



Estadística: Estadística unidimensional

Matemáticas I

1.º Bachillerato

Contenidos

Estadística

Estadística unidimensional

1. Introducción

Actualmente existen multitud de datos sobre cualquier asunto del que queramos tratar. Es muy común escuchar y leer en los medios de comunicación un sinfín de estudios de universidades y organismos totalmente desconocidas, sobre los temas más variopintos.

[Enlace a recurso reproducible >> https://www.youtube.com/embed/iDraWUqQYIk](https://www.youtube.com/embed/iDraWUqQYIk)

Vídeo del Ministerio de Cultura y Deporte alojado en [Youtube](#)

En nuestra vida cotidiana solemos echar mano de la estadística con cierta frecuencia para apoyar nuestros argumentos. En el siguiente vídeo del siempre ácido e inteligente Dr. House, tenemos una buena muestra.

[Enlace a recurso reproducible >> https://www.youtube.com/embed/NVXnbwxQH_M](https://www.youtube.com/embed/NVXnbwxQH_M)

Vídeo de Pilar Gallego alojado en Youtube

Pero en este tema veremos que la Estadística es bastante más que eso.



Curiosidad

-

Esta rama de las matemáticas se remonta a la antigüedad. Se dice que la **Estadística** o **Ciencia del Estado** es tan antigua como la escritura. Surge del interés de los poderes públicos por el conocimiento de los recursos de los que se disponían, ya fueran humanos o materiales.



Los romanos, maestros de la organización política, fueron quienes mejor supieron sacarle partido a la estadística. Cada cinco años realizaban un censo de la población y sus empleados públicos tenían la obligación de anotar nacimientos, defunciones y matrimonios, sin olvidar los recuentos periódicos del ganado y de las riquezas contenidas en las tierras conquistadas.

Imagen en [artelista](#). Licencia [CC](#)

Uno de estos empadronamientos, del que seguro has oído hablar, es el que se realizaba cuando Jesucristo nació.

2. Conceptos básicos

Siguen las altas temperaturas en el centro y el sur

tiempo.flv



Es un hecho que en verano hace calor.

Si nos fijamos bien, cuando el hombre del tiempo habla de las temperaturas mínimas nos está dando datos recogidos durante la noche, es decir, reales, y sin embargo, cuando habla de las alertas nos está dando datos de una predicción. Nos presenta el tiempo, por tanto, mezclando datos reales con predicciones.



Importante

La **estadística** es la rama de las matemáticas que se encarga de **recolectar** y **organizar** datos con el objeto de **inferir** conclusiones sobre ellos.

Para realizar el estudio de las temperaturas mínimas que aparecía en el video del tiempo, primero se ha decidido qué ciudades eran las que se iban a mostrar, el conjunto de todas ellas es lo que se considera **población** del estudio, y cada una de las ciudades se les denomina **individuo** del estudio.



Importante

Población: conjunto de elementos o individuos sobre los que queremos recabar datos.

Individuo: cada elemento de la población.



Imagen de foxspain en [Pixabay](#), CC by 2.0

El proceso de toma de información acerca de los individuos de una población puede realizarse mediante la elaboración de un censo o mediante la extracción de una muestra.

Censo: recoger la información deseada de cada uno de los individuos de la población, por ejemplo, los realizados en las ciudades. No siempre es posible realizar censos para cualquier consulta ya que sería muy costoso.

Muestreo: recoger la información deseada de una parte de la población llamada **muestra**. Evidentemente, es fundamental seleccionar adecuadamente la muestra para que la información sea comparable a la obtenida mediante el censo.



Importante

La **Estadística** se divide en dos ramas:

Estadística inferencial o inductiva: trata las condiciones bajo las cuales las conclusiones que se obtienen de analizar la información muestral, son válidas.

Estadística descriptiva o deductiva: trata solamente de describir y analizar un grupo dado de datos sin sacar conclusiones acerca de la población que los ha generado.

Durante esta unidad nos centraremos básicamente en la estadística descriptiva.

