INSTITUTO de ENSEÑANZAS a DISTANCIA de ANDALUCÍA

2º de Bachillerato

Matemáticas Aplicadas a las Ciencias Sociales II

Contenidos

Estadística inferencial: Parámetros estadísticos

1. Medidas de centralización

Sede del Instituto Nacional de Estadística
Imagen de -Merce- con licencia Creative Commons

En su origen la Estadística estuvo asociada a los Estados, para ser utilizada por los gobiernos y cuerpos administrativos generalmente con fines censales y recaudatorios. La obtención de datos al servicio de la administración del estado continúa actualmente a través de los trabajos que se llevan a cabo en organismos de estadística nacionales e internacionales. En España disponemos del Instituto Nacional de Estadística, si hay un lugar donde entiendan de encuestas, sin duda, es éste.

La realización de encuestas tiene un trabajo a posteriori, que es el de analizar y simplificar la información recogida y, a partir de una muestra, intentar generalizar los resultados a toda la población. Con estos resultados podremos hacer predicciones que ayudarán a la toma de decisiones.

En este punto vamos a estudiar cómo sintetizar esa información obtenida. Para conseguirlo vamos a buscar un valor que exprese el centro de las observaciones y sea representativo de toda una muestra. Esto es muy complicado, pues se trata de resumir en un sólo valor todos los datos referidos a una variable. No obstante, si lo conseguimos, este número nos permitirá explicar de forma rápida y concisa el comportamiento de una población con respecto a un determinado aspecto.

Importante

Se llaman **medidas de centralización** a los parámetros que indican el valor hacia el que tienden a situarse los datos de una distribución estadística.

Para saber más

En el siguiente vídeo podemos ver cual es la función hoy en día del Instituto Nacional de Estadística de España.

Curiosidad

Aquí tienes otro vídeo elaborado por el INE titulado "Un día en cifras"

1.1. Media muestral

Imagen de Iowapolitics.com con licencia Creative Commons

Nuestra empresa TisBet Survey ha recibido el encargo de realizar un estudio sobre el nivel de aceptación de un determinado líder político. A cada una de las 600 personas consultadas se le pidió que asignara una calificación con un número natural de 0 a 10, obteniéndose los siguientes resultados.

Puntuación	0	1	2	3	4	5	6	7	8	9	10
N.º de personas	12	38	62	97	160	130	65	18	11	6	1

La aceptación media del líder vendrá dada por la suma de todas las puntuaciones obtenidas dividida por las 600 opiniones.

Para el cálculo de la aceptación media nos ayudaremos de la construcción de la siguiente tabla, donde:

La primera columna es el valor de los posibles valores de la variable.

La segunda columna es la frecuencia absoluta de cada valor, es decir, el número de encuestados que han respondido un valor determinado.

Cada celda de la tercera columna corresponde al producto de las celdas, que están en la misma fila, de las dos primeras columnas.

x_i	f_i	$x_i \cdot f_i$
0	12	$0 \cdot 12 = 0$
1	38	$1 \cdot 38 = 38$
2	62	$2 \cdot 62 = 124$
3	97	291
4	160	640
5	130	650
6	65	390
7	18	126
8	11	88
9	6	54
10	1	10
	N=600	**2411**

En la última fila calcularemos la suma de todos los elementos de esa columna. Así, en la columna f_i la suma corresponderá con el número de personas entrevistadas. En la columna $x_i \cdot f_i$ sumamos cada valor multiplicado por el número de personas que lo han elegido, esto corresponde a la suma de todas las calificaciones obtenidas.

La aceptación media de este líder vendrá dada por la media aritmética de estos resultados. Es decir, el cociente entre la suma de todas las calificaciones obtenidas, **2411** (última fila, tercera columna) y el número de personas encuestadas, **600** (que se corresponde con la suma de los valores de la segunda columna).

$$Aceptacion \;\; media = \frac{2411}{600} = 4{,}02$$

Importante

Se llama **Media Muestral** (Media aritmética) de una variable estadística al cociente entre la suma de todos los valores obtenidos en la muestra y el tamaño de la muestra. La media muestral se representa por $\bar{x}$. Recuerda que la media poblacional se representa por la letra griega **μ**

