

Recordatorio de Estadística

1. Tabla de frecuencia.
2. Medidas de tendencia central.
3. Medidas de dispersión.
4. Medidas de posición.

Tabla de frecuencias.

Tabla de frecuencias

La distribución de frecuencias o tabla de frecuencias es una ordenación en forma de tabla de los datos estadísticos, asignando a cada dato su frecuencia correspondiente.

Existen 4 tipo de frecuencias:

1. **Frecuencia absoluta (f_i):** es el número de veces que aparece un determinado valor en un estudio estadístico.
1. **Frecuencia relativa (n_i):** es el cociente entre la frecuencia absoluta de un determinado valor y el número total de datos. A veces se expresa en porcentaje.
1. **Frecuencia absoluta acumulada (F_i):** es la suma de las frecuencias absolutas de todos los valores inferiores o iguales al valor considerado. En este caso, lo normal es ordenar los valores de menor a mayor.
1. **Frecuencia relativa acumulada (N_i):** es la suma de las frecuencias relativas de todos los valores inferiores o iguales al valor considerado. A veces se expresa en porcentaje.

Consideremos un ejemplo: Estudiantes contestan a la pregunta de cuántos hermanos tienen.

Las respuestas son: 1,1,2,0,3,2,1,4,2,3,1,0,0,1,2

Cantidad hermanos	f_i	n_i	F_i	N_i
0	3	3/15	3	3/15
1	5	5/15	8	8/15
2	4	4/15	12	12/15
3	2	2/15	14	14/15
4	1	1/15	15	15/15
TOTAL	15	1		

Medidas de
tendencia central.

Medidas de tendencia central

Las medidas de tendencia central son números que muestran (o intentan) donde está el centro de los datos.

Las más conocidas son: la media (o promedio), la mediana y la moda.

- » Suponiendo que los valores son representados por x_i y n es el total de datos, la mediana se calcula como:

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{1}{n} \sum_{i=1}^n x_i$$

- » La moda es el valor que más se repite entre todos los datos disponibles.
- » La mediana dependerá si el número total de valores es par o impar.

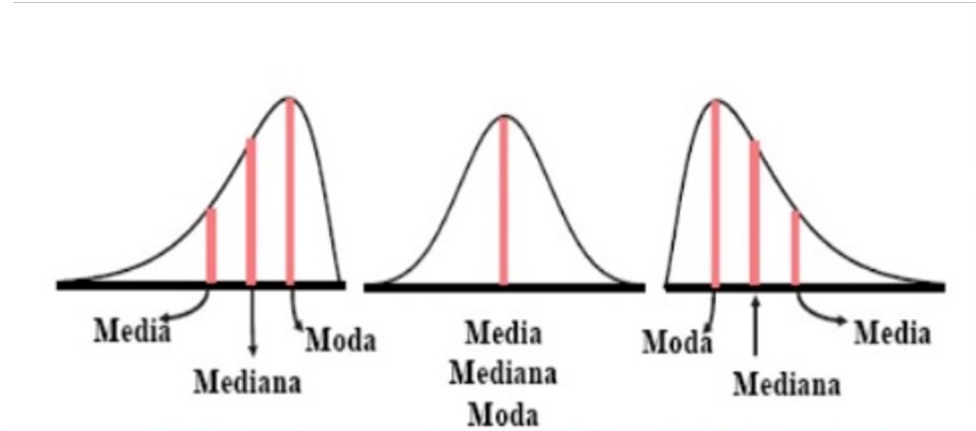
IMPAR

1, 1, 1, 1, 1, 1, **2**, 2, 2, 2, 3, 3, 4
Mitad inferior Mediana Mitad superior

PAR

1, 1, 1, 1, 1, **1, 2**, 2, 2, 3, 3, 4
Valores inferiores Valores intermedios Valores superiores

Gráficamente, podríamos ver cantidades así:



¿Qué pasa si hay valores muy alejados del centro de los datos?

Medidas de dispersión.

Las medidas de dispersión tratan, a través del cálculo de diferentes fórmulas, de entregar un valor numérico que ofrezca información sobre el grado de variabilidad de una variable.

