1. Medidas de centralización

Medidas estadísticas

Medidas de centralización

Hemos visto cómo el estudio del conjunto de los datos mediante la estadística permite realizar representaciones gráficas, que informan sobre ese conjunto. Además de los gráficos es conveniente resumir dichos datos en un solo número, que nos describan de una manera sencilla el comportamiento y las características de los datos estudiados.

Esos números que resumen los datos se llaman **medidas de centralización**. Hay varios, nosotros vamos a estudiar tres: la media aritmética, la moda y la mediana

Usaremos como ejemplos los cuatro problemas tipo utilizados en el tema anterior.

Vamos a profundizar lo comentado en el vídeo

2. La Moda

Medidas estadísticas

La Moda

Se llama moda al valor de la variable de mayor frecuencia

La moda se puede calcular de todos los tipos de variables aleatorias Si hay más de un valor con mayor frecuencia hay más de una moda

Ejemplo 1
De cien personas entrevistadas en la calle 80 llevan falda; 13 pantalones largos y 7 pantalones cortos. ¿Cuál es la moda?
La moda seria llevar falda

Ejemplo 2
Si las notas de un alumno son 3, 4, 4, 4, 6, 8, 8, 8,10 la nota 4 tiene de frecuencia absoluta 3 y la nota 8 tiene de frecuencia 8 por tanto hay dos modas 4 y 8

Calculemos la moda en los ejemplos tipos:

Ejemplo 1:

x_i	n_i
O	12
A	7
B	4
AB	2
	25

La moda es el grupo sanguíneo O

Ejemplo 2

x_i	n_i
1	4
2	7
3	8
4	4
5	2
	25

La moda es 3 paquetes de folios

Autoevaluación

3. La mediana

Medidas estadísticas

La mediana

La mediana de una colección de datos ordenados de menor a mayor es el valor que está en medio, es decir que la mitad de los datos son mayores que él y la otra mitad son menores que él, si hay un número impar de datos; si el número de datos es par, la mediana es la media aritmética entre los dos valores centrales.

Por lo tanto, la mediana sólo se puede calcular con variables estadísticas de tipo cuantitativo, que se pueden ordenar

Ejemplo 1: dados los datos: 3, 4, 2, 5, 3, 6, 3, 6, 5,
son nueve datos; los ordenamos de menor a mayor:
2, 3, 3, 3 ,4, 5, 5, 6, 6;
el dato del medio sería $\frac{9}{2} = 4,5$ el que ocupa el lugar 5º y la mediana es por tanto su valor 4
Si los datos son pares por ejemplo los 10 números siguientes 11, 24, 19, 7, 23, 14, 15, 20, 20, 11 los ordenamos de menor a mayor 7, 11, 11, 14, 19, 20, 20, 23, 24 la mitad de los datos $\frac{10}{2} = 5$ esta entre el 5º y el 6º el valor del que ocupa el lugar 5º es 15 y el valor del que ocupa el valor 6º es 19 por tanto la mediana será la media entre 15 y 19 es decir: $\frac{15+19}{2} = \frac{34}{2} = 17$

Autoevaluación

4. La media aritmética

Medidas estadísticas

La media aritmética

Se aplica solamente a variables estadísticas del tipo cuantitativo

La media aritmética o simplemente media la representaremos por una equis con una barra encima: $\bar{x}$

Se corresponde con la idea de repartir todo lo que hay en partes iguales para todos.

Por ejemplo, si cuatro amigos se reunen y cuentan el dinero que llevan, podríamos encontrar: uno tiene 50 €, el segundo lleva 40 €, el tercero tiene 55 € y el cuarto 55 €; por tanto, si sumamos todo el dinero que llevan, juntaríamos 200€; y si tuviesen que repartirlo en parte iguales, tocarían a 200/4=50 € cada uno; ésta es la idea de la media

Por tanto definiremos:

Dados *n* valores $x_1, x_2, x_3, \ldots x_n$ la media será la suma de los valores: $x_1 + x_2 + x_3 + \ldots + x_n$ y dividida entre el número n de valores:

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \ldots + x_n}{n}$$

Abreviadamente se escribe como $\bar{x} = \frac{\sum x_i}{n}$ donde $\sum$ es el signo que se usa en matemáticas para indicar suma cuando se trata de muchos números:

$$\sum x_i = x_1 + x_2 + x_3 + \ldots + x_n$$

Ejemplos
Un alumno saca las siguientes notas en matemáticas: 5, 4, 6, 7, 4, 8, 2, 5, 6, 6,
¿Qué nota le corresponde?
Calculamos la media $\bar{x} = \frac{5+4+6+7+4+8+2+5+6+6}{10} = \frac{53}{10} = 5{,}3$