Si X es una variable estadística que toma los valores x_1, x_2, x_3, ... x_n con frecuencias absolutas f_1, f_2, f_3, ... f_n , respectivamente, la media de la variable X viene dada por la siguiente expresión:

$$\bar{x} = \frac{x_1 \cdot f_1 + x_2 \cdot f_2 + \ldots + x_n \cdot f_n}{f_1 + f_2 + \ldots + f_n} = \frac{\sum_{i=1}^{n} x_i \cdot f_i}{\sum_{i=1}^{n} f_i} = \frac{\sum_{i=1}^{n} x_i \cdot f_i}{N}$$

Ejercicio resuelto

Se está haciendo un estudio sobre alimentación y sobrepeso en adolescentes. Para ello se ha escogido a una muestra de 25 personas de entre 15 y 16 años que son clientes habituales de locales de comida rápida. Los datos sobre el peso en kilogramos obtenidos son los siguientes:

Imagen de pointshoot con licencia Creative Commons

49,5	57,0	65,0	60,0	74,5	47,0	50,6	49,5	64,3
45,2	48,4	58,1	66,5	76,1	48,8	56,2	66,5	77,0
41,0	46,0	54,2	63,8	46,0	51,5	68,2		

Divide la población en 6 intervalos de clase de longitud 6 y halla el peso medio de esta muestra.

Mostrar retroalimentación

En el primer ejemplo, la aceptación del político, realizábamos la media de una variable aleatoria discreta (sus posibles valores eran 0, 1, 2, ...10).

Ahora estamos trabajando con una variable aleatoria continua. Aquí las posibilidades son infinitas. Por ello lo que hacemos es agrupar a la población en clases, cada una de las cuales está caracterizada por un intervalo de clase y una marca de clase, si a y b son los extremos del intervalo, la marca de clase será (a+b)/2.

Empezamos calculando los intervalos de clase. En este caso son 6 y empiezan en 41,0 y terminan en 77,0, con una amplitud de 6 unidades. Serán [41'0, 47'0), [47'0, 53'0), [53'0, 59'0), [59'0, 65'0), [65'0, 71'0) y [71'0, 77'0]

Vamos a representarlos en una tabla parecida a la del primer ejercicio. Ahora la frecuencia absoluta (f_i) será el número de observaciones que pertenecen a cada intervalo

Intervalos de clase	x_i	f_i	$x_i \cdot f_i$
[41'0, 47'0)	44	4	176
[47'0, 53'0)	50	7	350
[53'0, 59'0)	56	4	224
[59'0, 65'0)	62	3	186
[65'0, 71'0)	68	4	272
[71'0, 77'0)	74	3	222
		25	1430

El peso medio vendrá dado por la suma de todos los pesos dividido por el número de estos. En este caso:

$$Peso\ medio = \frac{\sum_{i=1}^{6} x_i \cdot f_i}{N} = \frac{1430}{25} = 57{,}2$$

En la siguiente página puedes encontrar ejercicios de cálculo de la media aritmética para practicar.

Para saber más

La media arítmetica no es la única media que podemos calcular. Hay una variación de esta en la que a cada valor se le da un peso. A esta media se le llama **Media ponderada**.

La otra media en la que en vez de sumar se multiplica se llama **Media geométrica**.

1.2. Moda y mediana

Imagen de **José Goulão** con licencia Creative Commons

¿Recuerdas el ejemplo anterior de la valoración de nuestro líder político? Volvamos a ver los resultados obtenidos en la encuesta:

Puntuación	0	1	2	3	4	5	6	7	8	9	10
N.º de personas	12	38	62	97	160	130	65	18	11	6	1

Si bien la media es el valor central más representativo de una muestra, no es el único valor central que podemos calcular. También podemos ver cuál es el dato que más se repite, en nuestro caso, cuál es la puntuación que más personas han elegido para valorar al líder político. En el ejemplo este valor es 4. Este dato es la moda de la variable estadística.

Importante

Se llama **moda** de una variable estadística al valor de la variable que presenta mayor frecuencia absoluta. La moda se representa por $\mathbf{M_0}$.

Si la variable es discreta, el cálculo de la moda no presenta ninguna dificultad, únicamente observamos las frecuencias absolutas en la tabla, vemos cuál es la mayor y la moda será el valor de la variable correspondiente a dicha frecuencia. Puede que haya más de un valor máximo, en este caso la distribución será bimodal (si hay dos), trimodal (si hay tres), ...

