

Actividad N°. 1 – Primer Curso de Grado en Ciencia de Datos – Estadística**1. Respuestas a las preguntas sobre las preinscripciones:**

Resuelve los siguientes problemas y sube un archivo con las soluciones al aula virtual.

1. La siguiente tabla presenta las preinscripciones a distintas universidades españolas. Responde a las siguientes cuestiones:
 - a. ¿Es una muestra o una población?
 - b. ¿Cuál es la media de las preinscripciones?
 - c. ¿Cuál es la mediana de las preinscripciones?
 - d. ¿Cuál es el rango de las preinscripciones?
 - e. Calcula la desviación estándar.

Universidad	Preinscripciones
Valencian International University	2.152
Universidad Europea de Valencia	17.298
Universidad de Cantabria	26.666
Universidad Politécnica de Valencia	33.347
Universidad de Alicante	17.529
Universidad Miguel Hernández	27.706
Universidad Jaume I	16.924
Universidad de Valencia	56.387
Universidad Complutense de Madrid	25.223
Universidad Autónoma de Barcelona	21.500
Universidad Oberta de Catalunya	4.630
Universidad de la Laguna	16.762
Universidad de Sevilla	13.813

a. ¿Es una muestra o una población?

La tabla presentada es una muestra. Esto se debe a que no contiene datos de todas las universidades españolas, sino un conjunto específico de universidades seleccionadas para este análisis. En estadística, una muestra es un subconjunto de una población completa y se utiliza para hacer inferencias sobre la población general.

b. ¿Cuál es la media de las preinscripciones?

La media de las preinscripciones es el promedio aritmético de todas las preinscripciones reportadas. Se calcula sumando todas las preinscripciones y dividiendo por el número total de universidades en la muestra. En este caso, la media es de aproximadamente 21.533 preinscripciones.

c. ¿Cuál es la mediana de las preinscripciones?

La mediana es el valor central de un conjunto de datos ordenados de menor a mayor. Si el número de datos es impar, la mediana es el valor del medio. En este caso, la mediana de las preinscripciones es 17.529 preinscripciones, lo que indica que la mitad de las universidades tienen más de 17.529 preinscripciones y la otra mitad tiene menos.

d. ¿Cuál es el rango de las preinscripciones?

El rango es la diferencia entre el valor máximo y el mínimo de la muestra, y proporciona una medida simple de la dispersión de los datos. En este caso, el rango de las preinscripciones es 54.235 preinscripciones, lo que indica la variabilidad entre la universidad con más preinscripciones y la que tiene menos.

e. Calcula la desviación estándar.

La desviación estándar es una medida de la dispersión o variabilidad de un conjunto de datos. Cuanto mayor es la desviación estándar, más dispersos están los datos respecto a la media. En este caso, la desviación estándar de las preinscripciones es aproximadamente 13.636 preinscripciones, lo que refleja la variabilidad de las preinscripciones entre las universidades de la muestra.

2.

2. El director de recursos humanos de la Universidad Europea de Valencia quiere hacer un estudio relacionado con el tiempo que los empleados invierten en llegar al trabajo para estimular a sus empleados a utilizar transporte público. Para ello hace una encuesta entre los empleados en los que recoge los siguientes datos:

- Transporte público (tiempo requerido): 23, 25, 25, 30, 31, 31, 32, 33, 35, 36, 37, 42.
- Coche particular (tiempo requerido): 32, 32, 33, 34, 37, 37, 38, 38, 38, 39, 40, 44.
 - a. Calcula la mediana, el primer y tercer cuartil del tiempo de desplazamiento de los empleados que utilizaron el transporte público.
 - b. Ídem para lo que utilizaron el coche particular.
 - c. Compara los tiempos de ambos grupos.

a. Mediana, primer y tercer cuartil del tiempo de desplazamiento para los empleados que utilizaron el transporte público:

- Mediana: 31.5 minutos
- Primer cuartil: 28.75 minutos
- Tercer cuartil: 35.25 minutos

b. Mediana, primer y tercer cuartil del tiempo de desplazamiento para los empleados que utilizaron el coche particular:

- Mediana: 37.5 minutos
- Primer cuartil: 33.75 minutos
- Tercer cuartil: 38.25 minutos

c. Comparación de los tiempos de ambos grupos:

Los empleados que utilizaron el transporte público tuvieron una mediana de tiempo de desplazamiento de 31.5 minutos, que es menor que la mediana de 37.5 minutos de los que usaron coche particular. Los tiempos de desplazamiento en transporte público muestran una menor dispersión entre el primer y tercer cuartil en comparación con los del coche particular, lo que sugiere que el tiempo de desplazamiento en coche es más consistente y tiene una menor variación.

3.