Propiedades de la media

Ejemplo: Supongamos que las notas de un alumno son 5, 6, 6 y 7 la media será: $\frac{5+6+6+7}{4} = \frac{24}{4} = 6$

<4>si a cada nota le sumamos 3 puntos la notas serían 8, 9, 9, 10 y la media sería $\frac{8+9+9+10}{4} = \frac{36}{4} = 9$

que es la media anterior 6 más 3 que hemos aumentado a cada nota individual. Por tanto podemos decir

Si a todos los valores se les suma o resta una misma cantidad, la media queda sumada o restada en dicha cantidad

Ejemplo: Supongamos que los sueldos de cuatro empleados en euros son 1000, 1500, 1800 y 2000 . La media será $\frac{1000+1500+1800+2000}{4}=\frac{6300}{4}=1575$ €

Para no utilizar números tan altos, podíamos dividir cada cantidad por mil y decir que los sueldos son 1;1,5; 1,8; 2 en miles de euros y la media será $\frac{1+1,5+1,8+2}{4}=\frac{6,3}{4}=1,575$ € que es la misma media que antes dividida por 1000. Por tanto podemos decir:

Si a todos los números se les multiplica o divide por un mismo número la media queda multiplicada o dividida por dicho numero

Calculo de la media con datos agrupados

Cuando los datos están agrupados la formula de la media se calcula mediante la expresión $\bar{x}=\frac{\sum x_i \cdot n_i}{n}$

Usaremos los ejemplos 2, 3 y 4 del tema anterior. Copiaremos los datos que necesitamos de la tabla de frecuencia

Ejemplo 2

x_i	n_i	$x_i \cdot n_i$
1	4	4
2	7	14
3	8	24
4	4	16
5	2	10
	25	68

$\bar{x}=\frac{68}{25}=2,72$ paquetes de folios

Ejemplo 3

Intervalos	x_i	n_i	$x_i \cdot n_i$
(18 - 22]	20	2	40
(22 - 26]	24	14	336
(26 - 30]	28	12	336
(30 - 34]	32	12	384
(34 - 38]	36	14	504
(38 - 42]	40	6	240
		60	1840

$\bar{x}=\frac{1840}{60}=30,67$ años

Ejemplo 4

Intervalos	x_i	n_i	$x_i \cdot n_i$
(4 - 5]	4,5	2	9
(5 - 6]	5,5	8	44
(6 - 7]	6,5	10	65
(7 - 8]	7,5	5	37,5
(8 - 9]	8,5	5	42,5
		30	198

$$\bar{x} = \frac{198}{30} = 6,6 \text{ €}$$

Autoevaluación

5. Medidas de dispersión

Medidas estadísticas

Medidas de dispersión

Al grado en que los datos tienden a extenderse alrededor de la media se le llama variación o dispersión de los datos.

Evidentemente sólo se pueden calcular en variables estadísticas cuantitativas

Vamos a profundizar en lo visto en el vídeo

6. Rango o recorrido

Medidas estadísticas

Rango o recorrido

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

Ejemplo

Los siguientes datos son los tiempos de duración en segundos de 50 conversaciones

125, 65, 80, 97, 325, 400, 98, 74, 90, 120, 240, 85, 370, 135, 78, 326, 282, 145, 192, 64, 108, 324, 207, 183, 94, 62, 315, 217, 192, 106, 78, 89, 207, 70, 69, 402, 68, 108, 361, 304, 273, 181, 91, 107, 404, 315, 125, 106, 176, 207

Cual es el rango

El valor mayor es 404 y el menor es 62; por tanto el rango es 404 - 62 = 342

7. Desviación media

Medidas estadísticas

Desviación media

Desviación media de una serie de valores es la media aritmética de las diferencias en valor absoluto de cada uno de los valores y la media.

Desviación media = D.M. = $\frac{\sum |x_i - \bar{x}|}{N}$

Ejemplo

Si los datos son 2; 4; 5, 5, La media será $\bar{x} = \frac{2+4+5+5}{4} = \frac{16}{4} = 4$

Y la D.M. = $\frac{|2-4|+|4-4|+|5-4|+|5-4|}{4} = \frac{2+0+1+1}{4} = \frac{4}{4} = 1$

La desviación media es una indicación de cómo están agrupados los datos: si estuviesen muy cercanos unos a otros, situados de forma muy sucesiva, la desviación media sería pequeña; si los datos estuviesen muy lejos unos de otros, o muy desagrupados, por ejemplo formando dos o tres grupos de datos separados entre sí, la desviación media sería grande.

