	X	0.306 (0.021)
NEG	—	–0.144 (0.032)	—	—	0.205 (0.225)
SN	—	—	—	—	–0.009 (0.016)
EN	—	—	—	—	–0.006 (0.007)
HN	—	—	—	—	–0.280 (0.065)
constante	0.390 (0.075)	0.459 (0.076)	0.411 (0.084)	Y	0.411 (0.082)
R-cuadrado	0.335	0.341	0.332	0.321	0.347
SCR	610.0	605.1	555.7	Z	600.0
n	2,408	2,408	2,135	273	2,408

Se pide lo siguiente (sea lo más completo y explícito que pueda en su respuesta, y utilice en su caso las tablas que acompañan este ejercicio)

1. ¿Qué significado tiene la variable constante en el modelo 5?

2. Calcular los coeficientes V, W, X e Y de la regresión del Modelo 4 (no calcular los errores estándar), y Z, el SCR perdido, explicando tus cálculos.

3. Dar una interpretación del coeficiente NEG de la regresión del Modelo 2.

4. Realizar un test tipo F del poder explicativo de NEG, SN, EN, y HN en la regresión del Modelo 5. Utilice la tabla adecuada de entre las que se le han proporcionado.

5. Dar una interpretación a los coeficientes de NEG, y HN en la regresión del Modelo 5.

6. Explicar si un test tipo t sobre el coeficiente NEG en la regresión del Modelo 2 es suficiente para mostrar que las ecuaciones de salarios son diferentes para trabajadores de raza blanca y negra.

8

Tribunal de la Oposición al Cuerpo Superior de Estadísticos del Estado

Pruebas selectivas para el ingreso en el Cuerpo Superior de Estadísticos del Estado. Orden ECC/1517/2015, de 16 de Julio (BOE 27/07/2015).

7. La investigadora realiza dos regresiones adicionales, una en la que la variable dependiente está formada por los residuos de la regresión del Modelo 5 (RESID), y otra por su cuadrado (RESID2). Las variables explicativas son las de la regresión del Modelo 5. Los R-cuadrado de dichas respectivas regresiones son los siguientes:

   R-cuadrado (RESID) = 0.25

   R-cuadrado (RESID2) =0.26

   Teniendo en cuenta esta información, indique

   (i) para qué sería necesario llevar a cabo un análisis de los residuos;

   (ii) qué tipo de contraste podría plantear;

   (iii) calcúlelo, si es posible, con la información disponible; en caso contrario indique por qué no se podría llevar a cabo.


EX LIBRIS
CUERPOS D NUMEN
NUMERUS ESTADISTICA
INSTITUTO NACIONAL
DE
ESTADISTICA

10

Tribunal de la Oposición al Cuerpo Superior de Estadísticos del Estado

Pruebas selectivas para el ingreso en el Cuerpo Superior de Estadísticos del Estado. Orden ECC/1517/2015, de 16 de Julio (BOE 27/07/2015).

5.- Demografía.

1) Calcular los siguientes parámetros a partir de la tabla de datos (en miles) de un país, referidos a un año T cualquiera, que figura al final del enunciado.

   a) Tasas de migración exterior con destino el extranjero.

   b) Probabilidades de migrar al extranjero.

   c) Sedentarios de la tabla en relación con la migración exterior.

   d) Añada otro parámetro básico que pueda establecerse en una tabla de migraciones.

2) Calcule dos medidas agregadas relativas a la migración exterior a partir de los datos de la tabla. Comente sus resultados.

3) Qué medida de la intensidad inmigratoria exterior por grupos de edades puede elaborarse con los datos que suministran. Explique algunos elementos de esta medida: a) sus diferencias con una tabla de migración y b) su sentido probabilístico, así como otras consideraciones que resulten pertinentes.

4) Comente las principales magnitudes de la evolución demográfica para el total de la población mencionada durante el periodo.

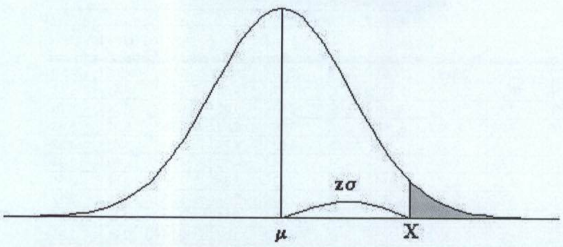
	Población al incio del año	Población al final del año	Defunciones	Migraciones entre provincias	Emigraciones hacia el extranjero	Inmigraciones desde el extranjero
0-4 años	2.422,8	2.320,4	1,5	24,6	25,0	18,0
5-9 años	2.440,5	2.478,1	0,2	21,1	21,6	13,8
10-14 años	2.226,7	2.267,6	0,2	16,1	20,4	16,0
15-19 años	2.165,6	2.140,7	0,4	18,2	23,7	23,4
20-29 años	5.343,3	5.121,9	1,5	108,9	125,7	79,4
30-39 años	7.761,9	7.484,1	3,8	137,9	163,9	57,1
40-49 años	7.522,8	7.547,3	10,8	71,5	90,0	34,3
50-59 años	6.079,9	6.212,0	23,7	37,3	43,2	22,0
60-64 años	2.502,3	2.492,8	17,6	13,5	11,7	9,8
65 y más años	8.262,1	8.442,9	330,9	40,1	22,6	17,3

11

EX LIBRIS
INSTITUTO NACIONAL
DE
ESTADISTICA

TABLA 1: DISTRIBUCIÓN NORMAL

Áreas bajo la curva normal



Ejemplo:

$$Z = \frac{X - \mu}{\sigma}$$

P [Z > 1] = 0.1587
P [Z > 1.96] = 0.0250

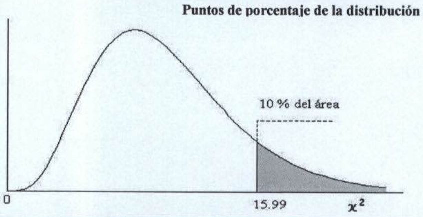
Desv. normal x	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641
0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010

TABLA 2: DISTRIBUCIÓN F

	F Snedecor (valores críticos al 5%)											
	Grados de libertad en el numerador											
g.l denominador	1	2	3	4	5	6	7	8	9	10	12	15
1	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5	241.9	243.9	245.9
2	18.51	19.00	19.16	19.25	19.3	19.33	19.35	19.37	19.38	19.4	19.41	19.43
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.70
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.86
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.51
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.01
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.85
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.72
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.53
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.46
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.40
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.42	2.35
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.38	2.31
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34	2.27
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.31	2.23
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.20
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.25	2.18
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.23	2.15
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.20	2.13
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.18	2.11
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.16	2.09
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.15	2.07
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20	2.13	2.06
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.12	2.04
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18	2.10	2.03
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.09	2.01
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.00	1.92
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.92	1.84
90	3.95	3.10	2.71	2.47	2.32	2.20	2.11	2.04	1.99	1.94	1.87	1.79
120	3.92	3.07	2.68	2.45	2.29	2.17	2.09	2.02	1.96	1.91	1.83	1.75

