

Cuestión 1 En el contexto de la gestión de la recogida de datos, explique brevemente qué son y para qué sirven los paradataos.

Cuestión 2 Bajo un muestreo estratificado (aleatorio simple sin reemplazamiento) sea $n_{dh} = \sum_{k \in h} z_{dk}$, $h : 1, 2, \dots, L$, el número de elementos de la muestra extraída en el estrato h que pertenece al dominio d , donde z_{dk} toma el valor 1 si la unidad k pertenece al dominio d y toma el valor 0 en caso contrario, y $p_{dh} = \frac{n_{dh}}{n_h}$ la proporción de elementos de la muestra extraída en el estrato h que pertenecen al dominio d .

Se pide:

- (i) ¿Cuál es el estimador de Horvitz-Thompson del total de número de elementos que pertenecen al dominio d , \hat{N}_d ?
- (ii) Escriba la expresión para la varianza de dicho estimador, $V(\hat{N}_d)$.

Cuestión 3 Describa brevemente en qué consiste la depuración selectiva.

Cuestión 4 Defina un muestreo sistemático. Comente brevemente la problemática de la estimación de la varianza con una muestra sistemática.

Cuestión 5 En el caso de un muestreo con probabilidades proporcionales al tamaño (sin reemplazamiento) de tamaño fijo n , se pide:

- (i) ¿Cuál es la probabilidad de inclusión (de pertenecer a la muestra s) de un elemento k denotada por π_k ?
- (ii) Demuestre que la varianza del estimador de Horvitz-Thompson del estimador del total Y , cuando y_k es proporcional a π_k , es cero, donde y_k es el valor que toma la variable Y en la unidad k .

Cuestión 6 En el contexto de un diseño muestral $p(\cdot)$ que a cada muestra s le hace corresponder la probabilidad de que dicha muestra sea seleccionada, $p(s)$, para todo s del conjunto posible de muestras, se pide:

- (i) Defina la probabilidad α en el caso de un intervalo de confianza para el parámetro θ , denotado por $IC_\theta(s)$, con un nivel de confianza $1 - \alpha$.
- (ii) Sea el siguiente intervalo de confianza para el parámetro θ :

$$IC_\theta(s) = [\hat{\theta} - z_{\alpha/2} \sqrt{\hat{V}(\hat{\theta})} , \hat{\theta} + z_{\alpha/2} \sqrt{\hat{V}(\hat{\theta})}]$$

¿Qué condiciones se deben cumplir para que $IC_\theta(s)$ tenga un nivel de confianza de aproximadamente $1 - \alpha$?

Cuestión 7 Defina qué es una celda sensible en un contexto de control de secreto estadístico. Indique tres métodos para evitar la aparición de celdas sensibles en las tablas estadísticas.

Cuestión 8 Comente en qué consiste el “record linkage” y describa brevemente dos aplicaciones de este método.

Cuestión 9 Explique para qué se usa el estándar SDMX.

Cuestión 10 Enuncie los subprocesos de la fase 5 (Procesar) del “Generic Statistical Business Process Model” (GSBPM).

Cuestión 11 Para una distribución de frecuencias de una variable bidimensional (X, Y) se han obtenido las rectas de regresión (a) $8x + y = 1$, (b) $2x + 4y = 3$.

- (i) Razone si la recta (a) es o no la recta de regresión de y sobre x .
- (ii) Estime el coeficiente de correlación de X e Y .

Cuestión 12 Para estudiar la incidencia de una enfermedad en un territorio, se considera este dividido en 100 cuadrículas de igual área y condiciones epidemiológicas homogéneas. En las condiciones actuales, se esperan tres enfermos en cada cuadrícula por término medio. Además, la variable aleatoria X número de pacientes en una cuadrícula se puede suponer que sigue una distribución de

Poisson.

- (i) Escriba la función de masa de probabilidad de X .
- (ii) ¿Cuál es la probabilidad de no encontrar ningún enfermo en una cuadrícula?
- (iii) ¿Cuál es la probabilidad de encontrar al menos cuatro en una cuadrícula?
- (iv) ¿Cuál es el número esperado de cuadrículas en las que se encontrarán cuatro enfermos?

Nota: Puede dejar los cálculos indicados en función de e^{-3} .

Cuestión 13 Los ingresos por hogar en una ciudad siguen aproximadamente una distribución normal de media 45000 euros y desviación típica 8000 euros. Se seleccionan al azar 100 hogares. Se tiene acceso a una rutina de sintaxis $pnorm(q)$ que dado q da la probabilidad en la distribución normal estándar de obtener una observación menor que q . Indique la probabilidad de que:

- (i) La media de los ingresos sea menor que 48200.
- (ii) La media sea mayor que 45800.
- (iii) La diferencia entre las medias de dos muestras independientes de tamaño 100 sea mayor que 100 euros.

Cuestión 14 La altura en cm de las mujeres adultas de una población sigue una distribución normal de media 168 y desviación típica 6. Se tiene acceso a una rutina $qnorm$ tal que para un valor de probabilidad da el cuantil de nivel p de la distribución normal estándar.

- (i) Halle un intervalo simétrico de probabilidad 0.98 para la altura de tales mujeres.
- (ii) Se extraen al azar 100 mujeres. Para un nivel de probabilidad 0.88, calcule entre qué valores estará comprendida la altura media de la muestra.

Cuestión 15 Se desea probar si una vacuna nueva es más eficaz que otra pre-existente cuya efectividad es del 65%. Se vacuna a 60 personas elegidas al azar, resultando eficaz en 48 de ellas. Realizar el contraste clásico para un nivel de significación 0.03, supuesto que se tiene acceso a una rutina $qnorm$ tal que para un valor de probabilidad da el cuantil de nivel p de la distribución normal estándar.

Cuestión 16 Se desea estimar la tasa λ a partir de un experimento exponencial de tamaño n . Se dispone de la información a priori de que λ se distribuye según una distribución exponencial de media $1/3$. Dése un estimador puntual basado en la media de la distribución a posteriori de λ .

Cuestión 17 Un dispositivo permite detectar cierta enfermedad en el 80% de las personas con la misma, cometiendo errores con sólo el 5% de las personas que no la tienen. La incidencia de la enfermedad en la población es de 30 casos por cada 10000 habitantes.

- (i) Calcule la probabilidad de que una persona tenga la enfermedad si da positivo una vez en la prueba.
- (ii) Calcule la probabilidad de que una persona tenga la enfermedad si da positivo en n repeticiones independientes de la prueba.
- (iii) Indique cómo calcular n para que esta probabilidad sea mayor que 0.5.

Cuestión 18 ¿En qué consiste el formato little-endian? ¿En qué se diferencia del formato big-endian? Explique cómo se representaría el entero 25, en una máquina de 32 bytes, con el formato de little endian, en una palabra situada en la dirección de memoria 520.

Cuestión 19 Describa las siguientes funcionalidades, indicando el papel que juegan, en su caso, en un dispositivo portátil, en un mainframe o en un sistema de tiempo real.

- (i) Programación por lotes (Batch programming).
- (ii) Memoria virtual.
- (iii) Tiempo compartido.

Cuestión 20 El Departamento de Clientes de un Banco centraliza los datos de los clientes en una tabla con la siguiente cabecera:

IdCliente	Nombre	DirCliente	NumCuenta	TipoCuenta	Saldo	Sucursal	DirSucursal
-----------	--------	------------	-----------	------------	-------	----------	-------------

Explique los problemas que se aprecian en este diseño y aporte soluciones concretas.

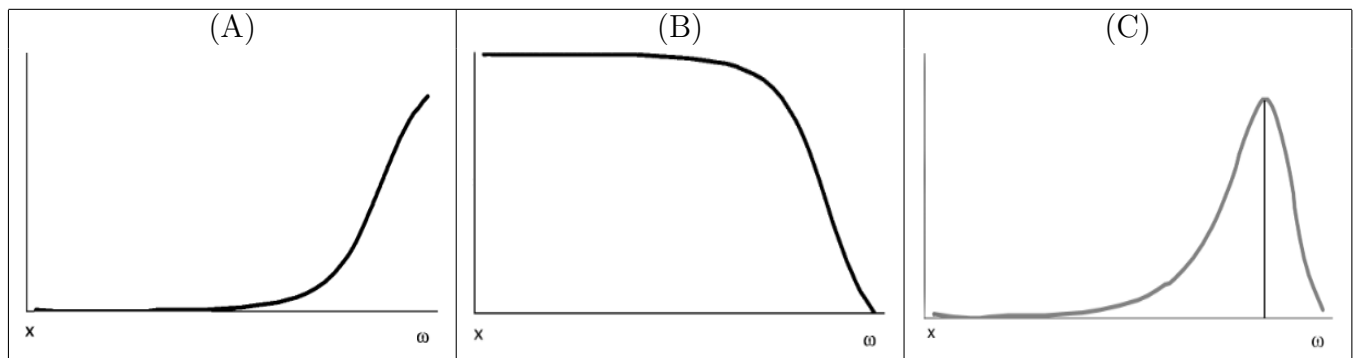
Cuestión 21 Describa los principios de partida doble y cuádruple en los que se basa el sistema de cuentas de la Contabilidad Nacional.

Cuestión 22 Describa qué es la tabla de origen-destino en el marco de la Contabilidad Nacional y cuál es su estructura básica.

Cuestión 23 Comente los enfoques generales para la elaboración de cuentas nacionales trimestrales.

Cuestión 24 Mencione y describa brevemente al menos 4 aspectos relacionados con la globalización económica que supongan un desafío para la medición estadística en el sistema de cuentas nacionales.

Cuestión 25 ¿Cuál de los siguientes gráficos se corresponde con una curva d_x de defunciones? Razone la respuesta.



Cuestión 26 Defina el indicador de edad media a la maternidad.

Cuestión 27 Explique brevemente en qué consiste el método de los “signos de vida” utilizado en los Censos de Población y Vivienda.