Si la variable es continua los valores se agrupan en intervalos, por lo que tendremos un **intervalo modal** que será el de mayor frecuencia absoluta. No obstante, si consideramos que los datos están uniformemente distribuidos podemos calcular un valor concreto para la moda que vendrá dado por la siguiente fórmula:

$$M_o = L_i + \frac{f_{M_o} - f_{M_o-1}}{(f_{M_o} - f_{M_o-1}) + (f_{M_o} - f_{M_o+1})} \cdot c$$

Donde L_i es el límite inferior del intervalo modal, las frecuencias absolutas corresponden al intervalo modal, al anterior y al siguiente y c es la amplitud de dicho intervalo.

Además de la moda, podemos calcular el valor que está justamente en el centro de todos los datos una vez que están ordenados. En nuestro ejemplo tenemos 600 datos, sería interesante conocer cuál es el valor que deja 300 datos menores que él y otros 300 mayores que él. Este dato es la mediana. Para calcularla nos vamos a ayudar de la tabla de frecuencias, a la que añadiremos la columna de la Frecuencia acumulada (F_i). Esta frecuencia se obtiene sumando todas las frecuencias absolutas hasta el dato en el que nos encontramos.

Imagen de IowaPolitics.com con licencia Creative Commons

x_i	f_i	F_i
0	12	$F_1=f_1=12$
1	38	$F_2=f_1+f_2=50$
2	62	$F_3=f_1+f_2+f_3=112$
3	97	209
4	160	369
5	130	499
6	65	564
7	18	582
8	11	593
9	6	599
10	1	600
	N=600	

Para el cálculo de la mediana lo primero que hay que hacer es construir la tabla de frecuencias anterior. A continuación, calculamos la mitad del número de datos $\frac{N}{2}=\frac{600}{2}=300$

- Si N es impar la mediana corresponderá al x_i correspondiente al primer F_i que supere ese valor.
- Si N es par:
  - Si exite algún valor de la columna F_i que es igual a $\frac{N}{2}$, la mediana es la media entre el x_i correspondiente a ese F_i y el siguiente, x_{i+1}.
  - En caso contrario, la mediana será el x_i que se corresponda con el primer valor de F_i que supere a $\frac{N}{2}$.

En nuestro caso la mediana es 4,porque 4 es el primer valor de la variable cuya frecuencia acumulada supera el valor de 300.

Importante

Se llama **mediana** de una variable estadística al valor de la variable tal que el número de observaciones menores que él es igual al número de observaciones mayores que él. La mediana de una variable estadística se represente por $\mathbf{M_e}$.

Hemos visto el cálculo de la mediana cuando la variable era discreta. Para el caso de la variable continua, si usamos intervalos se realiza de la misma forma y se obtiene el **intervalo mediano.** No obstante, si consideramos que los datos se distribuyen uniformemente se puede calcular el valor exacto de la mediana mediante la siguiente fórmula:

$$M_e=L_i+\frac{\frac{N}{2}-F_{M_e-1}}{f_{M_e}}\cdot c$$

Reflexiona

Los jugadores de un equipo de baloncesto se clasifican por la altura según la siguiente tabla

Alturas	Número de jugadores
[1'70,1'80)	3
[1'80,1'90)	4
[1'90,2'00)	5
[2'00,2'10)	3

Hallar la media, el intervalo modal, la moda, el intervalo mediano y la mediana

Mostrar retroalimentación

Empezamos con la tabla de frecuencias, calculando todas las columnas necesarias:

Alturas	Marcas de clase	f_i	$xi \cdot fi$	$x_i^2 \cdot f_i$	F_i
[1'70,1'80)	1'75	3	5'25	9'19	3
[1'80,1'90)	1'85	4	7'40	13'69	7
[1'90,2'00)	1'95	5	9'75	19'01	12
[2'00,2'10)	2'05	3	6'15	12'61	15
		15	28'55	54'50	

La media: $\overline{x} = \frac{28'55}{15} = 1'90$

El intervalo modal: [1'90, 2'00)

La moda: $M_o = 1'9 + \frac{(5-4)}{(5-4)+(5-3)} \cdot 0'1 = 1'9 + \frac{1}{1+2} \cdot 0'1 = 1'93$

El intervalo mediano: $\frac{N}{2} = \frac{15}{2} = 7'5$, por lo tanto el intervalo mediano es [1'90,2'00)

La mediana: $M_e = 1'9 + \frac{\frac{15}{2} - 7}{5} = 1'91$

En la siguiente página puedes encontrar ejercicios resueltos de cálculo de la media, mediana y moda de una serie de datos.