TABLA 3: DISTRIBUCIÓN χ^2

Puntos de porcentaje de la distribución χ^2



Ejemplo:
Para $\phi = 10$ grados de libertad

$P[\chi^2 > 15.99] = 0.10$

π / ϕ	0.995	0.99	0.975	0.95	0.9	0.75	0.5	0.25	0.1	0.05	0.025	0.01	0.005	π / ϕ
1	3.93E-05	1.57E-04	9.82E-04	3.93E-03	1.58E-02	0.102	0.455	1.323	2.71	3.84	5.02	6.63	7.88	1
2	1.00E-02	2.01E-02	5.06E-02	0.103	0.211	0.575	1.386	2.77	4.61	5.99	7.38	9.21	10.60	2
3	7.17E-02	0.115	0.216	0.352	0.584	1.213	2.37	4.11	6.25	7.81	9.35	11.34	12.84	3
4	0.207	0.297	0.484	0.711	1.064	1.923	3.36	5.39	7.78	9.49	11.14	13.28	14.86	4
5	0.412	0.554	0.831	1.145	1.610	2.67	4.35	6.63	9.24	11.07	12.83	15.09	16.75	5
6	0.676	0.872	1.237	1.635	2.20	3.45	5.35	7.84	10.64	12.59	14.45	16.81	18.55	6
7	0.989	1.239	1.690	2.17	2.83	4.25	6.35	9.04	12.02	14.07	16.01	18.48	20.3	7
8	1.344	1.647	2.18	2.73	3.49	5.07	7.34	10.22	13.36	15.51	17.53	20.1	22.0	8
9	1.735	2.09	2.70	3.33	4.17	5.90	8.34	11.39	14.68	16.92	19.02	21.7	23.6	9
10	2.16	2.56	3.25	3.94	4.87	6.74	9.34	12.55	15.99	18.31	20.5	23.2	25.2	10
11	2.60	3.05	3.82	4.57	5.58	7.58	10.34	13.70	17.28	19.68	21.9	24.7	26.8	11
12	3.07	3.57	4.40	5.23	6.30	8.44	11.34	14.85	18.55	21.0	23.3	26.2	28.3	12
13	3.57	4.11	5.01	5.89	7.04	9.30	12.34	15.98	19.81	22.4	24.7	27.7	29.8	13
14	4.07	4.66	5.63	6.57	7.79	10.17	13.34	17.12	21.1	23.7	26.1	29.1	31.3	14
15	4.60	5.23	6.26	7.26	8.55	11.04	14.34	18.25	22.3	25.0	27.5	30.6	32.8	15
16	5.14	5.81	6.91	7.96	9.31	11.91	15.34	19.37	23.5	26.3	28.8	32.0	34.3	16
17	5.70	6.41	7.56	8.67	10.09	12.79	16.34	20.5	24.8	27.6	30.2	33.4	35.7	17
18	6.26	7.01	8.23	9.39	10.86	13.68	17.34	21.6	26.0	28.9	31.5	34.8	37.2	18
19	6.84	7.63	8.91	10.12	11.65	14.56	18.34	22.7	27.2	30.1	32.9	36.2	38.6	19
20	7.43	8.26	9.59	10.85	12.44	15.45	19.34	23.8	28.4	31.4	34.2	37.6	40.0	20
21	8.03	8.90	10.28	11.59	13.24	16.34	20.3	24.9	29.6	32.7	35.5	38.9	41.4	21
22	8.64	9.54	10.98	12.34	14.04	17.24	21.3	26.0	30.8	33.9	36.8	40.3	42.8	22
23	9.26	10.20	11.69	13.09	14.85	18.14	22.3	27.1	32.0	35.2	38.1	41.6	44.2	23
24	9.89	10.86	12.40	13.85	15.66	19.04	23.3	28.2	33.2	36.4	39.4	43.0	45.6	24
25	10.52	11.52	13.12	14.61	16.47	19.94	24.3	29.3	34.4	37.7	40.6	44.3	46.9	25
26	11.16	12.20	13.84	15.38	17.29	20.8	25.3	30.4	35.6	38.9	41.9	45.6	48.3	26
27	11.81	12.88	14.57	16.15	18.11	21.7	26.3	31.5	36.7	40.1	43.2	47.0	49.6	27
28	12.46	13.56	15.31	16.93	18.94	22.7	27.3	32.6	37.9	41.3	44.5	48.3	51.0	28
29	13.12	14.26	16.05	17.71	19.77	23.6	28.3	33.7	39.1	42.6	45.7	49.6	52.3	29
30	13.79	14.95	16.79	18.49	20.6	24.5	29.3	34.8	40.3	43.8	47.0	50.9	53.7	30
40	20.7	22.2	24.4	26.5	29.1	33.7	39.3	45.6	51.8	55.8	59.3	63.7	66.8	40
50	28.0	29.7	32.4	34.8	37.7	42.9	49.3	56.3	63.2	67.5	71.4	76.2	79.5	50
60	35.5	37.5	40.5	43.2	46.5	52.3	59.3	67.0	74.4	79.1	83.3	88.4	92.0	60
70	43.3	45.4	48.8	51.7	55.3	61.7	69.3	77.6	85.5	90.5	95.0	100.4	104.2	70
80	51.2	53.5	57.2	60.4	64.3	71.1	79.3	88.1	96.6	101.9	106.6	112.3	116.3	80
90	59.2	61.8	65.6	69.1	73.3	80.6	89.3	98.6	107.6	113.1	118.1	124.1	128.3	90
100	67.3	70.1	74.2	77.9	82.4	90.1	99.3	109.1	118.5	124.3	129.6	135.8	140.2	100
Z_α	-2.58	-2.33	-1.96	-1.64	-1.28	-0.674	0.000	0.674	1.282	1.645	1.96	2.33	2.58	Z_α

Para $\phi > 100$ tómese $\chi^2 = \frac{1}{2}\left(Z_\alpha + \sqrt{2\phi - 1}\right)^2$. Z_α es la desviación normal estandarizada correspondiente al nivel de significancia y se muestra en la parte superior de la tabla.

E.2. Tercer examen del Cuerpo de Diplomados en Estadística del Estado

CDEE. Cuestiones 3er Ejercicio.

