

CUESTIÓN 1

Se quiere estimar el total de una característica Y en una población de tamaño N . Para ello, se tiene información de dos variables auxiliares X_1 , total poblacional de la variable X_1 y X_2 , total poblacional de la variable X_2 .

Se pide:

A) Sea el estimador sintético para el total de Y dado por la siguiente expresión:

$$\hat{Y}_R = \alpha \hat{Y}_{R1} + \beta \hat{Y}_{R2}$$

Donde $\alpha + \beta = 1$ e \hat{Y}_{Ri} es el estimador del total de Y por el método de la razón usando la variable auxiliar X_i , $i=1,2$.

(a) ¿Es el estimador \hat{Y}_R insesgado?

(b) Calcule el valor de α que minimice la varianza de \hat{Y}_R

B) Se quiere estimar la superficie media de 100 explotaciones agrarias de una comarca en 2023, media en hectáreas, que se denota por $\bar{Y} = Y/N$. De censos anteriores se conoce la superficie media en 2020 ($\hat{X}_1 = 1100$) y 2010 ($\hat{X}_2 = 900$).

Para ello se extrae una muestra aleatoria simple sin reemplazamiento de tamaño $n=40$. Los datos disponibles son:

- Estimación de las medias usando un estimador de expansión simple:

$$\hat{\bar{Y}} = 1050; \quad \hat{\bar{X}}_1 = 990; \quad \hat{\bar{X}}_2 = 910;$$

- Cuasidesviaciones muestrales:

$$\begin{aligned} s_y &= 975; & s_{x1} &= 901; & s_{x2} &= 820; \\ s_{x1y} &= 905; & s_{x2y} &= 850; & s_{x1x2} &= 800; \end{aligned}$$

(c) Calcule la estimación de la media de Y usando el estimador de razón \hat{Y}_{R1} junto a su error de muestreo.

(d) Calcule la estimación de la media de Y usando el estimador de razón \hat{Y}_{R2} junto a su error de muestreo.

(e) Calcule la estimación de la media de Y usando el estimador de razón \hat{Y}_R cuando α toma el valor obtenido a partir de la expresión del apartado A.2), junto a su error de muestreo.

CUESTIÓN 2

Suponga que las observaciones (x, y) satisfacen el modelo $y_i = \beta_0 + \sum_{j=1}^p x_{ij}\beta_j + \epsilon_i$ donde $\epsilon_1, \dots, \epsilon_n$ se distribuyen independientemente según una $N(0, \sigma^2)$.

(a) Escriba la verosimilitud de los datos.

(b) Suponga que a priori $(\beta_1, \dots, \beta_p)$ son independientes e idénticamente distribuidas según una distribución de densidad $p(\beta) = \frac{1}{2b} \exp(-|\beta|/b)$, con parámetro b común. Argumente que el estimador lasso surge como moda de la correspondiente distribución a posteriori.

(c) Suponga, ahora, que a priori $(\beta_1, \dots, \beta_p)$ son independientes e idénticamente distribuidas según una distribución normal de media 0 y varianza c . Argumente que el estimador de regresión *ridge* surge como moda de la correspondiente distribución a posteriori.

CUESTIÓN 3

Tenemos una base de datos con datos geográficos del mundo: contiene información sobre países, con su código correspondiente, la población total del país, la superficie y la renta per cápita. También contiene información sobre ciudades, con su nombre, un código, su población y el país al que pertenecen. Es posible que nuestra base de datos no contenga la población de alguna ciudad. En cambio, los demás datos (nombres de las ciudades y países, códigos, etc.) son obligatorios.

(a) Proporcione un ejemplo de un documento XML bien formado para este caso, e indique las reglas que cumple para que se considere un documento bien formado.

(b) Dé instrucciones SQL de creación de dichas tablas, sin olvidar vincularlas con las claves externas necesarias, en su caso.

(c) Elabore una consulta SQL para averiguar cuáles son las ciudades (5 máx.) más pobladas del país del mundo con la renta per cápita más alta registrado en nuestra base de datos.

