

INSTITUTO NACIONAL DE ESTADÍSTICA

EPIN



Índice de Precios del Trabajo (IPT)

Metodología

Madrid, 2020.

Índice

1	Objetivos	3
2	Fuentes de información utilizadas	4
3	Ámbitos de la operación estadística	6
4	Variables de clasificación	7
5	Metodología de cálculo	10
5.1	Fórmula de cálculo del índice	10
5.2	Periodo base	11
5.3	Periodo de referencia de las ponderaciones	11
5.4	Cálculo de la ganancia hora	11
6	Difusión	14

1. Objetivos

Un índice de precios del trabajo (IPT) es un indicador cuyo objetivo es medir el cambio en el precio de la mano de obra en el tiempo como consecuencia exclusivamente de las presiones del mercado laboral, es decir, sin que dicha medida esté afectada por cambios en la calidad y cantidad de trabajo realizado (por ejemplo, por cambios en la composición de la fuerza de trabajo, en el número de horas trabajadas, en las características de los asalariados, etc.) o, lo que es lo mismo, descontando el efecto composición¹.

Por analogía con el índice de precios al consumo que mide los cambios en el precio de una *cesta básica* estandarizada de bienes de consumo y servicios, el IPT debe medir los cambios en el precio de una *cesta básica* estandarizada de puestos de trabajo.

Para medir el cambio en el precio puro de un grupo de ítems entre dos períodos, deben cumplirse las dos condiciones siguientes: en primer lugar, el conjunto de elementos básicos debe ser el mismo, y en segundo lugar, las características de todos los ítems deben mantenerse inalteradas. Estas dos condiciones aplicadas al caso que nos ocupa exigen una caracterización de los puestos de trabajo muy detallada.

La forma más adecuada de elaborar un índice de precios del trabajo es mediante la organización de una encuesta específica, generalmente coyuntural, similar a la existente en Estados Unidos (Employment Cost Index, ECI), en la que trimestralmente se toman datos de los puestos de trabajo incluidos en la *cesta básica*.

La organización de tal encuesta tiene un coste elevado, por lo que en el momento actual es inabordable en España. Por ello, se ha optado por efectuar el seguimiento anual de una muestra de puestos de trabajo representativos de los que hay en el conjunto de la economía, aprovechando la información proporcionada por encuestas ya existentes, esto es, sin coste adicional de recogida de información.

¹La información existente actualmente en España, procedente de la Encuesta Trimestral de Coste Laboral, la Anual de Coste Laboral y las de Estructura Salarial del INE, consiste en series de resultados promedio del coste laboral y salarial por trabajador. Tales series recogen tanto la variación del coste como los cambios en la composición del empleo.

2. Fuentes de información utilizadas

Las Encuestas de Estructura Salarial han sido la fuente de información elegida para la elaboración del IPT por las siguientes razones fundamentales:

- Son una fuente de información ya existente.
- Su riqueza informativa: Disponen de datos de asalariados individuales, en especial de la variable ocupación, que es la central en este tipo de análisis.
- Su tamaño muestral permite el estudio de las características de los asalariados de forma muy detallada.
- Aunque no disponen de información de todos los componentes del coste laboral, sí la ofrecen sobre el más importante: el salario. También disponen de las horas normales de trabajo necesarias en el cálculo del Índice.

La Encuesta de Estructura Salarial es una investigación sobre la estructura y distribución de los salarios, de periodicidad cuatrienal (a partir del año 2002), que se realiza en todos los Estados miembros de la Unión Europea. Se utilizan criterios comunes de metodología y contenidos con el fin de obtener unos resultados comparables sobre el nivel, la estructura y distribución del salario entre sus estados miembros de acuerdo con los Reglamentos nº 530/1999 y nº 1738/2005.

La novedad principal que aporta frente a otras encuestas sobre esta materia es que se recogen los salarios en el cuestionario de forma individual y, junto a ellos, una gran cantidad de variables relacionadas con el asalariado. Gracias a esto es posible establecer relaciones entre el salario y algunas de esas variables que pueden contribuir a determinar su cuantía, como son el nivel de estudios alcanzado, la antigüedad, el tipo de contrato o la ocupación, entre otras.

A partir del año 2004 y para los años en los que no se realiza la encuesta cuatrienal, la Encuesta Anual de Estructura Salarial obtiene (para España, no es una encuesta europea) estimaciones de la ganancia bruta anual por asalariado clasificada por tipo de jornada, sexo, actividad económica y ocupación.