Para saber más

Además de las medidas de centralización existen otras medidas, las medidas de posición. Éstas, como ocurre con la mediana, nos dan el valor que deja un porcentaje de la población a su izquierda. La mediana puede considerarse también una medida de posición, ya que deja el 50% de la población a su izquierda.

posición, ya que deja el 50 % de la población a su izquierda.

Estas medidas son los **cuartiles**, los **deciles** y los **percentiles**. Realmente los cuartiles corresponden al percentil 25 y 75, pues dejan respectivamente el 25% y el 75% de la población a su izquierda. Vamos a ver en el siguiente vídeo como realizar el cálculo del percentil.

Comprueba lo aprendido

Responde Verdadero o Falso a las siguientes afirmaciones.

Los parámetros estadísticos permiten resumir y sintetizar un gran número de datos, en unos pocos números que proporcionan una idea lo más aproximada posible de toda la distribución.

Sugerencia

Verdadero Falso

Verdadero
Es justo para lo que sirven.

Siempre es posible calcular la media aritmética de una distribución estadística

Sugerencia

Verdadero Falso

Falso
No es posible cuando los datos son cualitativos.

Para el cálculo de la moda es necesario conocer el valor de todos los datos.

Sugerencia

Verdadero Falso

Falso
Para el cálculo de la moda sólo es necesario conocer el valor que aparece más veces.

La mediana es equivalente al percentil 50.

La mediana es equivalente al percentil 50.

Sugerencia

Verdadero Falso

Verdadero
La mediana deja el 50% de los datos a su izquierda, por lo tanto, corresponde con el valor de P_{50}

Se llaman parámetros de centralización a las medidas que suelen situarse hacia el centro del conjunto de datos ordenados.

Sugerencia

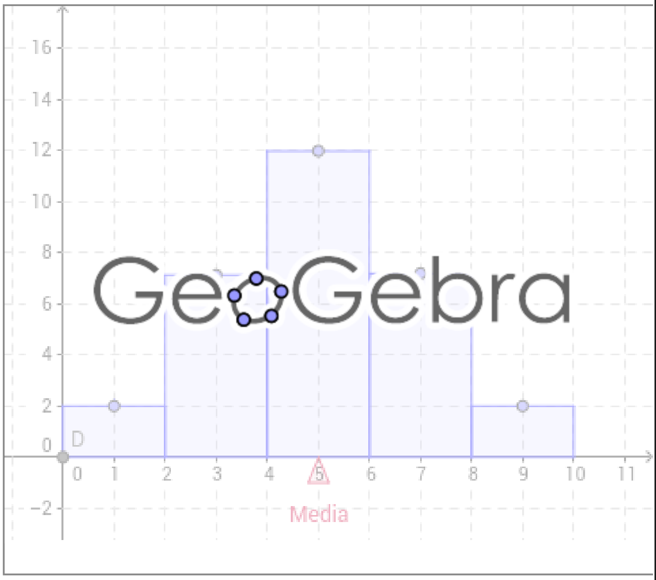
Verdadero Falso

Verdadero
Los parámetros de centralización suelen situarse hacia el centro del conjunto de datos ordenados

2. Medidas de dispersión

Imagen de ccarlstead con licencia Creative Commons

En el centro educativo "Benito V." están haciendo una comparativa entre las notas sacadas en la asignatura de Matemáticas Aplicadas a las CCSS II en los dos cursos de 2º de Bachillerato. Los resultados obtenidos puedes verlos en las siguientes gráficas, donde el eje X representa las calificaciones y el eje Y el número de alumnos que han obtenido dicha calificación:

2º Bachillerato A	2º Bachillerato B
	

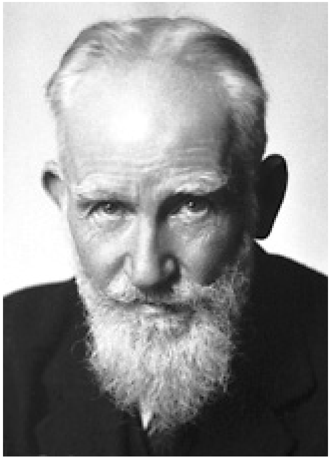
Puedes observar que para ambas clases la media es de las calificaciones es 5. Pero si sólo nos dicen que en ambas clases la nota media es de 5, la información es demasiado escueta. Es más, este parámetro es muy representativo en el grupo de 2º B, mientras que parece que no representa tanto a los resultados de 2ºA.