Siguiendo con el ejemplo del tiempo, una vez escogida la muestra de las capitales de provincia, el siguiente paso es plantearse las cuestiones que nos interesan sobre el tiempo en esas ciudades.

Por ejemplo, ¿qué temperatura mínima han tenido?, ¿qué tiempo hace?,...

Todas estas cuestiones que describen aspectos del tiempo en las ciudades escogidas en la muestra, es lo que se denominan **variables estadísticas**.

Si pensamos en las posibles respuestas de las variables estadísticas relacionadas con el tiempo, que pusimos anteriormente como ejemplos, nos encontramos:

Variable	Respuestas
¿Qué temperatura mínima han tenido?	menos de 15°, 15°, 16°, 17°, 18°, 19°, 20°, más de 20°
¿Qué tiempo hizo?	Nublado, lluvioso, soleado, ventoso

A cada una de las **posibles** respuestas de una variable estadística se les llama **modalidad**. Estas modalidades tendrán que estar definidas sin ambigüedad, de manera que cada individuo pueda representar una y sólo una de las modalidades de cada variable.



Importante

Llamaremos **variable estadística** a cada una de las características consideradas con el propósito de describir a cada individuo de la muestra escogida.

Según el tipo de modalidades, las variables se pueden clasificar en:

Variables cualitativas: son aquellas cuyas modalidades corresponden a cualidades, y no pueden expresarse con valores numéricos. Por ejemplo, las que corresponden a la variable ¿qué tiempo hizo?

Variables cuantitativas: son aquellas cuyas modalidades pueden expresarse mediante valores numéricos. Un buen ejemplo es el de la variable ¿qué temperatura mínima han tenido?

Las variables cuantitativas, a su vez, se clasifican en:

Variables discretas: toman valores aislados.

Variables continuas: toman cualquier valor en un intervalo numérico.



Ejercicio Resuelto

En el siguiente vídeo puedes practicar los tipos de variables que hemos visto.

[Enlace a recurso reproducible >> https://www.youtube.com/embed/wonmKS4Blk](https://www.youtube.com/embed/wonmKS4Blk)

Vídeo de matemóvil alojado en [Youtube](#)

3. Tablas de frecuencias



Imagen de Héctor Cen en [Pixabay](#). Licencia [by-nc-sa-2.0](#)



Imagen de Héctor Cen en [Pixabay](#). Licencia [by-nc-sa-2.0](#)

De una cantidad de datos sin orden, ¿podemos obtener alguna información?

Parece lógico pensar que ordenándolos y posteriormente encontrando semejanzas podamos agruparlos y obtener, de esta forma, una información que pueda describirnos la población escogida.

La manera habitual de trabajar es colocar los datos en tablas que indiquen el número de veces que se repite cada modalidad, estas tablas se llaman **tablas de frecuencias**.

Ahora bien no es lo mismo que una modalidad se repita 5 veces en un total de 50 datos que en un total de 500. Por tanto, hablaremos de frecuencias **absoluta** y **relativa**.

Por ejemplo, en una de las encuestas que aparecía en la autoevaluación del apartado anterior, la que se preguntaba por el presupuesto en euros que se disponía para Semana Santa, tuvo un total de 605 repuestas. En la siguiente tabla aparecen las frecuencias absolutas y relativas que tuvieron cada una de las respuestas posibles.

Variable	frecuencia absoluta	frecuencia relativa	frecuencia relativa (en %)
0	206	0,34	34
menos de 100	139	0,23	23
entre 100 y 300	139	0,18	18
más de 300	151	0,25	25
TOTAL	605	1	100



Importante



La **frecuencia absoluta** es el número de veces que se repite una modalidad de una variable en un estudio estadístico. Se suele representar por n_i .

La **frecuencia relativa** es el cociente entre la frecuencia absoluta (n_i) y el número total de datos (N). Se representa por f_i .

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

La suma de todas las frecuencias absolutas es igual al número total de datos (N).

Y por tanto, la suma de todas las frecuencias relativas es 1.

