MD.PlantillaTexto(04)Esp.dot

ESTADÍSTICA DESCRIPTIVA UNIDIMENSIONAL

UD000572_V(02)

ÍNDICE

MOTIVACIÓN

Numerosos estudios que se publican y que aparecen en televisión han tenido que realizarse mediante lo que se denomina estadística descriptiva. Gracias a esta ciencia podemos conocer, por ejemplo, cuál es el comportamiento medio de una población según el estudio de una variable.

En una encuesta de opinión, por ejemplo, es necesario realizar medias, varianzas, etc., para conocer cómo se comporta una población determinada ante la variable a estudiar.

Las medias, las varianzas, etc., forman parte de la Estadística Descriptiva Unidimensional que vamos a estudiar en esta unidad didáctica.

Además, a través de los gráficos, es posible ver en un golpe de vista el comportamiento de las variables.

PROPÓSITOS

Con el estudio de esta unidad, serás capaz de:

- Conocer los conceptos generales de una variable estadística.
- Construir tablas de frecuencias y tipos de representación de muestras.
- Establecer los parámetros de centralización en variables aleatorias unidimensionales.
- Interpretar los parámetros de dispersión en variables aleatorias unidimensionales.

PREPARACIÓN PARA LA UNIDAD

En esta unidad didáctica introduciremos los conceptos generales pertenecientes a la estadística (población, muestra, representaciones, parámetros, etc.). Si bien se trata de una unidad con ejercicios bastante mecánicos, los contenidos presentan una gran cantidad de fórmulas que el alumno deberá comprender antes de aplicarlas. El uso de tablas facilitará la labor de cálculo en la resolución de los problemas.

Es conveniente que anotes en una cartulina las fórmulas que aparecen aquí. Te serán de gran utilidad para la realización de los ejercicios.

1. CONCEPTOS BÁSICOS

La estadística es la rama de las matemáticas que proporciona métodos para el tratamiento sistemático de los datos, para inferir conclusiones de los mismos y tomar decisiones razonadas tras su análisis.

Existen varios métodos estadísticos que permiten analizar las variables que se pretenden estudiar. Son los siguientes:

1. **Estadística descriptiva:** se encarga de la recogida de datos, y del recuento, la tabulación y el diseño de gráficos.
2. **Análisis estadístico:** se sintetiza la información calculando las medidas estadísticas para resumirla (centralización, dispersión, etc.).
3. **Inferencia:** se obtienen conclusiones del trabajo realizado. Las conclusiones se valoran probabilísticamente y se indican los límites en que pueden oscilar dichos valores.

En estadística es importante conocer y diferenciar los conceptos de población y muestra:

- **Población:** conjunto de elementos que tiene una o más características comunes. Puede ser finito o infinito.
- **Muestra:** subconjunto representativo de la población, es decir, conjunto de individuos de una población al cual se quiere estudiar una determinada característica.

Nota

Si muestra y población coinciden, tendremos un **censo.**

Para seleccionar una muestra representativa han de respetarse criterios de dos tipos:

a) Criterios de carácter cualitativo: cómo se debe elegir la muestra de acuerdo con las características de la población. El tipo de muestreo a utilizar.

b) Criterios de carácter cuantitativo: tamaño de la muestra de acuerdo a las exigencias de precisión y a las características de la población.

El proceso mediante el cual se extrae una muestra representativa de la población se conoce con el nombre de muestreo aleatorio.

En el muestreo aleatorio cada individuo de la población tiene la misma probabilidad de ser incluido en la muestra. Las muestras así obtenidas se denominan muestras aleatorias.

Básico

Es importante destacar que la composición de la muestra debe estar en proporción con la composición de la población. Así, por ejemplo, si se desea elegir una muestra formada por 1.000 personas de una población en la que el 60% son mujeres, deberemos seleccionar para la muestra a 600 mujeres y a 400 hombres.

Se denomina **variable aleatoria** a una función que asigna un valor numérico a cada resultado del experimento aleatorio.

1.1. VARIABLES ESTADÍSTICAS

Las características de un individuo pueden ser diferentes, por lo que podemos clasificar las variables en:

- Variables discretas.
- Variables continuas.

Imprescindible

Los valores de las variables estadísticas suelen representarse por x_1, x_2..., en general x_i.

1.1.1. VARIABLES ESTADÍSTICAS DISCRETAS

Consideremos las siguientes variables estadísticas:

- Número de empleados de una empresa.
- Número de hijos de 50 familias.
- Número de acciones vendidas cada día en la Bolsa de Madrid durante todo un año.

Observamos que los valores posibles de estas variables son aislados. En este caso diremos que las variables estadísticas son discretas.

Atención

Una variable aleatoria es discreta cuando solo se puede tomar un número finito de valores de un conjunto, o un número infinito numerable de valores.

1.1.2. Variables estadísticas continuas

Consideremos las siguientes variables estadísticas:

- Presión sanguínea de varios enfermos.
- Diámetro de las ruedas de varios coches.
- Temperaturas registradas en un observatorio cada hora.

En todos estos ejemplos observamos que las variables son muy diferentes a las anteriores, ya que pueden tomar infinitos valores dentro de un intervalo. A estas variables las llamaremos variables estadísticas continuas.

Atención

Una variable aleatoria es la cualidad que se desea conocer en un estudio estadístico y no podemos saber a priori el valor que va a adoptar. Para hacernos una idea, una variable es cualquier cosa que se pueda medir, y que pueda tomar distintos valores.

1.2. FRECUENCIAS ABSOLUTAS Y RELATIVAS

Tras obtener los datos numéricos, necesitamos realizar el recuento y organizar la información. Para ello utilizaremos las frecuencias. A continuación, recordamos los distintos tipos de frecuencias, por ejemplo, la **frecuencia absoluta de un valor de la variable,** que es el número de veces que se repite dicho valor.