CUESTIÓN 1: El tiempo de retraso, medido en minutos, del AVE Madrid-Sevilla sigue una variable aleatoria continua con función de distribución:

$$F(x) = \begin{cases} 0 & si \quad x \leq -1 \\ k(x+1) + \frac{x^2-1}{2} & si \quad -1 < x \leq 0 \\ k(x+1) - \frac{x^2+1}{2} & si \quad 0 < x \leq 1 \\ 1 & si \quad x > 1 \end{cases}$$

1. Calcular el valor de k
2. Calcular la probabilidad de que el tren llegue antes de la hora prevista.
3. Calcular el tiempo esperado de retraso.
4. Calcular la probabilidad de que el tren llegue entre medio minuto de adelanto y un minuto de retraso.
5. Sabiendo que el tren ha llegado con retraso, calcular la probabilidad de que lo haya hecho menos de 15 segundos después de lo previsto.

CUESTIÓN 2: El tiempo de vida de un tipo de bombillas se sabe que es una variable aleatoria que sigue una distribución exponencial de parámetro $1/\Theta$ con $\Theta > 0$. Para estimar el tiempo medio de vida se extrae una muestra aleatoria simple de n bombillas y se consideran dos estimadores. Uno es la media muestral, $T_1(X_1, ..., X_n) = \frac{1}{n}\sum_{i=1}^{n} x_i$, y el otro se define como $T_2(X_1, ..., X_n) = min(X_1, ..., X_n)$. Se pide:

1. Calcular la media, el sesgo y la varianza del estimador T_1.
2. Calcular la media, el sesgo y la varianza del estimador T_2.
3. ¿Es T_2 insesgado? Si no lo es, define un estimador insesgado T_3 que sea función de él.
4. Entre los dos estimadores insesgados, T_1 y T_3 ¿cuál es más eficiente?

CUESTIÓN 3: Un fabricante de piezas de plomo tiene un stock de cien piezas clasificadas según su destino en 50 para Europa y 50 para el resto del mundo. Se desea estimar el peso total de las piezas y la única información disponible es relativa al peso (en toneladas) del envío en el año anterior tal y como viene en la tabla siguiente:

Destimo	*Media*	*Cuasivarianza*
Europa	6	4
Resto	4	9

Ignorando el factor de corrección para poblaciones finitas, se pide:

1

1. El fabricante decide seleccionar 10 piezas con probabilidades iguales y sin reemplazamiento ¿cuál es la varianza del estimador insesgado del peso total?
2. Un empleado de la fábrica le recomienda obtener estimaciones separadas según el destino de las piezas y deciden realizar un muestreo estratificado por destino y afijación proporcional ¿cuál es la varianza del nuevo estimador insesgado para el peso total?
3. Si el muestreo estratificado se llevase a cabo con afijación de mínima varianza ¿cuántas piezas se seleccionan según su destino?

CUESTIÓN 4: Dada la tabla adjunta de medidas de la cintura (en pulgadas) y del peso (en libras) de 10 individuos, se pide:

1. Calcular todos los momentos respecto al origen y a la media de órdenes 1 y 2 de la distribución bidimensional.
2. Calcular el coeficiente de correlación lineal de Pearson de ambas variables y comentar su significado.
3. Calcular los cuantiles c_p^{Peso} y c_p^{Cintura} de órdenes $p_i = 0, \frac{1}{3}, \frac{2}{3}, 1$ (i=0,..., 3) de sendas distribuciones unidimensionales.
4. Sean los intervalos $I_i^{\text{Peso}} = [c_{p_{i-1}}^{\text{Peso}}, c_{p_i}^{\text{Peso}})$ e $I_j^{\text{Cintura}} = [c_{p_{j-1}}^{\text{Cintura}}, c_{p_j}^{\text{Cintura}})$ con $i, j = 1, 2, 3$. Construir la tabla de contingencia del número de individuos con medidas de peso y cintura en cada par de intervalos $(I_i^{\text{Peso}}, I_j^{\text{Cintura}})$.
NB: $I_3^{\text{Peso}} = [c_2^{\text{Peso}}, c_3^{\text{Peso}}]$ e $I_3^{\text{Cintura}} = [c_2^{\text{Cintura}}, c_3^{\text{Cintura}}]$.
5. Calcular la distribución de frecuencias relativas del peso condicionada respecto a la cintura a partir de la tabla de contingencia anterior.

Cintura	32	36	38	33	39	40	41	35	38	38
Peso	175	181	200	159	196	192	205	173	187	188

CUESTIÓN 5: A partir de la información relativa al número total de asuntos judiciales resueltos en España (en miles), así como de las cantidades reconocidas a los trabajadores (en millones de euros) durante los años 2003, 2004 y 2005, que se muestra a continuación:

Materia objeto de la demanda	**Año 2003**		**Año 2004**		**Año 2005**	
	Cuantía	**Número**	**Cuantía**	**Número**	**Cuantía**	**Número**
Conflictos colectivos	43	1.6	94	2.1	1347	2
Despido	263	64	223	62	230	62
Reclamaciones por contrato de trabajo	249	147	200	139.4	213	127

Se pide calcular:

1. Los índices simples (base 2003 y en porcentaje) de las cantidades reconocidas a los trabajadores para las distintas materias objeto de la demanda.
2. La serie de números índice (base 2003 y en porcentaje) media aritmética ponderada de la cuantía de cantidades reconocidas a los trabajadores, utilizando como ponderaciones fijas el número de asuntos judiciales en 2003.
3. La repercusión de los conflictos colectivos en la variación del Índice de las cantidades reconocidas a los trabajadores afectados por asuntos judiciales resueltos en España, entre 2004 y 2005.
4. La participación porcentual de los conflictos colectivos en la variación del índice de las cantidades reconocidas a los trabajadores afectados por asuntos judiciales resueltos en España entre 2004 y 2005.
5. La suma de las participaciones en la variación del índice general.

CUESTIÓN 6: Utilizando los datos siguientes relativos a una economía imaginaria, calcule en términos del SEC2010:

1. El gasto en consumo final.
2. La formación bruta de capital fijo.
3. El saldo neto exterior de bienes y servicios.
4. El deflactor implícito del PIB para 2014 en base 2010 = 100
5. El valor añadido bruto a precios de mercado (o precios de adquisición).
6. Los impuestos netos sobre productos e importaciones.
7. Las subvenciones a los productos.
8. La remuneración de los factores de producción.
9. La remuneración de los asalariados.