Por último, añadir que multiplicando las frecuencias relativas por 100, obtenemos el **tanto por ciento** de esa modalidad en la muestra.

<https://www.geogebra.org/material/iframe/id/x6fpFwHA/width/756/height/373/border/888888>

En multitud de ocasiones nos interesa agrupar los datos de manera que sin perder información nos sean más útiles.

Volviendo al ejemplo del tiempo, es interesante conocer en cuántas provincias se puede dormir en verano, es decir, cuántas están por debajo del umbral del sueño.

Cogiendo los datos del video de la página anterior, que se refieren a las temperaturas de una sola noche, obtenemos la siguiente tabla:

Temperaturas	n_i	N_i
15°C	1	1
16°C	3	4
17°C	4	8
18°C	2	10
19°C	3	13



[Panneau-dormir](#) de Liquid_2003

con licencia [cc-by-sa-3.0](#)

20°C	1	14
21°C	5	19
22°C	6	25
23°C	2	27
24°C	2	29

Sabiendo que para poder conciliar el sueño necesitamos una temperatura inferior a 22°C, de la tabla anterior se deduce que sólo en 19 de las provincias estudiadas han dormido a pierna suelta.

La tercera columna de la tabla hace referencia al concepto de **frecuencia absoluta acumulada**.



Importante

La **frecuencia absoluta acumulada** es el resultado de sumar la frecuencia absoluta de una modalidad de la variable con todas las frecuencias de las modalidades anteriores.

$$N_i = n_1 + n_2 + \dots + n_i$$

En consecuencia, las **frecuencias relativas acumuladas** representa la proporción de individuos de una muestra que presentan alguna de las i primeras modalidades.

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

La *última* frecuencia absoluta acumulada es igual al número total de datos.

Y por tanto, la *última* frecuencia relativa acumulada es igual a 1.

Cuando la **variable es cuantitativa continua** se utiliza la técnica de construir **intervalos de clase**, en los cuales se agruparán los datos observados en una muestra. Así, cada intervalo es considerado como una modalidad, tomándose como frecuencia absoluta de cada modalidad el número de observaciones agrupadas en el intervalo correspondiente.

Una vez contruidos los intervalos de clase, se elige un representante de cada uno de ellos, llamado **marca de clase**, que normalmente es el **punto medio del intervalo**.

Pero veamos todo esto con el siguiente ejercicio resuelto en el que construiremos una tabla de frecuencias para una variable cuantitativa continua.



Caso práctico

En una biblioteca se ha realizado una encuesta entre los usuarios de los libros que se han leído en el último mes:

4,1,3,10,5,2,2,5,1,19,8,3,5,15,2,1,1,1,6,3,2,12,3,7,6,3,4,1,10,7,11,6,7,12,4,2,8,5,9,3,6,8,2,1,12,9,8,

Elaboraremos una tabla de frecuencias agrupando los valores en intervalos.

Elegiremos los intervalos de clases:

La modalidad de interés son los libros que leen los usuarios de esta biblioteca que van entre 1 libro y 19 libros. Podemos realizar intervalos de 5 en 5. Quedando:

Intervalos de clase

Libros leídos

[0,5)

[5,10)

[10,15)

[15, 20)

Calcularemos las marcas de clases: recordamos que son el punto medio del intervalo

Intervalos de clase Marca de clase

[0,5) 2,5

[5,10) 7,5

[10,15) 12,5

[15, 20) 17,5

Construiremos el resto de la tabla de frecuencias

Intervalos de clase	Marca de clase	n_i	f_i	N_i	F_i
[0,5)	2,5	32	0,427	32	0,427
[5,10)	7,5	32	0,427	64	0,854
[10,15)	12,5	8	0,106	72	0,960
[15, 20)	17,5	3	0,040	75	1



Comprueba lo aprendido

Ya sabes de la importancia que tiene en la actualidad seguir aprendiendo a lo largo de toda la vida. Podríamos afirmar que la formación continua está calando hondo cada vez más en nuestra sociedad.