1.2.1. Frecuencia absoluta

Consideremos el siguiente ejemplo: un profesor tiene anotadas las notas de un examen de sus 30 alumnos, que son las siguientes:

5	3	4	1	2	8	9	8	7	6
6	7	9	8	7	7	1	0	1	5
9	9	8	0	8	8	8	9	5	7

Es fácil observar que hay alumnos que tienen la misma nota. Existen, por tanto, valores de la variable que se repiten; por ejemplo, ha habido 7 alumnos que han obtenido un 8 en el examen. Decimos que la frecuencia absoluta de 8 es 7.

Se llama frecuencia absoluta del valor x_i, y la representamos por f_i, al número de veces que se repite dicho valor.

Se llama frecuencia acumulada del valor x_i, y la representamos por F_i, a la suma de las frecuencias absolutas de todos los valores anteriores a x_i, más la frecuencia absoluta de x_i:

$$F_i = f_1 + f_2 + \dots + f_i$$

1.2.2. Frecuencia relativa

La frecuencia absoluta no es suficiente para reflejar la intensidad con que se repite un determinado valor de la variable estadística. Por ejemplo, no es lo mismo obtener 3 veces un 5 en 10 lanzamientos de un dado que obtenerlo en 1.000 lanzamientos del mismo.

Se llama frecuencia relativa de un valor x_i, y la representamos por f_i, al cociente entre la frecuencia absoluta de x_i y el número total de datos que intervienen en la distribución:

$$f_i = \frac{n_i}{N}$$

Siendo N el número total de datos, o tamaño de la muestra.

Se llama frecuencia relativa acumulada del valor x_i, y se representa por f_i, al cociente entre la frecuencia absoluta acumulada de x_i y el número total de datos que intervienen en la distribución:

$$f_i = \frac{n_i}{N} = f_1 + f_2 + \ldots + f_i$$

Refuerzo

Recuerda lo que significa cada letra:

- x_i: variables.
- n: número de datos de la muestra.
- f_i: frecuencia relativa.

1.3. TABLA DE FRECUENCIAS PARA UNA VARIABLE ESTADÍSTICA DISCRETA

En el ejemplo anterior procederemos a realizar la tabla de frecuencias.

Un profesor tiene anotadas las notas de un examen de sus 30 alumnos, que son las siguientes:

5	3	4	1	2	8	9	8	7	6
6	7	9	8	7	7	1	0	1	5
9	9	8	0	8	8	8	9	5	7

Se observa que en este caso el número total de datos que constituyen el tamaño de la muestra es 30:

x_i	n_i	N_i	f_i	F_i
0	2	2	2/30	2/30
1	3	5	3/30	5/30
2	1	6	1/30	6/30
3	1	7	1/30	7/30
4	1	8	1/30	8/30
5	3	11	3/30	11/30
6	2	13	2/30	13/30
7	5	18	5/30	18/30
8	7	25	7/30	25/30
9	5	30	5/30	1
	30		1	

1.4. TABLA DE FRECUENCIAS PARA UNA VARIABLE ESTADÍSTICA CONTINUA

Si anotamos la duración en minutos de una colección de *compact-disc,* obtenemos los siguientes resultados:

67, 63, 57, 52, 45, 49, 53, 51, 59, 66, 68, 54, 48, 40, 43, 47, 53, 54,

44, 46, 42, 47, 43, 46, 41, 49, 44, 48, 43, 47, 42, 41, 47, 42, 44, 40

Observar que el número total de datos es N = 36, luego es aconsejable que agrupemos los datos en clases.

Básico

Para determinar el número de clases que debemos formar no existe un criterio idóneo para todos los casos, si bien uno de los criterios más sencillos es el de Norcliffe, que establece que el número de clases debe ser aproximadamente igual a la raíz cuadrada positiva del número de datos.

Por lo tanto, formaremos seis clases y trabajaremos con la marca de ellas. La marca es el punto medio de cada intervalo y se obtiene sumando ambos extremos del intervalo y dividiendo el resultado entre dos.

Veamos la tabla de frecuencias:

Clases	Marcas	n_i	N_i	f_i	F_i
[40, 45)	42.5	13	13	13/36	13/36
[45, 50)	47.5	11	24	11/36	24/36
[50, 55)	52.5	6	30	6/36	30/36
[55, 60)	57.5	2	32	2/36	32/36
[60, 65)	62.5	1	33	1/36	33/36
[65, 70)	67.5	3	36	3/36	1
		36		1	

+ Info

Distribución de carácter cualitativo.

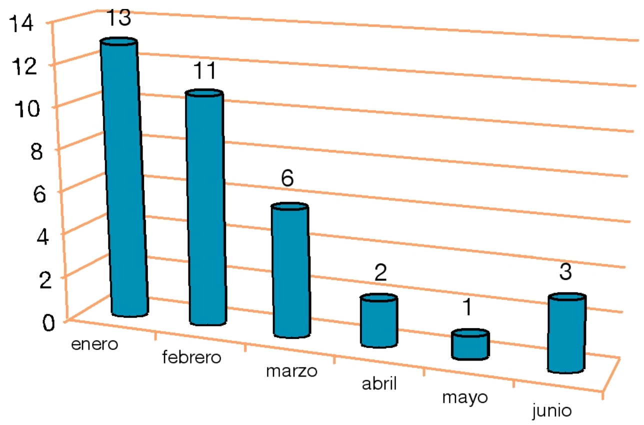
Los caracteres cualitativos no son medibles, luego no dan lugar a una variable estadística y, por tanto, no podemos calcular las frecuencias.

