

# ESTADÍSTICA PARA TOD@S: ADAPTACIÓN A LA DIVERSIDAD FUNCIONAL

## ESTADÍSTICOS DESCRIPTIVOS: LA MEDIANA



**UNIVERSIDAD PABLO DE OLAVIDE**

**ÁREA DE ESTADÍSTICA E IO**

**Autoras:**

Nieves Aquino Llinares

M<sup>a</sup> del Pilar Moreno Navarro

**UNIVERSIDAD PABLO DE OLAVIDE**

**Año 2021**

**ISBN: 978-84-09-34306-5**

## Introducción

Esta publicación forma parte de un proyecto más ambicioso que se denomina ***Estadística para tod@s: adaptación a la diversidad funcional*** y a través del cual las autoras, conscientes de la falta de material específico en estadística adaptado a diversas discapacidades, quieren poner su granito de arena para alcanzar una universidad más inclusiva, con la realización de píldoras formativas y publicaciones de apoyo al estudio accesibles a diversas discapacidades. En concreto este anexo es un complemento al vídeo formativo sobre la mediana que las autoras han realizado adaptado a personas con discapacidad auditiva. Además, dicho anexo se encuentra traducido a Braille, estando el documento EDICO, disponible para su consulta en la ONCE. La cápsula docente puede visualizarse en el siguiente enlace:

Cápsula formativa: [La Mediana](#)

### **1ª DIAPOSITIVA**

**Minuto: 00:10**

Esta píldora formativa pertenece a la serie “Estadística para tod@s: adaptación a la diversidad funcional”, cuyas autoras son Nieves Aquino Llinares y M<sup>a</sup> del Pilar Moreno Navarro, profesoras del área de Estadística e Investigación Operativa de la Universidad Pablo de Olavide y tiene como objetivo alcanzar una universidad más inclusiva.

### **2ª DIAPOSITIVA**

**Minuto: 00:30**

Concretamente, esta píldora formativa versa sobre la mediana, que es un estadístico descriptivo de posición central y nos indica el valor de la variable que divide a la distribución de forma que en cada una de las partes hay la misma cantidad de valores.

### **3ª DIAPOSITIVA**

**Minuto: 00:46**

Definimos la mediana como el valor que ocupa la posición central en un conjunto de datos una vez ordenados de menor a mayor.

Es decir, tal y como se observa en la figura, si tenemos 7 personas que hemos ordenado de menor a mayor según la variable estudiada, por ejemplo la altura, el valor de la mediana sería la altura de la persona que ocupa la cuarta posición pues dicho individuo deja a tres personas a su izquierda y a tres personas a la derecha.

No tiene sentido su cálculo para variables de tipo cualitativo nominal pues el orden de los valores de la variable es irrelevante.

### **4ª DIAPOSITIVA**

**Minuto: 01:25**

Por tanto, tras ordenar los datos de forma creciente, la mediana es el valor de la variable que deja a su izquierda y a su derecha el mismo n.º de datos, es decir, deja a su izquierda el 50% de los datos y a su derecha el otro 50% de datos.

**5ª DIAPOSITIVA****Minuto: 01:44**

La mediana se calcula de forma diferente dependiendo de si los datos están o no agrupados y en este último caso es algo más sencillo si  $n$  es impar.

Comencemos con el cálculo de la mediana cuando los datos no están agrupados.

En este caso la mediana es el valor de la variable que alcanza el individuo que ocupa la posición  $n$  más 1 dividido entre 2,  $(n+1)/2$ , una vez que se han ordenado los datos en orden creciente. En la imagen pueden observarse siete bombillas las cuales ha sido ordenadas según tamaño en orden ascendente. La mediana sería el tamaño de la bombilla que ocupa la posición  $(7+1)/2$ , es decir, la que ocupa la posición cuarta.

Si deseamos realizar el cálculo partiendo de unos datos agrupados según su tabla de frecuencias correspondiente debemos buscar el primer valor de la frecuencia absoluta acumulada,  $N_i$  (mayúscula) mayor o igual al total de individuos investigados,  $n$ , dividido entre 2 ( $n/2$ ). Si el valor de la frecuencia absoluta acumulada  $N_j$  es superior a  $n/2$  la mediana es  $x_j$ .

**6ª DIAPOSITIVA****Minuto: 03:00**

En el ejemplo 1 observamos las notas que obtiene Antonio en segundo de bachillerato. 8, 7, 4, 6, 5, 6, 7, 8 y 4. ¿Cómo procedemos a calcular la mediana?

En primer lugar se observa como el total de notas es  $n=9$ , impar, por lo que existe una posición central que es  $(n+1)/2 = 5$ .

En segundo lugar procedemos a ordenar los datos en orden creciente 4 4 5 6 6 7 7 8 y 8

Por último, buscamos el valor que ocupa la 5ª posición, que en este caso es 6 puntos.

Tras analizar las notas de Antonio, la mediana toma el valor de 6 puntos sobre 10.

**7ª DIAPOSITIVA****Minuto: 03:56**

Una vez motivado el concepto de la mediana, pasamos a explicar cómo se calcula a partir de una tabla de frecuencias, es decir, ¿cómo calcularíamos la mediana de las notas de Antonio si los datos hubiesen estado agrupados en una tabla de frecuencias?

