

FICHA L. Estadística

El objetivo de la estadística es tomar un número grande de datos (por ejemplo, las notas de un examen), y extraer de ellos una serie de parámetros que permitan, con unos pocos números, hacernos una idea de estos datos, sin necesidad de recurrir a la tabla completa. Por ejemplo, si tenemos una clase con notas 5, 7, 8, 7, 6, 5, 7, 9, 1, 10, 8, 7, 5, 6, 7, resulta difícil (aun siendo 15 datos) hacerse una idea de si la clase es o no buena, mala, dispersa, etc. Para ello, recurrimos a dos tipos de medidas: **medidas de centralización**, que determinan la zona media de los datos, y **medidas de dispersión**, que determinan cuánto se alejan los datos del centro. Usaremos el siguiente ejemplo para ilustrar la tabla que viene a continuación:

5, 6, 7, 5, 7, 6, 8, 7, 9, 4, 2, 7, 6, 5, 7, 8, 9

Ordenamos los datos y contamos cuántos hay de cada (f_i). Creamos las columnas correspondientes:

- F_i se obtiene sumando a la columna f_i el dato anterior
- h_i se obtiene dividiendo f_i entre el número total de datos (en este caso 17)
- H_i se obtiene sumando h_i el dato anterior.

x_i	f_i	$x_i f_i$	$x_i^2 f_i$	F_i	h_i	H_i
2	1	2	4	1	0.06	0.06
4	1	4	16	2	0.06	0.12
5	3	15	75	5	0.18	0.29
6	3	18	108	8	0.18	0.47
7	5	35	245	13	0.29	0.76
8	2	16	128	15	0.12	0.88
9	2	18	162	17	0.12	1
Σ	17	108	738		1	

Centralización	Media	$\bar{x} = \frac{\sum x_i f_i}{N}$	Promedio de x_i	$\bar{x} = \frac{108}{17} = 6.35$
	Moda	x_i con mayor f_i	El valor que más se repite	$Mo = 7$ (Se repite 5 veces)
	Mediana	Valor central	Ordenar los datos e ir tachando hasta que quede uno o dos (en este caso, hacer la media de los dos) Primer valor que supere 0.5 (50%) en H_i	$Me = 7$ (en H_i es 0.76 y el anterior es 0.47, así que es el primero en superarlo)

Nota: podríamos haber hecho la mediana, en vez de con la columna H_i , ordenando los valores:

2	4	5	5	5	6	6	6	7	7	7	7	7	7	8	8	9	9
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---

Al final, queda el 7 de en medio, por lo que la mediana es $Me = 7$ (Si hubiesen quedado dos números en el centro, se haría la media de los dos)

Dispersión	Varianza	$s^2 = \frac{\sum x_i^2 f_i}{N} - \bar{x}^2$	Sirve para calcular la desviación típica	$s^2 = \frac{738}{17} - 6.35^2 = 3.05$
	Desviación típica ¹	$s = \sqrt{s^2}$ $s = \sqrt{\frac{\sum x_i^2 f_i}{N} - \bar{x}^2}$	<p>Mide la distancia que hay entre los datos y la media</p> <ul style="list-style-type: none"> ▪ Si los datos son todos iguales, por ejemplo todos iguales a 6, la media será 6 y la desviación 0 (ninguno se desvía) ▪ Si los datos son, por ejemplo, la mitad 4 y la mitad 6, la desviación es 1 (se desvían un punto de la media, que es 5, aunque no haya ningún 5). ▪ Si los datos son 1,1,1,1,9,9,9,9, la desviación es 4 (se desvían 4 de la media) 	<p>$s = \sqrt{3.05} = 1.75$</p> <p>Es decir, los datos se desvían 1.75 puntos de la media. La mayor parte de los datos están entre</p> <p style="text-align: center;">$(\bar{x} - s, \bar{x} + s)$</p> <p>Es decir, entre:</p> <p style="text-align: center;">$(6.35 - 3.05, 6.35 + 3.05)$</p> <p>Operando, entre:</p> <p style="text-align: center;">$(3.32, 9.40)$</p>
	Coefficiente de variación	$CV = \frac{s}{\bar{x}}$	Mide sin unidades la desviación respecto a la media. Suele darse en porcentaje. Si es 0%, no hay desviación (el primer ejemplo). En el segundo ejemplo es 20%, hay muy poca desviación (todos más o menos iguales). En el tercero, hay un 80% de desviación (mucho desviación)	<p>$CV = \frac{s}{\bar{x}} = \frac{1.75}{6.35} = 0.27 \rightarrow 27\%$</p> <p>El CV nos dice que los datos están más o menos agrupados. Si fuese 0%, todos los datos serían iguales. Si fuese de un 100% habría una gran dispersión. Por encima del 50% podemos decir que la dispersión es alta. En este caso entendemos que es asumiblemente baja.</p>
	Rango	$R = x_{max} - x_{min}$	Diferencia entre el mayor y menor de los valores. Por ejemplo en 1,5,7,4,8,6 el rango sería $8-1=7$	<p>$R = 9 - 2 = 7$</p> <p>Que también nos da una idea de la dispersión de los puntos.</p>