1.5. REPRESENTACIONES GRÁFICAS

Según la naturaleza del carácter estudiado, utilizaremos uno u otro tipo de representaciones gráficas. Las principales son las siguientes:

- **Diagrama de barras.**

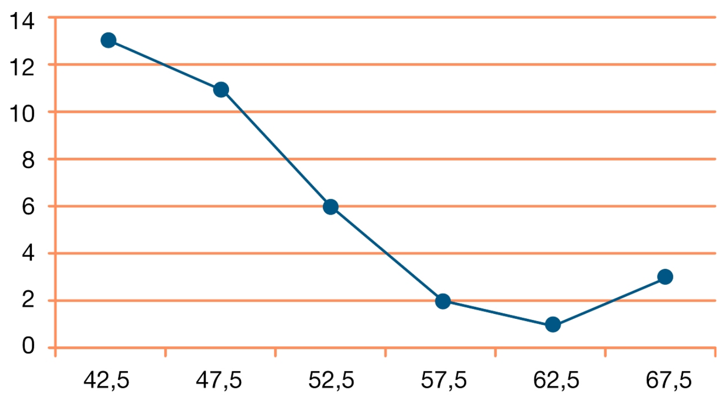
  Son muy útiles cuando queremos comparar datos cualitativos o datos cuantitativos de tipo discreto. Estos diagramas se representan en un eje de coordenadas, indicando en el eje OX los valores que toma la variable y en el eje OY las frecuencias absolutas o relativas.



- **Polígonos de frecuencias.**

  Los polígonos de frecuencias se forman uniendo los extremos de las barras mediante una línea quebrada.

  Se pueden representar polígonos de frecuencias relativas, absolutas, relativas acumuladas y absolutas acumuladas.

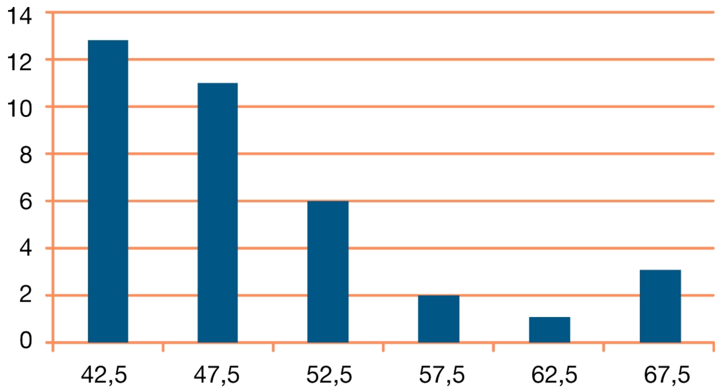


- **Histogramas.**

  Se suelen agrupar los datos obtenidos en intervalos de igual amplitud. Habitualmente se usan cuando la variable representada es continua, o si es discreta con gran cantidad de valores. En este caso, los valores se agrupan en clases o intervalos.

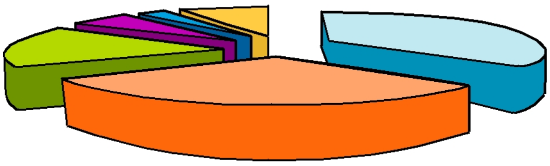
  Para construir un histograma se representan sobre el eje de abscisas los límites de las clases. Sobre dicho eje se construyen unos rectángulos que tienen como base la amplitud del intervalo y como altura, la frecuencia absoluta de cada intervalo, siempre que todos los intervalos tengan igual amplitud.

  En caso contrario, las alturas de los rectángulos han de ser calculadas teniendo en cuenta que sus áreas deben ser proporcionales a las frecuencias de cada intervalo.



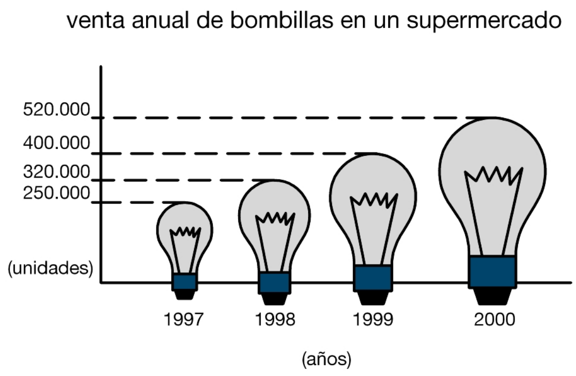
- **Diagrama de sectores.**

  En este tipo de diagrama se representan las frecuencias como sectores circulares. El ángulo central de cada sector ha de ser proporcional a la frecuencia absoluta correspondiente; en consecuencia, el área del sector circular será proporcional a la frecuencia absoluta.



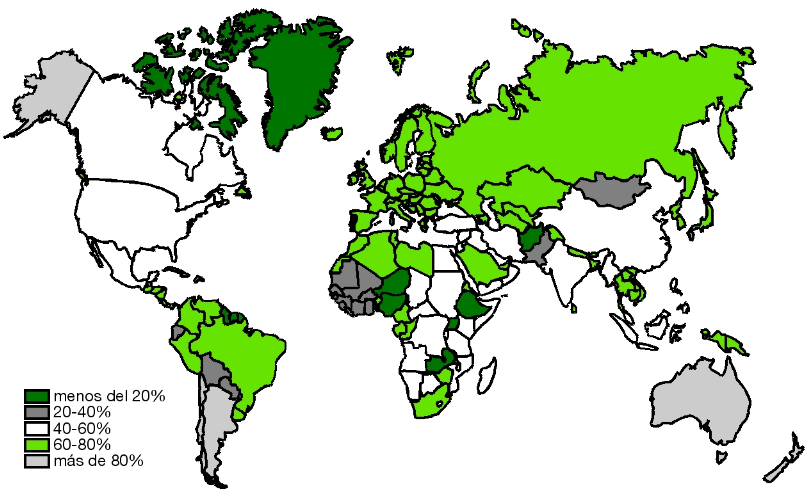
- **Pictogramas.**

  Un pictograma es un histograma donde las barras vienen representadas gráficamente por un dibujo característico de los datos que se están midiendo. Esta representación presenta cierta falta de precisión.