$x_i$	$n_i$	$N_i$
4	2	2
5	1	3
6	2	5
7	2	7
8	2	9
	<b><math>n=9</math></b>	

En esencia es igual que como lo hemos hecho antes pero con la ventaja de que ahora los datos o notas ya vienen ordenados en la tabla. Buscamos en la columna de frecuencias absolutas acumuladas,  $N_i$ , el primer valor mayor a  $n/2$ , es decir,  $N_i \geq n/2$ . El resultado de dividir el total de datos entre dos es 4,5 por lo que identificamos el primer valor de la columna de las frecuencias absolutas acumuladas superior o igual a 4,5. Se observa como dicha frecuencia absoluta acumulada es 5, y por tanto la mediana será el valor que toma la variable en la tabla para esa posición, que corresponde al valor de la variable igual a 6. Esto indica que la mediana toma el valor 6 puntos.

**8ª DIAPOSITIVA****Minuto: 05:05**

¿Qué ocurre si los datos no están agrupados pero el total de unidades investigadas es par?

En este caso la mediana es la media aritmética de los valores que ocupan las posiciones  $n/2$  y  $(n/2) + 1$  una vez ordenados los valores observados de forma ascendente. En la imagen se muestran 6 bombillas ordenadas por tamaño y la

mediana sería la media aritmética del tamaño de las bombillas que ocupan la tercera y cuarta posición.

Si trabajamos con los datos ya resumidos en una tabla de frecuencias, para calcular la mediana buscamos, igual que en el caso en el que  $n$  es impar, aquella frecuencia absoluta acumulada que superior a  $n$  entre 2 ( $N_i \geq n/2$ ). El valor de la variable para esa posición de la tabla de frecuencias nos dará la mediana pero si hay una coincidencia entre  $n/2$  y  $N_j$  entonces tendremos que calcular la media aritmética entre  $x_j$  y el valor siguiente  $x_{(j+1)}$  es decir, la mediana será:

$$(x_j + x_{j+1})/2$$

### 9ª DIAPOSITIVA

**Minuto: 06:22**

En el ejemplo 2 vamos a calcular la mediana de las valoraciones obtenidas por Susana en un concurso de pintura: 9 8 5 7 6 7 8 y 9 puntos.

Para calcular la mediana en primer lugar observamos que se han realizado 8 mediciones por parte de los jueces, lo que supone un número par, y, por tanto, existen dos posiciones centrales que son  $n/2 = 4$  y  $(n/2)+1 = 5$ . En segundo lugar ordenamos los datos de forma creciente: 5 6 7 7 8 8 9 y 9 y por último buscamos la valoración que ocupa la cuarta y quinta posición, 7 y 8 puntos respectivamente. En consecuencia, la mediana es la media aritmética entre 7 y 8 tomando un valor de 7'5 puntos,  $(7+8)/2$ .

Como interpretación de este estadístico podemos decir que el 50% de los jueces valoraron el trabajo de Susana con una valoración inferior o igual a 7'5 puntos y que el 50% de los jueces hizo una valoración de 7'5 puntos o superior.

### 10ª DIAPOSITIVA

**Minuto: 07:49**

Si en el ejemplo 2 conocemos las valoraciones dadas a Susana a partir de la tabla de frecuencias y debemos buscar el primer valor de la columna de frecuencias absolutas acumuladas,  $N_i$ , que supere o iguale el valor  $n/2$ , es decir, el primer valor de la columna  $N_i$  que supere  $8/2=4$ .

$x_i$	$N_i$	$N_i$
5	1	1
6	1	2
7	2	4
8	2	6
9	2	8
	<b>n=8</b>	

Cuando  $n/2$  coincide exactamente con uno de los valores  $N_i$ , la mediana es la media aritmética del valor de la variable que corresponde a  $n/2$  y el siguiente valor. Por tanto, la mediana es la media de 7 y 8, es decir, 7,5 puntos.

**11ª DIAPOSITIVA**

**Minuto: 08:35**

El caso más complicado para el cálculo de la mediana se da cuando tenemos una variable cuantitativa continua con los valores agrupados en intervalos. En este caso, en la columna  $N_i$ , buscamos el 1er valor mayor o igual al total de observaciones dividido entre 2, es decir, el primer  $N_i \geq n/2$ , que corresponde al intervalo donde pertenece la mediana, es decir, el intervalo comprendido entre  $L_{j-1}$  y  $L_j$ .