En esta encuesta la información se obtiene de la explotación conjunta del Fichero General de Afiliación de la Seguridad Social (SS) y de las declaraciones del Modelo 190: Resumen anual de Retenciones e Ingresos a Cuenta del IRPF de la Agencia Estatal de la Administración Tributaria (AEAT) y de las Haciendas Forales de Navarra, Bizkaia y Gipuzkoa, junto con las variables ocupación y tiempo de trabajo provenientes de la encuesta anexa a la Encuesta Trimestral de Coste Laboral del INE.

La encuesta anual y la cuatrienal están diseñadas de igual forma y sus ámbitos de estudio, periodos de referencia, etc. son los mismos. El procedimiento de selección aleatoria de unidades corresponde a un muestreo bietápico estratificado, donde las unidades de primera etapa son las cuentas de cotización a la Seguridad Social (la Tesorería General de la Seguridad Social, TGSS, facilita el marco de cuentas de cotización) mientras que las de segunda etapa son los asalariados.

Las unidades de primera etapa se estratifican por la actividad económica, la comunidad autónoma y por intervalos de tamaño (medido en número de asalariados).

Una vez seleccionadas las cuentas de cotización, la Tesorería General de la Seguridad Social obtiene la relación de asalariados que estuvieron de alta en ellas durante todo el mes de octubre del año de referencia. A partir de esta relación se seleccionan los asalariados, las unidades de segunda etapa.

En el caso de la encuesta anual, la muestra de primera etapa coincide con la muestra de la Encuesta Trimestral de Coste Laboral (ETCL). Junto con el envío de los cuestionarios de ETCL correspondientes a un trimestre determinado se incluye un cuestionario anexo con la muestra de asalariados seleccionados en cada cuenta de cotización e identificados por su número de Seguridad Social, solicitando información sobre la variable ocupación fundamentalmente.

La información de la ETCL y del cuestionario anexo se cruza con la procedente de las Administraciones Tributarias (tomada del Fichero Resumen anual de Retenciones e Ingresos a Cuenta del IRPF) para incorporar el dato de la ganancia anual de cada asalariado. Los tamaños muestrales son de unas 28.000 cuentas de cotización y más de 200.000 asalariados.

Las unidades muestrales de la encuesta anual, al coincidir con la muestra de ETCL, están divididas en cinco grupos de rotación, de manera que en el primer trimestre de cada año se reemplaza el grupo con mayor antigüedad, lo que supone una renovación anual del 20 % de la muestra. Sin embargo, en los años que se elabora la encuesta cuatrienal la muestra se selecciona de forma que no sea coincidente con la muestra de ETCL excepto en la parte exhaustiva (unidades de más de 500 asalariados).

3. Ámbitos de la operación estadística

Al utilizarse las Encuestas de Estructura Salarial para la elaboración del IPT, los ámbitos poblacional, geográfico y temporal así como la cobertura sectorial del Índice son los derivados de estas encuestas. Se detallan a continuación.

El **ámbito poblacional** está formado por todos los trabajadores por cuenta ajena (asalariados) que prestan sus servicios en centros de cotización y que hayan estado de alta en la Seguridad Social durante más de 2 meses durante el año, siendo uno de ellos el mes de octubre. La elección de este mes se debe a que tiene la ventaja de considerarse *normal* en todos los países de la UE, en el sentido de que está poco afectado por variaciones estacionales o por pagos de vencimiento superior al mes, como las pagas de Navidad.

Se excluyen los presidentes, miembros de consejos de administración y en, general, todo aquel personal cuya remuneración no sea principalmente en forma de salario, sino por comisiones o beneficios.

El **ámbito geográfico** abarca todo el territorio nacional, con resultados desagregados por comunidades autónomas. La información correspondiente a Ceuta y Melilla se proporciona conjuntamente con la de Andalucía. En cuanto a la **cobertura sectorial**, se investigan los centros de cotización, independientemente de su tamaño, cuya actividad económica esté encuadrada en los tres grandes sectores: la Industria, la Construcción y los Servicios.

Quedan excluidas de la encuesta las actividades agrícolas, ganaderas y pesqueras, personal doméstico y organismos extraterritoriales.

En el **ámbito temporal**, el periodo de referencia es el año natural.

4. Variables de clasificación

Las variables de clasificación que se utilizan para determinar los puestos de trabajo considerados en la elaboración del IPT son las siguientes:

- **Actividad:** las secciones de actividad según la CNAE-09 del ámbito muestral de las Encuestas de Estructura Salarial: B, C, D, E, F, G, H, I, J, K, L, M, N, O, P, Q, R y S.