- **Cartogramas.**

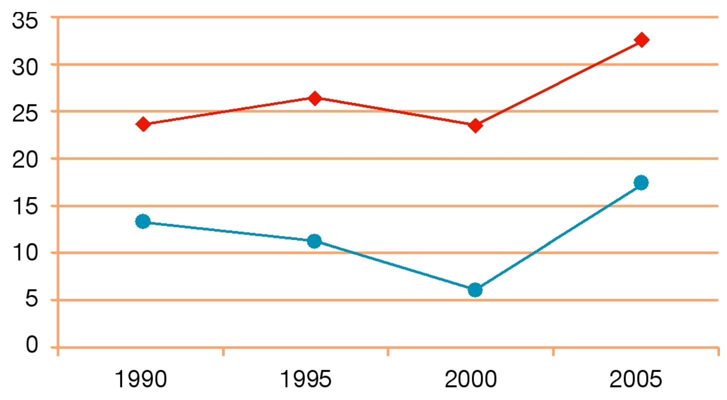
  Un cartograma es una representación gráfica de datos estadísticos que se realiza sobre un mapa. Se suelen utilizar estas representaciones para densidades demográficas, rentas per cápita, etc.



- **Diagramas lineales.**

  Se utilizan muy a menudo para mostrar las fluctuaciones de un determinado carácter estadístico con el paso del tiempo.

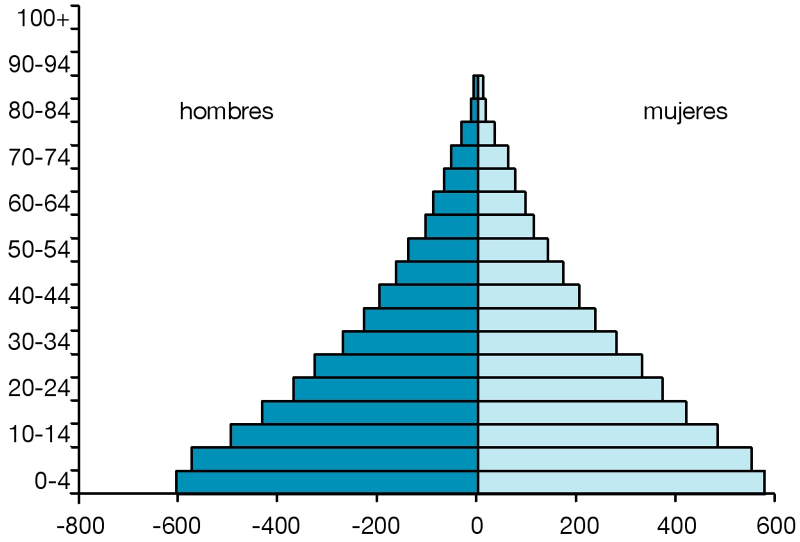
  Lo que nos interesa en este tipo de diagramas es, solamente, la altura de la línea referida a la base de este. Con frecuencia se aprovecha para representar sobre la misma escala varios diagramas lineales.



- **Pirámides de población**

Las pirámides de población se utilizan para estudiar conjuntamente la variable edad y el carácter estadístico cualitativo sexo.

La gráfica se obtiene representando en vertical el grupo de edad, y en horizontal, el número de personas por grupo y sexo. Para la modalidad mujer se toma el semieje de la derecha y para la modalidad hombre, el semieje de la izquierda.



2. MEDIDAS DE CENTRALIZACIÓN

Las tablas estadísticas y las representaciones gráficas dan una idea del comportamiento de una distribución. Sin embargo, es necesario simplificar ese conjunto de datos mediante unos valores numéricos llamados parámetros. Por tanto, el parámetro va a ser el valor numérico que caracterice la distribución.

Los parámetros o medidas de centralización tienden a situarse, en general, hacia el centro del conjunto de datos ordenados. También se les llama medidas de tendencia central o promedios.

Las medidas de centralización más importantes son la media aritmética (de tamaño), la moda (de frecuencia), la mediana (de posición) y los cuartiles, deciles y percentiles.

Recuerda

Recuerda que tienes a tu disposición un servicio de tutorías para resolver cualquier duda que te pueda surgir, bien mediante carta, Campus Virtual, fax o llamada telefónica.

2.1. MEDIA ARITMÉTICA

La media aritmética es el valor promedio de todos los valores de una colección de datos. La media aritmética de la variable X se representa por $\overline{X}$.

Básico

La media aritmética de una variable es la suma de los valores de dicha variable por sus frecuencias relativas, dividido por el número total de valores.

2.1.1. CÁLCULO DE MEDIA ARITMÉTICA

Sea X una variable estadística que toma los valores $x_1, x_2..., x_n$, con frecuencias absolutas $f_1, f_2..., f_n$, respectivamente. La media aritmética de la variable X viene dada por la siguiente expresión:

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

Si los datos son simples, las frecuencias serán $f_i = 1$ para todo $i = 1..., n$.

Si la variable X es continua, o aun siendo discreta, y por tratarse de muchos datos agrupados en clases, se toman como valores $x_1, x_2 ..., x_n$ las marcas de cada clase.

Ejemplo:

El número de horas dedicadas al estudio durante la semana de un estudiante de 2.º de Bachillerato es: 3,5, 5,5, 4, 6, 5 y 3. Calcular la media de horas estudiadas.

Por tratarse de datos simples (no repetidos), el valor de la media aritmética es el siguiente:

$$\bar{x} = \frac{3,5+5,5+4+6+5+3}{6} = 4,5$$

Caso práctico

Las calificaciones en Matemáticas de los 40 alumnos de una clase vienen dadas por la siguiente tabla:

Calificaciones	1	2	3	4	5	6	7	8	9	10
N.º de alumnos	2	2	4	5	8	9	3	4	3	0

Hallar la nota media de la clase.

Solución:

La primera fila de la tabla representa los valores x_i, mientras que la segunda corresponde a las frecuencias absolutas de cada valor.

Calcularemos la media aritmética mediante la fórmula:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i f_i}{\sum_{i=1}^{n} f_i} = \frac{212}{40} = 5,3$$

En la práctica resulta más fácil realizar los cálculos mediante una tabla que represente los valores de x_i, f_i, y los productos $x_i \cdot f_i$.