Necesitamos algo que nos mida si la media de una muestra es o no es representativa. De ello se van a encargar los parámetros de dispersión, que nos van a medir la desviación de los datos respecto a la media. Cuanto mayores sean los valores de los parámetros de dispersión, la media será menos representativa.

Importante

Se llaman **medidas de dispersión** a los parámetros que miden el nivel de concentración de los datos.

Curiosidad



Geroge Bernard Shaw en
Wikimedia Commons

Muchos son los detractores de la estadística. algunos se apoyan en la poca representatividad de la media en algunas ocasiones.

Al premio nobel de literatura del año 1925, Bernard Shaw, se le atribuye la siguiente frase:

"La estadística es una ciencia que demuestra que si mi vecino tiene dos coches y yo ninguno, los dos tenemos uno."

Este problema que se produce por la simplificación y la mala interpretación de los resultados, lo resolveremos mediante el cálculo de parámetros que nos permita medir el grado de dispersión de los datos que disponemos.

Aquí tienes otra frase mucho más acertada de este autor:

"Si tú tienes una manzana y yo tengo una manzana, e intercambiamos manzanas, tanto tú como yo seguimos teniendo una manzana cada uno.

Pero si tú tienes una idea y yo tengo una idea, e intercambiamos ideas, entonces ambos tenemos dos ideas".

2.1. Varianza y desviación típica

Imagen de Habladorcito con licencia Creative Commons

Vamos a calcular cuanto se desvían de la media los datos del grupo de 2ºA. Veamos de nuevo las calificaciones de nuestros dos grupos de alumnos:

2º Bachillerato A	
Calificaciones	N.º de Alumnos
[0,2)	7
[2,4)	6
[4,6)	4
[6,8)	6
[8,10]	7

Lo primero que haremos será calcular su media:

Calificaciones	Marcas de clase(x_i)	Nº de alumnos(f_i)	$x_i \cdot f_i$
[0,2)	1	7	7
[2,4)	3	6	18
[4,6)	5	4	20
[6,8)	7	6	42
[8,10]	9	7	63
		N=30	Suma =150

La media será: $\overline{x}=\dfrac{\sum_{i=1}^{5} x_i \cdot f_i}{N}=\dfrac{150}{30}=5$

Una vez que tenemos el valor de la media calcularemos la diferencia de cada elemento con respecto a este valor y calcularemos la media aritmética de estas diferencias. Nos ayudaremos de una tabla similar a la anterior donde añadiremos la columna $x_i - \overline{x}$.

Calificaciones	x_i	f_i	$x_i - \overline{x}$	$f_i \cdot (x_i - \overline{x})$
[0,2)	1	7	1-5=-4	-28
[2,4)	3	6	3-5=-2	-12

[4,6)	5	4	0	0
[6,8)	7	6	2	12
[8,10]	9	7	4	28
		N=30		Suma=0

Puedes observar que se compensan las sumas positivas con las negativas. Esto va a suceder siempre así y nos indicaría que a desviación siempre sería cero, cosa que no es cierta. Para resolver esta situación elevaremos al cuadrado esta diferencia. En matemáticas realizar esta operación tiene que ver con el cálculo de la distancia a un punto, y esto tiene cierta lógica, pues estamos calculando la distancia de los datos a la media. Ahora ya no se compensan los positivos con los negativos, ya que todos los valores serán positivos y, por lo tanto, ya no se anularán al sumarlos. Veamos como quedaría la tabla:

Calificaciones	x_i	f_i	$(x_i - \bar{x})^2$	$f_i \cdot (x_i - \bar{x})^2$
[0,2)	1	7	16	112
[2,4)	3	6	4	28
[4,6)	5	4	0	0
[6,8)	7	6	4	28
[8,10]	9	7	16	112
		N=30		Suma=280

La desviación será la media aritmética de todas las diferencias y vendrá dada por la suma de todos los $f_i \cdot (x_i - \bar{x})^2$, celda de color verde, dividida por el número total de elementos (N=30).

$$\frac{\sum_{i=1}^{5} f_i \cdot (x_i - \bar{x})^2}{\sum_{i=1}^{5} f_i} = \frac{280}{30} = 9'33$$

Importante

La **varianza** de una distribución estadística es la media aritmética de los cuadrados de las desviaciones respecto de la media. Se repreesenta por $\mathbf{s^2}$. (Recuerda que si la varianza es referida a una población en vez de a una muestra se representa por σ^2)

$$s^2 = \frac{\sum_{i=1}^{n} f_i \cdot (x_i - \bar{x})^2}{\sum_{i=1}^{n} f_i} = \frac{\sum_{i=1}^{n} f_i \cdot x_i^2}{\sum_{i=1}^{n} f_i} - \bar{x}^2$$

Cuanto más pequeña sea la varianza de una distribución mayor es el grado de representatividad de la media.