Datos a precios corrientes de 2014 (salvo indicación contraria):

Consumo individual efectivo	1.200
Formación neta de capital	155
Exportaciones de bienes y servicios	400
Rentas de la propiedad y de la empresa	700
VAB a precios básicos	2.100
Consumo colectivo efectivo	600
PIBpm a precios constantes de 2010	2.000
Consumo de capital fijo	20
Otros impuestos netos sobre la producción	220
Variación de existencias	5
Impuestos sobre productos e importaciones	175
Adquisiciones menos cesiones de objetos valiosos	5
Gasto del Estado en armamento	25
Importaciones de bienes y servicios	150

CUESTIÓN 7: Con los siguientes datos en millones de euros de la CNE-Base 2010 para la economía española en 2010, calcule:

1. El Valor Añadido Bruto a coste de los factores (VAB cf)
2. El Valor Añadido Bruto a precios de mercado (VAB pm)
3. El Valor Añadido Bruto a precios básicos (VAB pb)
4. El Producto Interior Bruto a precios de mercado (PIB pm)

Datos

Producción Total de bienes y servicios a precios básicos	2.038.315
Consumos intermedios	1.048.042
IVA que grava los productos	58.458
Otros impuestos ligados a la producción	14.654
Impuestos sobre los productos, excluidos el IVA y los impuestos sobre las importaciones	37.199
Otras subvenciones a la producción	12.095
Subvenciones a los productos	6.297
Impuestos y derechos sobre las importaciones (excluido el IVA)	1.640

CUESTIÓN 8: En una economía con tres sectores de actividad y en ausencia de impuestos, subvenciones e importaciones, construya la tabla input-output simétrica con los siguientes datos:

1. La utilización de los consumos intermedios por parte de los sectores productivos se realiza de la siguiente manera:
   - El sector1 consume 10€ procedente del sector2; 4€ del sector3 y 12€ del sector1.
   - El sector2 consume 25€ procedente del sector1; 12€ del sector2 y 8€ del sector3.
   - El sector3 consume 17€ procedente del sector 1; 20€ del sector2 y 15€ del sector3.
2. El gasto en consumo final ha sido de 53€ en productos del sector1; 53€ del sector2 y 41€ del sector3.
3. La formación bruta de capital ha sido 28€ de productos del sector1; 30€ del sector2 y 18€ del sector3.
4. Las exportaciones han sido 38€ de productos de sector1; 15€ del sector2 y 5€ del sector3.
5. El factor capital utilizado por cada sector ha sido: 60, 38 y 23, respectivamente.

Interprete el significado económico de cada una de las cifras que encontramos en la columna correspondiente al sector2 y en la fila relativa al sector3.

CUESTIÓN 9: Se tienen los siguientes datos extraídos de las tablas de nacimientos y de la población española. Se pide:

1. Representar en el esquema de Lexis los datos de la tabla "Nacimientos. Año 2013".
2. Calcular la tasa de fecundidad específica para la generación de 1989.
3. Calcular la tasa de fecundidad específica para las mujeres de 24 años.

Nacimientos. Año 2013.

Edad y generación de la madre	Total nacimientos
22 años: 1991	3.066
22 años: 1990	3.219
23 años: 1990	3.656
23 años: 1989	3.799
24 años: 1989	4.259
24 años: 1988	4.463
25 años: 1988	5.184

Población residente en España

	1 de enero de 2013			1 de enero de 2014		
Edad	Ambos sexos	Hombres	Mujeres	Ambos sexos	Hombres	Mujeres
22 años	483.688	245.742	237.947	471.980	240.541	231.439
23 años	500.067	253.050	247.016	482.902	244.467	238.435
24 años	516.666	260.090	256.576	497.608	250.963	246.645
25 años	532.641	267.563	265.078	512.336	257.292	255.043
26 años	552.538	276.701	275.837	526.425	263.779	262.647

CUESTIÓN 10: A partir de la información incluida en la siguiente tabla referida al total de España y la Comunidad Autónoma de Andalucía para 2013, calcular:

1. La tasa de mortalidad infantil.
2. La tasa de mortalidad infantil neonatal.
3. La tasa de mortalidad infantil neonatal temprana.
4. La tasa de mortalidad infantil neonatal tardía.
5. La tasa de mortalidad infantil post-neonatal.

	Defunciones en 2013												
	Todos los menores de 1 año	Menos de 24 horas	De 1 día	De 2 días	De 3 días	De 4 días	De 5 días	De 6 días	De 7 a 13 días	De 14 a 20 días	De 21 a 27 días	Nacimientos 2013	Población media 0 años en 2013
España	1.164	293	65	62	42	29	43	23	140	68	41	425.715	437.015
Andalucía	256	65	13	11	5	13	13	6	31	11	11	81.470	82.012

Apéndice F

Exámenes de problemas de 2014

F.1. Segundo examen del Cuerpo Superior de Estadísticos del Estado

Tribunal de la Oposición al Cuerpo Superior en Estadísticos del Estado

Pruebas selectivas para el ingreso en el cuerpo superior de estadísticos del estado. Orden ECC/1384/2014, de 24 de julio (BOE 30/07/2014).

SEGUNDO EJERCICIO

1.- Estadística

Sea $(X_1,...,X_n)$ una muestra aleatoria simple de la distribución:

$$f_\theta(x) = \theta \exp(-\theta x) \quad x > 0$$

a) Calcular estimador de máxima verosimilitud para $P(X > t)$.

b) Supongamos que se miden 10 componentes cuya duración sigue la distribución anterior y que si se suma el tiempo de funcionamiento de todas ellas se contabiliza un total de 1200 horas. A partir de esta muestra, estimar la probabilidad de que una componente vaya a durar menos de 130 horas.

c) Probar si el estadístico $T(X_1,...,X_n) = \dfrac{n-1}{\sum_{i=1}^{n} X_i}$ es centrado para θ

d) Probar si $\sum_{i=1}^{n} X_i$ es estadístico suficiente y completo. ¿Es $T(X_1,...,X_n)$ estimador centrado de uniformemente mínima varianza?

1

2.- <u>Muestreo</u>

Sea un diseño estratificado con muestreo aleatorio simple sin reemplazamiento en cada estrato. Se dispone de H estratos de tamaño N_h, h=1,...H. El objetivo es estimar la media poblacional $\overline{Y}$ de una cierta característica Y. Sean $\overline{X}_h$, h=1,...H las medias poblacionales de los estratos de una característica auxiliar X, que suponemos conocidos.

a) Para estimar la media poblacional $\overline{Y}$ se propone el siguiente estimador:

$$\hat{\hat{Y}}_D = \hat{\hat{Y}}_{HT} + \overline{X} - \hat{\hat{X}}_{HT}$$

donde $\overline{X}$ es la media poblacional de X y $\hat{\hat{Y}}_{HT}, \hat{\hat{X}}_{HT}$ son los estimadores de Horvitz-Thompson de las medias de Y y X, respectivamente. Se pide:

a.1 Demostrar que $\hat{\hat{Y}}_D$ es un estimador insesgado para $\overline{Y}$

a.2 Calcular la varianza de $\overline{Y}_D$

a.3¿Cuál es la asignación óptima del tamaño muestral en h,n_h, cuando se minimiza la varianza de $\hat{\hat{Y}}_D$?

