



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 2025, 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 29 de Diciembre de 2025)

Acceso Promoción Interna, CDEE

Cuestión 1. Una encuesta mensual dirigida a empresas tiene como objetivo recoger información sobre su actividad económica reciente.

La muestra ha sido seleccionada mediante un diseño probabilístico estratificado por tamaño de empresa según número de trabajadores (que se considerará dado a efectos de este ejercicio). El marco de selección dispone, para cada empresa, de un identificador único, de su razón social (*nombre oficial de una empresa para efectos legales y administrativos*), de su dirección postal, correo electrónico y de un número de teléfono de contacto.

- (a) Proponga de forma justificada un plan de recogida de datos y de gestión del trabajo de campo. Deberá especificar los canales y métodos de recogida, así como su organización en el tiempo, detallando la secuencia de contactos y la estrategia para la mejora de las tasas de cobertura/respuesta.
- (b) Se dispone de información auxiliar procedente de un registro administrativo que contiene variables de interés para la encuesta. Sin embargo, no existe un identificador único común entre ambas fuentes.

Ambas fuentes comparten las siguientes variables:

- razón social y dirección (en formato literal),
- código de actividad económica (numérico), aunque basado en clasificaciones de referencia distintas,
- fecha de inicio de la actividad de la empresa.

Proponga un procedimiento para integrar la información de ambas fuentes mediante un enlace probabilístico de registros. En particular:

- 1) Describa las tareas de preparación de los datos que serían necesarias antes del proceso de emparejamiento.
 - 2) Describa, de forma general, el método de enlace probabilístico, indicando los elementos necesarios para su aplicación y el criterio de decisión para determinar si dos registros corresponden a la misma unidad.
- (c) Tras el proceso de enlace de la encuesta con el registro administrativo, y una vez depurada e imputada la información, así como obtenidos los factores de elevación, se dispone de un fichero final de microdatos con información económica detallada de las empresas.
- 1) Describa cómo se evalúa el riesgo de identificación en estos microdatos, distinguiendo entre:
 - medidas del riesgo de identificación a nivel del fichero;
 - medidas del riesgo de identificación a nivel del registro.
 - 2) Mencione asimismo algunos métodos de control del secreto estadístico aplicables a microdatos.

Cuestión 2. Sea X_1, \dots, X_n una muestra aleatoria simple de una población con distribución exponencial de parámetro $\lambda > 0$:

$$f(x; \lambda) = \lambda e^{-\lambda x}, \quad x > 0$$

Se supone que el tamaño muestral es $n = 25$ y que $\sum_{i=1}^{25} X_i = 40$

- (a) Obtenga el estimador máximo-verosímil de λ . Indique si el estimador es invariante para estimar:

$$\theta = \frac{1}{\lambda}$$

- (b) Calcule el sesgo y el error cuadrático medio del estimador:

$$T = \frac{n-1}{\sum_{i=1}^n X_i}$$

sabiendo que:

$$E\left(\frac{1}{\sum X_i}\right) = \frac{\lambda}{n-1}, \quad \text{Var}\left(\frac{1}{\sum X_i}\right) = \frac{\lambda^2}{(n-1)^2(n-2)}$$

Indique si T es insesgado.

- (c) Se desea contrastar:

$$H_0 : \lambda = 0,5 \quad \text{frente a} \quad H_1 : \lambda > 0,5$$

- 1) Obtenga el estimador máximo-verosímil de λ .
- 2) Construya el test de razón de verosimilitudes.
- 3) Determine la región crítica en función de $\sum X_i$.
- 4) Interprete intuitivamente cuándo se rechazará H_0 .

- (d) Utilizando que:

$$2\lambda \sum_{i=1}^n X_i \sim \chi_{2n}^2$$

construya un intervalo de confianza del 95% para λ .

$$(\text{Datos: } \chi_{50,0,975}^2 = 71,42, \quad \chi_{50,0,050}^2 = 34,76 \quad \chi_{50,0,025}^2 = 32,36)$$

- (e) Responda razonadamente:

- 1) ¿Qué significa que un intervalo de confianza tenga probabilidad de cobertura 0,95?
- 2) Explique la relación entre intervalos de confianza y tests de hipótesis.
- 3) ¿Por qué un intervalo más corto no siempre es preferible?

Cuestión 3. Sea un sistema de representación de números reales en coma flotante de 16 bits, trabajando únicamente con números normalizados, con la siguiente estructura:

Campo	Bits	Descripción
Signo (s)	1 bit	0 = positivo, 1 = negativo
Exponente (e)	5 bits	Exponente en exceso-15 (sesgo = 15)
Mantisa (m)	10 bits	Parte fraccionaria de la mantisa normalizada (el "1." inicial está implícito)

- (a) Represente el número decimal $-13,625$ en el formato de 16 bits descrito anteriormente, indicando la cadena final completa de 16 bits y detallando razonadamente todos los pasos realizados.
- (b) Trabajando únicamente con números normalizados, la siguiente cadena de 16 bits se almacena en memoria siguiendo el formato descrito anteriormente:

0011000110100000

¿A qué valor decimal exacto corresponde?

- (c) Considere el siguiente fragmento de pseudocódigo, donde todas las variables son de tipo `Entero8bits`, con las siguientes características:
- enteros con signo de 8 bits,
 - representados en complemento a dos,
 - todas las operaciones aritméticas se realizan con desbordamiento módulo 256,
 - tras cada operación, el resultado se almacena nuevamente como entero de 8 bits.