$(L_{i-1}, L_i]$	$n_i$	$N_i$
$[L_0, L_1]$	$n_1$	$N_1$
$\vdots$	$\vdots$	$\vdots$
$(L_{j-2}, L_{j-1}]$	$n_{j-1}$	$N_{j-1}$
$(L_{j-1}, L_j]$	$n_j$	$N_j$
$\vdots$	$\vdots$	$\vdots$
$(L_{k-1}, L_k]$	$n_k$	$N_k$
	<b>n</b>	

A continuación, la mediana se calcula con la fórmula siguiente:

$$Me = L_{j-1} + \frac{\frac{n}{2} - N_{j-1}}{n_j} (L_j - L_{j-1})$$



Extremo inferior del intervalo donde se encuentra la mediana  $L_{j-1}$  más el cociente que se obtiene de considerar en el numerador el valor de  $n$  dividido entre 2 menos el valor de la frecuencia absoluta acumulada del intervalo justo anterior  $N_{j-1}$  y en el denominador la frecuencia absoluta del intervalo donde está la mediana  $n_j$ . Dicho cociente se multiplica por la amplitud del intervalo donde se encuentra la mediana ( $L_j$  menos  $L_{j-1}$ )

**12ª DIAPOSITIVA****Minuto: 09:49**

Como ventaja más destacable de la mediana se tiene que sólo influyen los valores centrales de la distribución y, por tanto, no es sensible a valores extremos, inconveniente que sí tenía la media aritmética.

La 2ª ventaja es que es la medida estadística más representativa en caso de variables que sólo admiten la escala ordinal.

**13ª DIAPOSITIVA****Minuto: 10:12**

Si nos centramos en los inconvenientes de la mediana podemos resaltar que no puede calcularse para variables cualitativas nominales y que en su cálculo no intervienen todos los valores de la variable.

**14ª DIAPOSITIVA****Minuto: 10:26**

A continuación, se muestran dos ejemplos de cálculo de la mediana de una distribución en función de si los datos están o no agrupados. Pincha en el icono correspondiente para acceder a ellos.

**Vídeo 1: Datos no agrupados****Minuto: 10:40**

En el siguiente ejemplo se observan los datos recogidos por un analista de riesgos que desea estudiar el número de partes de accidente que presentan al

año sus clientes con el objeto de hacer una comparación con las aseguradoras de la competencia.

$x_i = \text{n}^\circ \text{ de partes de accidente}$	$n_i = \text{n}^\circ \text{ de clientes}$	$N_i$
0	132	132
1	60	192
2	36	228
3	12	240
	<b>n = 240</b>	

En la tabla de datos se puede ver cómo hay 132 clientes que no presentaron ningún parte de accidente en el período estudiado. 60 clientes presentaron 1 parte, 36 clientes presentaron 2 partes y por último 12 clientes presentaron 3 partes en el año estudiado. Como se puede observar el total de clientes analizados asciende a 240 (par).

Se observa además como en la tabla de frecuencias existe ya calculada una nueva columna que indica la frecuencia absoluta acumulada,  $N_i$ . Dicha frecuencia acumulada se ha calculado sumando las frecuencias absolutas de cada categoría y las anteriores.

Para calcular la mediana buscamos el cliente que ocupa la posición  $240/2=120$  dentro de la tabla a través de la frecuencia absoluta acumulada ( $N_i$ ). Dicho cliente, el que ocupa la posición 120 se encuentra entre los 132 primeros, primer valor de la distribución, y corresponde a los clientes que no han dado en el año estudiado ningún parte por lo que la mediana es 0 partes de accidente.

¿Cómo se interpreta que la mediana sea 0 partes? El 50% de los clientes analizados presentan 0 partes y el 50% de los clientes presentan 0 o más partes. En este caso se observa como la distribución es asimétrica pues la mediana no ocupa una posición centrada en la distribución.

**Vídeo 2: Datos agrupados**

**Minuto: 13:04**

En el siguiente ejercicio se muestran los tiempos, en segundos, que tardaron 35 nadadores en realizar la prueba de 50 metros estilo espalada en una piscina de 25 metros.

$x_i$	Tiempo	N.º de nadadores ( $n_i$ )	$N_i$
23	[22 , 24]	7	7
26	(24 , 28]	12	19
29	(28 , 30]	10	29
31	(30 , 32]	6	35
		<b>n = 35</b>	

En esta ocasión la variable es cuantitativa continua y viene agrupada en intervalos. Se observa como 7 nadadores tardaron entre 22 y 24 segundos. 12 deportistas nadaron la prueba entre 24 y 28 segundos. 10 tardaron entre 28 y 30 segundos y por último, 6 nadadores tardaron entre 30 y 32 segundos. ¿Cómo podemos calcular la mediana?

Para calcular la mediana calculamos en primer lugar el individuo que ocupa la posición  $n/2=35/2=17'5$ . En segundo lugar buscamos la primera frecuencia absoluta acumulada que iguale o supere este valor. Nótese que la primera  $N_i$  que supera 17'5 es 19 por lo que la mediana pertenece al intervalo (24 , 28]. A partir de este intervalo calculamos la mediana con la siguiente fórmula:

$$M = x_{i-1} + \frac{\frac{n}{2} - N_{i-1}}{n_i} x_i = 24 + \frac{\frac{35}{2} - 7}{12} 4 = 27,50$$

Se concluye que la mediana es 27'5 segundos y que el 50% de los nadadores tardan en realizar la prueba 27'5 segundos o más.

**15ª DIAPOSITIVA**

**Minuto: 15:43**

Muchas gracias por la atención.

Esperamos que la serie Estadística para tod@s sea un granito más para alcanzar una Universidad más inclusiva.