(Se supone que el coste por unidad muestral no depende del estrato)

a.4 ¿En qué condiciones $\hat{\hat{Y}}_D$ sería preferible a $\hat{\hat{Y}}_{HT}$?

b) Bajo las mismas condiciones anteriores, consideramos ahora el siguiente estimador:

$$\hat{\hat{Y}}_b = \hat{\hat{Y}}_{HT} + b(\overline{X} - \hat{\hat{X}}_{HT})$$

donde b es un número real fijo. ¿Cuál es el valor óptimo para b?

c) Ahora se supone que no conocemos los valores poblacionales de X ni se dispone de la población estratificada. Entonces, en primer lugar se obtiene una muestra aleatoria simple sin reemplazamiento, S*, de tamaño n* entre las N unidades de la población y calculamos el estimador de Horvitz-Thompson $\hat{\overline{X}}^*$, de la media poblacional $\overline{X}$ .Después, se obtiene una submuestra aleatoria simple sin reemplazamiento de n unidades a partir de S*.

2

Se propone el siguiente estimador:

$$\hat{\bar{Y}}_c = \hat{\bar{Y}} + c(\hat{\bar{X}}^* - \hat{\bar{X}}) \quad donde$$

$$\hat{\bar{Y}} = \frac{1}{n}\sum_{i=1}^{n} y_i ; \hat{\bar{X}} = \frac{1}{n}\sum_{i=1}^{n} x_i ; \hat{\bar{X}}^* = \frac{1}{n^*}\sum_{i=1}^{n^*} x_i$$

Demostrar que:

$$Var(\hat{\bar{Y}}_c) = \frac{N - n^*}{Nn^*} S_y^2 + \frac{n^* - n}{nn^*} S_u^2 \quad donde$$

$$u_i = y_i - cx_i ;$$

$$S_y^2 = \frac{\sum_{i=1}^{N} (y_i - \bar{y})^2}{N-1} \quad ; S_u^2 = \frac{\sum_{i=1}^{N} (u_i - \bar{u})^2}{N-1}$$

3

3.- Economía

La estimación del Producto Interior Bruto a precios de mercado (PIB) de España para el año 2013 en la nueva base 2010 de la Contabilidad Nacional ha sido de 1.049,2 miles de millones de euros (mm €).Se conocen también los siguientes datos porramas de actividadrelativos al Valor Añadido Bruto a precios básicos (VAB) y a la Remuneración de los Asalariados (RA) para el citado año (mm €):

	VAB	RA
Agricultura, ganadería, silvicultura y pesca	26,6	4,3
Industria	168,6	82,3
Construcción	55,1	27,6
Servicios	708,2	376,1

Para el conjunto de la economía, se tiene además la siguiente información (mm €):

Otros impuestos menos subvenciones sobre la producción __________9,6

Gasto en consumo final ______________________________________814,5

Formación bruta de capital ___________________________________198,9

Remuneración de los asalariados residentes por empleadores no residentes ____ 2,3

Remuneración de los asalariados no residentes por empleadores residentes ____ 0,2

Impuestos menos subvenciones sobre la producción y las importacionespagados al resto del mundo ___________________________________–4,6

Rentas netas de la propiedad pagadas al resto de mundo___________13,8

a) Se pide hallar los siguientes agregados:

a.1 Excedente de explotación bruto/Renta mixta bruta de la economía.

a.2 Saldo de intercambios exteriores de bienes y servicios.

a.3 Renta nacional bruta

b) Sabiendo que la estimación del PIB a precios corrientes para el año 2012 fue de 1.055,2 y que la variación interanual en volumen del PIB entre 2012 y 2013 ha sido de –1,2%:

b.1 Calcular la tasa de variación interanual del deflactor implícito de la economía.

b.2 Si se sabe que la variación interanual en volumen de los tres primeros trimestres del año 2013 fue respectivamente de –2,2%, –1,7% y –1,0%, ¿se podría conocer la variación interanual aproximada para el cuarto trimestre de dicho año?

4.- <u>Econometría</u>

Considere la siguiente expresión de la curva de Phillips a corto plazo aumentada con expectativas:

$$y_t - E_{t-1}(y_t) = \beta(u_t - u_t^*) + e_t \quad (1)$$

donde y_t es la inflación en el periodo t y $E_{t-1}(y_t)$ es el valor esperado en t-1 para la inflación en t, u_t es la tasa de paro observada en t, y u^* se refiere a la tasa natural de paro. Haga el supuesto de que la esperanza de inflación en t-1 es precisamente la inflación observada en t-1, de manera que,

$$y_t - y_{t-1} = \beta_1 + \beta_2 u_t + e_t \quad (2)$$

a) Justifique si el modelo (2) está anidado en (1), o si el modelo (1) está anidado en (2), o si, por el contrario, no son modelos anidados.

b) Con los datos de una economía europea para el periodo abarcado entre el primer trimestre de 1975 y el último de 1990, se obtiene la siguiente estimación del modelo (2) para los parámetros β_1, β_2:

Variables	*Coeficiente*	*Desv. típica*	*Estadístico t*	*valor p*
Constante	1,4126	2,32829	0,6067	0,54629
"paro"	-0,205876	0,323353	-0,6367	0,52671

Media de la var. dependiente = -0,0432317
Desviación típica de la var. dependiente. = 3,46601
Suma de cuadrados de los residuos = 739,902
Desviación típica de los residuos = 3,48275
R^2 = 0,0066016
R^2 corregido = -0,00968362
Grados de libertad = 61
Estadístico de Durbin-Watson = 2,61265
Coef. de autocorr. de primer orden. = -0,339296
Log-verosimilitud = -166,99
Criterio de información de Akaike = 337,979
Criterio de información Bayesiano de Schwarz = 342,266
Criterio de Hannan-Quinn = 339,665

A partir del modelo estimado, estime la tasa natural de paro.

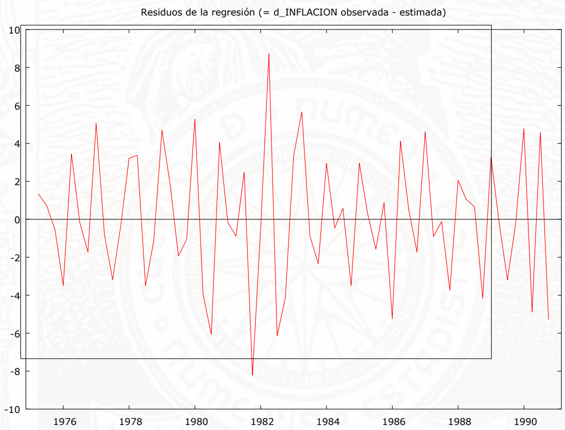
c) La tasa de paro estimada es una función no lineal de los parámetros estimados. Utilice el método delta para estimar su varianza. Para ello considere la siguiente matriz de covarianzas de los coeficientes estimados. Posteriormente, construya un intervalo de confianza del 95% para dicha estimación. Utilice el valor crítico de 2.