Ejercicio resuelto

Calcula la varianza de las calificaciones del curso de 2ºB. Recuerda la distribución de las notas:

2º Bachillerato B	
Calificaciones	Nº de Alumnos
[0,2)	2

[0,2)	2
[2,4)	7
[4,6)	12
[6,8)	7
[8,10]	2

¿En cuál de los dos cursos la media es más representiva?

Imagen de edans con Licencia Creative Commons

Mostrar retroalimentación

Para calcular la varianza de 2º B lo primero que hay que vamos a hacer es construir la tabla de frecuencias con las columnas que nos interesan. Para el cálculo de la media usaremos los valores $x_i \cdot f_i$. Para la varianza usaremos las columnas x_i^2 y $f_i \cdot x_i^2$ ya que vamos a aplicar la segunda parte de la fórmula de la varianza:

$$s^2 = \frac{\sum_{i=1}^{n} f_i \cdot x_i^2}{\sum_{i=1}^{n} f_i} - \overline{x}^2$$

Calificaciones	Marcas de clase(x_i)	Nº de alumnos(f_i)	$x_i \cdot f_i$	x_i^2	$f_i \cdot x_i^2$
[0,2)	1	2	2	1	2
[2,4)	3	7	21	9	63
[4,6)	5	12	60	25	300
[6,8)	7	7	49	49	343
[8,10]	9	2	18	81	162
		N=30	150		Suma=870

La media será: $\overline{x} = \frac{\sum_{i=1}^{5} x_i \cdot f_i}{N} = \frac{150}{30} = 5$

La varianza vendrá dada por: $s^2 = \frac{\sum_{i=1}^{5} f_i \cdot x_i^2}{\sum_{i=1}^{5} f_i} - \overline{x}^2 = \frac{870}{30} - 5^2 = 29 - 25 = 4$

En este caso la varianza es 4, mientras que en el primer caso la varianza de 2º A era 9'33. Por lo tanto, la varianza de las calificaciones es bastante menor en 2º B, por lo que los resultados en esta clase están más representados por el valor $\overline{x} = 5$.

Un inconveniente de la varianza es que no tiene las mismas unidades que los datos. Si estamos midiendo, por ejemplo, el número de hijos en familias de un determinado lugar, la varianza nos dará el resultado en número de hijos al cuadrado, mientras que su raíz cuadrada vendrá medida en número de hijos. Por ello se usa la raíz cuadrada de ese parámetro. Este nuevo valor tiene las mismas unidades que los datos, lo que facilita su interpretación y le da un sentido.

Importante

Se llama **desviación típica** de una variable estadística a la raíz cuadrada positiva de la varianza y se representa por **s** (recuerda que la desviación típica referida a una población se representa por σ).

La desviación típica es el medidor de la dispersión más usado en Estadística Descriptiva. Al igual que ocurre con la varianza, cuanto menor sea este valor más representativa será la media.

Reflexiona

En la siguiente tabla se muestran los resultados del estudio sobre el peso en alumnos de 2º de Bachillerato que se está llevando a cabo en nuestro centro "Benito V."

Peso en Kg.	[50,56)	[56,62)	[62,68)	[68,74)	[74,80)	[80,86]
Nº de Alumnos	12	25	25	20	12	6

Calcula la media, la varianza y la desviación típica de la distribución.

Mostrar retroalimentación

En el siguiente vídeo tienes la solución a la media:

En este puedes ver la solución de la varianza y la desviación típica:

Para saber más

Los ordenadores nos permiten calcular de forma más rápida las tablas de frecuencia que usamos habitualmente en este tema. Las hojas de cálculo van a ser las encargadas de realizar este trabajo. En este vídeo elaborado por Laureano Serrano te mostramos como calcular **la media** y la **desviación típica** con ayuda del programa **Calc de OpenOffice**.

2.2. Otras medidas de dispersión

Imagen del colegio "Benito V."

Además de las medidas de dispersión anteriores, existen otras tres que también son muy utilizadas.

