FUENTES ESTADÍSTICAS PARA SABER HACER UN ANÁLISIS CRÍTICO

La estadística como herramienta para analizar críticamente la información en la vida real

4ª SESIÓN

Francisco José Alegre Ansuategui

El abuso de la media aritmética

- Es innegable que, no solo en los medios de comunicación, sino en cualquier organismo, se tiende a hacer un uso abusivo de la media aritmética o el promedio para describir una situación.
- La mediana supone un indicador, en muchas ocasiones, mucho más fiable que la media de la centralización de la misma.
- Podrían complementarse perfectamente y su aporte no supondría ningún esfuerzo ni a nivel organizativo ni de gestión de recursos, no obstante, casi siempre es la media la protagonista.
- Incluso la moda tiene un papel más protagonista que la mediana.

¿Qué ocurre si nos quedamos únicamente con parámetros de centralización (media, mediana, moda..)?

- Perdemos visión global del análisis de la situación. No tenemos toda la información necesaria para realizar un análisis crítico.
- Podemos caer en el error de sacar conclusiones equivocadas por una precipitación a la hora de sacarlas.
- Los parámetros de centralización nos indican en torno a qué valor (centro) se distribuyen los datos, pero no los dispersos que están.
- Necesitamos otros elementos o parámetros estadísticos que nos permitan ampliar nuestra visión de la realidad.

Retomamos el ejemplo anterior:

- NOTAS DE LOS ALUMNOS
- 2, 2, 3, 3, 4, 7, 7, 7, 10

- Resultados:
- Media = 5
- Mediana = 4
- Moda = 7

- Lo que la gente saca de media es un 5 … pero no hay ni un 5 ni un 6
- El alumno en el centro de la distribución saca un 4
- La nota que más se repite es un 7

TEMA 3- LA NECESIDAD DE INCLUIR PARÁMETROS DE DISPERSÍÓN

DEBATE PREVIO:

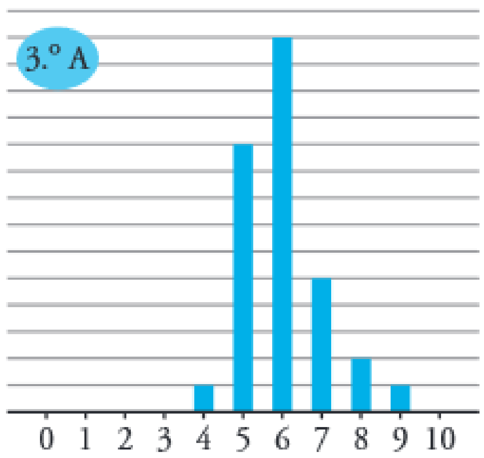
- ¿Qué parámetros conocéis?
- ¿Creéis que su uso es muy frecuente?

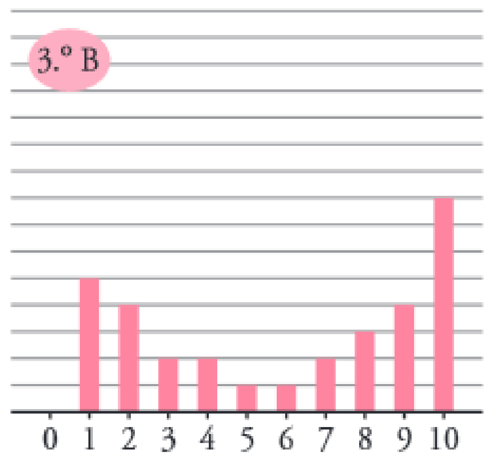
La media no es suficiente

Las gráficas de la derecha corresponden a las notas de dos clases. En ambas, la nota media es, aproximadamente, 6. La **media** es un parámetro que nos informa sobre el centro alrededor del cual se distribuyen los valores. Pero observa que, aun teniendo la misma media, estas distribuciones son muy distintas. Necesitamos otros parámetros que señalen esas diferencias.

Los parámetros de centralización nos indican en torno a qué valor (centro) se distribuyen los datos.

Los parámetros de dispersión nos informan sobre cuánto se alejan del centro los valores de la distribución.





Medidas de dispersión

Vamos a estudiar ahora parámetros que sirven para medir cómo de dispersos están los datos. En todos ellos, la idea clave es medir el grado de separación de los datos a la media.

Recorrido o rango

Es la diferencia entre el dato mayor y el menor. Es decir, es la longitud del tramo dentro del cual están los datos.

Desviación media

Es el promedio de las distancias de los datos a la media:

$$\mathrm{DM} = \frac{|x_1 - \bar{x}| + |x_2 - \bar{x}| + \ldots + |x_n - \bar{x}|}{n} = \frac{\Sigma |x_i - \bar{x}|}{n}$$

Varianza

Es el promedio de los cuadrados de las distancias de los datos a la media:

$$\text{Varianza} = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \ldots + (x_n - \bar{x})^2}{n} = \frac{\Sigma (x_i - \bar{x})^2}{n}$$

Esta fórmula es equivalente a la siguiente:

$$\text{Varianza} = \frac{x_1{}^2 + x_2{}^2 + \ldots + x_n{}^2}{n} - \bar{x}^2 = \frac{\Sigma x_i{}^2}{n} - \bar{x}^2$$

Desviación típica, σ

Es la raíz cuadrada de la varianza: $\sigma = \sqrt{\text{varianza}}$

A partir de ahora prestaremos especial atención a los parámetros media ($\bar{x}$) y desviación típica (σ). La información que da cada uno de ellos complementa a la del otro.

