Conceptos básicos de inferencia estadística | Semestre 2018-1

Tarea 2

Fecha de entrega: 31 de agosto

1. Suponer que se tienen las observaciones $x_1, \ldots, x_n$. Mostrar que la función

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

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.

2. Suponer que se tienen las observaciones $x_1, \ldots, x_n$, todas distintas. Mostrar que la función

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

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.

3. 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 `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.

4. 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 la desviación absoluta media; esta última definida como

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

   donde M es la mediana de las observaciones.

   b) Eliminar la última observación y calcular 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.

5. Graficar *a mano* (con el código del ejemplo de clase) un *boxplot* de los datos del inciso anterior. Utilizar $k = 1.1$. Responder lo siguiente.

   a) ¿Cuál es el valor de los estadísticos Q_1, Q_2, Q_3 e IQR?

   b) ¿Cuál es el tamaño del *bigote* derecho y por qué? y ¿cuántas observaciones extremas hay a la derecha?

   c) ¿Cuál es el tamaño del *bigote* izquierdo y por qué? y ¿cuántas observaciones extremas hay a la izquierda?

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

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

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

   Hint: utilizar la desigualdad de Cauchy-Schwarz.

7. Con los datos del ejercicio 3 responder lo siguiente.

   a) Calcular la correlación entre la educación del jefe del hogar y el ingreso corriente. ¿Existe alguna relación lineal entre estas dos variables? Si la respuesta es afirmativa, ¿de qué tipo es la relación?

   b) Calcular la correlación entre la edad y la educación del jefe del hogar. ¿Existe alguna relación lineal entre estas dos variables? Si la respuesta es afirmativa, ¿de qué tipo es la relación?

   c) Hacer un gráfico de disperción para cada par de variables de los incisos anteriores. ¿Los gráficos indican que hay alguna relación lineal entre cada par de variables?