B Industrias extractivas

C Industria manufacturera

D Suministro de energía eléctrica, gas, vapor y aire acondicionado

E Suministro de agua, actividades de saneamiento, gestión de residuos y descontaminación

F Construcción

G Comercio al por mayor y al por menor; reparación de vehículos de motor y motocicletas

H Transporte y almacenamiento

I Hostelería

J Información y comunicaciones

K Actividades financieras y de seguros

L Actividades inmobiliarias

M Actividades profesionales, científicas y técnicas

N Actividades administrativas y servicios auxiliares

O Administración Pública y defensa; Seguridad Social obligatoria

P Educación

Q Actividades sanitarias y de servicios sociales

R Actividades artísticas, recreativas y de entretenimiento

S Otros servicios

- **Tamaño del centro de trabajo (estrato):** toma 3 valores distintos, según el número de asalariados del centro:

1 de 1 a 49 asalariados

2 de 50 a 199 asalariados

3 200 o más asalariados

- **Región:** Comunidad Autónoma en la que se encuentra el centro de trabajo:
 - 01 Andalucía (incluye Ceuta y Melilla)
 - 02 Aragón
 - 03 Asturias (Principado de)
 - 04 Balears (Illes)
 - 05 Canarias
 - 06 Cantabria
 - 07 Castilla y León
 - 08 Castilla - La Mancha
 - 09 Cataluña
 - 10 Comunitat Valenciana
 - 11 Extremadura
 - 12 Galicia
 - 13 Madrid (Comunidad de)
 - 14 Murcia (Región de)
 - 15 Navarra (Comunidad Foral de)
 - 16 País Vasco
 - 17 Rioja (La)
- **Sexo:** mujeres y varones.
- **Ocupación:** los grandes grupos de la Clasificación Nacional de Ocupaciones (CNO-11): del 1 al 9. Los grupos 0 y 5, y 6 y 7 aparecen unidos por problemas de representación muestral en la Encuestas de Estructura Salarial.
 - 1 Directores y gerentes
 - 2 Técnicos y profesionales científicos e intelectuales
 - 3 Técnicos; profesionales de apoyo
 - 4 Empleados contables, administrativos y otros empleados de oficina
 - 5 y 0 Trabajadores de los servicios de restauración, personales, de protección y vendedores. Ocupaciones militares.
 - 6 y 7 Trabajadores cualificados en el sector agrícola, ganadero, forestal y pesquero. Artesanos y trabajadores cualificados de las industrias manufactureras y la construcción (excepto operadores de instalaciones y maquinaria)
 - 8 Operadores de instalaciones y maquinaria, y montadores
 - 9 Ocupaciones elementales

- **Tipo de contrato:** indefinido y temporal.
- **Tipo de jornada:** tiempo completo y tiempo parcial.
- **Edad del asalariado:** toma 5 valores distintos, determinados por los siguientes intervalos de edad:
 - 1 menos de 25 años
 - 2 de 25 a 34 años
 - 3 de 35 a 44 años
 - 4 de 45 a 54 años
 - 5 55 o más años
- **Nacionalidad del asalariado:** española y resto de nacionalidades.
- **Antigüedad del asalariado en el centro de trabajo:** toma 6 valores distintos, determinados por los siguientes intervalos:
 - 1 menos de 1 año de antigüedad
 - 2 de 1 a 3 años de antigüedad
 - 3 de 4 a 10 años de antigüedad
 - 4 de 11 a 20 años de antigüedad
 - 5 21 o más años de antigüedad

5. Metodología de cálculo

5.1. Fórmula de cálculo del índice

Para la elaboración del IPT se ha utilizado un índice de Laspeyres encadenado.

La elección de tal índice se debe a que pueden existir cambios en la estructura del empleo de un año a otro y este tipo de indicador permite introducirlos anualmente a través de las ponderaciones sin tener que esperar a un cambio de base como ocurre en un índice de base fija.

La formulación del IPT se detalla a continuación.

Se denotan con c los cruces de las variables que definen las características de los puestos de trabajo (detalladas en el punto 4). Es decir, se considera una estructura fija que es la constituida por los mencionados cruces, con el objetivo de ver cómo ha variado la ganancia por hora descontando el efecto composición.

Sea,

G_c^k la ganancia por hora de los asalariados en el cruce c en el año k ,

h_c^k las horas trabajadas por los asalariados en el cruce c en el año k .

$\omega_c^k = G_c^k \cdot h_c^k$ la ganancia anual de los asalariados en el cruce c en el año k .

