0.1. Tema 1 - Estadística descriptiva

Ejercicio 2: Demostrar que

$$\sum_{i=1}^{n} (x_i - \overline{x})^2 = \min_{a \in \mathbb{R}} \sum_{i=1}^{n} (x_i - a)^2$$

Definimos una función

$$g(a) = \sum_{i=1}^{n} (x_i - a)^2$$

, buscamos su derivada

$$g'(a) = -2 \sum_{i=1}^{n} (x_i - a)$$

e igualamos a cero:

$$-2 \sum_{i=1}^{n} (x_i - a) = 0$$

$$\sum_{i=1}^{n} x_i - \sum_{i=1}^{n} a = 0$$

$$n\overline{x} = na$$

$$\overline{x} = a$$

Esto quiere decir que la media muestral es el valor que minimiza la distancia con cada uno de los datos de la muestra.

Ejercicio 5: Determina si es verdadero o falso:

a) Si añadimos 7 a todos los datos de un conjunto, el primer cuartil aumenta en 7 unidades y el rango intercuartílico no cambia.

b) Si todos los datos de un conjunto se multiplican por -2, la desviación típica se dobla.

APARTADO A) *Añadir siete a todos los datos es una traslación, así que la distribución de los datos no cambia.*

APARTADO B) *Teniendo en cuenta que si multiplicamos todos los datos del conjunto por -2 la media también se multiplica por -2, y sustituyendo en la fórmula de la varianza:*

$$\sigma' = \sqrt{\frac{1}{n}\sum_{i=1} n(-2x_i)^2 - (-2\overline{x})^2} = \sqrt{\frac{1}{n}\sum_{i=1} 4\left(nx_i^2 - \overline{x}^2\right)} = \sqrt{4\sigma^2} = 2\sigma$$

Por lo tanto, la desviación típica sí se dobla.

APARTADO C) *Usando los cálculos del apartado anterior vemos que la varianza se multiplica por cuatro.*

APARTADO D) *Efectivamente: cambiar el signo haría una reflexión de los datos sobre el eje Y y la asimetría estaría orientada hacia el lado contrario.*

0.2. Tema 2 - Estadísticos

Ejercicio 4: Denotemos por

$$C_n = \int_{\mathbb{R}} (\mathbb{F}_n(t) - F(t))^2\, dF(t)$$

la llamada discrepancia de Cramer -Von Mises entre $\mathbb{F}_N$ y F.

a) ¿Se verifica necesariamente que $C_n \to 0$ c.s.?

b) Calcular la distribución asintótica de $D_N = \sqrt{n}(\mathbb{F}_n(t) - F(t)) \xrightarrow[n\to\infty]{d}$?

a)

$$\begin{aligned} C_n &= \int_{\mathbb{R}} (\mathbb{F}_n(t) - F(t))^2\, dF(t) \\ &= \int_{\mathbb{R}} \underbrace{(\mathbb{F}_n(t) \overset{2}{-} F(t))}_{\leq sup} f(t)dt \\ &\leq ||\mathbb{F}_n - F||_\infty^2 \underbrace{\int_{\mathbb{R}} f(t)dt}_{=1.\ f\ densidad} = ||\mathbb{F}_n - F||_\infty^2 \end{aligned}$$

Aplicando el teorema de Glivenko-Cantelli $||\mathbb{F}_n - F||_\infty^2 \to 0$

Tanto si es continua como si es discreta: $\int_{\mathbb{R}} 1dF(t) = 1$

b) *Recordamos que* $\mathbb{F}_n(t) = \frac{1}{n}\sum_{i=1}^n \underbrace{1_{(-\infty,t]}(X_i)}_{Y_i}$

Queremos aplicar el teorema central del límite (que es lo que usaremos siempre que queramos hallar funciones de distribución).

TCL: $\sqrt{n}(\overline{X}-\mu) \xrightarrow[n\to\infty]{d} N(0,\sigma)$

Llamando $Y = 1_{-\infty,t]}(X)$. *Calculamos* $E(Y) = P\{X \leq t\} = F(t)$, *porque* Y *es una Bernoulli.*

$$D_N = \sqrt{n}(\mathbb{F}_n(t) - F(t)) = \sqrt{n}(\overline{Y}$$