Veamos un ejemplo sencillo para aclarar cuál es el sentido de la desviación media:

Tenemos estos datos: 50, 50

- la media es (50 + 50)/ 2 = 50;
- la desviación media será [| 50 - 50 | + | 50 - 50 |] / 2= 0

Si ahora consideramos los datos: 25, 75

- la media es (25 + 75) / 2 = 50; es la misma media que con los datos anteriores, pero...
- la desviación media será [| 50 - 25 | +| 75 - 25 |] / 2 = 37,5

Es decir, con la misma media, la desviación media es muy diferente, porque los datos son muy distintos, están situados de forma muy distinta.

Cálculo con datos agrupados

Ejemplo 2

Retomamos un ejercicio anterior, el de los paquetes de folios. La media calculada antes era: $\bar{x}$ = 2,72

x_i	n_i	$\lvert x_i - \bar{x} \rvert$	$\lvert x_i - \bar{x} \rvert \cdot n_i$
1	4	1,72	6,88
2	7	0,72	5,04
3	8	0,28	2,24
4	4	1,28	5,12

5	2	2,285	4,56
	25		23,84

D.M. = $\frac{23,84}{25} = 0,95$ paquetes de folios

Ejemplo3

La media calculada antes era $\bar{x} =$ 30,67

Intervalos	**x_i**	**n_i**	$\lvert x_i - \bar{x}\rvert$	$\lvert x_i - \bar{x}\rvert \cdot n_i$
(18 - 22]	20	2	10,67	21,33
(22 - 26]	24	14	6,67	93,33
(26 - 30]	28	12	2,67	32,00
(30 - 34]	32	12	1,33	16,00
(34 - 38]	36	14	5,33	74,67
(38 - 42]	40	6	9,33	56,00
		60		293,33

D.M. = $\frac{293,33}{60} = 4,89$ años

Ejemplo 4

La media calculada antes era $\bar{x} =$ 6,6

Intervalos	**x_i**	**n_i**	$\lvert x_i - \bar{x}\rvert$	$\lvert x_i - \bar{x}\rvert \cdot n_i$
(4 - 5]	4,5	2	2,1	4,2
(5 - 6]	5,5	8	1,1	8,8
(6 - 7]	6,5	10	0,1	1
(7 - 8]	7,5	5	0,9	4,5
(8 - 9]	8,5	5	1,9	9,5
		30		28

D.M. = $\frac{28}{30} = 0,93$ €

8. Varianza y Desviación típica

Medidas estadísticas

Varianza y Desviación típica

Varianza de una serie es la media aritmética de las diferencias al cuadrado de cada uno de los valores y la media; es decir, una vez hallada la media, se halla las diferencias entre cada valor individual y la media, esas medidas se elevan al cuadrado; una vez hecho esto, se suman todas y se dividen entre el número de datos

Varianza = $s^2 = \dfrac{\sum (x_i - \bar{x})^2}{N}$

haciendo operaciones se demuestra que la varianza es también igual a

Varianza = $s^2 = \dfrac{\sum x_i^2}{N} - \bar{x}^2$

fórmula que hace los cálculos más sencillos y que será la que utilizaremos

Ejemplo: Si los datos son 2; 4; 5; 5

La media será $\bar{x} = \dfrac{2+4+5+5}{4} = \dfrac{16}{4} = 4$

Y la $s^2 = \dfrac{(2-4)^2+(4-4)^2+(5-4)^2+(5-4)^2}{4} = \dfrac{4+0+1+1}{4} = \dfrac{6}{4} = 1{,}5$

O bien

$s^2 = \dfrac{2^2+4^2+5^2+5^2}{4} - 4^2 = \dfrac{4+16+25+25}{4} - 16 = \dfrac{70}{4} - 16 = 17{,}5 - 16 = 1{,}5$

Como las diferencias están elevadas al cuadrado la varianza tiene las mismas unidades que la media pero elevada al cuadrado, con lo cual no se pueden comparar y utilizaremos su raíz cuadrada que recibe el nombre de desviación típica y se representa por **s**

s = $\sqrt{\text{Varianza}}$

Tal y como se comentó con la desviación media, la desviación típica da idea de cómo se distribuyen los datos iniciales, si están más o menos repartidos de forma homogénea, o si bien están situados formando grupos separados y/o alejados de la media.

