



MINISTERIO
DE ECONOMÍA, COMERCIO
Y EMPRESA

INE
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 25 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, resto

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 parados 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 parados.

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.

- 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.
- Un valor atípico (outlier) siempre corresponde a un error en los datos, por lo que debe corregirse durante el proceso de depuración.
- Los errores sistemáticos son especialmente problemáticos en la producción estadística porque pueden trasladarse a los agregados y provocar sesgos en los resultados publicados.
- 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.

Inferencia y Modelización Estadísticas

Pregunta 9. 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, \dots, X_{16} . Se realizan las siguientes operaciones: se calcula \bar{X} , se define el estadístico $T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$, 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 10. 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, \dots, X_n de tiempos entre llamadas. Obtenga el estimador del tiempo medio entre llamadas y justifique sus repuestas.

Pregunta 11. Sea X una variable aleatoria con distribución Bernoulli(p). Se considera el contraste:

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

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 12. Sea X_1, \dots, X_n una muestra aleatoria de una distribución $U(0, \theta)$ con $\theta > 0$ desconocido. Sea $X_{(n)} = \max\{X_1, \dots, X_n\}$ 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

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

Pregunta 13. 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 ($\mu\text{g}/\text{m}^3$) 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 14. 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 $\mathbb{E}[Y | X]$, $\mathbb{E}[Y^2 | X]$ y $\text{Var}(Y | X = x)$ para $x = 1$. Justifique sus repuestas.

Pregunta 15. 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 16. 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 17. Realice una clasificación de la memoria de un sistema de computación en función de la forma de acceso a la misma.

Pregunta 18. Defina el núcleo (kernel) de un sistema operativo y enumere sus funciones principales.

Cuentas Nacionales

Pregunta 19. Describa cómo afectaría al Ahorro bruto del Sector Hogares cada una de las siguientes medidas, indicando todas las operaciones, saldos y cuentas del sector alteradas:

- a) Una bajada de tipos del IVA en alimentación.
- b) Una subida de tipos marginales de la tarifa del IRPF.

Pregunta 20. Justificando la respuesta, señale si son Verdaderas (V) o Falsas (F) las siguientes afirmaciones sobre las Cuentas Trimestrales:

- a) Las cuentas trimestrales son un compromiso entre el rigor y la amplitud de las cuentas anuales, y la rapidez y oportunidad de los indicadores coyunturales.
- b) Las series de agregados y operaciones trimestrales ajustadas de estacionalidad y de efectos de calendario están limpias de todo aquello que pueda enmascarar los movimientos relevantes a corto y largo plazo de las series como los movimientos repetitivos explicables o las fluctuaciones extraordinarias.
- c) El conjunto de procedimientos utilizados para la elaboración de las cuentas trimestrales suele ser una mezcla de métodos directos e indirectos donde el énfasis se pone en un lado u otro dependiendo del tiempo del que se disponga para elaborarlas.

Pregunta 21. Atendiendo a los criterios contemplados en el SEC 2010, indique en qué operación/es se registran los siguientes flujos:

- a) Beca de estudios concedida por una universidad pública.
- b) Pensión de jubilación.
- c) Adquisición de una autocaravana.
- d) Impuesto sobre Bienes Inmuebles.
- e) Ayuda de una Administración Pública al propietario de una vivienda destruida por una dana.
- f) Propina.
- g) Cotización a la seguridad Social de un autónomo.
- h) Entrada a un museo.
- i) Compra de un billete de avión a una compañía extranjera.
- j) Impuesto sobre sucesiones y donaciones.

Pregunta 22. Defina el empleo equivalente a tiempo completo en el marco de la Contabilidad Nacional (CN). La última información publicada muestra que en 2024 la Productividad por puesto de trabajo equivalente a tiempo completo creció a un ritmo inferior a la Productividad por hora efectivamente trabajada, 0,6 % frente a 1,2 %. Explique qué factores podrían explicar este comportamiento.

Demografía

Pregunta 23. Dibuje un diagrama de Lexis e identifique en el mismo las principales líneas.

- a) ¿Qué tipos de análisis demográfico permite hacer esta herramienta gráfica? Mencione y describa los más relevantes.
- b) Explique qué líneas del diagrama delimitan cada uno de los distintos tipos de análisis demográfico.
- c) ¿Qué individuos quedan contenidos entre dos líneas paralelas de cada dirección principal?
- d) ¿Cómo se representan los flujos en un diagrama de Lexis? ¿Y si se refieren a edades exactas?
- e) ¿Y dónde se sitúan los stocks?

