

Ayudantia 1. Mat 044

Record: Sea x_1, \dots, x_n una m.a (n)

datos: $\vec{x} = (x_1, \dots, x_n)$; $x_1 < \dots < x_n$

mean(\vec{x}) = $\frac{1}{n} \sum_{i=1}^n x_i$; Propiedades.

mean es lineal; sea $\vec{x}, \vec{y} \in \mathbb{R}^n$; $\alpha \in \mathbb{R}$

$$\text{mean}(\alpha \cdot \vec{x} + \vec{y}) = \alpha \text{mean}(\vec{x}) + \text{mean}(\vec{y})$$

$$\text{mean}(\alpha \cdot \vec{1}) ; \vec{1} = (1, \dots, 1) \in \mathbb{R}^n$$

$$\text{mean}(\alpha \vec{1}) = \sum_{i=1}^n \alpha \left(\frac{1}{n}\right) = \frac{n\alpha}{n} = \alpha$$

$$\text{Var}(\vec{x}) = s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

$$\text{Var}(\alpha \cdot \vec{x}) = \alpha^2 \cdot \text{Var}(\vec{x})$$

$$\text{Var}(\vec{x} + \alpha \cdot \vec{1}) = \text{Var}(\vec{x})$$

$$\boxed{\text{Var}(\vec{x} + \vec{y}) \neq \text{Var}(\vec{x}) + \text{Var}(\vec{y})}$$

Problema 1

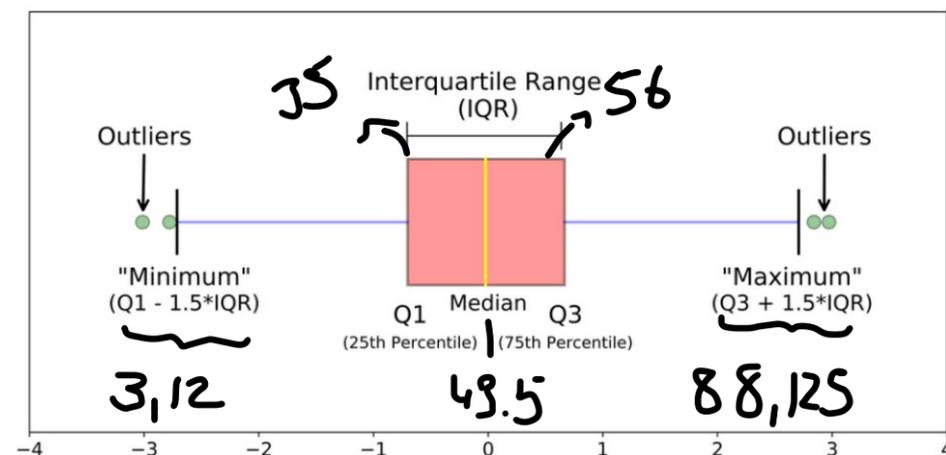
Considere las notas del Certamen 1 de MAT044 del año pasado en cada uno de los dos paralelos que se dictaron (P100 y P101):

Paralelo	n	min	P25	mediana	media	P75	max	s
P100	40	19	35	43.5	45.4	56.25	76	12.8
P101	13	21	34	39.0	41.8	43.00	78	17.8

← Duro.
Estándar
(s^2 para Var)

1. Describa la distribución de las notas obtenidas en el Certamen 1 por los estudiantes del paralelo P100.
2. El profesor del paralelo P100 ha decidido sumar los puntos obtenidos en una tarea en clases al Certamen 1. Suponga que todos los alumnos obtuvieron 5 puntos en la tarea. Sea X: nota del Certamen 1 (incorporando puntos obtenidos en la tarea) del paralelo P100. Obtenga la media, desviación estándar y CV de X.
3. El profesor del paralelo P101 ha decidido incorporar la asistencia a clases al Certamen 1, como un factor que pondera dicha nota. Suponga que todos los alumnos tienen un factor igual a 1.1. Sea Y: nota del Certamen 1 (ponderada por el factor) del P101. Calcule el CV de Y.
4. Luego de las decisiones de los profesores, ¿Cuál paralelo es el que tuvo la mayor variabilidad en las notas del Certamen 1?

1.



P100: IQR = 21,25 ; IQR = $Q_3 - Q_1$

• No hay datos atípicos.

