

Ejercicios teóricos

Responder tres ejercicios

1. Sean $x_1, \dots, x_n, y_1, \dots, y_n$ observaciones. Mostrar las siguientes igualdades

$$a) \sum_{i=1}^n (x_i - \bar{x}_n)^2 = \sum_{i=1}^n x_i^2 - n\bar{x}_n^2,$$

$$b) \sum_{i=1}^n (x_i - \bar{x}_n)(y_i - \bar{y}_n) = \sum_{i=1}^n x_i y_i - n\bar{x}_n \bar{y}_n.$$

2. En el ejercicio anterior, suponer que se definen $z_i = a + bx_i$, para $i = 1, \dots, n$, con a y b constantes. Mostrar las siguientes igualdades

$$a) \bar{z}_n = a + b\bar{x}_n,$$

$$b) s_z^2 = b^2 s_x^2.$$

donde $\bar{x}_n = \frac{1}{n} \sum_{i=1}^n x_i$, $\bar{z}_n = \frac{1}{n} \sum_{i=1}^n z_i$, $s_x^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x}_n)^2$ y $s_z^2 = \frac{1}{n-1} \sum_{i=1}^n (z_i - \bar{z}_n)^2$.

3. Suponer que se tienen las observaciones x_1, \dots, x_n . Mostrar que la función

$$f(y) = \sum_{i=1}^n (x_i - y)^2, \quad y \in \mathbb{R}$$

alcanza su mínimo cuando $y = \bar{x}_n$, donde $\bar{x}_n = n^{-1} \sum_{i=1}^n x_i$.

En el resultado anterior se muestra que el promedio minimiza la suma de las desviaciones al cuadrado.

4. Suponer que se tienen las observaciones x_1, \dots, x_n , todas distintas. Mostrar que la función

$$g(y) = \sum_{i=1}^n |x_i - y|, \quad y \in \mathbb{R}$$

alcanza su mínimo cuando y es igual a la mediana de las observaciones.

Hint: considerar por separado los casos cuando n es par e impar.

En el resultado anterior se muestra que la mediana minimiza la suma de las desviaciones absolutas.

5. Sean $\{x_1, \dots, x_n\}$ y $\{y_1, \dots, y_n\}$ dos conjuntos de datos con el mismo número de observaciones. Mostrar que

$$a) r_{xy} \in [-1, 1],$$

$$b) r_{xy} = \pm 1 \text{ si y sólo si existen constantes } a \text{ y } b \text{ tales que } y_i = a + bx_i, \text{ para } i = 1, \dots, n.$$

Hint: utilizar la desigualdad de Cauchy-Schwarz.

Ejercicios prácticos

Responder todos los ejercicios

6. Del archivo de descripción de las bases de microdatos de la Encuesta Intercensal 2015 del INEGI (disponible [aquí](#)), seleccionar 9 variables de la base de datos de viviendas (hoja TR_Viviendas) y 9 variables de la base de datos de personas (hoja TR_personas), de manera que se incluyan: 3 nominales, 3 ordinales y 3 en escala de intervalo o de razón.

No considerar las variables de identificación: llaves, claves geográficas, identificadores, variables del diseño de muestreo. Tampoco tomar en cuenta las categorías nulo, blanco por pase, no respondió, no sabe, no especificado, etc.

7. El archivo `muestraENIGH.csv` (disponible en [aquí](#)) contiene una muestra de una de las bases de microdatos de la Encuesta Nacional de Ingresos y Gastos en los Hogares (ENIGH) realizada por el INEGI en 2014. Cada registro contiene la información de un hogar.

Utilizar la información en el archivo anterior para responder las siguientes preguntas. Se puede usar como referencia el documento *Descripción de la bases de datos* del INEGI (disponible ([disponible aquí](#))), la tabla a consultar es `CONCENTRADOHOGAR`.

- ¿Cuántas variables y cuántas observaciones contiene el archivo?
 - Calcular el promedio y la mediana del número de integrantes del hogar, ¿son iguales estos dos estadísticos?
 - Calcular el promedio y la mediana de las edades de los jefes del hogar, ¿son iguales estos dos estadísticos?
 - Graficar un histograma y la densidad estimada de las edades de los jefes del hogar.
 - Calcular la proporción de jefes del hogar con y sin **educación básica completa**, ¿qué proporción es mayor?
 - Calcular el promedio del ingreso corriente en los hogares según la escolaridad del jefe del hogar, considerar únicamente las categorías educación básica incompleta y educación básica completa. ¿Parece haber alguna diferencia en los ingresos?
 - Graficar en un mismo plano *boxplots* del ingreso corriente de los hogares para las categorías educación básica incompleta y educación básica completa. ¿Parece haber alguna diferencia en los ingresos?
 - Graficar las densidades estimadas del ingreso corriente en los hogares según la escolaridad del jefe del hogar. De nuevo, considerar únicamente las categorías educación básica incompleta y educación básica completa. ¿En que categoría hay mayor variabilidad en los ingresos?
8. Utilizar la base muestra de microdatos de la ENIGH 2014 del archivo `muestraENIGH.csv` (disponible en [aquí](#)), para responder lo siguiente.

- a) Calcular el coeficiente de asimetría del ingreso corriente. ¿La distribución del ingreso es simétrica?
- b) Crear una nueva variable que sea el logaritmo natural (base e) del ingreso corriente y calcular la su coeficiente de asimetría. ¿Es simétrica la distribución de esta nueva variable?
- c) Calcular la curtosis del logaritmo del ingreso. ¿La forma del esta variable es similar a la distribución normal o presenta alguna desviación?
- d) En un mismo plano, graficar la densidad estimada del logaritmo del ingreso (con las opciones por defecto de \mathbb{R}) y la curva de una densidad normal con media y desviación estándar las calculadas de los datos. ¿El gráfico anterior refleja las conclusiones obtenidas en los incisos anteriores?
- e) Repetir los incisos anteriores para la edad de los jefes del hogar.
- f) Calcular la covarianza y la correlación entre la edad del jefe del hogar y el ingreso corriente. ¿Parece haber alguna relación lineal entre estas dos variables? Si la respuesta es afirmativa, ¿de qué tipo es la relación?

9. Considerar el siguiente conjunto de 12 observaciones

$$\{0.0, 2.3, 1.9, 3.9, 0.6, 2.7, 5.8, 1.8, 3.2, 1.7, 1.1, 21.5\}$$

- a) Calcular los estadísticos: promedio, mediana, desviación estándar y el siguiente estadístico

$$MAD = \frac{1}{n} \sum_{i=1}^n |x_i - m|,$$

donde m es la mediana de las observaciones.

- b) Eliminar la última observación y calcular nuevamente los estadísticos del inciso anterior.
- c) ¿Qué estadísticos tuvieron menos cambios al eliminar la última observación?

En el resultado anterior se comprueba que la mediana y la desviación absoluta media son estadísticas más *robustas* que la media y la desviación estándar, es decir, son menos susceptibles al efecto de observaciones atípicas.