¹ En esta notación usamos \bar{x} para la media, y s para la desviación (*Standard deviation*), pero también es frecuente utilizar letras griegas μ para la media y σ para la desviación. Estas letras son, en griego, $\mu = m$ de media y $\sigma = s$

Ejercicio 1

Se han estudiado dos clases, siendo los resultados los mostrados en la tabla:

S2X	3	7	7	10	4	9	10	7	6	8
S2Y	0	8	5	5	9	0	1	5	8	3

Rellena las siguientes tablas y calcula los parámetros media, desviación, CV, mediana, moda y rango. Añade un comentario sobre cómo ves a cada clase a la vista de los resultados.

Realiza un diagrama de frecuencias. Para ello dibuja en un eje $x - y$ la variable x_i de cada tabla en el eje de las x , y la frecuencia f_i de cada tabla en el eje y . Busca en internet cómo se dibuja un diagrama de frecuencias o histograma.

S2X						
x_i	f_i	$x_i f_i$	$x_i^2 f_i$	F_i	h_i	H_i
Σ						

Parámetros:

S2Y						
x_i	f_i	$x_i f_i$	$x_i^2 f_i$	F_i	h_i	H_i
Σ						

Parámetros:

Ejercicio 2. Se tienen dos experimentos, uno en que se han medido unas masas de 1, 1, 5, 10 y 10 kg, y otro en que las mediciones son 4,4,6,6,7kg. Determina para cada experimento los parámetros correspondientes. A la hora de presentar resultados, compara ambos experimentos.

A							B						
x_i	f_i	$x_i f_i$	$x_i^2 f_i$	F_i	h_i	H_i	x_i	f_i	$x_i f_i$	$x_i^2 f_i$	F_i	h_i	H_i
Σ							Σ						

Cálculo de parámetros:

Ejercicio 3. Se ha preguntado a 200 familias por la cantidad que son capaces de ahorrar cada mes, siendo el resultado el siguiente (se muestra en el diagrama de barras la frecuencia sobre cada una de las barras)

NOTA: puedes hacer el ejercicio en €, o cambiar la variable a “cientos de €”. El resultado no cambia.

Calcula el ahorro medio, la desviación y el coeficiente de variación de estas personas.



x_i	f_i	$x_i f_i$	$x_i^2 f_i$	h_i	H_i	Parámetros:										
100																
200																
300																
400																
500																
600																
700																
800																
900																
1000																
Σ																

Soluciones:

(NOTA1: algunos cuartiles y medianas están calculados con Excel, por lo que darán un valor distinto al que se obtiene calculando a mano.)

(NOTA2: se muestra en algunos ejercicios más información, por si se quiere hacer como ejercicio extra)

Ej1	Media	Desv	CV	Mediana	Moda	Rango
S2X	7.1	2.21	31%	7	7	7
S2Y	4.4	3.17	72%	5	5	9

Ej2	Media	Desv	CV	Mediana	Moda	Rango
A	5.4	4.03	75%	5	1	9
B	5.4	1.20	22%	6	4	3

Ej3	Media	Desv	CV	Mediana	Moda	Rango
Ahorro	604.00	237.45	39%	600	500 600	900