La definición de la fórmula básica de Laspeyres que debe utilizarse para calcular el IPT durante el año j , para el año base k es la siguiente.

$$IPT_{j(k)} = \frac{\sum_c G_c^j h_c^k}{\sum_c G_c^k h_c^k} = \frac{\sum_c \frac{G_c^j}{G_c^k} G_c^k h_c^k}{\sum_c G_c^k h_c^k} = \frac{\sum_c \frac{G_c^j}{G_c^k} \omega_c^k}{\sum_c \omega_c^k},$$

donde $\frac{\omega_c^k}{\sum_c \omega_c^k}$ son las ponderaciones para calcular el índice y $\frac{G_c^j}{G_c^k}$ es el índice simple o elemental en el cruce c .

La definición del coeficiente de enlace del año j con el año $j + 1$ es,

$$L_{j,j+1} = \frac{\sum_c G_c^{j+1} h_c^j}{\sum_c G_c^j h_c^j}.$$

Entonces, la fórmula del índice de Laspeyres encadenada para el año j con referencia al año k es,

$$IPT_{j(k)} = 100 \cdot L_{k,k+1} \cdot L_{k+1,k+2} \cdots L_{j-2,j-1} \cdot IPT_{j(j-1)}.$$

Además, se observa que el coeficiente de enlace $L_{j,j+1}$ es equivalente al índice del año $j + 1$ con base el año j , es decir que $L_{j,j+1} = IPT_{j+1(j)}$.

Por tanto, el índice de Laspeyres queda como sigue,

$$IPT_{j(k)} = 100 \cdot IPT_{k+1(k)} \cdot IPT_{k+2(k+1)} \cdots IPT_{j-1(j-2)} \cdot IPT_{j(j-1)},$$

La fórmula para calcular el índice encadenado para el año j de una combinación Q de cruces c , siendo el año base k es la siguiente,

$$IPT_{j(k),Q} = 100 \cdot IPT_{k+1(k),Q} \cdot IPT_{k+2(k+1),Q} \cdots IPT_{j-1(j-2),Q} \cdot IPT_{j(j-1),Q}.$$

5.2. Periodo base

El año 2016 es el periodo base, por lo que todos los índices que se calculen estarán referidos a ese año, y ese año el valor del índice es 100.

5.3. Periodo de referencia de las ponderaciones

La estructura de las ponderaciones establece la importancia de cada puesto de trabajo respecto al resto de ellos basado en el coste salarial realizado por el empleador. Para cada año, la medición de las ponderaciones se realiza a partir de las ganancias por hora y las horas trabajadas por los trabajadores asociados a cada tipo de puesto de trabajo en el año anterior.

5.4. Cálculo de la ganancia hora

Para obtener una estimación más robusta de la ganancia hora en los distintos puestos de trabajo definidos por las variables de clasificación, se utiliza un modelo de regresión hedónica. No es posible incorporar directamente datos observados, dado que hay puestos de trabajo que aparecen/desaparecen de un año a otro y que la representación muestral de algunos puestos es escasa.

El modelo de regresión utilizado en el cálculo de la ganancia hora es el siguiente.

Para cada año j , la ganancia por hora G para un asalariado que pertenece a un cruce c de las variables explicativas del modelo, verifica:

$$l_c^j = \ln G_c^j = x_c' \beta^j + \varepsilon_c^j, \quad (1)$$

donde,

x_c' es un vector de dimensión $(1 \times p)$ cuyos elementos son iguales a 0 o 1 en función de las características que definen al cruce c , en cuanto a los efectos principales y a las interacciones se refiere,

β^j es un vector de dimensión $(p \times 1)$ que contiene los parámetros desconocidos del modelo,

ε_c^j es la componente aleatoria del modelo para la celda c en el año j .

El vector β^j define el efecto proporcional sobre la ganancia por hora de las p variables dicotómicas incluidas en x'_c . Los p parámetros desconocidos incluyen la constante y los coeficientes de las variables dicotómicas asociadas a los efectos principales y a las interacciones del modelo de regresión.

Para cada r categorías posibles que tiene un efecto principal, el modelo incluye $(r - 1)$ parámetros y si la interacción tiene $(r \times s)$ posibles combinaciones de valores, el modelo incluiría $(r - 1) \times (s - 1)$ parámetros.

Las componentes aleatorias verifican:

$$E[\varepsilon_c^j] = 0; \quad Var[\varepsilon_c^j] = \sigma_j^2; \quad Cov[\varepsilon_c^j, \varepsilon_d^j] = 0 \quad \forall c \neq d.$$

Una vez definido el modelo, se debe estimar el vector de parámetros β^j con la información disponible cada año. Para ello se formula el modelo (1) en notación matricial de la siguiente manera:

$$L^j = X^j \beta^j + \varepsilon^j$$

donde,

L^j es un vector de dimensión $(n_j \times 1)$ que contiene los n_j elementos del año j denotados anteriormente por $l_c^j \forall j = 1, \dots, n_j$. Es decir, contiene tantas filas como asalariados tenemos en la muestra del año j .