• Muestra tiene una ligera tendencia a la derecha.

obs: Un mejor boxplot sería con los bigotes iguales al max y min, pero no cambiaría la descripción del conjunto de datos.

2. El profesor del paralelo P100 ha decidido sumar los puntos obtenidos en una tarea en clases al Certamen 1. Suponga que todos los alumnos obtuvieron 5 puntos en la tarea. Sea X : nota del Certamen 1 (incorporando puntos obtenidos en la tarea) del paralelo P100. Obtenga la media, desviación estándar y CV de X .

Paralelo	n	min	P25	mediana	media	P75	max	s
P100	40	19	35	43.5	45.4	56.25	76	12.8

Sea Z_i : la nota original
 Sea X_i : la nota i -ésima más 5 puntos.

$$\begin{aligned} \bullet \text{mean}(\vec{X}) &= \text{mean}(\vec{Z} + 5 \cdot \vec{1}) \\ &= \text{mean}(\vec{Z}) + 5 \\ &= 45.4 + 5 \\ &= 50.4 \end{aligned}$$

$$\begin{aligned} \bullet \text{Var}(\vec{X}) &= \text{Var}(\vec{Z} + 5 \cdot \vec{1}) \\ &= \text{Var}(\vec{Z}) \end{aligned}$$

$$\Rightarrow \text{sd}(\vec{X}) = \text{sd}(\vec{Z}) = s = 12.8$$

$$\bullet \text{CV}(\vec{X}) = \frac{s_x}{\bar{x}} = \frac{12.8}{50.4} = 0,25 //$$

3. El profesor del paralelo P101 ha decidido incorporar la asistencia a clases al Certamen 1, como un factor que pondera dicha nota. Suponga que todos los alumnos tienen un factor igual a 1.1. Sea Y : nota del Certamen 1 (ponderada por el factor) del P101. Calcule el CV de Y .

Paralelo	n	min	P25	mediana	media	P75	max	s
P100	40	19	35	43.5	45.4	56.25	76	12.8
P101	13	21	34	39.0	41.8	43.00	78	17.8

Z : Notas originales ; Y : Notas con la pond.

$$\begin{aligned} \bullet \text{mean}(Y) &= \text{mean}(1.1 \cdot Z) \\ &= 1.1 \text{mean}(Z) \\ &= 1.1 \cdot 41.8 \\ &= 45.98 \end{aligned}$$

$$\begin{aligned} \bullet \text{Var}(Y) &= \text{Var}(1.1 \cdot Z) \\ &= (1.1)^2 \text{Var}(Z) \end{aligned}$$

$$\begin{aligned} \Rightarrow \text{sd}(Y) &= (1.1) \text{sd}(Z) \\ &= 1.1 \cdot 17.8 \\ &= 19.58 \end{aligned}$$

$$\bullet \text{CV}(Y) = \frac{s_y}{\bar{y}} = \frac{19.58}{45.98} = 0,42$$

4. Luego de las decisiones de los profesores, ¿Cuál paralelo es el que tuvo la mayor variabilidad en las notas del Certamen 1?

P101 pues $\text{CV}(Y) > \text{CV}(X)$

Problema 3

Los empleados de una compañía británica reciben sus salarios mensuales en libras esterlinas (£). Una división de la compañía será reubicada en Francia por un año, donde sus salarios serán pagados en euros (Eu). Asuma que una libra esterlina es igual a 1.27 euros. Mientras los empleados están en Francia, cada uno recibirá también un bono mensual de 325 euros. La siguiente tabla contiene información sobre los salarios mensuales originales en Gran Bretaña.

Mínimo	£ 800
Primer cuartil	£ 1250
Mediana	£ 1470
Tercer cuartil	£ 2250
Máximo	£ 4500
Promedio	£ 2025
Desviación Estándar	£ 475

- Encuentre el promedio y la desviación estándar de los salarios mensuales en euros luego de reubicarse en Francia.
- Basado en los salarios en Gran Bretaña, ¿Existe algún salario atípico en los datos? Justifique su respuesta.
- ¿Existe alguna evidencia de asimetría en los datos? Justifique su respuesta.