Para justificar lo que acabamos de decir, observa los datos que aparecen en la siguiente tabla extraída de la [Encuesta sobre la Participación de la Población Adulta en las Actividades de Aprendizaje >> Documento de descarga](#) (EADA) 2007, realizada por el Instituto Nacional de Estadística (INE), en la población española de edad comprendida entre 25 y 75 años.

Algunos datos aparecen en blanco. ¿Podrías completarlos?

Tramos de edad (en años)	Marca de clase	Número de personas que participan en actividades formativas, n_i	Frecuencia absoluta acumulada, N_i	Frecuencia relativa acumulada, F_i
[25, 35)	30	3.018.026	3.018.026	0,37
[35, 45)	<input type="text"/>	2.433.843	<input type="text"/>	<input type="text"/>
[45, 55)	50	1.624.953	7.076.822	0,87
[55, 65)	<input type="text"/>	<input type="text"/>	7.882.197	0,96
[65, <input type="text"/>)	70	298.697	<input type="text"/>	<input type="text"/>
TOTAL		8.180.894		

No es tan complicado, sólo es cuestión de sumar y restar. El dato que falta en la columna de frecuencias relativas acumuladas, está redondeado a la segunda cifra decimal.



Curiosidad

La frecuencia relativa es la proporción de datos en los que aparece una modalidad. En cierto modo, se corresponde con la *probabilidad* de que se dé esa modalidad. Esto es lo que se conoce como *enfoque frecuentista de la probabilidad*.

[Enlace a recurso reproducible >> http://www.youtube.com/embed/bfHN2oS2KBw](http://www.youtube.com/embed/bfHN2oS2KBw)

Como has podido ver C3PO, el androide más famoso del cine, es un entusiasta de los datos y les insinúa a sus compañeros que sólo 1 nave de 3750, ha conseguido pasar el campo de asteroides, es decir, hace un enfoque frecuentista de la probabilidad.

4. Gráficos

Siempre se ha dicho : "*Vale más una imagen que mil palabras*". Y en estadística también es cierto. Los datos que hemos ordenado antes en tablas también los podemos representar gráficamente, para ver de una forma más visual la información.

A continuación, se hará un recorrido por los gráficos estadísticos más usuales, a través de ejemplos.

Representemos gráficamente el ejemplo que vimos con GeoGebra del dado. Al ver el gráfico, podemos intuir que las frecuencias de cada resultado tienden a ser iguales.

<https://www.geogebra.org/material/iframe/id/nez4Aw7/width/756/height/373/border/888888>



Importante

Un **diagrama de barras** es una representación gráfica de ejes coordenados donde a cada modalidad se le asigna una barra y la altura de cada barra es la frecuencia absoluta de la modalidad.

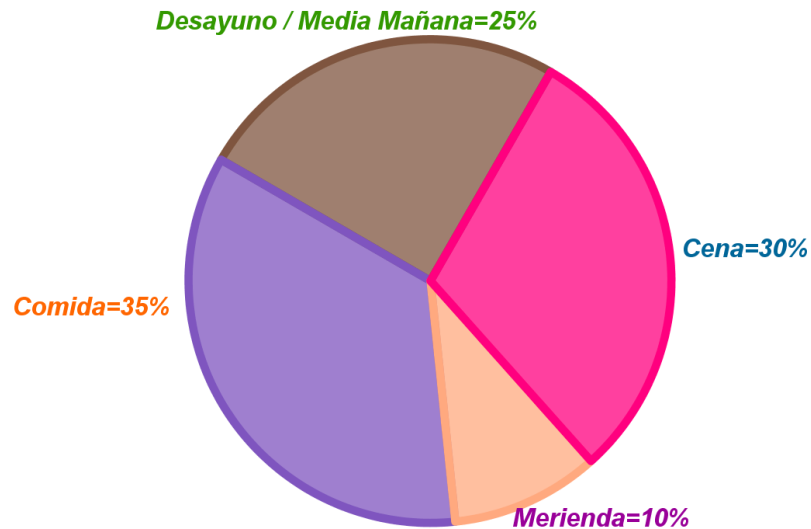
Un **diagrama de sectores** es una representación gráfica circular donde a cada modalidad le corresponde un sector circular de amplitud proporcional a la frecuencia.