X^j es una matriz de dimensión $(n_j \times p)$, cuyos elementos son iguales a 0 o a 1. Cada fila representa a un asalariado y cada columna contiene las p características que definen a dicho asalariado.

β^j es un vector de dimensión $(p \times 1)$ que contiene los parámetros desconocidos del año j . Incluye la constante y los coeficientes asociados a los efectos principales y las interacciones del modelo.

ε^j es un vector de dimensión $(n_j \times 1)$ que contiene las n_j componentes aleatorias del modelo en el año j . Este vector verifica:

$$E[\varepsilon^j] = \mathbf{0}; \quad Var[\varepsilon^j] = \sigma_j^2 I_{n_j \times n_j}.$$

Como los datos utilizados en el modelo de regresión proceden de muestras extraídas de una población total con un determinado diseño muestral, calculamos el

estimador de β^j por el método de mínimos cuadrados ponderados por los pesos de muestreo (MCOW):

$$\widehat{\beta}^j = (X'^j W^j X^j)^{-1} X'^j W^j L^j,$$

donde W^j es una matriz diagonal de dimensión $n_j \times n_j$ que contiene los pesos de muestreo.

Y la varianza del estimador es:

$$Var[\widehat{\beta}^j] = \sigma_j^2 (X'^j W^j X^j)^{-1} X'^j W^{j2} X^j (X'^j W^j X^j)^{-1} = V^j,$$

donde la matriz V^j tiene dimensión $(p \times p)$.

La ganancia por hora estimada correspondiente a cada cruce se obtiene a partir de la fórmula (1), con lo que la ganancia estimada en el cruce c el año j es la siguiente:

$$\widehat{G}_c^j = \exp\{x'_c \widehat{\beta}^j\}. \quad (2)$$

El problema de este estimador es que tiene sesgo, pues aunque el estimador β^j sí es insesgado, al calcular su exponencial esta característica se pierde. Para corregir este sesgo utilizamos una variación del estimador propuesto por Shaarawi y Viveros (1997) es el siguiente,

$$\widehat{G}_c^j = \exp\{x'_c \widehat{\beta}^j - \frac{1}{2} x'_c \widehat{V}^j x_c + \frac{1}{2} \widehat{\varphi}_j^2 \widehat{\sigma}_j^2\}. \quad (3)$$

El estimador (3) corrige sustancialmente el sesgo del estimador (2), asumiendo la normalidad de los errores ε_c^j .

Para obtener la estimación de la varianza que aparece en las expresiones anteriores, se definen los residuos e_c^j como la diferencia entre los logaritmos de la ganancia por hora observada y la estimada:

$$e_c^j = l_c^j - x'_c \widehat{\beta}^j.$$

La varianza $\widehat{\sigma}_j^2$ se estima con la media de cuadrados residual:

$$\widehat{\sigma}_j^2 = \frac{1}{n_j - p} \sum_c^{n_j} (e_c^j)^2.$$

Siendo la estimación de φ_j^2 :

$$\widehat{\varphi}_j^2 = 1 - \frac{\widehat{\sigma}_j^2}{2(n_j - p)} - \frac{\widehat{\sigma}_j^4}{3(n_j - p)^2}. \quad (4)$$

No obstante, en (4) las varianzas del numerador son menores que 1 y al dividir las por el tamaño muestral resulta ser $\widehat{\varphi}_j^2 \approx 1$, por lo que la expresión (3) queda finalmente como sigue,

$$\widehat{G}_c^j = \exp\{x'_c \widehat{\beta}^j - \frac{1}{2} x'_c \widehat{V}^j x_c + \frac{1}{2} \widehat{\sigma}_j^2\}.$$

6. Difusión

Se publican datos para el total nacional y por comunidades autónomas (las Ciudades Autónomas de Ceuta y de Melilla se proporcionan junto con Andalucía).

En cuanto a las variables de clasificación de los puestos de trabajo, se publica información desagregada por:

- Sección de actividad
- Estrato (tamaño del centro de trabajo)
- Sexo
- Ocupación
- Tipo de contrato
- Tipo de jornada
- Edad
- Nacionalidad
- Antigüedad

También se facilitarán las ponderaciones utilizadas en cada año desagregadas por Comunidad Autónoma, Sección de actividad y tamaño del centro de trabajo.