



MINISTERIO
DE ECONOMÍA, COMERCIO
Y EMPRESA

INE
Instituto Nacional de Estadística

Oposición al Cuerpo Superior de Estadísticos de Estado

Segundo Ejercicio

Convocatoria de la oferta pública de empleo de 2024

Resolución de 22 de diciembre de 2024, de la Subsecretaría, por la que se convocan procesos selectivos para ingreso, por el sistema de acceso libre y promoción interna, en el Cuerpo Superior de Estadísticos del Estado. (BOE 31 de Diciembre de 2024)

Acceso Promoción Interna, CDEE

Cuestión 1. Se considera una población finita de tamaño N , compuesta por unidades ordenadas de acuerdo con una variable de interés definida por:

$$y_k = k, \quad \text{para } k = 1, 2, \dots, N.$$

donde los parámetros poblacionales de esta variable son:

- Media: $\mu_y = \frac{N+1}{2}$,
- Varianza: $\sigma_y^2 = \frac{(N-1)(N+1)}{12}$.

Además, se supone que $N = n \cdot a$, con $n, a \in \mathbb{N}$.

El objetivo es analizar la varianza del estimador de Horvitz-Thompson de la media poblacional bajo distintos esquemas de muestreo. Para ello, se plantea:

- (a) Suponiendo que se selecciona una muestra aleatoria simple sin reemplazamiento de tamaño n , deduzca la expresión de la varianza del estimador de Horvitz-Thompson de la media poblacional en este caso.
- (b) La población se divide en n estratos consecutivos de tamaño a , de acuerdo al orden según la variable de interés, extrayéndose una unidad aleatoriamente por estrato. Deduzca la varianza del estimador de Horvitz-Thompson de la media poblacional en el muestreo estratificado descrito, sabiendo que la varianza poblacional dentro de cada estrato es $\sigma_i^2 = \frac{a^2-1}{12}, \forall i = 1, \dots, n$.
- (c) Suponer que se selecciona una muestra de tamaño n mediante muestreo sistemático con probabilidades iguales aplicado sobre la población ordenada de acuerdo a la variable de interés. Obtenga la varianza del estimador de Horvitz-Thompson de la media poblacional en este caso.
- (d) Compare los tres planes de muestreo y justifique cuál resulta más adecuado en la situación planteada.

Observación:

Sea $n \in \mathbb{N}$. Entonces se verifican las siguientes identidades:

- La suma de los primeros n números naturales está dada por:

$$\sum_{k=1}^n k = \frac{n(n+1)}{2}$$

- La suma de los cuadrados de los primeros n números naturales está dada por :

$$\sum_{k=1}^n k^2 = \frac{n(n+1)(2n+1)}{6}$$

Cuestión 2. Un investigador está analizando la duración (en horas) de cierto tipo de bombillas LED. Se asume que la duración de las bombillas sigue una distribución exponencial con parámetro λ . Se toma una muestra aleatoria simple de tamaño $n = 10$, con los siguientes datos (en horas):

$$\{88, 102, 96, 91, 110, 85, 99, 107, 95, 100\}$$

El fabricante afirma que el tiempo medio de vida es de al menos 100 horas.

- (a) Obtenga en el contexto de este problema, el valor del parámetro λ y qué significado tiene el mismo.
- (b) Formule el contraste de hipótesis para comprobar si los datos contradicen la afirmación del fabricante utilizando un contraste de razón de verosimilitudes.
- (c) Calcule la función de verosimilitud bajo la hipótesis nula y bajo la alternativa (en este caso usa el estimador máximo verosímil)
- (d) Calcule el estadístico de razón de verosimilitudes y determine si se rechaza H_0 al nivel $\alpha = 0,05$.
- (e) Interprete el resultado en el contexto del problema.

Cuestión 3. .

- (a) Sea un sistema de representación binario de números enteros con **8 bits en complemento a dos**. Indicar detalladamente como realizar la operación: **45 – 84** en este sistema, así como el resultado final.
- (b) Se quiere expresar el número entero **-315** en un sistema binario en complemento a dos, ¿cuál es el número **mínimo** de bits que se necesitan?
- (c) Se quiere representar el número **6.25** en un sistema binario normalizado de números **punto flotante de 32 bits**. El primer bit es el del signo, los siguientes 8 bits representan el exponente (con un sesgo de 127) y los restantes bits la mantisa. Indicar detalladamente los pasos para representarlo y el resultado final .

Cuestión 4. Las autoridades municipales desean conocer la utilización de un nuevo tramo del carril bici.

Para ello, se empleó un muestreo por conglomerados sin reemplazamiento y con probabilidades iguales. En concreto, se seleccionaron aleatoriamente 4 semanas del total de 52 que componen el año. Durante cada una de las semanas seleccionadas, se registró diariamente el número de bicicletas que circularon por un punto específico del nuevo tramo del carril bici.

