

FORMULARIO DE ESTADÍSTICA

ESTADÍSTICA UNIDIMENSIONAL

Medidas de tendencia central:

Media:

$$\bar{x} = \frac{\sum f_i x_i}{n}$$

Mediana:

Se ordenan los datos. Si hay un número par de datos la mediana es la media de los dos datos centrales; si el número de datos es impar, la mediana es justamente el dato central.

Moda:

Valor más frecuente de la variable.

Interpretación: análisis de los datos

Supongamos que estamos estudiando el número de vuelos semanales que realizan 10 pilotos. Los datos obtenidos son los siguientes:

Nº de vuelos	0	1	2	3
Frecuencia absoluta	2	4	3	1

La media es 1,3, y nos indica, que por término medio, el número de vuelos es de 1,3, es decir, que *por término medio estos pilotos vuelan entre 1 y 2 veces por semana*.

La moda es 1, lo que nos indica que lo más frecuente es que vuelen 2 veces por semana.

Y por último, la mediana es 1, lo que nos dice que *hay tantos pilotos que vuelan 1 o más veces, como pilotos que lo hacen 1 vez o menos*.

Medidas de posición:

Son valores de la variable que nos informan del lugar que ocupa un dato dentro del conjunto ordenado de valores.

Cuartiles: (Q_1 , Q_2 y Q_3)

Son medidas que dividen el conjunto de datos ordenados en cuatro partes iguales, es decir, en cada una de esas partes está el 25 % de los datos recogidos en el estudio.

Percentiles: (P_1, \dots, P_{99})

Son medidas que dividen el conjunto de datos ordenados en cien partes iguales.

Para su **cálculo**, se calcula el tanto por ciento correspondiente del número total de datos, y dicho valor se mira en la columna de la frecuencias absolutas acumuladas. Si no aparece (que será lo más normal), se toma el inmediato superior, y se ve que valor de la variable le corresponde.

Interpretación: análisis de los datos

Para comprar zapatillas a los miembros de una peña de bolos, se les he preguntado por la talla de calzado que usan y los resultados son los siguientes:

Nº de calzado	35	36	37	38	39	40	41	42
Frecuencia absoluta	7	13	20	37	42	50	23	8

El primer cuartil es $Q_1 = 38$ y lo que nos dice es que el 25 % de los miembros de la peña utilizan una talla de calzado menor o igual que 38.

El segundo cuartil es $Q_2 = 39$ (que coincide con la mediana) y lo que nos dice es que el 50 % de miembros usa una talla de calzado menor o igual que 39 y el otro 50 % mayor o igual.

El tercer cuartil es $Q_3 = 40$ que nos dice que el 75 % de los miembros del club de bolos usa una talla de calzado menor o igual que 40.

Medidas de dispersión:

Permiten conocer el grado de agrupamiento de los datos en torno a las medidas de centralización.

Rango

Es la diferencia entre el mayor valor de la variable y el menor valor de la variable.

Varianza:

$$\sigma^2 = \frac{\sum f_i (x_i - \bar{x})^2}{n} \longrightarrow \sigma^2 = \frac{\sum f_i x_i^2}{n} - \bar{x}^2$$

Desviación típica:

$$\sigma = +\sqrt{\sigma^2}$$

Interpretación: análisis de los datos

Supongamos que estamos estudiando el número de aciertos de 100 alumnos en una prueba de 30 preguntas. Los resultados obtenidos se recogen en la siguiente tabla:

Aciertos	x_i	f_i
[0,5)	2,5	3
[5,10)	7,5	10
[10,15)	12,5	25
[15,20)	17,5	38
[20,25)	22,5	16
[25,30]	27,5	8
Total		100

En este caso el rango es 30, y por tanto, no nos proporciona ninguna información.

La varianza es $\sigma^2 = 33,79$ y la desviación típica es $\sigma = 5,81$, que son relativamente grandes, lo que nos dice que los datos presentan una agrupación relativamente pequeña respecto de la media.