Pregunta 24. A continuación, se presenta un extracto de una tabla de mortalidad, relativa al año 2024, obtenida a partir del estudio de los fenómenos demográficos que realiza el Instituto Nacional de Estadística (INE). Calcule los valores faltantes aplicando las fórmulas adecuadas para cada caso.

	m_x	q_x	p_x	l_x	d_x	L_x	T_x	e_x
0 años	3,01453	3,00646	996,99354	100.000,00	300,65	99.732,20	8.400.971,85	84,01
1 año	0,18869	0,18867	999,81133	99.699,35	18,81	99.690,15	8.301.239,65	83,26
2 años	0,14397	0,14396	999,85604	99.680,54	14,35	99.672,97	8.201.549,50	82,28
3 años	0,12623	0,12623	999,87377	99.666,19	12,58	99.659,58	8.101.876,53	81,29
4 años	0,08891	0,08890	999,91110	99.653,61	8,86	99.648,78	8.002.216,94	80,30
5 años	0,09699	0,09699				99.640,14		

Cuadro 1: Unidades: %1000, Años, Personas, Años-persona

Nota: m_x = tasa de mortalidad; q_x = riesgo de muerte; p_x = probabilidad de supervivencia; l_x = supervivientes; d_x = defunciones teóricas; L_x = población estacionaria; T_x = tiempo por vivir; e_x = esperanza de vida.

Pregunta 25. .

- a) ¿Qué dos fenómenos demográficos pueden estudiarse a partir del flujo de nacimientos? Describa qué se mide en cada uno de ellos.
- b) Indique las fórmulas para calcular:
 - La tasa específica de fecundidad por edad
 - La edad media al nacimiento del primer hijo

Pregunta 26. Consultando el Padrón Municipal de Habitantes, se conoce la población residente en España a 1 de enero de cada año durante el período 2017-2022. Además, utilizando el Movimiento Natural de la Población, se obtienen las cifras de nacimientos y defunciones producidas durante el mismo período.

Año	Población	Nacimientos	Defunciones
2022	47.475.420	329.251	464.417
2021	47.385.107	337.380	450.744
2020	47.450.795	341.315	493.776
2019	47.026.208	360.617	418.703
2018	46.722.980	372.777	427.721
2017	46.572.132	393.181	424.523

Cuadro 2: Unidades: Personas

A partir de dicha información, calcule el saldo migratorio y la tasa de migración neta para el quinquenio 2017-2021.

Pregunta 27. La siguiente tabla recoge la distribución por edad (grupos quinquenales) y sexo en 2025 de la población residente en la Comunidad de Madrid nacida en Marruecos y Colombia.

Cuadro 3: Unidades: Personas

	Marruecos			Colombia		
	Total	Hombres	Mujeres	Total	Hombres	Mujeres
Todas las edades	100.939	56.916	44.023	199.760	89.917	109.843
De 0 a 4 años	284	147	137	1.761	859	902
De 5 a 9 años	1.263	652	611	5.341	2.802	2.539
De 10 a 14 años	2.083	1.101	982	6.674	3.428	3.246
De 15 a 19 años	2.906	1.792	1.114	7.650	3.851	3.799
De 20 a 24 años	6.142	3.859	2.283	14.453	7.087	7.366
De 25 a 29 años	8.683	5.218	3.465	25.999	12.774	13.225
De 30 a 34 años	9.576	5.565	4.011	26.772	13.210	13.562
De 35 a 39 años	11.272	6.174	5.098	21.702	10.341	11.361
De 40 a 44 años	12.196	6.872	5.324	20.645	9.350	11.295
De 45 a 49 años	11.116	6.662	4.454	18.863	8.031	10.832
De 50 a 54 años	9.364	5.631	3.733	15.486	6.142	9.344
De 55 a 59 años	7.616	4.301	3.315	12.863	4.937	7.926
De 60 a 64 años	6.019	3.088	2.931	9.337	3.401	5.936
De 65 a 69 años	4.774	2.431	2.343	5.753	1.918	3.835
De 70 a 74 años	3.464	1.666	1.798	3.234	962	2.272
De 75 a 79 años	1.980	911	1.069	1.756	467	1.289
De 80 a 84 años	1.182	500	682	923	219	704
De 85 a 89 años	638	224	414	382	100	282
De 90 a 94 años	306	105	201	143	34	109
De 95 a 99 años	68	17	51	18	3	15
100 y más años	7	0	7	5	1	4

Calcule para cada procedencia:

- Índice de envejecimiento
- Índice de juventud
- Tasa de dependencia
- Tasa de masculinidad