Matriz de covarianzas de los coeficientes

constante	“paro”	
5,42093	-0,739369	constante
	0,104557	“paro”

d) Un investigador independiente, con datos para las mismas variables y la misma economía, pero para el periodo 1950-2000, obtiene una estimación de la tasa natural de 5.46%. Contraste si la obtenida en este ejemplo es significativamente diferente y, en su caso, trate de explicar la diferencia o semejanza.

e) Los residuos del modelo estimado (modelo (2)) son los siguientes:



Sobre estos residuos se realiza un contraste estadístico de Breusch y Godfrey. El resultado del mismo es

Contraste de Breusch-Godfrey para autocorrelación hasta el orden 4estimaciones MCO
utilizando las 59 observaciones 1976:2-1990:4
Variable dependiente: ehat

VARIABLE	COEFICIENTE	DESV.TÍP.	ESTAD T	VALOR P
const	1,52799	1,97027	0,776	0,44148
PARO	-0,206071	0,276687	-0,745	0,45970
ehat_1	-0,695497	0,139759	-4,976	<0,00001 ***
ehat_2	-0,763386	0,165642	-4,609	0,00003 ***
ehat_3	-0,340544	0,163138	-2,087	0,04167 **
ehat_4	-0,119786	0,148330	-0,808	0,42295

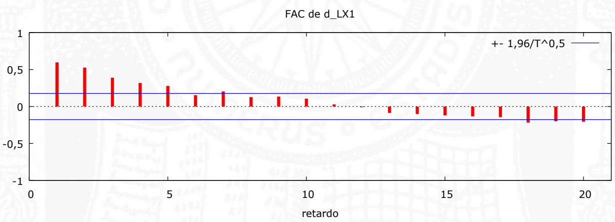
R-cuadrado = 0,426673
Estadístico de contraste: LMF = 10,249645, con valor p = P(F(4,53) > 10,2496) = 3,2e-006

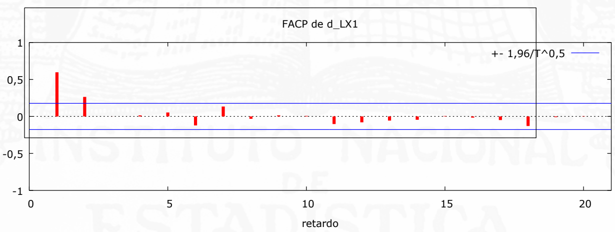
Estadístico alternativo: TR^2 = 25,173689, con valor p = P(Chi-cuadrado(4) > 25,1737) = 4,64e-005

Ljung-Box Q' = 20,3203 con valor p = P(Chi-cuadrado(4) > 20,3203) = 0,000432

Explique (i) en qué consiste el contraste (ii) qué conclusión obtiene sobre el modelo, y (iii) conteste justificadamente a la siguiente pregunta: ¿los resultados obtenidos eran esperados a la luz del gráfico de los residuos del modelo?

f) Considere ahora que es pertinente modelizar los residuos. Para ello realiza un análisis de las funciones de autocorrelación, que proporcionamos a continuación:





Y responda a las siguientes cuestiones: ¿Cómo modelizaría los residuos? ¿Qué modelo plantearía ahora para estimar el modelo (2)?

8

g) Un colega le proporciona la siguiente estimación del modelo (2)

VARIABLE	COEFICIENTE	DESV.TÍP.	ESTAD T	VALOR P
const	1,34076	0,683277	1,962	0,05462
PARO	-0,191344	0,0955573	-2,002	0,05001

Estimaciones de los coeficientes AR:

e_1	-0,678453	0,136016	-4,988	<0,00001
e_2	-0,740201	0,160427	-4,614	0,00002
e_3	-0,314694	0,157269	-2,001	0,05034
e_4	-0,100201	0,144001	-0,696	0,48946

Suma de los coeficientes AR = -1,83355
Estadísticos basados en los datos rho-diferenciados:
Suma de cuadrados de los residuos = 419,857
Desviación típica de los residuos = 2,71402
R-cuadrado = 0,424775
R-cuadrado corregido = 0,414683
Grados de libertad = 57
Estadístico de Durbin-Watson = 1,85502
Coef. de autocorr. de primer orden. = 0,0245641
Criterio de información de Akaike (AIC) = 287,215
Criterio de información Bayesiano de Schwarz (BIC) = 291,37
Criterio de Hannan-Quinn (HQC) = 288,837

Responda justificadamente a las siguientes cuestiones: ¿Considera que su colega está utilizando la información del correlograma proporcionada en el apartado anterior? ¿Diría que el modelo ha mejorado respecto del primer modelo estimado en el apartado b)? ¿De qué manera se le ocurre mejorar el modelo? ¿Puede decir que se trata de una estimación de mínimos cuadrados generalizados?

9

5.- Demografía

La oficina estadística de un país facilita el siguiente repertorio de informaciones para un año T todo él referido a la población de mujeres (cifras en miles):

Todos los datos referidos a la población femenina en el año T							
				Migraciones exteriores		Migraciones interiores	
Cantidades en miles	Población a 1 de Enero	Nacimientos por edad de la madre	Defunciones por edad del difunto	Flujo emigración con destino al extranjero	Flujo inmigración con origen en el extranjero	Flujo de migraciones entre provincias	Población a 31 de Diciembre
TOTAL	23.710,2	206,6	190,6	239,2	146,1	232,1	23.633,7
0-4 años	1.172,5	-	0,7	12,0	8,6	11,8	1.123,8
5-9 años	1.184,1	-	0,1	10,5	6,7	10,3	1.200,4
10-14 años	1.080,6	0,1	0,1	9,9	7,5	7,9	1.102,2
15-19 años	1.051,1	4,3	0,1	11,0	11,4	9,3	1.038,5
20-29 años	2.645,4	53,7	0,4	58,7	43,5	56,0	2.538,2
30-39 años	3.795,6	134,4	1,3	64,8	26,7	61,2	3.681,0
40-49 años	3.705,0	14,1	3,7	34,6	16,1	29,5	3.719,6
50-59 años	3.067,9	-	7,6	19,5	11,6	16,8	3.134,2
60-64 años	1.286,7	-	5,2	6,0	5,1	6,3	1.281,0
65 y más años	4.721,3	-	171,4	12,2	8,9	23,0	4.814,8

Se consideran como migraciones interiores todos y cada uno de los cambios de residencia permanente que traspasan la frontera de una provincia y tienen como destino otra provincia del mismo país, sin distinción de orden de la migración u otra tipología cualquiera.