2.1.2. OBSERVACIONES A LA MEDIA ARITMÉTICA

1. La media aritmética es la medida de centralización más usada.
2. Se calcula fácilmente y tiene en cuenta todos los datos de la distribución.
3. Es una medida muy sensible; es decir, si la distribución posee valores extremos, excepcionalmente raros y poco significativos, estos alteran el valor de la media.
4. Puede ser útil eliminar los valores extremos y calcular la media aritmética con el resto de los valores. Estaríamos calculando una media aritmética truncada.
5. No siempre se puede calcular la media aritmética. En estos casos recurriremos a la mediana o a la moda.

2.2. MODA

La moda de una variable es simplemente el valor más repetido de dicha variable en el estudio estadístico, es decir, el que tiene mayor frecuencia absoluta. Se representa por Mo.

Recuerda

La moda es la medida de centralización que se define como el valor que más se repite de los valores de una variable.

La moda puede no ser única, puede haber varios valores de la variable con la mayor frecuencia. En este caso se dirá que la distribución es bimodal, trimodal, etc.

2.2.1. CÁLCULO DE LA MODA

Dependiendo de si las frecuencias son absolutas o relativas la manera de calcular la moda se realizará así:

Para variables discretas: en este caso el cálculo resulta muy sencillo, pues basta mirar en la columna de las frecuencias absolutas y ver cuál es la mayor. El valor de la variable correspondiente a dicha frecuencia será nuestra moda.

Ejemplo

Calcula la moda de la serie:

5, 7, 7, 8, 5, 5, 3, 1.

El valor de la moda es 5.

Para variables continuas: en caso de que los datos se encuentren agrupados en intervalos es fácil determinar la clase modal (clase con mayor frecuencia), pero el valor del intervalo que tiene mayor frecuencia no se conoce. Como primera aproximación podemos tomar como moda la marca de la clase modal; pero si deseamos mayor exactitud aplicaremos la siguiente expresión:

$$M_o = L_i + c \cdot \frac{D_1}{D_1 + D_2}$$

Donde:

- L_i = límite inferior de la clase modal.
- C = amplitud de los intervalos.
- D_1 = diferencia entre la frecuencia absoluta de la clase modal y la frecuencia absoluta de la clase anterior.
- D_2 = diferencia entre la frecuencia absoluta de la clase modal y la frecuencia absoluta de la clase siguiente.

Caso práctico

En el curso de 1.º de ESO de un instituto se han registrado en la siguiente tabla las calificaciones de una prueba de nivel de lengua realizada a 40 alumnos:

Calificaciones	1	2	3	4	5	6	7	8	9	10
N.º de alumnos	2	2	4	5	8	9	3	4	3	0

Calcula la moda.

Solución:

Podemos observar que $M_o = 6$, ya que es la calificación que ha obtenido un mayor número de alumnos. Esta distribución, por tener una sola moda, se llama unimodal.

2.2.2. Observaciones de la moda

1. Pueden existir distribuciones que no tengan moda esto ocurre cuando las frecuencias de todos los datos son iguales.
2. La moda es menos representativa que la media aritmética, pero en algunas ocasiones es más útil que esta; por ejemplo, cuando se trata de una distribución con datos cualitativos.
3. En la moda no intervienen todos los datos de la distribución.
4. Aunque es una medida de tendencia central, no siempre tiene por qué situarse en la zona central; es frecuente encontrar la moda próxima a los valores extremos de la distribución.

2.3. MEDIANA

La mediana de una variable estadística es el punto medio de un conjunto ordenado de observaciones de dicha variable. Es decir, el número de datos que preceden a la mediana es igual al número de datos que la siguen. La mediana se representa por M o Me.

Recuerda

La mediana es la medida de centralización que nos dice en un serie ordenada dónde se encuentra su mitad.

2.3.1. CÁLCULO DE LA MEDIANA Y EJEMPLOS

Veamos cómo se calcula la mediana según hablemos de una variable estadística discreta o continua.

En la **variable estadística discreta** se ordenan los datos de menor a mayor valor, y la mediana será el término que ocupa el lugar central.

Si el número de datos es impar, el valor central de la variable es único.

Ejemplo

La mediana de la siguiente serie estadística es:

2, 3, 5, **6** , 9, 11, 12.

Luego **M = 6,** por ser este el valor central.

Si el número de datos es par no existe término central, sino dos términos centrales. Calcularemos la semisuma de los dos valores, aunque obtengamos un valor que no pertenezca al conjunto de datos.

Ejemplo

En la serie estadística:

2, 3, 5, **6 , 9,** 11, 12, 13.

Los valores centrales son **6 y 9,** luego la mediana es:

$$M = \frac{6+9}{2} = 7{,}5$$

Reto

Realiza el siguiente ejercicio.

Consideremos la serie estadística:

169, 150, 162, 155, 157, 153, 164, 153, 170, 167, 172, 167.

¿Cuál es la mediana?

Solución:

Si ordenamos la serie de menor a mayor, tenemos:

150, 153, 153, 155, 157, **162, 164,** 167, 167, 169, 170, 172.

Al ser un número par de datos tenemos dos valores centrales 162 y 164. Calculando la media de ellos obtenemos la mediana M = 163.

La **variable estadística continua** se calcula de forma análoga a la moda. Podemos obtener fácilmente la clase mediana (intervalo donde se alcancen la mitad de los datos), pero si queremos un valor más exacto aplicaremos la siguiente expresión:

$$M = L_i + c \cdot \frac{\left(\frac{N}{2}\right) - F_{i-1}}{f_i}$$

Donde el significado de las variables implicadas es:

- L_i = Límite inferior del intervalo
- c = Amplitud del intervalo.
- N = Número total de datos.
- F_{i-1} = Frecuencia absoluta acumulada del intervalo anterior a la clase mediana.
- f_i = Frecuencia absoluta de la clase mediana.