1. X_i : Salario en GB. Y_i : Salario en Francia.

$$Y_i = 1.27 \cdot X_i + 325$$

$$\begin{aligned} \Rightarrow \text{mean}(\vec{Y}) &= \text{mean}(1.27\vec{X} + 325 \cdot \vec{1}) \\ &= 1.27 \text{mean}(X) + 325 \\ &= 1.27 \cdot 2025 + 325 \\ &= 2896.75 \end{aligned}$$

$$\begin{aligned} \Rightarrow \text{Var}(\vec{Y}) &= \text{Var}(1.27\vec{X} + 325 \cdot \vec{1}) \\ &= \text{Var}(1.27\vec{X}) \\ &= (1.27)^2 \cdot \text{Var}(\vec{X}) \end{aligned}$$

$$\begin{aligned} \Rightarrow \text{sd}(Y) &= 1.27 \cdot \text{sd}(X) \\ &= 1.27 \cdot 475 \\ &= 603.25 \end{aligned}$$

2. $IQR = Q_3 - Q_1 = 1000$
 $LI = Q_1 - 1.5 \cdot IQR = -250$
 $LS = Q_3 + 1.5 \cdot IQR = 3750$

Mínimo	£ 800
Primer cuartil	£ 1250
Mediana	£ 1470
Tercer cuartil	£ 2250
Máximo	£ 4500
Promedio	£ 2025
Desviación Estándar	£ 475

Resp: Si, es 4500 el atípico pues $\max > LS$

3. Si hacia la derecha pues $Me < \bar{X}$.

Problema 2

Considere el siguiente conjunto de datos que corresponde a la temperatura de un conjunto de metales en Kelvin:

X_i	Temperatura [K]	Frecuencia	F
50	[0 - 100)	50	$\frac{50}{171} \rightarrow 29\%$
150	[100 - 200)	41	$\frac{41}{171} \rightarrow 24\%$
250	[200 - 300)	78	$\frac{78}{171} \rightarrow 46\%$
350	[300 - 400)	2	$\frac{2}{171} \rightarrow 1\%$

- Realice un análisis descriptivo del conjunto de datos, teniendo especial cuidado en el tratamiento y conceptos de datos acumulados. Comuníquese adecuadamente sus resultados.
- ¿Existe alguna evidencia de asimetría en los datos? Justifique su respuesta.
- Construya un box-plot del conjunto de datos. ¿Qué dificultades trae trabajar con datos agrupados? Justifique.

1. Promedio: n : cantidad de datos
 K : cantidad de clases.

$$\begin{aligned} \bar{X} &= \frac{1}{n} \sum_{i=1}^K x_i \cdot f_i \\ n &= \sum_{i=1}^K f_i \end{aligned}$$

$$n = 171 ; K = 4$$

$$\bar{X} = 168.71$$

2. Mediana

$$M_e = l_{M_e} + \frac{\frac{n}{2} - N_{M_e}^-}{n_{M_e}} I_{me}$$

l_{M_e} = límite inferior del intervalo de la clase mediana

n = número total de datos

$N_{M_e}^-$ = frecuencia absoluta acumulada hasta la clase anterior a la clase mediana

n_{M_e} = es la frecuencia de la clase mediana

I = ancho del intervalo de la clase mediana.

Clase mediana

1. 50% de los datos.

$$n/2 = 85.5$$

2. Ordenar F_i

3. Ubicar $n/2$ en algún F_i

$$l_{Me} = 100$$

$$N_{me} = 50$$

$$n_{Me} = 41$$

$$I = 100$$

$$\Rightarrow Me = 186.58$$

Moda:

$$\text{Moda} = l_{m_o} + \frac{\Delta_1}{\Delta_1 + \Delta_2} I_{m_o}$$

donde:

m_o = clase modal (clase con mayor frecuencia absoluta)

l_{m_o} = límite inferior del intervalo de la clase modal $\rightarrow 200$

$\Delta_1 = n_{m_o} - n_{m_o}^-$

$\Delta_2 = n_{m_o} - n_{m_o}^+$

n_{m_o} = es la frecuencia de la clase modal $\rightarrow 78$

$n_{m_o}^+$ = es la frecuencia de la clase posterior a la clase modal $\rightarrow 2$

$n_{m_o}^-$ = es la frecuencia de la clase anterior a la clase modal $\rightarrow 41$

I = ancho del intervalo de la clase modal. $\rightarrow 100$

$$\text{Moda} = 200 + \frac{37}{37+76} \cdot 100$$

$$= 232.74$$

Temperatura [K]	Frecuencia
[0 - 100)	50
[100 - 200)	41
[200 - 300)	78
[300 - 400)	2

$$\Delta_1 = 37$$

$$\Delta_2 = 76$$