Una de las medidas que se suele calcular en primer lugar cuando hacemos un estudio de los datos, es la diferencia entre los datos mayor y menor de una muestra. La diferencia de estos datos nos indica la amplitud del intervalo al que van a pertenecer todos los datos.

Según sea esta diferencia, pequeña o grande teniendo como referencia el objeto de estudio, podremos deducir directamente la posible dispersión de los datos.

Además, si los datos pertenecen a una variable continua nos va a ayudar a decidir la amplitud de los intervalos en los que podemos subdividir la muestra.

Importante

Se llama **rango** o **recorrido** de una distribución a la diferencia entre el mayor y el menor valor de la variable estadística.

El problema que tiene esta medida es que si existe un valor muy grande o muy pequeño alejado del resto de los datos, su valor se dispara y no expresa la realidad de la distribución. Por eso en algunas ocasiones suele usarse el **rango intercuartílico** que es la diferencia entre el tercer y el primer cuartil (Q_3 - Q_1) para evitar estos desajustes.

¿Recuerdas del punto anterior, el ejemplo de las calificaciones de la asignatura de matemáticas, cuando calculábamos la desviación de los datos respecto de la media y nos salía cero? Lo solucionábamos elevando esas diferencias al cuadrado. Otra posible solución es calcular el valor absoluto de esta diferencia. Veamos como se hace ayudándonos de la tabla de frecuencias:

Calificaciones	x_i	f_i	$\lvert x_i - \bar{x} \rvert$	$f_i \cdot \lvert x_i - \bar{x} \rvert$
[0,2)	1	7	4	28
[2,4)	3	6	2	12

[4,6)	5	4	0	0
[6,8)	7	6	2	12
[8,10]	9	7	4	28
		N=30		Suma=80

La desviación media vendrá dada por la suma de las diferencias absolutas de todos los valores con respecto a la media, celda de color rosa, dividido por el número total de elementos (N=30).

$$\frac{\sum_{i=1}^{5} f_i \cdot |x_i - \bar{x}|}{\sum_{i=1}^{5} f_i} = \frac{80}{30} = 2'67$$

Importante

Se lama **desviación media** de una variable estadística, y se representa por $D_{\bar{x}}$, a la media aritmética de los valores absolutos de las desviaciones respecto de la media.

$$D_{\bar{x}} = \frac{\sum_{i=1}^{n} f_i \cdot |x_i - \bar{x}|}{\sum_{i=1}^{n} f_i}$$

Reflexiona

El número de hijos de 10 familias , seleccionadas aleatoriamente, es el siguiente.

5, 2, 0, 1, 1, 2, 1, 3, 3, 1

Calcula el rango y la desviación media de la distribución.

Imagen de Oneras con licencia Creative Commons

Mostrar retroalimentación

En primer lugar, calculamos el rango:

R=5-0=5.

Por lo tanto, el rango es 5 hijos.

Para la desviación media nos ayudamos de la tabla de frecuencias con las columnas que nos interesan:

x_i	f_i	$x_i \cdot f_i$	$\|x_i - \bar{x}\|$	$f_i \cdot \|x_i - \bar{x}\|$
0	1	1	2	2
1	4	4	1	4
2	2	4	0	0
3	2	6	1	2

5	1	5	3	3
	N=30	Suma=20		Suma=11

Calculamos en primer lugar la media:

$$\overline{x}=\frac{\sum_{i=1}^{5} x_i \cdot f_i}{N}=\frac{20}{10}=2$$

Por lo que la desviación media es igual a

$$D_{\overline{x}}=\frac{\sum_{i=1}^{5} f_i \cdot |x_i-\overline{x}|}{N}=\frac{11}{10}=1'1$$

En nuestro ejemplo de las calificaciones de matemáticas, vamos a intentar responder a la siguiente pregunta: ¿En cuál de los dos grupos de matemáticas el resultado es más homogéneo?

Recuerda los datos que ya hemos calculado: la media y la desviación típica. A partir de ellos vamos a calcular un nuevo parámetro que no dependa de ninguna unidad.

$$\overline{x}_{2A}=5,\ s_{2A}=3'05$$
$$\overline{x}_{2B}=5,\ s_{2B}=2$$

En este caso dividiremos la desviación típica entre la media:

En la clase de 2ºA el resultado sería: $\frac{s_{2A}}{\overline{x}_{2A}}=\frac{3{,}05}{5}=0'61$. Mientras que en 2ºB quedaría: $\frac{s_{2B}}{\overline{x}_{2B}}=\frac{2}{5}=0'4$.