Consideremos el ejemplo visto anteriormente referido al test de satisfacción aplicado a 88 trabajadores.

Para calcular la mediana formamos la tabla estadística con las frecuencias absolutas acumuladas F_i:

Clases	f_i	F_i
[38, 44)	7	7
[44, 50)	8	15
[50, 56)	15	30<44
[56, 62)	25	55>44
[62, 68)	18	73
[68, 74)	9	82
[74, 80)	6	88

El primer intervalo cuya frecuencia absoluta acumulada exceda la mitad del número de datos clase mediana. En este caso, la clase mediana es [56, 62).

Así pues, se tiene:

$L_i = 56$, $c = 6$, $N/2 = 44$, $F_{i-1} = 30$, y $f_i = 25$

Y sustituyendo en la expresión de la mediana:

$$M = 56 + 6 \cdot \frac{44 - 30}{25} = 59{,}36$$

2.3.2. Observaciones de la mediana

1. La mediana es muy útil cuando existe algún dato muy extremo que invalida la media, o cuando los datos están agrupados en clases y alguna de ellas es abierta.

2. La mediana es el primer parámetro de centralización que depende del orden de los datos y no de su valor. Como consecuencia de la definición, deja el 50% de los datos menores o iguales a ella a la izquierda, y el otro 50 % de los valores mayores a la mediana a la derecha.

+ Info

Relación entre moda, mediana y media.

Si al construir el polígono de frecuencias se observa que la distribución es simétrica o ligeramente asimétrica, es posible comprobar experimentalmente la siguiente relación:

Media – moda = 3· (media – mediana).

Reto

Realiza el siguiente ejercicio relacionando:

[1]Moda	Valor que se encuentra en medio en una serie ordenada[a]
[2]Mediana	Suma de valores dividido por número de valores[b]
3Media	Valor que más se repite[c]

Solución:

1 c, 2 a, 3 b.

2.4. CUANTILES

En el estudio de la mediana hemos podido observar que, una vez ordenados los datos de la distribución, aquella los divide en dos partes iguales.

Análogamente existen otros parámetros llamados cuantiles, que dividen a la distribución en función de otras cuantías. Los más importantes son los cuartiles, quintiles, deciles y percentiles.

2.4.1. CUARTILES

Los cuartiles son tres valores que dividen a la serie de datos en cuatro partes iguales. Se representan como Q_1, Q_2 y Q_3, y se llaman primer, segundo, y tercer cuartil, respectivamente.

2.4.2. QUINTILES

Los quintiles son cuatro valores que dividen a la serie de datos en cinco partes iguales. Se representan como K_1, K_2, K_3 y K_4, y se llaman primer, segundo, tercer y cuarto quintil, respectivamente.

2.4.3. DECILES

Los deciles son 9 valores que dividen a la serie de datos en 10 partes iguales. Se representan como D_1, D_2..., D_9, y se llaman primer, segundo..., y noveno decil, respectivamente.

2.4.4. PERCENTILES

Los percentiles son 99 valores que dividen a la serie de datos en 100 partes iguales. Se representan como P_1, P_2 ..., P_{99} , y se llaman primer, segundo y ..., nonagésimo noveno percentil, respectivamente.

2.4.5. OBSERVACIONES A LOS CUANTILES

1. A los cuantiles también se les llama parámetros de posición.
2. El cuartil primero coincide con el percentil de orden 25, y el cuartil tercero coincide con el percentil de orden 75.

3. MEDIDAS DE DISPERSIÓN

Una vez calculada la media de una distribución de datos, podemos estudiar las medidas de dispersión de dicha distribución. Estas medidas nos permiten analizar lo agrupados o no que se encuentran los valores de la distribución. Así, con estos datos, podemos concluir si existen diferencias importantes entre los valores de la distribución y la media de dichos valores.

Básico

Las medidas de dispersión solo tienen sentido para variables cuantitativas.

A continuación, veremos las medidas de dispersión más importantes: el rango o recorrido, la desviación media, la varianza y la desviación típica.

3.1. RANGO O RECORRIDO

El rango es la diferencia entre el valor más alto y el más bajo que toma la variable. Por ejemplo, si los valores de una variable son 5, 3, 1, 4, 6, 3 y 2, el rango es la diferencia entre el valor más alto, 6, y el más bajo, 1, luego el rango es 6 – 1 = 5.

3.1.1. OBSERVACIONES AL RANGO

1. Como ya hemos visto en el ejemplo, el rango o recorrido es un parámetro que se calcula de forma sencilla, sin utilizar complejas expresiones matemáticas.

2. Si el rango es pequeño, los valores centrales de la distribución tienen gran importancia, puesto que será más fácil que estos se agrupen a su alrededor.

3. Dado que el rango es la diferencia entre el valor más alto y el más bajo que toma la variable, solo depende de los valores extremos que toma, lo cual representa un gran inconveniente puesto que basta que uno de los extremos se aleje mucho del punto en el que se concentran el resto para que el rango se vea afectado.

4. El inconveniente que acabamos de plantear podemos corregirlo utilizando solo unos valores concretos de la distribución, esto es, lo que hace el rango intercuartílico, el cual indica el 50% de los valores que se encuentran en la mitad de la distribución de los mismos, basándose en los valores intermedios y no en los extremos. Se calcula como:

Rango intercuartílico: $Q = Q_3 – Q_1$

Donde Q_1 es el punto en la distribución por debajo del cual queda el 25% de los casos. Se conoce como primer cuartil. Q_3 es el punto por debajo del cual queda el 75% de los casos. Se conoce con el nombre de segundo cuartil.

Otro parámetro que podemos utilizar es el rango entre percentiles:

$P = P_{90} – P_{10}$

3.2. DESVIACIONES RESPECTO A LA MEDIA

El valor de la media de una distribución podríamos interpretarlo como el valor esperado de dicha variable. Así, podríamos suponer que cualquier otro valor de la distribución sería "esperable" que estuviese cerca de este.