Sin embargo, a diferencia de la desviación media

En el ejemplo anterior la desviación típica s = $\sqrt{1{,}5} = 1{,}22$

Propiedades de la desviación típica

1. Si a los datos anteriores les sumamos 4 tendremos 6, 8, 9, 9 ya sabemos que la media será 4+4 = 8 calculemos ahora la desviación típica

$s^2 = \dfrac{6^2+8^2+9^2+9^2}{4} - 8^2 = \dfrac{36+64+81+81}{4} - 64 = \dfrac{262}{4} - 64 = 65{,}5 - 64 = 1{,}5$

s = $\sqrt{1,5} = 1,22$

Por tanto:

Si a una serie de datos se les suma o resta una misma cantidad la varianza y la desviación típica no varían

2. Si a los datos anteriores los multiplicamos por 10 tendremos 20, 40, 50, 50 ya sabemos que la media será 4 (10 = 40 Calculemos ahora la desviación típica

$s^2 = \dfrac{20^2 + 40^2 + 50^2 + 50^2}{4} - 40^2 = \dfrac{7000}{4} - 1600 = 1750 - 1600 = 150$

por tanto s = $\sqrt{150} = 12,2$

Por tanto

Si a una serie de datos los multiplicamos o dividimos por la misma cantidad la desviación típica queda multiplicada o dividida por dicha cantidad

Cálculo con datos agrupados

Ejemplo 2

La media calculada antes era: $\bar{x}$ =2,72

x_i	n_i	$x_i^2 n_i$
1	4	4
2	7	28
3	8	72
4	4	64
5	2	50
	25	218

$s^2. = \dfrac{218}{25} - 2,72^2 = 8,72 - 7,40 = 1,32$ paquetes de folios^2

s = $\sqrt{1,32} = 1,15$ paquetes de folios

Ejemplo 3

La media calculada antes era $\bar{x} =$ 30,67

Intervalos	x_i	n_i	$x_i^2 n_i$
(18 - 22]	20	2	800
(22 - 26]	24	14	8064
(26 - 30]	28	12	9408
(30 - 34]	32	12	12288
(34 - 38]	36	14	18144

(38 - 42]	40	6	9600
		60	58304

$s^2 .= \frac{58304}{60} - 30{,}67^2 = 971{,}73 - 940{,}65 = 31{,}08$ años^2

$s = \sqrt{31{,}08} = 5{,}57$ años

Ejemplo 4

La media calculada antes era $\bar{x} = 6{,}6$

Intervalos	**x_i**	**n_i**	**$x_i^2 n_i$**
(4 - 5]	4,5	2	40,5
(5 - 6]	5,5	8	242
(6 - 7]	6,5	10	422,55
(7 - 8]	7,5	5	281,25
(8 - 9]	8,5	5	361,25
		30	1347,5

$s^2 = \frac{21347{,}58}{30} - 6{,}6^2 = 44{,}92 - 43{,}56 = 1{,}36$ €^{2}

$s = \sqrt{1{,}36} = 1{,}16$ €

Autoevaluación

9. Coeficiente de variación

Medidas estadísticas

Coeficiente de variación

Las medidas de dispersión nos dan una idea de cuánto se alejan los valores de la media, pero no nos permite saber de dos distribuciones cuál es la que se separa más o menos de su media, unas veces porque no están en las mismas unidades (en los ejemplos 2, 3 y 4, ¿cuál es la distribución más dispersa?), y otras porque no es lo mismo separarse 10 metros si la media es 20 metros, que separarse 10 metros si la media es 100 metros.

Para ello se utiliza no la dispersión absoluta sino la relativa.

De estas mediadas utilizaremos la llamada **Coeficiente de variación** C.V. que es el cociente entre la desviación típica y la media C.V. = $\frac{s}{\bar{x}}$

Este coeficiente se suele dar en tantos por ciento. Señalamos que el coeficiente de variación no tiene unidades y que no es fiable cuando la media es un número próximo a cero.

Dadas dos series de datos aquella que tenga mayor coeficiente de variación es la más dispersa

Además se considera que si el coeficiente de variación es mayor que 30% la media no es representativa del conjunto de los datos y es conveniente usar otra medida de centralización

Vamos a calcular el coeficiente de variación en los ejemplos que estamos utilizando

Ejemplo 2

C.V. = $\frac{1,15}{2,72} = 0,42$ si lo multiplicamos por 100 obtenemos el 42%

Ejemplo 3

C.V. = $\frac{5,57}{30,67} = 0,18$ si lo multiplicamos por 100 obtenemos el 18%

Ejemplo 4

C.V. = $\frac{1,16}{6,6} = 0,178$ si lo multiplicamos por 100 obtenemos el 18%

Podemos concluir que las distribuciones de los ejemplos 3 y 4 son igualmente dispersas y la del ejemplo 2 el más dispersa, y que la media es representativa en los ejemplos 3 y 4 pero no lo es mucho en el ejemplo 2

Autoevaluación

10. Números índices

Medidas estadísticas

Números índices

Otra medida usada en estadística y que encontrarás en muchos periódicos y noticias en la televisión son los que se denominan **números índices;** habrás oído hablar del índice de precios de consumo, el índice del coste de la vida, el índice de la bolsa, etc...