La amplitud en grados se calcula con la fórmula $\frac{360}{N} \cdot n_i$

Estos dos tipos de gráficos se utilizan para variables *cualitativas* y *cuantitativas discretas*.

Para ilustrar el concepto de diagrama de sectores se ha escogido un ejemplo sobre nutrición. Donde también se han calculado los porcentajes que representa cada modalidad.

Aportes de nutrientes de cada comida del día



Comprueba lo aprendido

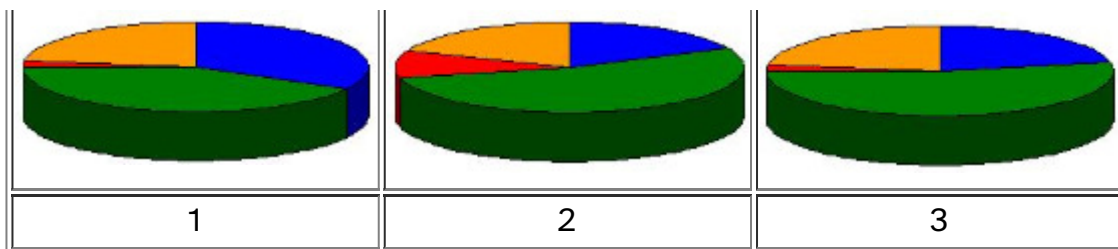
En una encuesta se pregunta a los lectores sobre qué harían para reducir las muertes por violencia de género.

En la siguiente tabla aparecen las frecuencias relativas, en tanto por ciento, de las respuestas que dieron los lectores.

Respuestas	Frecuencia relativa en tanto por ciento
Publicar las listas de maltratadores	22
Promover la educación en igualdad	53
Aumentar el presupuesto	2
Introducir cambios en la ley	23

Indica qué número de los siguientes diagramas de sectores pertenece a la tabla anterior:

--	--	--



Además, escribe el color que corresponde a las siguientes opciones: a "Promover la educación en igualdad" le pertenece el y a "Aumentar el presupuesto" el .



Importante

El **histograma** es una representación gráfica de ejes coordenados mediante barras de una variable *cuantitativa continua* cuyos valores se han agrupado en intervalos. La base de cada barra se corresponde con la amplitud del intervalo y la altura, con la frecuencia absoluta, si las bases son iguales. En caso de que las amplitudes de los intervalos sean distintas se calcula la densidad.

Uniando, mediante segmentos de recta, los puntos medios de las bases superiores de cada rectángulo del histograma y del diagrama de barras, se obtiene la representación gráfica llamada **polígono de frecuencias**.



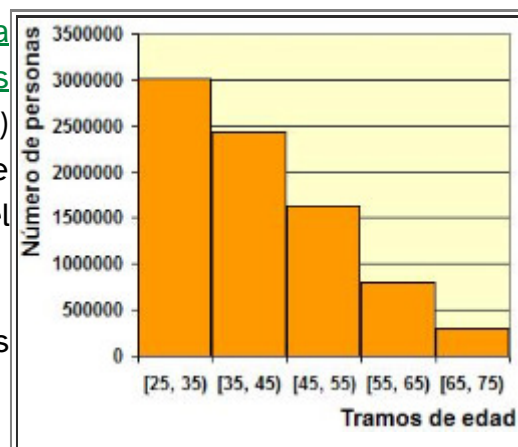
Reflexiona

Volvemos a hablar de la [Encuesta sobre la Participación de la Población Adulta en las Actividades de Aprendizaje](#) >> [Documento de descarga](#) (EADA) 2007, realizada por el Instituto Nacional de Estadística (INE), que ya mencionamos en el apartado anterior.