Podemos ver que el cociente es menor en el grupo de 2ºB.

Cuando se quiere comparar el grado de dispersión de dos distribuciones que no vienen dadas en las mismas unidades o que las medias no son iguales se utiliza el coeficiente de variación de Pearson (CV) que se define como el cociente entre la desviación típica y la media aritmética.

CV representa el número de veces que la desviación típica contiene a la media aritmética y por lo tanto cuanto mayor es CV mayor es la dispersión y menor la representatividad de la media.

Importante

Se llama **coeficiente de variación**, y se representa por **CV**, al cociente entre la desviación típica y la media aritmética.

$$CV=\frac{s}{\overline{x}}$$

Ejercicio resuelto

Las calificaciones de Blanca y Gonzalo en los exámenes de esta unidad son las siguientes:

Calificaciones de Blanca: 4, 5, 6, 6, 7, 8.

Calificaciones de Gonzalo: 5, 6, 7, 7, 8, 9.

¿Cuál de los dos alumnos tiene sus calificaciones más concentradas?

Mostrar retroalimentación

Tendremos que calcular el coeficiente de variación en ambos casos. Lo primero que hacemos es calcular la media de los dos:

$\overline{x}_B = 6$

$\overline{x}_G = 7$

Para la desviación típica calculamos la varianza y hallamos su raíz cuadrada:

$s_B = \sqrt{2}$

$s_G = \sqrt{2}$

Para el coeficiente de variación de Blanca tenemos que: $CV = \frac{s_B}{\overline{x}_B} = \frac{\sqrt{2}}{6} = 0'24$, mientras que el coeficiente de variación de Gonzalo es $CV = \frac{s_G}{\overline{x}_G} = \frac{\sqrt{2}}{7} = 0'20$. Por lo que las calificaciones de Gonzalo están más concentrados que las de Blanca

Resumen

Importante

Se llama **Media Muestral** (Media aritmética) de una variable estadística al cociente entre la suma de todos los valores obtenidos en la muestra y el tamaño de la muestra. La media muestral se representa por $\overline{x}$. Recuerda que la media poblacional se representa por la letra griega **μ**

Si X es una variable estadística que toma los valores x_1, x_2, x_3, ... x_n con frecuencias absolutas f_1, f_2, f_3, ... f_n , respectivamente, la media de la variable X viene dada por la siguiente expresión:

$$\overline{x}=\frac{x_1\cdot f_1+x_2\cdot f_2+...+x_n\cdot f_n}{f_1+f_2+...+f_n}=\frac{\sum_{i=1}^{n}x_i\cdot f_i}{\sum_{i=1}^{n}f_i}=\frac{\sum_{i=1}^{n}x_i\cdot f_i}{N}$$

Importante

Se llama **moda** de una variable estadística al valor de la variable que presenta mayor frecuencia absoluta. La moda se representa por $\mathbf{M_0}$.

Importante

Se llama **mediana** de una variable estadística al valor de la variable tal que el número de observaciones menores que él es igual al número de observaciones mayores que él. La mediana de una variable estadística se represente por $\mathbf{M_e}$.

Importante

La **varianza** de una distribución estadística es la media aritmética de los cuadrados de las desviaciones respecto de la media. Se repreesenta por $\mathbf{s^2}$. (Recuerda que si la varianza es referida a una población en vez de a una muestra se representa por σ^2)

$$s^2 = \frac{\sum_{i=1}^{n} f_i \cdot (x_i - \overline{x})^2}{\sum_{i=1}^{n} f_i} = \frac{\sum_{i=1}^{n} f_i \cdot x_i^2}{\sum_{i=1}^{n} f_i} - \overline{x}^2$$

Cuanto más pequeña sea la varianza de una distribución mayor es el grado de representatividad de la media.

Importante

Se llama **desviación típica** de una variable estadística a la raíz cuadrada positiva de la varianza y se representa por **s** (recuerda que la desviación típica referida a una población se representa por σ).

La desviación típica es el medidor de la dispersión más usado en Estadística Descriptiva. Al igual que ocurre con la varianza, cuanto menor sea este valor más representativa será la media.

Importante

Se llama **coeficiente de variación**, y se representa por **CV**, al cociente entre la desviación típica y la media aritmética.

$$CV = \frac{s}{\overline{x}}$$