A la vista del anterior enunciado resuelva los siguientes asuntos, formulando y razonando cada uno de los extremos:

a) Calcule las tasas de migración interior para cada grupo de edades de mujeres.

b) Calcule las funciones de una tabla de migración interior femenina de acuerdo a los datos del mencionado año T, considerando que en el registro de las migraciones la incidencia de migraciones repetidas realizadas por un mismo individuo durante dicho período se considera empíricamente irrelevante:

b.1 Las probabilidades de migrar dentro del país entre provincias diferentes.

b.2 Sedentarios de la tabla de migración interprovincial.

b.3 Migraciones de la generación ficticia de la tabla.

c) De acuerdo a la anterior tabla de migración: ¿cuál es la intensidad del fenómeno demográfico migración interior interprovincial?

d) ¿Cuál es la edad media a la migración?, ¿bajo qué hipótesis se calcula?

e) Otros comentarios generales a la vista de los flujos demográficos que se presentan.

F.2. Tercer examen del Cuerpo de Diplomados en Estadística del Estado

INSTITUTO NACIONAL DE ESTADISTICA

Tribunal de la Oposición al Cuerpo de Diplomados en Estadística del Estado

Pruebas selectivas para el ingreso en el Cuerpo de Diplomados en Estadística del Estado. Orden ECC/1384/2014, de 24 de julio (BOE 30/07/2014).

TERCER EJERCICIO

1. Sea X una variable aleatoria que representa la producción semanal de ácido nítrico de una fábrica medido en miles de m3. La función de densidad de esta variable viene dada por: $f(x) = k(1-x)^3$ si $x \in (0,1)$. Calcular:

   a) El valor de k para que $f(x)$ sea función de densidad.

   b) La capacidad de los depósitos para que la probabilidad de que los depósitos se desborden sea 0,001.

   c) Si el tiempo de vida, medido en meses, de los peces está relacionado con la producción de ácido nítrico según la siguiente función: $V = 12e^{-x}$, calcular la probabilidad de que un pez viva más de tres meses y la vida media de los peces.

---

2. Sea (X_1, X_2) una muestra aleatoria simple de tamaño 2 de la variable aleatoria X que sigue una distribución Normal con media cero y varianza $1/\theta$ siendo θ un parámetro desconocido. Consideramos la siguiente función de la muestra

   $T(X_1, X_2) = 1/2(X_1^2 + X_2^2)$

   Se pide:

   a) Calcular la distribución de $2\theta T(X_1, X_2)$.

   b) Hallar un intervalo de confianza basado en $2\theta T(X_1, X_2)$ para el parámetro θ y para un nivel $1-\alpha$ con $\alpha \in (0,1)$.

---

3. Los taxis en servicio de una ciudad están numerados del 1 al N y se desea conocer cuántos taxis hay. Para ello se observa una muestra de n taxis y se apuntan sus números. Se pide:

   a) Obtener un estimador por el método de los momentos. ¿Es insesgado?.

   b) Obtener un estimador por el método de máxima verosimilitud.

---

4. Sean los siguientes valores de altura (en centímetros) y de peso (en kilogramos) de un conjunto de 20 personas:

Altura	164	164	166	167	169	169	169	169	170	170	171	171	171	173	173	173	174	174	175	177
Peso	68	70	69	69	69	70	69	67	71	69	70	71	76	71	70	68	70	74	71	70

1

Se pide:

a) El valor de la media de la distribución de alturas.

b) El valor de la desviación típica de la distribución de alturas.

c) El valor de la mediana de la distribución de alturas.

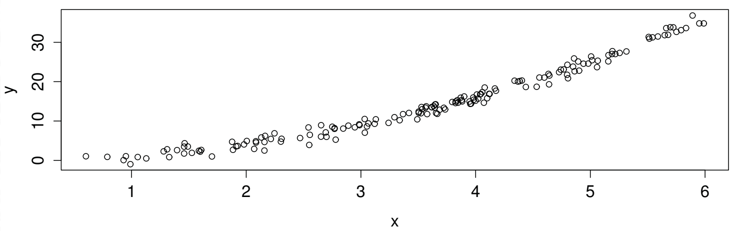
d) Agrúpense los datos en intervalos con una longitud razonadamente escogida y calcúlese la mediana de la distribución de datos agrupados.

e) El intervalo que contiene el 40 % central de la distribución de valores agrupados de alturas construida en el punto anterior.

f) Sean las variables binarias A y P que indican, respectivamente, si una persona tiene una altura superior a la mediana de las alturas y un peso superior a la mediana de los pesos. ¿Son A y P independientes? Demuéstrese.

---

5. Sean x e y variables cuantitativas con $n = 162$ pares de valores representados en la siguiente gráfica:



Se tiene el siguiente conjunto de cantidades:

$$\sum_{k=1}^{n} x_k = 592.4786 \qquad \sum_{k=1}^{n} y_k = 2469.627$$

$$\sum_{k=1}^{n} x_k^2 = 2469.245 \qquad \sum_{k=1}^{n} x_k y_k = 11111.01$$

$$\sum_{k=1}^{n} x_k^3 = 11093.92 \qquad \sum_{k=1}^{n} x_k^2 y_k = 52449.87$$

$$\sum_{k=1}^{n} x_k^4 = 52351.37 \qquad \sum_{k=1}^{n} x_k^3 y_k = 256396.9$$

$$R^2_{y\cdot x} = 0.9745407, \quad R^2_{y\cdot xx^2} = 0.9950063 \qquad \sum_{k=1}^{n} y_k^2 = 51455.63$$

$$X^{(1)} = [\mathbb{1}\ \mathbf{x}] \quad H^{(1)} = X^{(1)} \cdot (X^{(1)T} X^{(1)})^{-1} X^{(1)T} \qquad \mathbf{y}^T(\mathbb{I}_n - H^{(1)})\mathbf{y} = 759.0528$$

$$X^{(2)} = \left[\mathbb{1}\ \mathbf{x}\ \mathbf{x}^2\right] \quad H^{(2)} = X^{(2)} \cdot (X^{(2)T} X^{(2)})^{-1} X^{(2)T} \qquad \mathbf{y}^T(\mathbb{I}_n - H^{(2)})\mathbf{y} = 150.4276$$

Se pide:

2

a) Los coeficientes de la recta $y = \beta_0 + \beta_1 x$ ajustada por mínimos cuadrados.

b) La varianza residual del modelo anterior.

c) La varianza residual del modelo $y = \beta_0 + \beta_1 x + \beta_2 x^2$ ajustado por mínimos cuadrados.

d) A partir de los valores de las varianzas residuales obtenidos anteriormente, arguméntese qué modelo es más apropiado.

e) En un diagrama con ejes de abscisas los valores predichos $\hat{y}_k$ y de ordenadas los residuos e_k, indíquese cualitativamente cómo se espera que sea la relación entre $\hat{y}_k$ y e_k.