Podemos decir que un número índice es un número que muestra los cambios de una variable en función del tiempo. Es una medida relativa a un valor llamado **base**, y suele darse en tantos por ciento

Vamos a ver como se calculan

Tomemos como ejemplo la variación de las hectáreas quemadas en España en el periodo de tiempo de 1995 al 2002, dichos datos se reflejan en la tabla siguiente

Años	1995	1996	1997	1998	1999	2000	2001	2002
Miles de hectáreas	121	153	90	170	100	140	55	30

En primer lugar, se elige un año base, normalmente el primero (en nuestro ejemplo 1995) y su valor se llama valor base (en nuestro ejemplo, 121); a este valor base pasa a ser 100 (el 100%)

Para obtener la variación de los demás datos se realiza una proporción

$$\frac{\text{Valor base}}{100} = \frac{\text{Valor a calcular}}{\text{indice}}$$

así para 1996 tendríamos

$$\frac{121}{100} = \frac{153}{\text{indice}} \Rightarrow \text{Índice} = \frac{153 \times 100}{121} = 126{,}45$$

Su variación sería 126,45 - 100 = 26,45

Para 1997 tendríamos

$$\frac{121}{100} = \frac{90}{\text{indice}} \Rightarrow \text{Índice} = \frac{90 \times 100}{121} = 74{,}38$$

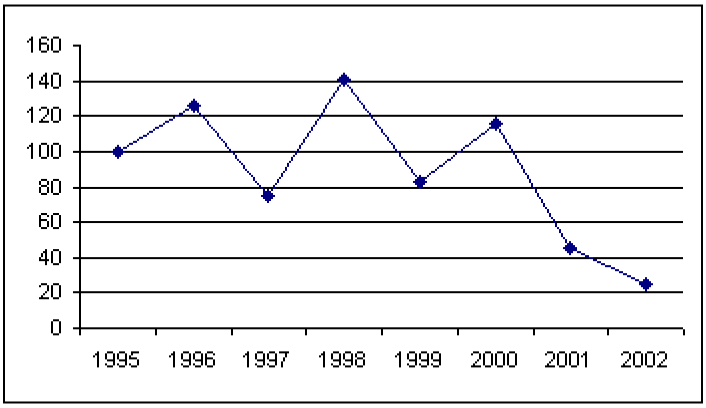
Su variación sería 74-38 - 100 = - 25,62

Continuando hasta 2002 tendríamos la siguiente tabla:

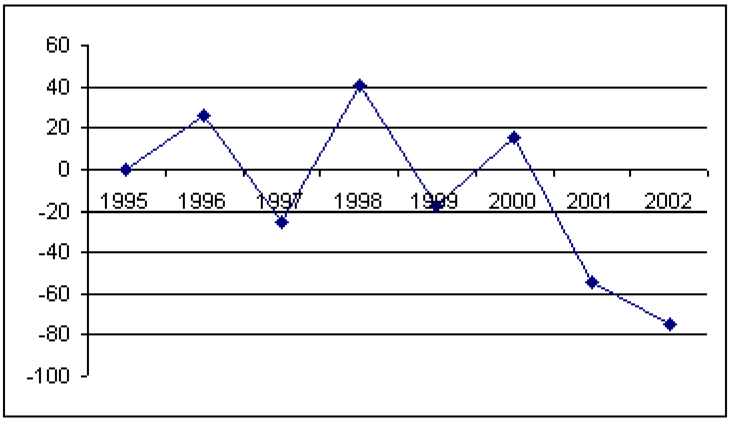
Años	1995	1996	1997	1998	1999	2000	2001	2002
Miles de hectáreas	121	153	90	170	100	140	55	30
Índice	100	126,45	74,38	140,50	82,64	115,70	45,45	24,79
Variación	-	26,45	-25,62	40,50	-17,36	15,70	-54,45	-75,21

Se pueden presentar gráficamente dichas variaciones

Grafica de números índices



Gráfica de variación del índice



Autoevaluación