3. Un avión recorre las distancias d_1 , d_2 y d_3 a las velocidades v_1 , v_2 y v_3 (en km/h) respectivamente.

a. Demuestra que la velocidad promedio está dada por la siguiente expresión:

$$\frac{d_1 + d_2 + d_3}{V} = \frac{d_1}{v_1} + \frac{d_2}{v_2} + \frac{d_3}{v_3}$$

- b. Lo que acabamos de hallar es una media armónica ponderada. Encuentra cuánto vale esa media (V) teniendo en cuenta que:

$$d_1 = 2.500 \text{ km}$$

$$d_2 = 2.500 \text{ km}$$

$$d_3 = 500 \text{ km}$$

$$v_1 = 500 \text{ km/h}$$

$$v_2 = 400 \text{ km/h}$$

$$v_3 = 250 \text{ km/h}$$

a. Demostración de la fórmula de la velocidad promedio

La fórmula de la velocidad promedio para un conjunto de distancias d_1 , d_2 , d_3 recorridas a diferentes velocidades v_1 , v_2 , v_3 está dada por:

$$V = (d_1 + d_2 + d_3) / (d_1/v_1 + d_2/v_2 + d_3/v_3)$$

Demostración:

El tiempo total T_{total} que el avión tarda en recorrer todas las distancias es la suma de los tiempos individuales:

$$T_{\text{total}} = d_1/v_1 + d_2/v_2 + d_3/v_3$$

La velocidad promedio se define como la distancia total dividida por el tiempo total:

$$V = (d_1 + d_2 + d_3) / T_{\text{total}} = (d_1 + d_2 + d_3) / (d_1/v_1 + d_2/v_2 + d_3/v_3)$$

b. Cálculo de la velocidad promedio

- $d_1 = 2,500$ km, $d_2 = 2,500$ km, $d_3 = 500$ km

- $v_1 = 500$ km/h, $v_2 = 400$ km/h, $v_3 = 250$ km/h

Sustituimos estos valores en la fórmula para obtener la velocidad promedio:

$$V = (2,500 + 2,500 + 500) / (2,500/500 + 2,500/400 + 500/250)$$

Velocidad promedio = 415.09 km/h

4.

4. Proporciona seis términos estadísticos diferentes para el punto de equilibrio o valor central en una curva de frecuencias en forma de campana simétrica.

Supongamos que tenemos el siguiente conjunto de datos ordenados:

3, 5, 7, 9, 11, 13, 15, 17, 19

1. Media aritmética:

La media aritmética se calcula sumando todos los valores y dividiendo entre el número total de datos:

$$\text{Media} = (3 + 5 + 7 + 9 + 11 + 13 + 15 + 17 + 19) / 9 = 99 / 9 = 11$$

2. Mediana:

La mediana es el valor en el medio del conjunto de datos cuando están ordenados. Dado que hay 9 datos (un número impar), el valor en la posición 5 es la mediana:

$$\text{Mediana} = 11$$

3. Moda:

En este conjunto de datos, todos los valores aparecen una sola vez, por lo que no hay una moda específica.

4. Segundo cuartil:

El segundo cuartil coincide con la mediana:

$$Q2 = 11$$

5. Percentil 50:

El percentil 50 es también igual a la mediana:

$$P50 = 11$$

6. Decil 5:

El decil 5 corresponde al 50% de los datos, que es el valor de la mediana:

$$D5 = 11$$

5.

5. Imagina que trabajas en una multinacional de dispositivos electrónicos. El departamento de ventas ha observado que estas aumentan cuando se invierte más dinero en mejorar el *software* del producto. En base a esto, la empresa decide encargar un estudio estadístico y recoge la siguiente información:

X	Y
0,2	0,5
0,5	0,7
1	1,8
1,2	2,8
2	3,2
2,7	3,3
3	3,3
3,4	3,3
3,6	3,4

Donde:

- X = inversión en mejorar el *software* (millones de €).
- Y = venta de dispositivos electrónicos (millones de €).

Se pide calcular los siguientes estadísticos descriptivos de la variable "ventas de dispositivos electrónicos": media, mediana, varianza, desviación estándar y coeficiente de asimetría de Fisher.

Resultados y explicación de los estadísticos descriptivos:

A continuación, se presentan los cálculos de los estadísticos descriptivos para la variable 'ventas de dispositivos electrónicos':

1. Media:

La media es el promedio de todas las observaciones y proporciona una medida del valor central de los datos.

Media = 2.56 millones de euros

2. Mediana:

La mediana es el valor que divide el conjunto de datos en dos partes iguales. En este caso, la mediana de las ventas es:

Mediana = 3.25 millones de euros

3. Varianza:

La varianza mide la dispersión de los datos respecto a la media. Una varianza alta indica una mayor dispersión.

Varianza = 1.29 millones de euros

4. Desviación estándar:

La desviación estándar es la raíz cuadrada de la varianza y proporciona una medida de la dispersión en las mismas unidades que los datos.

Desviación estándar = 1.14 millones de euros

5. Coeficiente de asimetría de Fisher:

El coeficiente de asimetría de Fisher mide la simetría de la distribución de los datos. Un valor negativo indica asimetría hacia la izquierda (cola más larga a la izquierda).

Coeficiente de asimetría de Fisher = - 1.24