El gráfico de la derecha corresponde a los datos que aparecen en dicha encuesta.

Contesta a las siguientes preguntas.

a) ¿Qué tipo de gráfico estadístico es?



- b) ¿Cómo interpretas las alturas que van teniendo los rectángulos al aumentar los tramos de edad?
- c) ¿La altura de los rectángulos indican frecuencias absolutas o relativas?
- d) Si representáramos el polígono de frecuencia, ¿qué tipo de línea nos aparecería dibujada?

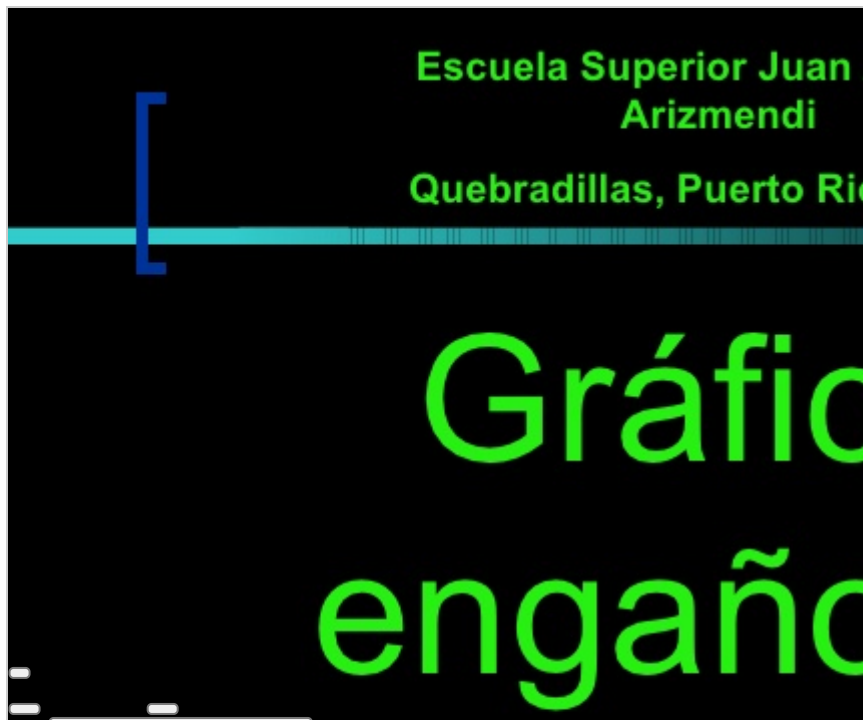
Seguro que tus respuestas son muy parecidas a las siguientes.

- a) Es un histograma.
- b) A más edad, menor número de personas que participan en actividades de formación.
- c) Frecuencias absolutas.
- d) Casi una línea recta que descende.

Para terminar el tema, no podemos olvidar que es fundamental estar atentos y ser críticos con la información que recibimos. Y ser mucho más cautos si esta información nos llega en formato gráfico, ya que son muchas las ocasiones en que se intentan tergiversar los datos representados.

Lo puedes comprobar en esta presentación:

http://www.slideshare.net/slideshow/embed_code/32932



[Gráficas Engañosas](#) from [jaa_math](#)



Curiosidad

Cuando se realizan representaciones correspondientes a edades de población, cambiamos el eje Y por el eje X para obtener las llamadas **pirámides de población**.

Por tanto, se trata de dos histogramas horizontales.

Además de los vistos hasta ahora existen otros muchos tipos de gráficos que sirven para representar información, como los **pictogramas** que son dibujos alusivos al carácter que se está estudiando y cuyo tamaño es proporcional a la frecuencia que representan y los **cartogramas** que se realizan sobre mapas, en los que aparecen indicados sobre las distintas zonas cantidades o colores de acuerdo con el carácter que representan.