Además, el valor mostrado por `print` se interpreta siempre como entero con signo en complemento a dos.

```
a = 70
b = a + a
print(b)

b = b + a
print(b)
```

Explique, razonadamente, cual es el valor impreso en cada `print`.

Cuestión 4. Se selecciona una muestra de tamaño n mediante muestreo aleatorio simple sin reposición a partir de una población U de tamaño N . Sea y_k la variable de estudio asociada a la unidad k .

Una vez seleccionada la muestra, la unidad k responde con probabilidad θ_k .

Debido a la falta de respuesta, únicamente se observa información para una submuestra efectiva $r \subset s$, formada por las unidades respondientes.

Se considera el estimador del total poblacional

$$Y = \sum_{k \in U} y_k$$

dado por

$$\widehat{Y}_0 = \frac{N}{n} \sum_{k \in r} y_k.$$

Se pide:

- (a) Obtener la esperanza y el sesgo del estimador \widehat{Y}_0 . Comentar e interpretar los resultados obtenidos.
- (b) Proponer de forma razonada un estimador alternativo que permita corregir el sesgo debido a la falta de respuesta.
Justificar la propuesta a partir del mecanismo conjunto de selección y respuesta y discutir las principales dificultades prácticas asociadas a su implementación.
- (c) Supóngase ahora que, mediante información auxiliar, las unidades de la población pueden clasificarse en grupos con distintos comportamientos de respuesta.
Bajo la hipótesis de que la falta de respuesta es aleatoria dentro de cada grupo, proponer un estimador del total poblacional y justificar razonadamente sus propiedades.

Cuestión 5. Un investigador desea estudiar el gasto mensual en transporte público (Y , en euros) de una muestra de individuos en función de:

- la edad (X_1 , en años),
- el tipo de residencia (X_2), donde

$$X_2 = \begin{cases} 0 & \text{zona rural} \\ 1 & \text{zona urbana} \end{cases}$$

Se propone el siguiente modelo lineal:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i$$

y se supone inicialmente que:

$$\varepsilon_i \sim N(0, \sigma^2)$$

Tras el ajuste por mínimos cuadrados se obtiene:

$$\hat{Y}_i = 12 + 0,8X_{1i} + 15X_{2i}$$

junto con la siguiente información:

Parámetro	Estimación	Error estándar
$\hat{\beta}_0$	12	4
$\hat{\beta}_1$	0.8	0.25
$\hat{\beta}_2$	15	X

con $n = 40$.

Además:

$$\hat{\sigma}^2 = 64$$

y

$$(X'X)_{33}^{-1} = 0,10$$

(a) Obtenga la expresión de la esperanza condicionada

$$E(Y | X_1, X_2)$$

e interprete:

- 1) el efecto de la variable cuantitativa,
- 2) el efecto de la variable cualitativa,
- 3) el significado de β_0 .

(b) Calcule e interprete el gasto esperado para:

- 1) una persona de 30 años residente en zona rural,
- 2) una persona de 30 años residente en zona urbana.

Interprete la diferencia entre ambas predicciones.

(c) Contraste al 5% si el tipo de residencia tiene efecto significativo sobre el gasto medio en transporte.

- 1) Formule las hipótesis.
- 2) Construya el estadístico de contraste.
- 3) Indique su distribución bajo H_0 .
- 4) Obtenga la conclusión.

Dato:

$$t_{0,975,37} \approx 2,03$$

(d) Construya un intervalo de confianza aproximado del 95% para β_2 .

Interprete el resultado en términos del problema.

(e) Suponga ahora que la variable respuesta no fuese continua sino binaria:

$$Y_i = \begin{cases} 1 & \text{usa transporte público diariamente} \\ 0 & \text{en otro caso} \end{cases}$$

Responda razonadamente:

- 1) ¿Por qué el modelo lineal normal puede resultar inadecuado?
- 2) ¿Qué componente del modelo cambia en un modelo lineal generalizado?
- 3) Indique un modelo más apropiado y la distribución asumida para Y_i .

Cuestión 6. En una oficina de estadística se desea estudiar el gasto mensual en alimentación de los hogares. Para ello, se dispone de un presupuesto de 12.000 euros para realizar una encuesta.

La información disponible en el marco muestral permite clasificar a las familias en dos grupos: el 60 % dispone de correo electrónico y el 40 % restante no dispone de correo electrónico.

El coste de realización de cada entrevista es de 4 euros si se realiza por web y de 16 euros si se realiza mediante entrevista personal. Además, por estudios anteriores se sabe que las cuasivarianzas del gasto mensual en alimentación son similares en ambos grupos.

Teniendo en cuenta que se desea realizar un muestreo aleatorio estratificado sin reemplazamiento, determinar cuántas familias deben ser entrevistadas en cada grupo si:

- a) Todas las entrevistas se realizan presencialmente. Indicar de manera razonada si el diseño es autoponderado.
- b) Las familias con correo electrónico son entrevistadas por web y las familias sin correo electrónico son entrevistadas presencialmente. Indicar si el diseño es autoponderado.
- c) Supóngase ahora que el objetivo principal del estudio es estimar

$$\Delta = \bar{Y}_1 - \bar{Y}_2.$$

Proponer un estimador insesgado de Δ , obtener su varianza y determinar el reparto óptimo de la muestra bajo la misma restricción presupuestaria. Comparar dicho reparto con el obtenido en el apartado b), justificando razonadamente las diferencias observadas.