Sin embargo, esto no es así, puesto que como sabemos, si un valor de la distribución se aleja mucho del resto, esto afectará al valor de la media. Por ejemplo, si tenemos los datos 2, 3, 2, 2,5 y 12, la media será 4,3, que se aleja de los valores 2 y 3, cerca de los que se encuentran cuatro de los cinco valores.

Así, es interesante estudiar, además de la media, las desviaciones de cada valor de la variable respecto a esta. Las desviaciones respecto a la media se calculan como $d_i = x_i - \bar{x}$, y hay tantas como valores toma la variable.

Nota

La desviación respecto de la media no tiene una única medida sino tantas como valores tenga la variable.

3.2.1. OBSERVACIONES

1. Las desviaciones respecto a la media son útiles para estudiar la proximidad de cada valor que toma la media a este parámetro.
2. Así, las desviaciones respecto a la media pueden ser positivas, si el valor de la variable es mayor que la media, negativas, si el valor de la variable es menor que la media, o nulas, si el valor de la variable coincide con la media.
3. En consecuencia, por la propia definición de las desviaciones respecto a la media, la suma de estas respecto a la media de todos los valores que toma la variable es cero. Por tanto, este parámetro no es útil para estudiar la dispersión. Para solucionar este problema introduciremos más adelante dos parámetros nuevos: la desviación media y la varianza.

3.3. DESVIACIÓN MEDIA

Como acabamos de comentar, las desviaciones respecto a la media suman cero, por lo que no nos son útiles como medida de dispersión. Sin embargo, si utilizamos los valores absolutos de las desviaciones respecto a la media, obtendremos un valor positivo a partir del cual podremos calcular un nuevo parámetro, la desviación media.

3.3.1. DEFINICIÓN DE DESVIACIÓN MEDIA

Definimos la desviación media como el valor medio de los valores absolutos de las desviaciones respecto a la media.

3.3.2. CÁLCULO DE LA DESVIACIÓN MEDIA

Para calcular la desviación media utilizaremos la siguiente expresión:

$$D_{\bar{x}} = \frac{f_1 \cdot |x_1 - \bar{x}| + f_2 \cdot |x_2 - \bar{x}| + \ldots + f_n \cdot |x_n - \bar{x}|}{f_1 + f_2 + \ldots + f_n}$$

Siendo x_1, x_2..., x_n, los valores que toma la variable estadística y f_1, f_2..., f_n, las frecuencias absolutas de dichos valores.

Reto

Realiza el siguiente ejercicio.

Una desviación puede ser:

a) Positiva.
b) Negativa.
c) Nula.
d) Todas son correctas.

Solución:

Una desviación puede ser positiva, negativa o nula.

3.4. VARIANZA Y DESVIACIÓN TÍPICA

Otra forma de subsanar el inconveniente que plantean las desviaciones respecto a la media si suman cero, es elevar al cuadrado estos valores. Los nuevos valores serán la base para el cálculo de la varianza y la desviación típica.

3.4.1. CÁLCULO DE LA VARIANZA Y LA DESVIACIÓN TÍPICA

Definiremos la varianza como el valor medio de los cuadrados de las desviaciones respecto a la media de los valores que toma la variable estadística. Se calculará usando la siguiente expresión:

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

Donde x_1, x_2..., x_n, son los valores que toma la variable estadística y f_1, f_2..., f_n, las frecuencias absolutas de dichos valores.

La media aritmética $(\bar{x})$ se calcula como:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{x_1 + x_2 + \ldots + x_n}{n}$$

O equivalentemente:

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

A partir de la definición de la varianza podemos definir la desviación típica, ya que para calcular este nuevo parámetro basta hacer la raíz cuadrada de la varianza. Así, la expresión que nos servirá para realizar este cálculo será:

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

O equivalentemente:

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

Reto

Realiza el siguiente ejercicio.

¿Cómo se denomina a la media de los cuadrados de las desviaciones?

Solución:

Varianza.

3.4.2. EJERCICIOS RESUELTOS

Emplearemos los mismos ejemplos que en el apartado dedicado a las medidas de centralización.

El número de horas dedicadas al estudio durante una semana de un estudiante de 2.º de Bachillerato son 3,5, 5,5, 4, 6, 5 y 3. Calcular el rango, la varianza y la desviación típica.

Ya calculamos anteriormente la media $\bar{x} = 4,5$

Observar:

- Rango: 6 – 3 = 3.
- Varianza: $s^2 = 7/6 = 1,16$.
- Desviación típica: $s = \sqrt{1,16} = 1,08$

Si aplicamos un test de satisfacción en el trabajo a 88 empleados de una empresa, obtenemos los siguientes resultados:

Puntuaciones	Número de trabajadores
[38, 44)	7
[44, 50)	8
[50, 56)	15
[56, 62)	25
[62, 68)	18
[68, 74)	9
[74, 80)	6

Calcula el rango, la varianza, la desviación típica y la desviación media.

Rango: 80 – 38 = 42

Sabemos que $\bar{x} = 59,14$

Completaremos la tabla anterior con las marcas de cada clase y los productos correspondientes:

Clases	Marcas de clase x_i	f_i	$x_i \cdot fi$	$x_i{\cdot}^2 \cdot f_i$	$\vert x_i \cdot$ $\vert$	f_i (x_i-)
[38, 44)	41	7	287	11.767	18,14	126,98
[44, 50)	47	8	376	17.672	12,14	97,12
[50, 56)	53	15	795	42.135	6,14	92,1
[56, 62)	59	25	1.475	87.025	0,14	3,5
[62, 68)	65	18	1.170	76.050	5,86	105,48
[68, 74)	71	9	639	45.369	11,86	106,74
[74, 80)	77	6	462	35.574	17,86	107,16
		88	5.204	315.592		639,08

Varianza:

$$s^2 = 315.592/88 - (59{,}14)^2 = 88{,}73$$

Desviación típica:

$$s = \sqrt{88{,}73} = 9{,}4$$

Desviación media:

$$D_{media} = 639{,}08/88 = 7{,}26$$

3.4.3. OBSERVACIONES A LA VARIANZA

1. Estos dos nuevos parámetros dependen de los valores de la distribución y del valor de la media.
2. Dado que el cálculo de la varianza utiliza los valores de las desviaciones respecto a la media elevadas al cuadrado, este parámetro no viene expresado en las mismas unidades que la variable estadística estudiada. Así, por ejemplo, si la variable estadística se mide en metros, la unidad en la que se mide la varianza es el metro cuadrado. En consecuencia, la desviación típica sí se medirá en las mismas unidades que la variable estudiada puesto que se calcula como la raíz cuadrada de la varianza. Por esta razón, la desviación típica es un parámetro más interesante a la hora de extraer conclusiones del estudio estadístico.