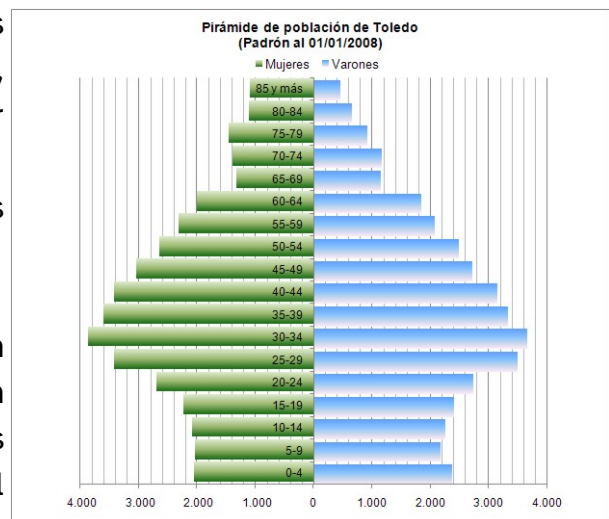
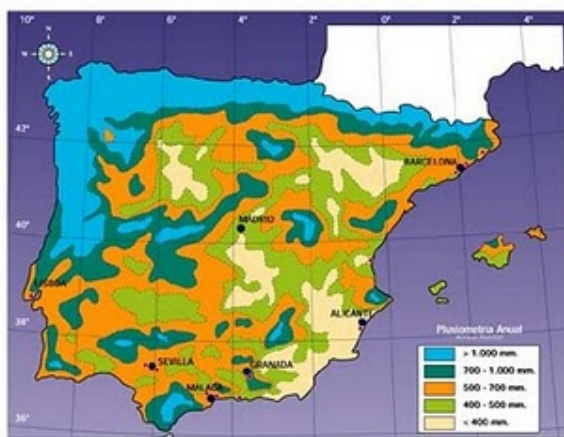
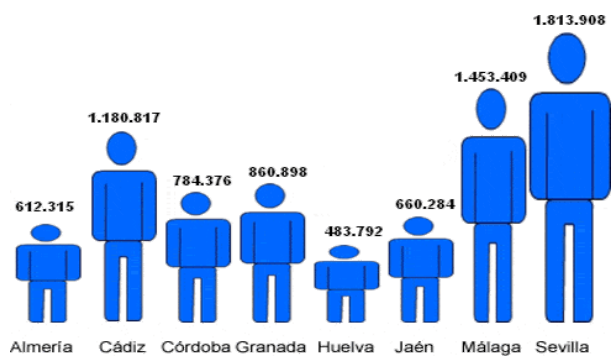


Imagen en [Wikimedia Commons](#). Licencia [GNU Free Documentation License](#)



Resumen



Importante

La **estadística** es la rama de las matemáticas que se encarga de **recolectar** y **organizar** datos con el objeto de **inferir** conclusiones sobre ellos.

Población: conjunto de elementos o individuos sobre los que queremos recabar datos.

Individuo: cada elemento de la población.



Importante

Llamaremos **variable estadística** a cada una de las características consideradas con el propósito de describir a cada individuo de la muestra escogida.

A cada una de las **posibles** respuestas de una variable estadística se les llama **modalidad**.

Según el tipo de modalidades, las variables se pueden clasificar en:

Variables cualitativas: son aquellas cuyas modalidades corresponden a cualidades, y no pueden expresarse con valores numéricos. Por ejemplo, las que corresponden a la variable ¿qué tiempo hizo?

Variables cuantitativas: son aquellas cuyas modalidades pueden expresarse mediante valores numéricos. Un buen ejemplo es el de la variable ¿qué temperatura mínima han tenido?

Las variables cuantitativas, a su vez, se clasifican en:

Variables discretas: toman valores aislados.

Variables continuas: toman cualquier valor en un intervalo numérico.



Importante

La **frecuencia absoluta** es el número de veces que se repite una modalidad de una variable en un estudio estadístico. Se suele representar por n_i .

La **frecuencia relativa** es el cociente entre la frecuencia absoluta (n_i) y el número total de datos (N). Se representa por f_i .

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

La suma de todas las frecuencias absolutas es igual al número total de datos (N).