Ejercicio resuelto

Obtener las medidas de dispersión de la siguiente distribución de notas:

2, 4, 4, 4, 5, 7, 9, 9, 10

RECORRIDO: $10 - 2 = 8$ MEDIA: $\bar{x} = 6$

DESVIACIÓN MEDIA: $$\text{DM} = \frac{|2-6| + |4-6| + |4-6| + \ldots}{9} = \frac{22}{9} = 2{,}44$$

VARIANZA: $$\text{Var} = \frac{(2-6)^2 + (4-6)^2 + (4-6)^2 + \ldots}{9} = \frac{64}{9} = 7{,}11$$

o bien: $$\text{Var} = \frac{2^2 + 4^2 + 4^2 + 4^2 + 5^2 + \ldots}{9} - 6^2 = \frac{388}{9} - 36 = 7{,}11$$

DESVIACIÓN TÍPICA: $\sigma = \sqrt{\text{varianza}} = \sqrt{7{,}11} = 2{,}67$

Ejercicio propuesto:

- Calcula la varianza, la desviación típica y el recorrido de la siguiente distribución:

9, 3, 8, 8, 9, 8, 9, 18

$$\bar{x} = \frac{9+3+8+8+9+8+9+18}{8} = 9$$

$$\sigma^2 = \frac{(9-9)^2+(3-9)^2+(8-9)^2+(8-9)^2+(9-9)^2+(8-9)^2+(9-9)^2+(18-9)^2}{8} = 15$$

¿Qué unidades tienen estos parámetros?

- El recorrido tiene unidades (por ejemplo, euros)
- La varianza tiene unidades elevadas al cuadrado (por ejemplo, euros elevados al cuadrado)
- La desviación típica tiene unidades (por ejemplo euros)

¿Por qué la desviación típica?

La varianza tiene un grave inconveniente. Imagina que estamos tratando con una distribución de estaturas dadas en cm. La media vendría dada en cm, pero la varianza vendría en cm^2 (es decir, una superficie en lugar de una longitud). Por eso, extraemos su raíz cuadrada, obteniendo la desviación típica que, en nuestro ejemplo, sí sería una longitud dada en cm.

EL coeficiente de variación

Coeficiente de variación

En estadística, cuando se desea hacer referencia a la relación entre el tamaño de la media y la variabilidad de la variable, se utiliza el **coeficiente de variación**.

Su fórmula expresa la desviación estándar como porcentaje de la media aritmética, mostrando una mejor interpretación porcentual del grado de variabilidad que la desviación típica o estándar. Por otro lado presenta problemas ya que a diferencia de la desviación típica este coeficiente es variable ante cambios de origen. Por ello es importante que todos los valores sean positivos y su media dé, por tanto, un valor positivo. A mayor valor del coeficiente de variación mayor heterogeneidad de los valores de la variable; y a menor C.V., mayor homogeneidad en los valores de la variable. Suele representarse por medio de las siglas **C.V.**

Se calcula:

$$C_V = \frac{\sigma}{|\bar{x}|}$$

Se calcula:

$$C_V = \frac{\sigma}{|\bar{x}|}$$

Donde σ es la desviación típica, y $\bar{x}$ es la Media. Se puede dar en porcentaje calculando:

$$C_V = \frac{\sigma}{|\bar{x}|} \cdot 100$$

Propiedades y aplicaciones [editar]

- El coeficiente de variación no posee unidades.
- El coeficiente de variación es típicamente menor que uno. Sin embargo, en ciertas distribuciones de probabilidad puede ser 1 o mayor que 1.
- Para su mejor interpretación se expresa como porcentaje.
- Depende de la desviación típica, también llamada "desviación estándar", y en mayor medida de la media aritmética, dado que cuando ésta es 0 o muy próxima a este valor el C.V. pierde significado, ya que puede dar valores muy grandes, que no necesariamente implican dispersión de datos.

PROBLEMA 1 - ¿A qué empresa le encargamos las piezas?

Una empresa aeronáutica necesita una pieza fabricada en una aleación muy especial para uno de sus proyectos. Sólo dos empresas fabrican este tipo de piezas. Puesto que el precio que fijan ambas empresas por sus productos es el mismo, se realiza un estudio estadístico de estos objetos muy detallado. Así pues, se realizan impactos sobre 150 objetos de cada empresa para determinar el número de impactos que necesita hasta romperse.

Estos fueron los resultados obtenidos:

Empresa A

IMPACTOS	1	2	3	4	5	6	7	8	9	10
FRECUENCIA	15	20	35	40	11	10	7	5	4	3

Empresa B

IMPACTOS	1	2	3	4	5	6	7	8	9	10
FRECUENCIA	13	12	42	35	15	11	10	6	5	1

Ejemplos de uso del coeficiente de variación en la vida real

PROBLEMA 2 - ¿A qué jugadora fichamos?

Dos equipos se debaten entre fichar a una jugadora u otra. Para ello, realizan un análisis estadístico de los puntos anotados por ambas jugadoras en los 15 partidos de competición. Los puntos anotados para estas dos jugadoras (A y B) en cada partido fueron los siguientes:

A	13	12	17	18	11	15	15	17	18	24	15	12	11	10	12
B	23	8	15	17	13	21	12	16	13	11	19	8	6	12	11

EJERCICIO PROPUESTO

- Obtén el coeficiente de variación para cada uno de los problemas anteriores, indica las conclusiones más destacables que pueden obtenerse a partir de ellos.