+ Info

Puedes ver en el siguiente vídeo una explicación de la forma de cálculo del parámetro estadístico varianza.

(Tienes el vídeo en el Campus virtual).

CONCLUSIONES

En esta unidad didáctica hemos visto cómo se puede estudiar una variable unidimensional a través de las medidas de centralización y dispersión. Gracias a estos indicadores se pueden obtener conclusiones sobre el comportamiento que tiene la variable estudiada.

¿Serías capaz de conocer cuál es el comportamiento de una variable gracias a las medidas de centralización y dispersión?

RECAPITULACIÓN

- **Medidas de centralización.**
  - De tamaño: media aritmética.
  - De frecuencia: moda.
  - De posición:
    - Mediana.
    - Cuantiles.
- **Medidas de dispersión.**
  - Rango.
  - Rango intercuartílico.
  - Desviaciones respecto a la media.
  - Desviación media.
  - Varianza.
  - Desviación típica.

AUTOCOMPROBACIÓN

1. **La varianza verifica lo siguiente:**

   a) No viene afectada por cambios de escala.

   b) Viene afectada por cambios de origen.

   c) Viene afectada por cambios de origen y escala.

   d) Depende del valor de la media.

2. **¿Qué tipo de representación gráfica emplea rectángulos y se utiliza para comparar datos cuantitativos agrupados en clases?**

   a) Diagramas de barras.

   b) Diagramas lineales.

   c) Histogramas.

   d) Todas las anteriores menos una.

3. **Indica la afirmación verdadera respecto de la media aritmética:**

   a) Es única pero no siempre existe.

   b) Si cambiamos las unidades de la variable no varía.

   c) Si suprimimos un dato no varía.

   d) Tiene en cuenta todos los datos de la distribución.

4. **Señala la propiedad falsa para la varianza y la desviación típica:**

   a) Se usan todos los datos para sus cálculos.

   b) Los valores extremos influyen mucho en ellas.

   c) Las dos se expresan en las mismas unidades que la variable.

   d) Junto con la media nos dan una idea del lugar en el que se encuentran la mayoría de los datos.

5. **La principal medida de posición no central es:**

   a) El error cuadrático medio.

   b) Los cuartiles.

   c) La media.

   d) La mediana.

6. **¿Cuál de las siguientes respuestas no es válida refiriéndose a la mediana?**

   a) Los valores extremos no influyen en ella.

   b) No se puede usar para comparar variables medidas en unidades distintas.

   c) No es tan sencilla de calcular como la media.

   d) No depende del orden de los datos.

7. **Referente a la moda, es falso que:**

   a) Es única y existe siempre.

   b) Los valores extremos apenas influyen en ella.

   c) No se usan todos los datos para calcularla.

   d) Si unimos varios conjuntos de datos la moda no es, en general, la moda ponderada del conjunto de las modas.

8. **La marca de clase de un intervalo es:**

   a) La diferencia entre los extremos del intervalo.

   b) Cualquier punto del intervalo.

   c) El punto del intervalo que más se repite.

   d) El punto medio del intervalo.

9. ¿Cuál de las siguientes gráficas no se utiliza para variables discretas con muy pocos datos?

a) El histograma.

b) El diagrama de barras.

c) Los gráficos por sectores.

d) El diagrama lineal.

10. Si queremos representar el índice de evolución del IPC en España, lo más recomendable es utilizar:

a) Un sector circular.

b) Un polígono de frecuencias absolutas.

c) Un diagrama lineal.

d) Un diagrama de barras.

SOLUCIONARIO

1.	d	**2.**	c	**3.**	d	**4.**	c	**5.**	b
6.	d	**7.**	a	**8.**	d	**9.**	a	**10.**	c

PROPUESTAS DE AMPLIACIÓN

Existen más medidas de centralización y dispersión, además de las que te acabamos de indicar en esta unidad didáctica.

Dado el número que hay y su importancia, hemos decidido indicar las más comunes y útiles para los diferentes análisis estadísticos que se pueden realizar.

No obstante, si deseas más información, puedes acudir a páginas web especializadas en estadística descriptiva

Te proponemos que compruebes si tienes materiales complementarios, o clases grabadas dentro de la unidad. Si es así, descárgalos para ampliar la información sobre el tema y **recuerda marcar he terminado.**

Te proponemos también que entres **en la sección de agenda** y compruebes qué clases en directo y/o talleres tienes disponibles, para complementar tus estudios, o tu preparación a la hora de afrontar los exámenes.

BIBLIOGRAFÍA

- VV.AA. EULER. *Matemáticas I*. Madrid: S.M, 2001.
- VV.AA. EULER. *Matemáticas II* aplicadas a las ciencias sociales. Madrid: S.M, 2000.
- COLERA, J y otros. *Matemáticas. En tus manos.* Madrid: Anaya, 2002.
- BESCÓS, E y PEÑA, Z**.** *Proyecto Exedra. Matemáticas I.* Madrid: Oxford, 2001.
- VV.AA. *Matemáticas I.* Madrid: Edelvives, 2003.
- VV.AA. *Matemáticas II.* Madrid: Edelvives, 2003.