Y por tanto, la suma de todas las frecuencias relativas es 1.

Por último, añadir que multiplicando las frecuencias relativas por 100, obtenemos el **tanto por ciento** de esa modalidad en la muestra.



Importante

La **frecuencia absoluta acumulada** es el resultado de sumar la frecuencia absoluta de una modalidad de la variable con todas las frecuencias de las modalidades anteriores.

$$N_i = n_1 + n_2 + \dots + n_i$$

En consecuencia, las **frecuencias relativas acumuladas** representa la proporción de individuos de una muestra que presentan alguna de las i primeras modalidades.

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

La *última* frecuencia absoluta acumulada es igual al número total de datos.

Y por tanto, la *última* frecuencia relativa acumulada es igual a 1.

Cuando la variable es **cuantitativa continua** se utiliza la técnica de construir **intervalos de clase**, en los cuales se agruparán los datos observados en una muestra. Así, cada intervalo es considerado como una modalidad, tomándose como frecuencia absoluta de cada modalidad el número de observaciones agrupadas en el intervalo correspondiente.

Una vez contruidos los intervalos de clase, se elige un representante de cada uno de ellos, llamado **marca de clase**, que normalmente es el **punto medio del intervalo**.



Importante

Un **diagrama de barras** es una representación gráfica de ejes coordenados donde a cada modalidad se le asigna una barra y la altura de cada barra es la frecuencia

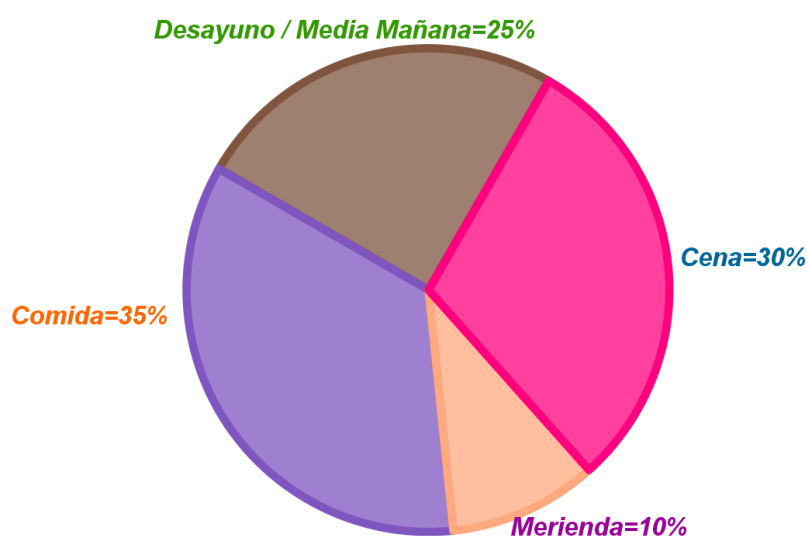
absoluta de la modalidad.

Un **diagrama de sectores** es una representación gráfica circular donde a cada modalidad le corresponde un sector circular de amplitud proporcional a la frecuencia.

La amplitud en grados se calcula con la fórmula $\frac{360}{N} \cdot n_i$

Estos dos tipos de gráficos se utilizan para variables *cualitativas* y *cuantitativas discretas*.

Aportes de nutrientes de cada comida del día



Importante

El **histograma** es una representación gráfica de ejes coordenados mediante barras de una variable *cuantitativa continua* cuyos valores se han agrupado en intervalos. La base de cada barra se corresponde con la amplitud del intervalo y la altura, con la frecuencia absoluta, si las bases son iguales. En caso de que las amplitudes de los intervalos sean distintas se calcula la densidad.

Uniando, mediante segmentos de recta, los puntos medios de las bases superiores de cada rectángulo del histograma y del diagrama de barras, se obtiene la representación gráfica llamada **polígono de frecuencias**.

Aviso legal

Las páginas externas no se muestran en la versión imprimible

<http://www.juntadeandalucia.es/educacion/permanente/materiales/index.php?aviso#space>