MINISTERIO
DE ECONOMÍA, COMERCIO
Y EMPRESA

Instituto Nacional de Estadística

Oposición al Cuerpo Superior de Estadísticos de Estado

Primer Ejercicio

Convocatoria de la oferta pública de empleo de 2025

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

Producción Estadística Oficial: Principios Básicos del Ciclo de Producción de Operaciones Estadísticas

Pregunta 1. Explique formalmente los conceptos de muestreo probabilístico y diseño muestral, formulando este último mediante el espacio de muestras posibles S y la función $p(s)$, $s \in S$.

Pregunta 2. En un estudio mediante muestreo aleatorio simple sin reemplamamiento se desea estimar una proporción poblacional.

a) Suponiendo que se dispone, a partir de estudios previos, de un valor aproximado de la proporción poblacional p, obtenga una expresión aproximada del tamaño muestral necesario para con un error de muestreo fijado E.

b) Cuando no se dispone de ninguna información previa sobre la proporción poblacional, indique qué valor debe utilizarse en la expresión del apartado anterior y explique qué criterio matemático justifica esa elección, teniendo en cuenta que dicho valor debe proporcionar un tamaño muestral suficientemente grande para asegurar la precisión requerida.

Pregunta 3. En el muestreo por conglomerados sin submuestreo, suponiendo conglomerados de igual tamaño, la varianza bajo este diseño puede expresarse en función de la correspondiente bajo muestreo aleatorio simple. Explique esta relación e interprete en qué situaciones ambos diseños podrían presentar una eficiencia similar.

Pregunta 4. Considere un diseño de muestreo con probabilidades proporcionales al tamaño.

a) Justifique, mediante una demostración matemática, el incentivo para utilizar este tipo de diseño en términos de la varianza del estimador de Horvitz-Thompson.

b) Describa el procedimiento de construcción de las probabilidades de inclusión cuando no es posible imponer una proporcionalidad estricta a la variable de tamaño.

Pregunta 5. Explique el concepto de equilibrio entre riesgo de identificación y utilidad en el control del secreto estadístico y describa cómo puede formularse este como un problema de optimización.

Pregunta 6. Defina qué se entiende por paradatos en la recogida de datos estadísticos, indique cuál es el objetivo de su utilización en la gestión del proceso de recogida y cite tres ejemplos de paradatos.

Pregunta 7. Represente mediante un esquema un paso de producción estadística estándar desglosado mostrando la relación entre el flujo de datos y el flujo de metadatos en el marco del GSBPM y el GSIM, e interprete los principales elementos que intervienen en dicho esquema.

Pregunta 8. Indique si las siguientes afirmaciones son verdaderas o falsas y justifique su respuesta en cada caso.

a) La depuración de datos estadísticos tiene como objetivo corregir todos los errores presentes en los microdatos antes de la publicación de los resultados, con el fin de garantizar la calidad de las estimaciones publicadas.

b) Un valor atípico (outlier) siempre corresponde a un error en los datos, por lo que debe corregirse durante el proceso de depuración.

c) Los errores sistemáticos son especialmente problemáticos en la producción estadística porque pueden trasladarse a los agregados y provocar sesgos en los resultados publicados.

d) En el enfoque de imputación completa frente a la falta de respuesta, la estimación de los parámetros poblacionales se realiza aplicando el mismo estimador que se utilizaría si no existiera falta de respuesta, tratando los valores imputados como si fueran observaciones reales.

Pregunta 9. Defina población objetivo y marco muestral y explique el papel que desempeñan en la aparición de los errores de cobertura en una encuesta.

Pregunta 10. En la integración de información procedente de distintos ficheros, es necesario determinar qué registros corresponden a una misma unidad estadística. Explique brevemente cómo se aborda este problema mediante técnicas de record linkage y describa el objetivo de las técnicas de bloqueo y el principio en el que se basan. Ilustre su respuesta con un ejemplo.

Inferencia y Modelización Estadísticas

Pregunta 11. Sea X el tiempo (en minutos) que tarda un empleado en atender a un cliente. Se sabe que

$$X \sim \mathcal{N}(\mu = 10,\ \sigma^2 = 4).$$

Se extrae una muestra aleatoria simple de $n = 16$ observaciones independientes $X_1, X_2, \ldots, X_{16}$. Se realizan las siguientes operaciones: se calcula $\bar{X}$, se define el estadístico $T = \dfrac{\bar{X} - \mu_0}{S/\sqrt{n}}$, siendo S^2 la varianza muestral corregida, y se extraen dos muestras independientes de tamaños $n_1 = 16$ y $n_2 = 25$ de la misma población. Indique qué distribución sigue $\bar{X}$, T y S_1^2/S_2^2.

Pregunta 12. Un servicio de atención telefónica registra el tiempo (en minutos) entre llamadas consecutivas. Se modela como una variable aleatoria X con distribución exponencial de parámetro $\lambda > 0$ siendo esta la *tasa de llegada* de llamadas (llamadas por minuto). Se dispone de una muestra aleatoria simple $X_1, X_2, \ldots, X_n$ de tiempos entre llamadas. Obtenga el estimador del tiempo medio entre llamadas y justifique sus repuestas.

Pregunta 13. Sea X una variable aleatoria con distribución Bernoulli(p). Se considera el contraste:

$$H_0 : p = \tfrac{1}{4} \qquad \text{frente a} \qquad H_1 : p = \tfrac{3}{4}.$$

Se toman $n = 2$ observaciones independientes X_1, X_2 y se define el estadístico $T = X_1 + X_2$. Se propone la región crítica:

$$C = \{T = 2\}.$$

Calcule el tamaño del test, la probabilidad del error de tipo II y la potencia.

Pregunta 14. Sea $X_1, \ldots, X_n$ una muestra aleatoria de una distribución $\mathrm{U}(0, \theta)$ con $\theta > 0$ desconocido. Sea $X_{(n)} = \max\{X_1, \ldots, X_n\}$ el máximo muestral.

Se puede demostrar que la variable

$$Q = \frac{X_{(n)}}{\theta}$$

tiene una función de distribución $F_Q(q) = q^n$ para $q \in [0, 1]$.

Estudia si $Q = X_{(n)}/\theta$ es una cantidad pivotal para θ, y construya un intervalo de confianza para θ al nivel $1 - \alpha$ partiendo de

$$P\left(a \leq \frac{X_{(n)}}{\theta} \leq 1\right) = 1 - \alpha, \qquad \text{donde } a = \alpha^{1/n}.$$

Pregunta 15. Considere los dos estudios siguientes.

Estudio A. Un investigador quiere estudiar el efecto de tres dosis de un fertilizante (baja, media, alta) sobre el rendimiento de una cosecha. Asigna aleatoriamente cada dosis a varias parcelas de terreno en condiciones controladas de riego y temperatura.

Estudio B. Una epidemióloga analiza la relación entre el nivel de contaminación atmosférica (μg/m^3) y la tasa de hospitalización respiratoria en 50 ciudades distintas, usando registros históricos.

Para cada estudio: (i) clasifique cada estudio como experimental u observacional; (ii) escriba el modelo lineal correspondiente de forma esquemática (componente sistemática + término error); (iii) explique qué recoge el término error en cada estudio y su relevancia para hacer inferencia. Justifique sus respuestas.

Pregunta 16. Sean X e Y dos variables aleatorias que toman valores en $\{0, 1\}$, con distribución conjunta dada por la siguiente tabla:

	$Y = 0$	$Y = 1$
$X = 0$	0,1	0,4
$X = 1$	0,2	0,3

Calcula $\mathbb{E}[Y \mid X]$, $\mathbb{E}[Y^2 \mid X]$ y $\mathrm{Var}(Y \mid X = x)$ para $x = 1$. Justifique sus repuestas.

Pregunta 17. En el modelo lineal $\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, el estimador de mínimos cuadrados $\hat{\boldsymbol{\beta}}$ se obtiene minimizando $\|\mathbf{y} - X\boldsymbol{\beta}\|^2$. La solución requiere resolver las ecuaciones normales. Indique (i) cuáles son dichas ecuaciones, (ii) cuándo tienen solución única, y (iii) en caso de que no sean únicas, conteste: ¿el vector de valores ajustados $\hat{\mathbf{y}} = X\hat{\boldsymbol{\beta}}$ tampoco lo sería? Justifique sus respuestas.

Pregunta 18. En el modelo lineal normal $\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ con $\boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \sigma^2 I)$, dado un nuevo vector de covariables $\mathbf{x}_0$, se pueden construir dos tipos de intervalos al nivel $1 - \alpha$: un intervalo de confianza para la respuesta media $\mathbb{E}[y_0]$ y un intervalo de predicción para nueva observación individual y_0. Ambos se centran en $\mathbf{x}_0^\top \hat{\boldsymbol{\beta}}$. Entonces, (i) explique la fuente de incertidumbre de cada intervalo, (ii) analice qué le ocurre a cada intervalo cuando crece el tamaño muestral, y (iii) conteste a esta pregunta: Si los errores no son normales, ¿siguen siendo válidos estos intervalos en muestras grandes? Justifique sus respuestas.

Almacenamiento y Modelos de Datos

Pregunta 19. Realice una clasificación de la memoria de un sistema de computación en función de la forma de acceso a la misma.

Pregunta 20. Defina el núcleo (kernel) de un sistema operativo y enumere sus funciones principales.