(d) Elabore una consulta SQL para eliminar de nuestra base de datos el país de menor superficie. Esta operación debe evitar dejar inconsistencias en la base de datos con respecto a las ciudades del país eliminado.

CUESTIÓN 4

Diseñe una función que simule una v. a. exponencial, cuya función de densidad es la siguiente:

$$f(x) = \lambda e^{-\lambda x}, \text{ para } x \geq 0$$

Para la simulación, podemos suponer definida una función $U(0, 1)$, esto es, una función que simula una v. a. real, uniformemente distribuida en el intervalo $[0, 1]$.

El diseño de este algoritmo puede realizarse en pseudocódigo o en algún lenguaje imperativo como C/C++, Java, Python o R.

CUESTIÓN 5

Una población está formada por $M = 3000$ viviendas (unidades elementales) agrupadas en 30 municipios de tamaños desiguales ($M_i, i = 1, 2, \dots, 30$). Se quiere estimar la proporción de viviendas que pertenecen a una cierta clase y para ello se usa un muestreo de conglomerados con submuestreo.

En ambas etapas se emplea un procedimiento de selección con probabilidades iguales sin reposición.

En la primera etapa se seleccionan 3 municipios con los siguientes valores de M_i : 10; 50; 100.

En la segunda etapa, realizada con una fracción de muestreo de $f_{2i} = 4/M_i$, se obtienen en los 3 municipios muestrales los valores 2; 2; 3 respectivamente para el número de viviendas que pertenecen a la clase.

Se pide:

- (a) Una estimación insesgada de la mencionada proporción y su error de muestreo.
- (b) Una estimación, usando el estimador de la razón al tamaño, de la proporción mencionada y su error de muestreo.
- (c) Comentar ventajas e inconvenientes del estimador de (b) frente al estimador de (a).

CUESTIÓN 6

Se modeliza la ocurrencia diaria de lluvia en otoño en cierta estación meteorológica mediante una cadena de Markov con dos estados (1) no llueve y (2) llueve. Se han recogido los siguientes datos de lluvia en 49 días:

2222222222 1121111111 2211112221 2111112111 111111111

Se pide:

- (a) Defina una matriz de transición para la cadena.
- (b) Estime mediante máxima verosimilitud las entradas de la matriz de la cadena.
- (c) Suponga distribuciones a priori uniformes para los elementos de la diagonal de la matriz.
 - Calcule las distribuciones a posteriori para los elementos de la cadena.
 - Calcule las probabilidades de transición esperadas a posteriori.
 - Calcule la probabilidad predictiva a posteriori de lluvia para el día 50.
- (d) Calcule la probabilidad predictiva a posteriori de que no llueva los días 50 y 51.

CUESTIÓN 7

Una agencia de viajes necesita desarrollar un sistema para gestionar información sobre pisos turísticos, incluyendo el país, ciudad y dirección en que se ubican, así como otras características tales como número de dormitorios y baños, disponibilidad de cocina, etc.

(a) Defina XML, diga qué reglas debe cumplir un documento bien formado y proporcione un ejemplo de un documento XML bien formado para el ejemplo anterior.

(b) Se desea diseñar una aplicación web para gestionar la información de nuestra agencia. Describa los conceptos de frontend, backend, lado del cliente y lado del servidor, apoyándose en nuestra agencia de viajes. En cada uno de estos conceptos, mencione alguna tecnología que tendría sentido usar.

CUESTIÓN 8

Deseamos obtener la mediana de una secuencia de números enteros. Los números enteros no están ordenados. La secuencia puede ser un vector o una lista, posiblemente de gran tamaño, por lo que nuestro algoritmo ha de ser lo más eficiente posible.

El diseño de este algoritmo puede realizarse en pseudocódigo o en algún lenguaje imperativo como C/C++, Java, Python o R. No estará permitido usar ninguna función predefinida que ordene dicha secuencia. Adicionalmente, se pide discutir la complejidad en tiempo de dicho algoritmo, o al menos un boceto de dicho cálculo.