---

6. Las empresas del sector informático de cierta región facturaron durante los años 2001, 2002 y 2003 las cantidades que se indican y a los precios que figuran en la siguiente tabla:

		Ordenadores de sobremesa		Ordenadores portátiles				
t	Año	p_{1t}	q_{1t}	p_{2t}	q_{2t}	$\sum_{i=1}^{2} p_{it}q_{i0}$	$\sum_{i=1}^{2} p_{i0}q_{it}$	$\sum_{i=1}^{2} p_{it}q_{it}$
0	2001	750	30	1100	15	39000	39000	39000
1	2002	805	31	1150	20	41400	45250	47955
3	2003	820	40	1175	25	42225	57500	62175

N.B.: p_{it} y q_{it} denotan precio (en euros) por unidad y cantidad vendida, respectivamente, del producto i en el período de tiempo t.

a) Construir, con base 2001, los índices de precios y cantidades de Laspeyres, Paasche y Fischer para el año 2002.

b) Calcular, con base 2001, el índice de valor para el año 2002 a partir de los índices anteriores.

c) Hallar la repercusión de los ordenadores portátiles en la variación del índice de precios de Laspeyres entre los años 2002 y 2003. Conocemos que la ponderación de los ordenadores portátiles es de 42,31 % y es constante en el tiempo.

---

7. Conociendo los siguientes saldos de una Balanza de Pagos (MBP6), en m.m. de euros:

Exportaciones de bienes FOB	166,0
Exportaciones de servicios	25,7
Importaciones de bienes FOB	162,9
Importaciones de servicios	6,4
Transferencias de capital	3,7
Adquisición / Enajenación de activos no financieros no producidos	0,3
Rentas primarias y secundarias	-16,7
Errores y omisiones	1,9

Calcular:

3

a) El saldo de la Cuenta de Bienes y Servicios.

b) El saldo de la Cuenta Corriente y explicar el significado económico de ese saldo.

c) El saldo de la Cuenta de Capital.

d) La Capacidad (+) Necesidad (-) de financiación y su significado económico.

e) El saldo de la Cuenta Financiera.

---

8. A partir de la siguiente información sobre una economía

Importación de bienes y servicios(M)	3600
Exportación de bienes y servicios (X)	2900
Impuestos sobre productos	250
Formación bruta de capital (FBC)	2000
Consumo intermedio	11000
Gasto en consumo individual (CF individual)	6900
Gasto en consumo colectivo (CF colectivo)	3580
Consumo de capital fijo	150
Rentas de la propiedad recibidas del exterior	220
Rentas de la propiedad pagadas al exterior	230
Rentas de los asalariados recibidas del exterior	160
Rentas de los asalariados pagadas al exterior	180
Impuestos netos sobre producción e importaciones	900
Remuneraciones de asalariados	8500
Transferencias corrientes netas del exterior	35
Transferencias de capital netas del exterior	185

Calcule:

a) El PIB a precios de mercado.

b) El Excedente de explotación / Renta mixta bruto.

c) El Producto interior neto a precios de mercado.

d) La Renta Nacional Bruta a precios de mercado.

e) La Renta Nacional Disponible Bruta a precios de mercado.

f) La capacidad o necesidad de financiación de esta economía.

4

9. Se dispone de la siguiente tabla de nacimientos abreviada y de las cifras de población.

Nacimientos. Año 2013.

Total	425.494
De 15 a 19 años	8.816
De 20 a 24 años	32.251
De 25 a 29 años	78.929
De 30 a 34 años	155.810
De 35 a 39 años	120.720
De 40 a 44 años	27.313
De 45 a 49 años	1.655

Población residente. Mujeres.

Edad	1 de enero de 2013	1 de enero de 2014
De 15 a 19 años	1.051.130	1.038.329
De 20 a 24 años	1.201.940	1.166.856
De 25 a 29 años	1.443.413	1.372.353
De 30 a 34 años	1.811.757	1.711.917
De 35 a 39 años	1.983.801	1.970.685
De 40 a 44 años	1.886.012	1.890.390
De 45 a 49 años	1.818.983	1.829.884

Calcular, siempre que sea posible:

a) Edad media a la maternidad.

b) Tasa global de fecundidad.

c) Tasa global de natalidad.

d) Tasa de fecundidad específica para el grupo 15-19 años.

10. El número de nacimientos de varones en una determinada región en 2013 fue de 42.104. Además, se tienen estos datos de defunciones y población de varones:

	Población		**Defunciones**
	Varones		Varones
Edad	01-ene-2013	01-ene-2014	Año 2013
0 años	42.634	41.539	137
1-4 años	191.260	184.086	27
5-9 años	249.865	252.464	28
10-14 años	230.342	233.090	17
15-19 años	226.043	224.299	51
20-24 años	253.177	248.243	99

Rellene todos los datos del siguiente extracto de tabla de mortalidad abreviada. Suponer que las defunciones se reparten de forma homogénea, es decir, $a_x = 0,5$ para todas las edades.

Edad	Supervivientes a edad exacta x	Defunciones teóricas con edad $(x, x+n)$	Riesgo o probabilidad de muerte con edad $(x, x+n)$	Tasas específicas de mortalidad $m(x, x+n)$	Población estacionaria ${}_nL_x$
0	100.000				
1					
5					
10					
15					
20					

6

Apéndice G

Galeria de imágenes

G.1. Autoría y referencias bibliográficas de las imágenes portada de cada capítulo

1. Imágenes de la portada del capítulo **??** (Año 2022):

   - Dibujo del número pi con los dígitos coloreados:
   https://www.reddit.com/r/dataisbeautiful/comments/7apa1y/2500_digits_of_pi_coloured_oc/

   - Imágenes de los simpsons:

   1) Imagen (2,2): https://experimentalmath.info/blog/category/disclaimer-and-copyright/ El copyright es de David H. Bailey ©2023.

   2) Imagen (1,2): http://www.numberphile.com/

   3) Imagen (2,1): https://t.me/Crypto_Trendingz

   4) Imagen (1,1): https://www.pinterest.es/pin/859483910111063672/
   drawthesimpsons.tumblr.com

2. Imagen de la portada del capítulo **??** (Año 2022):

   La imagen la hemos tomado de la web:

- https://www.revistadearte.com/2022/01/23/los-maestros-antiguos-cara-a-cara-en-el-museo-picasso-malaga/

Los derechos de esta imagen son de:

Pablo Picasso (1881-1973)

Busto de hombre

Mougins, 8 noviembre 1970

Óleo sobre contrachapado

96,5 x 57,5 cm /

Fundación Almine y Bernard Ruiz-Picasso para el Arte, Madrid

©FABA Foto: Soko-studio

©Sucesión Pablo Picasso, VEGAP, Madrid, 2021

EX LIBRIS
CUERPOS D · NUMEN · ESTADISTICA · NUMERUS
INSTITUTO NACIONAL
DE
ESTADISTICA

πάντα ρεῖ

Heráclito

9 788409 557516