Los datos recogidos se presentan en la siguiente tabla:

Semana	Lunes	Martes	Miércoles	Jueves	Viernes	Sábado	Domingo	Total Semana
1	20	18	15	17	22	25	30	147
2	12	10	13	11	14	15	17	92
3	14	16	12	13	15	20	21	111
4	15	13	10	12	14	17	18	99

Se pide:

- (a) Una estimación insesgada del total de usuarios anuales del nuevo tramo del carril bici y su error de muestreo relativo o coeficiente de variación. Calcule también un intervalo de confianza del 95 % para dicho total. Considerar $z_{\frac{\alpha}{2}} = 2$.
- (b) Dichas autoridades estarían interesadas en estimar el número medio diario de usuarios del carril bici en un año futuro utilizando el mismo diseño muestral. Con el objetivo de garantizar que la varianza del estimador del número medio de usuarios diarios sea menor que uno, se pide determinar el número mínimo de semanas que deben seleccionarse.

Cuestión 5. Sean los siguientes modelos

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u \quad (\text{Modelo 1})$$

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_2^2 + u \quad (\text{Modelo 2})$$

$$\log(y) = \beta_0 + \beta_1 x_1 + \beta_2 \log(x_2) + u \quad (\text{Modelo 3})$$

donde las variables y, x_1, x_2, u son aleatorias tales que los modelos están correctamente definidos.

Se pide:

- (a) Considere que $\mathbb{E}(u|x_1, x_2) = 0$, y obtenga el efecto parcial de x_2 sobre $E(y|x_1, x_2)$ para los modelos 1 y 2.
- (b) Considere que $\mathbb{E}(u|x_1, x_2) = 0$, e interprete el significado de β_1 y β_2 en el modelo 3 (justifique la respuesta).
- (c) Obtenga, para el Modelo 1, $\mathbb{E}(u)$ sabiendo que $\mathbb{E}(u|x_1, x_2) = k$, $k \in \mathbb{R}$
- (d) Considere que $\mathbb{E}(u|x_1, x_2) = 0$, obtenga (si es posible con la información facilitada) y para el Modelo 1 las siguientes covarianzas: $\text{cov}(u, x_1)$, $\text{cov}(u, x_2^2)$.
- (e) Explique el significado y alcance de $\mathbb{E}(u|x_1, x_2) = 0$ en el contexto de la modelización.

Cuestión 6. El ayuntamiento de una ciudad tiene como objetivo estudiar el nivel de actividad de sus comercios locales y para ello ha diseñado una muestra probabilística. Cada comercio seleccionado ha respondido a un cuestionario y finalmente se dispone de los siguientes datos:

Variable	Tipo	Descripción
ID_Comercio	Alfanumérico	Código único de cada comercio. El identificador puede ser cualquier cadena de caracteres que asegure su unicidad, de 4 posiciones
Mes	Entero (1–12)	Mes de referencia de la actividad, donde 1 = Enero, 2 = Febrero, ..., 12 = Diciembre.
Actividad	Binaria (0,1)	Indica si el comercio ha tenido actividad durante el mes (1 = Sí, 0 = No).
Ventas_Tienda (€)	Numérica	Ventas realizadas en el local físico. El valor debe ser 0 o un número positivo.
Ventas_Online (€)	Numérica	Ventas realizadas a través de internet. El valor debe ser 0 o un número positivo.
Ventas_Totales (€)	Numérica	Total de ventas del mes,incluyendo las realizadas en tienda física y online . El valor debe ser 0 o un número positivo.
Gastos (€)	Numérica	Gastos del negocio durante el mes, incluyendo los gastos relacionados con los empleados (salarios, seguros sociales, etc.), alquiler, suministros, entre otros. El valor debe ser 0 o un número positivo.
Empleados	Entero	Número de empleados contratados. El valor debe ser 0 o un número entero positivo.

Se pide:

- (a) Establecer 5 controles de validación (*edits*) para los datos recogidos en la encuesta, cubriendo al menos uno de cada uno de los siguientes tipos: rango, balance y bivariante.
Se debe:
 - Escribir la regla de validación como una condición lógica o en pseudocódigo, determinando si es de rango, balance y bivariante.
 - Proporcionar un ejemplo donde se incumpla la regla y explicar brevemente por qué sería un error.
- (b) En los datos recogidos en la encuesta, se ha observado la existencia de valores faltantes en algunas variables de ciertos cuestionarios. Suponiendo que la información recogida corresponde a cinco comercios (registros), se requiere imputar los valores faltantes (NA) utilizando el método indicado y los datos disponibles en la siguiente tabla.

ID	Mes	Actividad	Ventas_Tienda	Ventas_Online	Ventas_Totales	Gastos	Empleados
C001	1	1	1200	300	1500	1100	3
C002	1	1	1000	NA	NA	850	2
C003	1	0	0	0	0	0	0
C004	1	1	1200	400	1600	NA	4
C005	1	1	1000	400	1400	1300	5

Métodos de imputación a aplicar:

- **C002 - Ventas_Online:** imputación por la media.
- **C004 - Gastos:** vecino más próximo o cercano.

Indicar para cada caso el valor imputado y los pasos llevados a cabo para obtener dicho valor.

- (c) Suponga que se está valorando la posibilidad de difundir un fichero de microdatos que incluya las variables ventas totales , número de empleados e identificador del comercio para su explotación por parte de usuarios externos. Proponga al menos dos medidas concretas de control del secreto estadístico que podrían aplicarse sobre dichas variables antes de la difusión del fichero.