Ejercicios 1

Rodrigo Zepeda

28 de enero de 2020

1. Calidad de la información

Del Lohr capítulo 1 sección 1.7 (en comunidad) los ejercicios 6 al 15

2. Análisis Exploratorio de Datos

1. Obtén una expresión en fórmula matemática (como hicimos en clase) para los siguientes estadísticos para unos datos observados $\mathcal{S} = \{x_1, x_2, \dots, x_n\}$ donde por simplicidad asumiremos que si $\mathcal{S}$ puede ser ordenado entonces $x_1 \leq x_2 \leq x_n$:

   a) El valor numérico o categoría menos común en $\mathcal{S}$

   b) El número de veces que la categoría o el número ℓ fue observado en $\mathcal{S}$ (podría ser 0).

   c) La media de las diferencias entre las x_i quitando la de x_k consigo misma

   d) Este cálculo de R para una S dada como vector numérico:

   ```
   #S es la muestra; S <- c(x1,x2, ..., xn)
   datos_nuevos <- c()
   for (i in 1:length(S)){
       datos_nuevos <- c(datos_nuevos, S[i]^i)
   }
   mean(datos_nuevos) #Este valor es el que me interesa
   ```

2. **Pseudocódigo**. Escribe el pseudocódigo de funciones (una por inciso) que hagan lo siguiente para una muestra $\mathcal{S}$ dada por `S <- c(x1,x2, ..., xn)`

   a) Calcule la moda.

   b) Enliste todos los valores de `S` que están después de la mediana (sin incluirla). Considera que hay al menos un valor después de la mediana y distinto a ella.

   c) Calculen el rango.

d) Calculen el MAD (desviación mediana absoluta).

e) Implemente una densidad kernel con núcleo triangular dado por:

$$K(x) = (1 - |x|)\mathbb{I}_{[-1,1]}(x)$$

y con h dada por el usuario.

3. Demuestra que:

$$\sigma_{\mathcal{S}}^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \mu_{\mathcal{S}}^2$$

4. Demuestra que la media aritmética es mayor que la geométrica; es decir:

$$\mu_{\mathcal{S}} = \frac{1}{n}\sum_{i=1}^{n} x_i \geq \left(\prod_{i=1}^{n} x_i\right)^{1/n}$$

Hint:

a) Primero demuestra que $f(x) = e^{x-1} - x$ es estrictamente convexa con mínimo cuando $x = 1$

b) Demuestra que $x \leq e^{x-1}$

c) Usa dicha desigualdad para probar que:

$$\prod_{i=1}^{n} \frac{x_i}{\mu_{\mathcal{S}}} \leq \exp\Big(\sum_{i=1}^{n} \big[\frac{x_i}{\mu_{\mathcal{S}}} - 1\big]\Big) \tag{1}$$

d) Concluye despejando

5. Suponiendo que en unos datos observados (numéricos) $\mathcal{S} = \{x_1, \dots, x_n\}$ se tiene que las x toman el valor n_1 a_1 veces, el valor n_2, a_2 veces y el valor n_k, a_k veces ($k \leq n$, $a_j > 0$ y $\sum_j n_j = n$)

a) Demuestra que:

$$\mu_S = \frac{1}{\sum_{\ell=1}^{k} a_\ell} \cdot \sum_{j=1}^{k} n_j a_j \tag{2}$$

b) Redefine la varianza usando las n_k, a_k

6. Sea n impar y f una función estrictamente creciente.

a) Demuestra que si x_* es la mediana de $\mathcal{S} = \{x_1, x_2, \dots, x_n\}$ entonces $f(x_*)$ es la mediana de $\tilde{\mathcal{S}} = \{f(x_1), f(x_2), \dots, f(x_n)\}$.

b) ¿Qué pasa si n es par y f es una función estrictamente creciente y lineal dada por $f(x) = ax + b$?

c) ¿Qué pasa si n es par y f es una función estrictamente creciente general?

7. Sea ϕ una función convexa. Demuestra que:

$$\phi\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) \leq \frac{1}{n}\sum_{i=1}^{n}\phi(x_i) \tag{3}$$

Hint Recuerda que si ϕ es convexa, para $0 \leq \alpha \leq 1$ se tiene que:

$$\phi\big(\alpha x + (1-\alpha)y\big) \leq \alpha\phi(x) + (1-\alpha)\phi(y) \tag{4}$$

8. De la tabla 1 calcula todo lo que se te pide de la variable calificaciones:

Persona	Calificación
1	10
2	5
3	6.75
4	10
5	6.5
6	3.25
7	6
8	9
9	7.5
10	100

Cuadro 1: Tabla de calificaciones de un salón

a) La media
b) La mediana
c) La moda
d) La varianza
e) Una tabla de calificaciones normalizadas
f) La MAD
g) El rango
h) El rango intercuartílico
i) El cuantil 0.6
j) El total
k) La cantidad de valores mayores o iguales a 6
l) La curtosis
m) La asimetría
n) La desviación estándar
ñ) La media truncada en $d_1 = 2$ y $d_2 = 8$
o) Mínimo y máximo

p) Elaborar un histograma con 3 `bins`

q) Elaborar un boxplot

r) La función de distribución acumulada empírica (y graficarla)

s) La función de densidad kernel evaluada en $x = 7$ con un kernel Epanechnikov y ancho de banda $h = 0.5$.