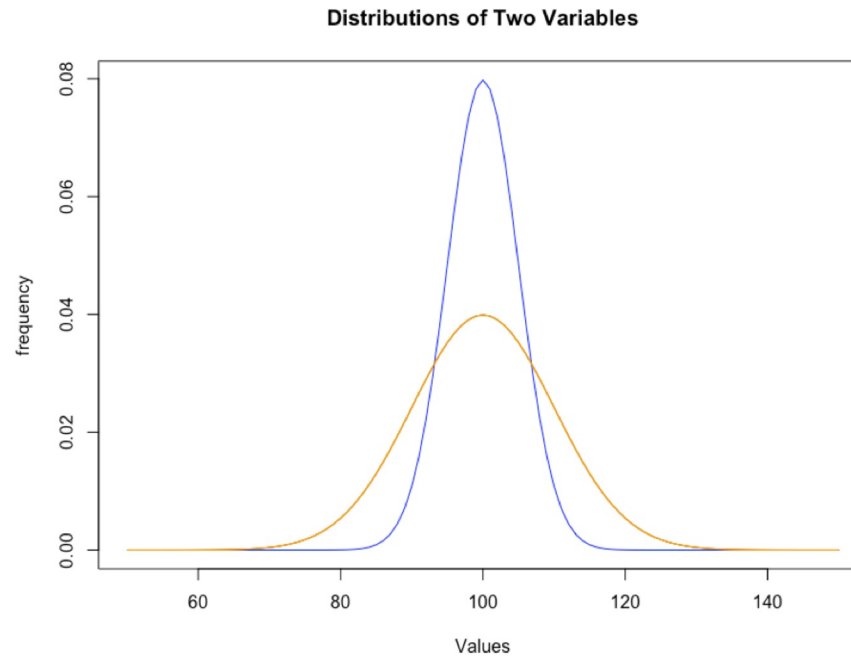
MEDIDAS DE DISPERSIÓN

VARIANZA	DESVIACIÓN ESTÁNDAR
$\sigma^2 = \frac{\sum_1^N (x_i - \bar{X})^2}{N}$	$\sigma = \sqrt{\frac{\sum_1^N (x_i - \bar{X})^2}{N}}$

- **X** → Variable sobre la que se pretenden calcular la varianza.
- **x_i** → Observación número i de la variable X. i puede tomará valores entre 1 y n.
- **N** → Número de observaciones.
- **\bar{X}** → Es la media de la variable X.

RANGO ESTADÍSTICO	COEFICIENTE DE VARIACIÓN
$R = Máx_x - Mín_x$	$CV = \frac{\sigma_x}{ \bar{X} }$
<ul style="list-style-type: none"> • R → Es el rango. • Máx → Es el valor máximo de la muestra o población. • Mín → Es el valor mínimo de la muestra o población estadística. • x → Es la variable sobre la que se pretende calcular esta medida. 	<ul style="list-style-type: none"> • X → Variable sobre la que se pretenden calcular la varianza. • σ_x → Desviación típica de la variable X. • \bar{x} → Es la media de la variable X en valor absoluto con $\bar{x} \neq 0$.

¿Cuál curva tiene mayor y menor varianza?



¿Qué podemos decir de la mediana y la moda?

Medidas de posición.

Las medidas de posición son indicadores estadísticos que permiten resumir los datos en uno solo, o dividir su distribución en intervalos del mismo tamaño, por lo tanto, sirven para medir y para dividir.

Los más comunes son:

El cuartil (Q_i): Es uno de los más utilizados y divide la distribución en cuatro partes iguales. Así, existen tres cuartiles. Los valores inferiores de la distribución se sitúan por debajo del primero (Q_1). La mitad o mediana son los valores menores iguales al cuartil dos (Q_2) y los superiores son representados por el cuartil tres (Q_3).

El decil (D_i): Estamos ante un cuartil que divide los datos en diez partes iguales. Existen nueve deciles, de D_1 a D_9 . El D_5 se corresponde con la mediana. Por su lado, los valores superiores e inferiores (equivalentes a los diferentes cuartiles) se sitúan en puntos intermedios entre estos.

El percentil (P_i): Por último, este cuartil divide la distribución en cien partes. Hay 99 percentiles. Tiene, a su vez, una equivalencia con los deciles y cuartiles.

Gráficamente podemos ver estas medidas así:

