

# Somos lo que comemos. Las personas y la salud: Haciendo estudios estadísticos

---



INSTITUTO de ENSEÑANZAS a DISTANCIA de ANDALUCÍA

## ESPAD Nivel II Ámbito Científico Tecnológico Contenidos

### Somos lo que comemos. Las personas y la salud: Haciendo estudios estadísticos

#### Sedentarismo, hábitos alimenticios y obesidad

Según la **Encuesta Nacional de Salud del año 2017** el **INE(Instituto Nacional de Estadística)**, en los grupos de edad de 18 a 64 años, un 44,3% de hombres y un 30,0% de mujeres padecen sobrepeso y un 18,2% de hombres de 18 y más años y un 16,7% de mujeres padecen obesidad.

En 2017 se redujo el volumen de compra de frutas frescas para el hogar un 3,5%, la reducción en valor fue en menor medida (-0,9%), debido al aumento del precio medio del 2,7%, situándose en 1,46 €/kg. (Informe Consumo Alimentario del Ministerio Agricultura, Pesca y Alimentación).

Ahora plantéate lo siguiente:

- ¿Has entendido la información?
- ¿Cómo sabe el INE y el Ministerio de Agricultura todo eso si a ti ni a mí nos han preguntado? ¿A quiénes preguntan?
- ¿Qué cuentas han hecho los del INE para saber toda esa información?

En este tema **aprenderás cosas relacionadas con la Estadística**, una parte de las Matemáticas que se ocupa de estudiar **grandes cantidades de datos** sobre un tema concreto. Verás desde cómo se hace una **encuesta** hasta cómo hacer **cuentas** con los **datos** y sacar **conclusiones**, todo ello sobre hábitos de alimentación saludable.

¿Te parece interesante? Pues entonces, adelante, que la puerta está abierta...



Fotografía en flickr de [jmalдона](#). Licencia [cc](#)

# 1. ¿Cómo se empieza un estudio estadístico como el de la nota de prensa?

---



Imagen en [pixabay](#) de [Merio](#). Dominio Público

Debemos tener en cuenta:

- ¿Qué queremos saber?
- ¿De quién queremos saber la información?
- ¿Cómo obtenemos los datos?

## 1.1. Lo primero es decidir qué quiero saber



Imagen  
en MEC -ITE . Licencia cc

Por ejemplo, según el informe 'Estado de situación sobre el desayuno en España' presentado por la Fundación Española de Nutrición (FEN) en 2018, "el 85% de la población desayuna habitualmente, el 11% lo hace de forma irregular y un 5% directamente no suele desayunar. Entre estos últimos, destacan los adolescentes –más entre la población femenina– y los adultos – más entre la población masculina–."

Sin embargo el desayuno es algo fundamental para nuestra salud.

Nosotros **queremos estudiar el tiempo dedicado a desayunar**, ya que es difícil hacer un desayuno adecuado con prisas.

*Importante*

### Empezando a adquirir vocabulario

A la característica o cualidad que queremos estudiar la llamaremos **variable estadística**

*Ejercicio resuelto*

### Otros ejemplos de variable estadística

El "tiempo dedicado a desayunar" sería una variable estadística. Te ofrecemos otros ejemplos:

- La marca de cereales para el desayuno.
- El número de horas de sueño.
- El color de los ojos.
- La estatura.
- El número de libros leídos el último mes.

**Pero no todas las variables estadísticas son iguales.** Las hay básicamente de dos clases, según el tipo de datos que estudiemos:

**Variables cuantitativas:** son aquellas que pueden medirse numéricamente. Por ejemplo, "el tiempo dedicado a desayunar", (que puede expresarse en número de minutos).

Piensa ahora en qué diferencia hay entre medir el número de veces que se comen legumbres a la semana y el peso. ¿A que uno no come 3'4567 veces legumbres? ¡o come 3 veces o 4! Pero sí podemos pesar 69'658 kg. Por eso se distingue entre variables cuantitativas continuas y discretas:

- **Variable discreta:** cuando no puede tomar ningún valor entre dos valores consecutivos. Un ejemplo sería el número de hijos. ¡Tenemos 1, 2 o 3...pero nunca 1'7!
- **Variable continua:** cuando puede tomar cualquier valor dentro de un intervalo, así que debemos **agruparlas en intervalos**. Por ejemplo, como son demasiadas las posibles respuestas si preguntamos a muchas personas su peso, agrupamos las respuestas en intervalos: El número de personas que pesan "de 50 hasta 60 kilogramos", "de 60 hasta 70 kilogramos", etc.



Variable cuantitativa discreta: N° hijos

Fotografía en Flickr de [Jacinta Iluch valero](#). Licencia [cc](#)



Variable cuantitativa continua:  
Peso en kg

Fotografía en Flickr de [Hernando](#). Licencia [cc](#)

**Variables cualitativas:** no todas las cosas que podríamos estudiar estadísticamente pueden expresarse con números. En algunos casos se trata de cualidades no medibles numéricamente. Por ejemplo, "la marca de cereales para el desayuno" se expresará de forma no numérica, con el nombre de la marca ("Estadifibra", "Matechoco", etc).



Variable Cualitativa: Marca de cereales  
Fotografía en Flickr de [Amor8](#). Licencia [cc](#)

*Importante*

### Recuerda:

Hay variables estadísticas **cuantitativas** (que pueden ser discretas o continuas) y **cualitativas**.

*Actividad de lectura*

- a. De la lista de variables que has visto en los ejemplos del "Ejercicio resuelto", señala aquellas que son cuantitativas e indica cuáles de ellas son continuas.
- b. Escribe un ejemplo, distinto a los vistos, de variable cualitativa y otro de variable cuantitativa. Pero que estén relacionadas con el desayuno.

#### **Mostrar retroalimentación**

- a. Son cuantitativas el número de horas de sueño, la estatura y el número de libros leídos el último mes. De estas variables, sólo la estatura es de tipo continuo.
- b. Aunque podríamos idear muchos ejemplos, como variable cuantitativa podemos escoger el número de calorías consumidas en el desayuno, y como variable cualitativa el tipo de café que nos gusta (sólo, con leche, cortado, capuchino, etc).

## 1.2. ¿De quién quiero saber la información?

Lo segundo que debemos decidir es **¿de quién o qué queremos saber la información?** Es decir ¿a qué personas vamos a preguntar o en qué objetos vamos a medir la característica que queremos estudiar?

*Importante*



Fotografía en Flickr de [Frado76](#). Licencia [cc](#)

- El grupo de personas u objetos (individuos) en el que vamos a estudiar la variable estadística se llama **población (P)**.
- El número de individuos de la población se llama **tamaño poblacional (N)**.

*Ejercicio resuelto*

### En nuestra nota de prensa...

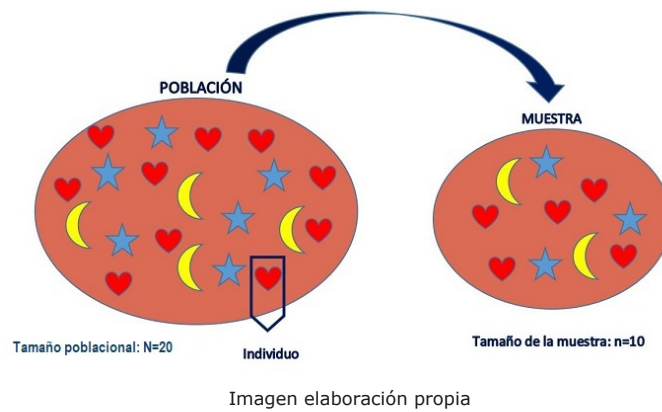
En el caso de nuestra nota de prensa, la **población P** es el **conjunto de todos los españoles** y **N** es el **número de habitantes de España**.

Pero... ¿A que a ti no te preguntó nada, nadie del INE? Ni a nosotros tampoco. Entonces... ¿cómo se obtuvo la información? Imagina el coste en tiempo, dinero y personal que supondría preguntar a todos los españoles sobre sus hábitos... ¡sería gigantesco!, por eso, lo que se hace es **seleccionar una muestra** que sea representativa de toda la población.

*Importante*

- Una **muestra** es una **parte de la población** sobre la que estudiaremos la variable estadística

- Estadística.
- El número de individuos de la muestra es el **tamaño muestral** o de la muestra, **n**.



El objetivo es **extender las conclusiones que se obtengan con la muestra a TODA la población**. Así, preguntando a un grupo de españoles, en vez de a todos, el INE extrae conclusiones sobre la población española total (y se ahorra una pasta...).

Pero pensemos un poco...

## Actividad de lectura

### Autoevaluación

- ¿Deberíamos tener cuidado con el número de individuos de la muestra?
- ¿Influirán las características de los individuos de la muestra en las conclusiones finales?

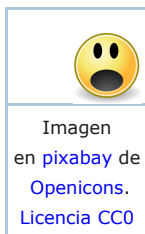
### Mostrar retroalimentación

- Claro que hay que tener cuidado con el número. Si seleccionamos demasiados no tendremos ninguna ventaja... ya puestos sería mejor tomar toda la población. Y si son demasiado pocos... quizá no sean representativos de la población y los datos no informen de lo que realmente ocurre.
- Claro que sí, no es lo mismo preguntar hábitos de alimentación a deportistas que a diabéticos, por ejemplo.

Por tanto, **la selección de una muestra adecuada es fundamental** si queremos que lo que estudiemos en ella pueda extenderse a toda la población.

## Curiosidad

### Curiosidad



**Si tuvieras que escoger una muestra de toda la población de Cádiz para averiguar, con un margen de error pequeño, qué tanto por ciento de gaditanos comen legumbres al menos dos veces por semana, ¿A cuántas personas diferentes le preguntarías? ¿a 1000?, ¿a 5000?, ¿a 10000?...**

¡Ten en cuenta que la población de Cádiz a 1 de enero de 2019 era según el INE de 116.027 habitantes!

¿Quieres sorprenderte?, pues sigue leyendo...

**¡Bastaría con preguntarle a 383 personas!** y eso con un error de tal solo el 5%, es decir, que estamos casi seguros en un 95% (No podemos estar seguros al 100%, ya que no vamos a preguntarles a todos los gaditanos). Increíble, ¿verdad? Eso sí, bien escogidos.

Pero lo más curioso es que si Cádiz tuviese 100 habitantes, entonces el tamaño de la muestra tendría que ser de 80 personas. Para 1000 habitantes se necesita una muestra de 278 personas. Para 10000 habitantes la muestra debe ser de 370 ¡Qué poco va aumentando! Para 100000, sólo se necesitan 383.

Y eso, ¿por qué es así? Porque hay un fórmula matemática de estadística que lo deduce y que como asusta un poco no voy a enseñártela. En su lugar, seguro que te es más fácil utilizar esta [calculadora](#).



## 1.3. ¿Cómo obtengo los datos?

Una vez seleccionada la muestra, hay dos formas de **obtener la información** que necesitamos:

→ **Obtención indirecta:** los datos están recogidos y se consultan:

La **ventaja** es que icasi todo el trabajo está hecho! **Pero...**

- Puede que los datos no estén actualizados (¿y si el INE usase información de 1985 para el estudio que hemos visto?).
- Puede que los datos no se ajusten a nuestro estudio. Si queremos estudiar el tiempo que un grupo de niños dedican a desayunar y los datos disponibles se refieren a adultos ide poco nos sirven!

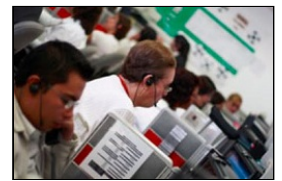


Imagen en Pixabay de  
Peggy\_Marco.Licencia CC0

→ **Obtención directa:** los datos se observan **directamente sobre los individuos**, específicamente para el estudio.

Seguramente los datos se ajustarán a nuestros deseos pero costará más obtenerlos.

- A veces se hará midiendo la característica en los individuos (por ejemplo, la estatura...)
- A veces se hará mediante cuestionarios que los encuestados responden...



Fotografía en freeimages  
de C. ChavezLic.  
Freeimages

**Debemos tener en cuenta que las preguntas de nuestra encuesta pueden influir en nuestra investigación.**

### *Ejercicio resuelto*



Imagen en [MEC-ITE](#) Licencia  
CC

**Por ejemplo:**

**¿Son iguales estas dos preguntas para un cuestionario?**

1. ¿Qué prefiere desayunar, yogurt desnatado o churros con chocolate?
2. ¿Qué suele desayunar con mayor frecuencia?

Es evidente que **las dos preguntas causarán un efecto muy diferente en el entrevistado**. Con la primera las opciones del encuestado son muy reducidas (es bastante probable que el individuo hubiese dado otra respuesta) y se ve claramente una intención oculta: separar falsamente a la población en dos grupos ficticios que no se corresponden con la realidad.

#### **Mostrar retroalimentación**

Por tanto es preferible la segunda pregunta, aunque con ella se corre un riesgo: puede que las respuestas obtenidas sean demasiado diferentes entre sí y la información sea poco útil... ¿y si todos responden una cosa diferente? ¿Podríamos extraer alguna conclusión?



## *Importante*

---

Podemos obtener la información por **observación directa o indirecta**.

## 2. ¿Qué hacemos con los datos que hemos tomado?

Vamos a suponer el siguiente estudio estadístico: queremos saber el tiempo que nuestros vecinos dedican a desayunar. Para ello hacemos un sorteo y **seleccionamos** para nuestra muestra **10 vecinos al azar de los 50** que hay en la vecindad y...

¡Les preguntamos...claro!

En este estudio tenemos:

- Variable aleatoria:  $X$  = tiempo dedicado al desayuno (en minutos). Es una variable cuantitativa discreta.
- $N = 50$ ,  $n = 10$ .
- Los individuos encuestados han dado las respuestas: **0,0,9,0,5,5,9,5,15,0.**

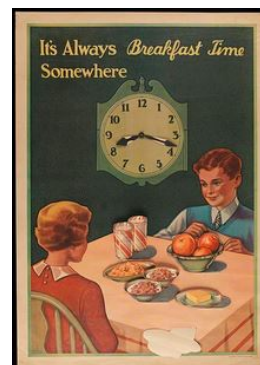


Imagen en [Wikimedia commons](#). Licencia [cc](#)

**Pero esta información no nos sirve demasiado.** Vemos que se han dado los siguientes cuatro valores de la variable:  $x_1=0$ ,  $x_2=5$ ,  $x_3=9$ ,  $x_4=15$ . ¿Se repite alguno más que otro? ¿Y si estudiásemos cuántas veces nos han respondido cada valor?

### Nota:

En matemáticas para representar un elemento en general que pertenezca a un conjunto de elementos ordenados, se usa el subíndice  $i$ . Así una valor genérico del conjunto  $x_1, x_2, \dots, x_4$ , se representará como  $x_i$ , donde el subíndice  $i$  puede tomar los valores 1, 2, ..., 4 (o los que haya, claro)

### Importante

La **frecuencia absoluta** de un valor  $x_i$  de la variable es el número de veces que se ha observado dicho valor y se representa  $n_i$

Si sumas todas las frecuencias absolutas debes obtener el valor de  **$n$**

Por ejemplo, la frecuencia absoluta del valor  $x_1=0$  es 4 (4 vecinos han respondido 0 minutos).

Es decir  **$n_1=4$** . Sólo hay que saber contar...

### Comprueba lo aprendido

### Autoevaluación

Ahora indica tú la frecuencia absoluta del resto de los valores de la variable:

Recuerda, sólo hay que contar las veces que se repite cada valor, las veces que nos han respondido cada cantidad de minutos.



$n_2 = \square$

$n_3 = \square$



$$n_4 = \square$$

y todas las frecuencias absolutas deben sumar  $n = \square$

**Enviar**

Imagen elaboración propia

Vamos encaminados a elaborar una **TABLA DE FRECUENCIAS**, donde aparecerá **la información recogida de manera ORGANIZADA**. El siguiente paso es hallar otro tipo de frecuencias.

Hemos visto que 4 de los 10 vecinos no desayunan, casi la mitad. ¿Y si fuesen 4 vecinos de 4.566.123? ¡Prácticamente todos desayunarían! ¿A que la cosa cambia?

Las frecuencias absolutas **hay que "relativizarlas"**...

*Importante*

La **frecuencia relativa** de un valor de la variable es su frecuencia absoluta dividida por el número de observaciones. Para el valor  $x_i$  se representa  $f_i$ .

Así, como  $n_1 = 4$ , entonces  **$f_1 = 4:10 = 0.4$** . (la frecuencia absoluta del valor  $x_1=0$  dividida entre 10 los vecinos encuestados).

NOTA: La frecuencia relativa nos da la misma información que un "tanto por ciento" o %, pero en este caso expresado en "tanto por uno".

En nuestro ejemplo los vecinos respondieron que no desayunaban con una frecuencia relativa de 0.4. Es decir, el 40% de los vecinos no desayunan.

*Comprueba lo aprendido*

### Autoevaluación

Ahora indica tú la frecuencia relativa de cada valor de la variable:

$$f_2 = \square$$

$$f_3 = \square$$

$$f_4 = \square$$

y las frecuencias relativas deben sumar  $\square$

**Enviar**

## Curiosidad

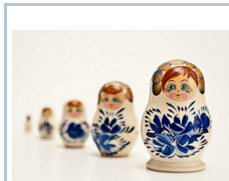
### Y todavía puede ser útil usar más tipos de frecuencias...

Tenemos las frecuencias absolutas (y relativas) acumuladas **para las que tenemos que ordenar los valores** de menor a mayor:

La **frecuencia absoluta (o relativa) acumulada** de un valor es la suma de todas las frecuencias absolutas (o relativas) de todos los valores MENORES O IGUALES QUE DICHO VALOR.

Se representan como las anteriores pero con mayúscula.

Por ejemplo, la frecuencia absoluta acumulada del valor 9 minutos es la suma de las frecuencias absolutas de los valores 0, 5 y 9.



Fotografía en Flickr  
de M.  
Bohigas. Licencia cc

## Comprueba lo aprendido

### Autoevaluación

¿Serías capaz tu solito/a de calcular las frecuencias absolutas y relativas ACUMULADAS de los valores de nuestra variable? Seguro que sí.

Fíjate en el valor de la última frecuencia en ambos casos. ¿Te suena de algo?

**Frecuencias absolutas acumuladas:**  $N_1 = \square$  ;  $N_2 = \square$  ;  $N_3 = \square$  ;  $N_4 = \square$

**Frecuencias relativas acumuladas:**  $F_1 = \square$  ;  $F_2 = \square$  ;  $F_3 = \square$  ;  $F_4 = \square$

**Enviar**

Y todos estos numeracos **¿para qué los hacemos?** Pues con un único objetivo: **para tener los datos bien ordenaditos**, porque con todos estos números podemos **hacer la TABLA DE FRECUENCIAS** de nuestra variable.

## Importante

### Para hacer una tabla de frecuencias...

Se pone una primera columna con los valores de la variable y después una columna con cada tipo de frecuencia (y la suma o total en las no acumuladas), en nuestro caso:

Valores de la variables	Frecuencias	Frecuencias	Frecuencias absolutas	Frecuencias relativas
-------------------------	-------------	-------------	-----------------------	-----------------------

(minutos dedicados a las variables): $x_i$	absolutas: $n_i$	relativas: $f_i$	acumuladas: $N_i$	acumuladas: $F_i$
$x_1=0$	$n_1=4$	$f_1=0,4$	$N_1=4$	$F_1=0,4$
$x_2=5$	$n_2=3$	$f_2=0,3$	$N_2=7$	$F_2=0,7$
$x_3=9$	$n_3=2$	$f_3=0,2$	$N_3=9$	$F_3=0,9$
$x_4=15$	$n_4=1$	$f_4=0,1$	$N_4=10$	$F_4=1$
total:	<b>10</b>	<b>1</b>		

Aunque, utilizando los símbolos con los que se representa cada cosa, basta escribir esto:

$x_i$	$n_i$	$f_i$	$N_i$	$F_i$
0	4	0,4	4	0,4
5	3	0,3	7	0,7
9	2	0,2	9	0,9
15	1	0,1	1	1
total	10	1		

## Actividad de lectura

### Autoevaluación

- Extrae alguna conclusión de la tabla anterior
- Realiza la tabla de frecuencias asociada al siguiente estudio...

- $X$  = nº de piezas de fruta consumidas al día.
- $P$  = habitantes de Nofrutataun.
- $N$  = 500
- $n$  = 25
- Respuestas: 2, 3, 4, 2, 1, 2, 0, 0, 3, 1, 2, 0, 1, 1, 0, 2, 0, 1, 4, 2, 3, 0, 1, 1, 2



Imagen en [pixabay](#) de [romanov](#). Licencia CC0

### Mostrar retroalimentación

- Algunas conclusiones que podríamos sacar son las siguientes:
  - El 40% (4 de 10 encuestados) de los vecinos NO desayuna. Es la respuesta más frecuente (repetida).
  - Sólo 1 de 10 dedica más de 10 minutos al desayuno (sólo el 10% de los vecinos).
  - 7 de los 10 vecinos encuestados dedica 5 minutos o menos a su desayuno (esto lo hace el 70% de los vecinos).
  - Por tanto, parece claro que el tiempo dedicado es por lo general insuficiente para un desayuno de calidad.
- La tabla quedaría así:

$x_i$	$n_i$	$f_i$	$N_i$	$F_i$
0	6	0,24	6	0,24
total	25	1	12	0,52

1	7	0.28	13	0.52
$x_i$	$n_i$	$f_i$	$N_i$	$F_i$
2	7	0.28	20	0.8
3	3	0.12	23	0.92
4	2	0.08	25	1
total	25	1		

### 3. ¿Cómo podemos ver los datos gráficamente?

Los datos estadísticos se pueden "ver" en dibujos. Sí sí, en serio. ¿A que si lo logramos será genial? Vamos allá.

#### Diagrama de Barras:

Un diagrama de barras de una variable estadística se hace de forma muy sencilla. Por pasos, tras hacer la tabla de frecuencias:

- Primero dibujamos **dos ejes** de coordenadas.
- Después colocamos los **valores de la variable en el eje x**. Deben ser valores de variable discreta.
- Por último levantamos una **barra sobre cada valor**. ¿Hasta qué altura? Hasta lo que indiquen **las frecuencias** (absolutas o relativas).

**Por ejemplo**, en el caso anterior, a partir de la tabla podemos obtener el diagrama de barras siguiente:

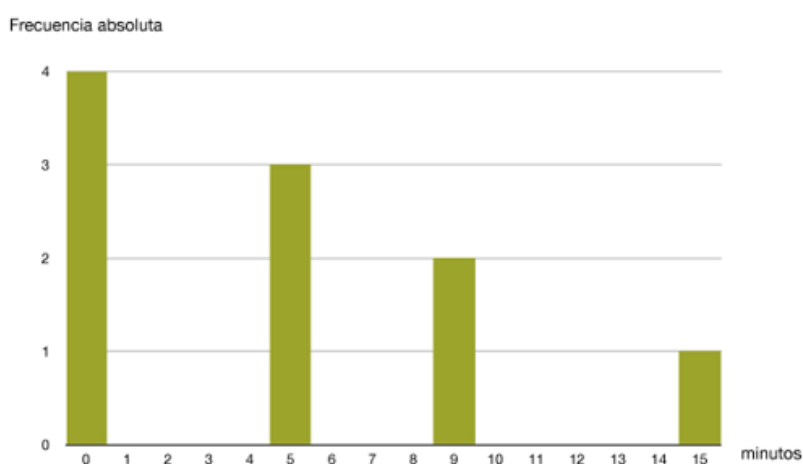


Imagen elaboración propia

*Importante*

#### Así que

Un diagrama de barras se construye dibujando barras sobre los valores que midan lo que indiquen las frecuencias absolutas (o relativas).

#### Diagrama de sectores

Un diagrama de sectores es un círculo dividido en sectores cuya área será tanto mayor cuanto mayor sea la frecuencia del valor que representa.

Dibujamos un **círculo**.

Lo dividimos en tantos **sectores circulares** ("quesitos") como valores queramos representar. Pero claro, al valor con mayor frecuencia le corresponderá una parte más grande ¿verdad?

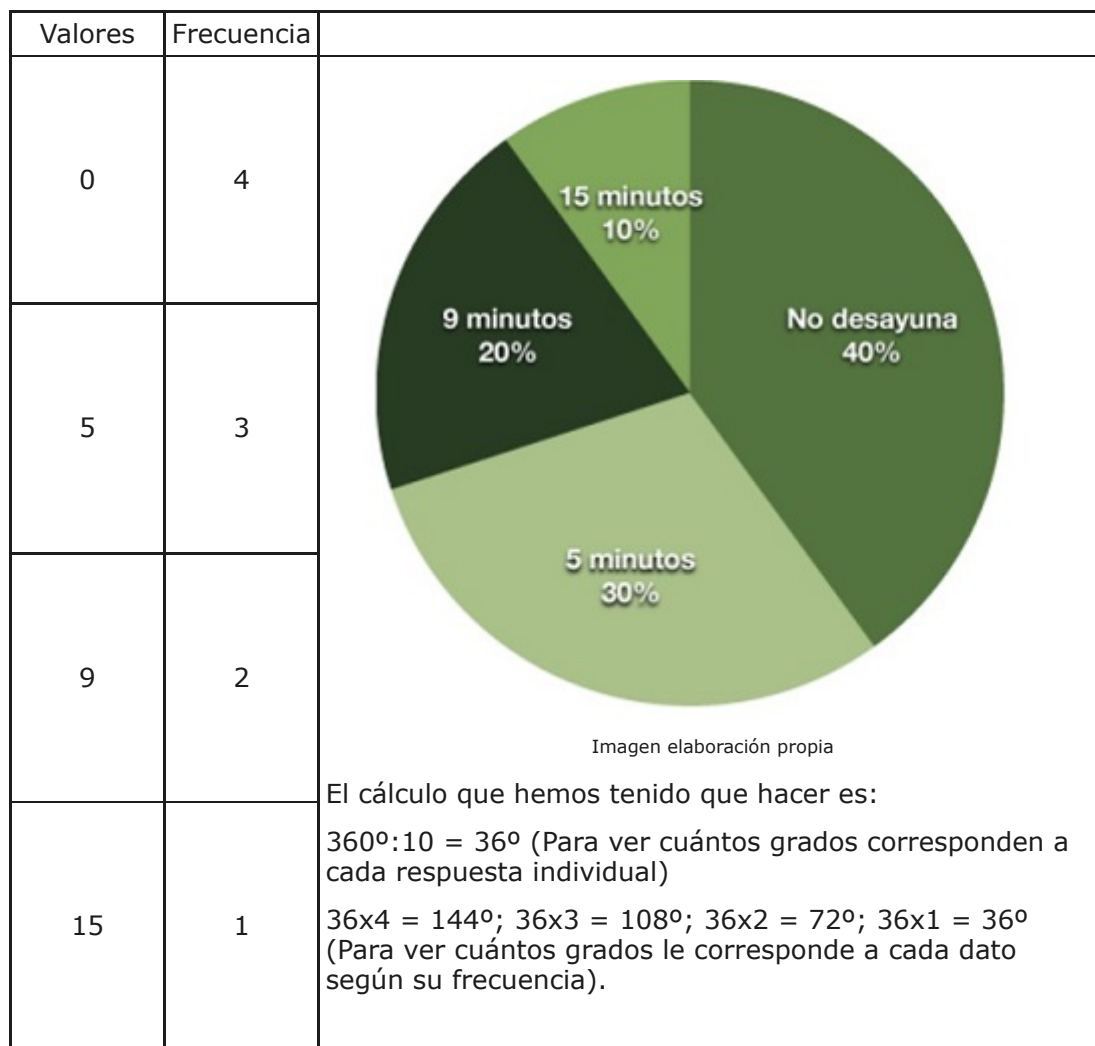
Para cada valor su sector tendrá **un ángulo PROPORCIONAL** a su frecuencia. ¿Que te has asustado? Vamos a ver cómo hacerlo, que es fácil, sólo hay que hacer **una "regla de tres"**. El círculo abarca 360°, ¿verdad? Pues bien:

- Repartimos los 360° entre el total de respuestas que, en nuestro ejemplo, es 10  $\rightarrow 360^\circ : 10 = 36^\circ$  **grados para respuesta**.



- Si cada respuesta se representa con un sector ("quesito") de  $36^\circ$  el valor 0 que se repitió **en cuatro respuestas** se representará con un sector de  $4 \bullet 36 = 144^\circ$ . Y así con el resto de valores.

Haciendo esto, obtenemos el siguiente **diagrama de sectores**:



## Importante

### Recuerda:

Para saber el ángulo que tiene que ocupar cada sector, dividimos  $360^\circ$  entre el número de respuestas y multiplicamos lo obtenido por la frecuencia absoluta de cada valor.

## Reflexiona

A la hora de hacer una representación gráfica, no hay que olvidar que su objetivo es facilitar la comprensión de la información.

Por eso hay que tener cuidado a la hora de elegir el tipo de gráfico y el diseño del mismo.

Los gráficos circulares presentan una serie de inconvenientes, por los que no siempre es

aconsejable su uso.

En este enlace se te explican algunos de ellos:



### Las tartas son para el postre: 5 razones por las que no uso gráficos circulares

Observa ahora el gráfico inferior. ¿Qué problema de los que se mencionan en el texto del enlace presenta?

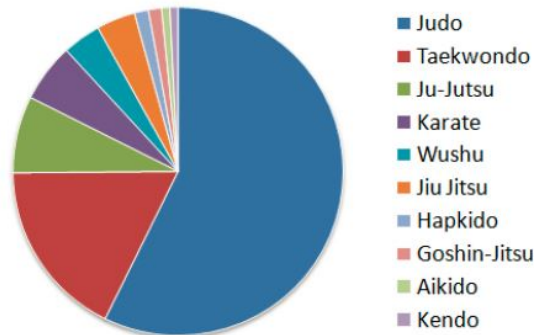


Imagen en [Wikimedia commons](#) de [Hoosic06](#) . Licencia [cc](#)

### Mostrar retroalimentación

Si te fijas, no se pueden determinar con exactitud los porcentajes y tampoco es posible comparar con precisión los tamaños de cada sector.

## Pictograma

En este caso **representamos cada valor con un dibujo alusivo** cuyo TAMAÑO dependerá de la frecuencia absoluta (o relativa) del valor.

Por ejemplo, ya que estamos con el tiempo dedicado a desayunar, representaremos cada valor con una taza, que será más grande para las respuestas más frecuentes y menor para las respuestas menos frecuentes.



Imagen en pixabay de [OpenClipart-Vectors](#). Licencia [CC0](#)

## Curiosidad

### Pero cuidado

Si con tamaño nos referimos al área, entonces a doble frecuencia no corresponde doble

tamaño, sino 4 veces el tamaño, porque la relación es cuadrática (un área se calcula elevando a cuadrado una longitud) y el dibujo obtenido no se corresponderá con los datos.

Por tanto, tenemos que dejar claro a qué nos referimos con tamaño. Podemos quedar en que a doble, triple... frecuencia corresponda doble, triple... altura del dibujo, por ejemplo.

## Histograma

Un histograma es muy **similar a un diagrama de barras**, solo que se usa **para variables cuyos datos se agrupan en intervalos**. Por ejemplo, si preguntamos a nuestros vecinos si están apuntados a algún gimnasio o club deportivo, la tabla de frecuencias por intervalo de edades puede ser:

Valores de la variable (edad en años) $I_i$	Frecuencias absolutas $n_i$	Frecuencias relativas $f_i$	Frecuencias absolutas acumuladas $N_i$	Frecuencias relativas acumuladas $F_i$
15-24	20	0,18	20	0,18
25-34	41	0,37	61	0,55
35-44	29	0,26	90	0,81
45-54	15	0,14	105	0,95
55-64	4	0,04	109	0,99
65-74	1	0,01	110	1
	110	1		

¿Has visto algo nuevo en esta tabla? Seguro que te has dado cuenta de que en la primera columna no aparecen datos concretos de años, sino los intervalos en que agrupamos los valores y las frecuencias absolutas se refieren al número de personas cuya edad está dentro del intervalo correspondiente.

Pues bien, para hacer el histograma:

Primero dibujamos dos **ejes de coordenadas**.

Después colocamos **los intervalos** en el eje X.

Por último **levantamos un rectángulo sobre cada intervalo**. ¿Hasta qué edad? Pueden ocurrir dos cosas:

- Que todos los intervalos tengan la **misma amplitud** (como en nuestro caso, que todos "van de 10 en 10")
- Que haya intervalos más amplios que otros. Este caso lo dejaremos de momento.

En el primer caso, la altura del rectángulo será **la indicada por la frecuencia** (absoluta o relativa) correspondiente.

En nuestro ejemplo quedaría así:



Elaboración propia

## Para saber más

### Para saber más...

Los intervalos en los que se agrupan los valores de una variable continua o discreta con demasiados valores se llaman **intervalos de clase**.

El valor representante de un intervalo de clase se llama **marca de clase** y suele coincidir con el valor medio del intervalo (el dato que queda justo en medio del mismo). En el ejemplo de los vecinos que están apuntados a un gimnasio o club deportivo, la marca del intervalo [15-24] es 19,5 años. Si no lo ves claro, primero suma  $15 + 24 = 39$  y después divide  $39:2 = 19,5$ . Ya lo tienes

## Ejercicio resuelto

### Autoevaluación

Realiza los gráficos estadísticos que puedas para la variable trabajada en la anterior autoevaluación. Te recordamos la tabla de frecuencias:

$x_i$	$n_i$	$f_i$	$N_i$	$F_i$
0	6	0.24	6	0.24
1	7	0.28	13	0.52
2	7	0.28	20	0.8
3	3	0.12	23	0.92
4	2	0.08	25	1
total	25	1		

### Mostrar retroalimentación

Como los valores no están agrupados en intervalos, podríamos hacer los gráficos siguientes:

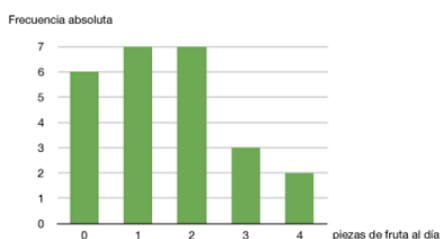


DIAGRAMA DE BARRAS(imagen elaboración propia)

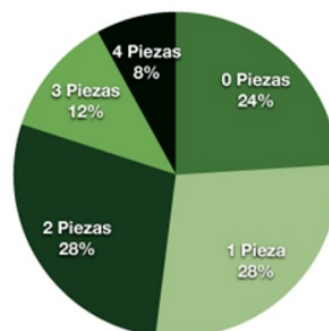


DIAGRAMA DE SECTORES (imagen elaboración propia)





PICTOGRAMA(imagen elaboración propia)

## Diagrama de burbujas

Otro gráfico muy utilizado es el de burbujas. Muy parecido a los pictogramas pero situados en unos ejes.

En el siguiente ejemplo se refleja la información de una encuesta que se ha hecho a la totalidad de los niños de un colegio.

Se pretendía saber cuántas veces a la semana consumen habitualmente golosinas. Y estas fueron sus respuestas:

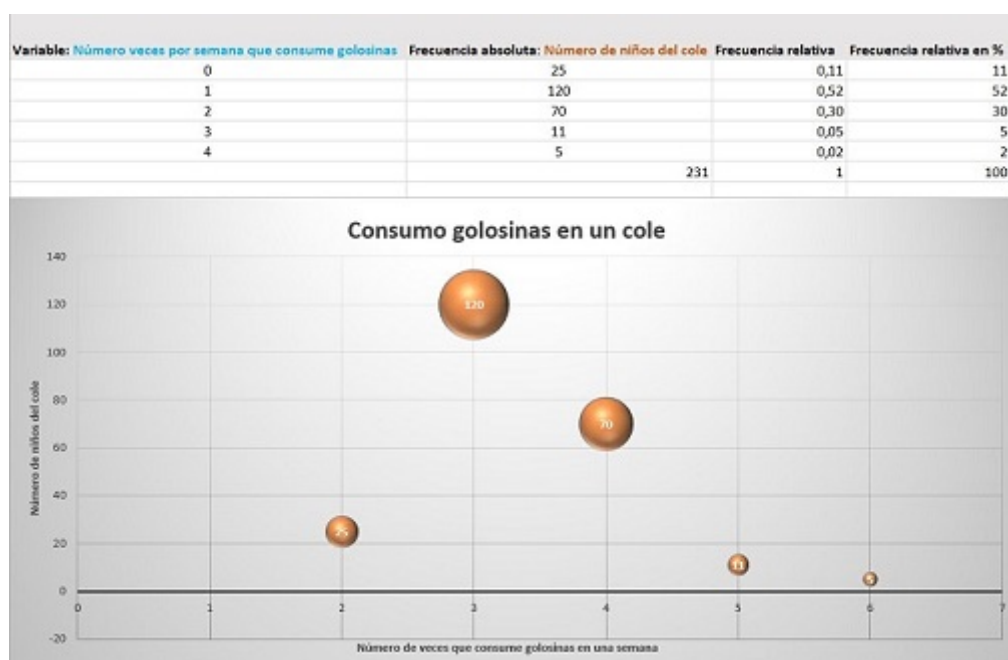



Imagen elaboración propia

¿No te parece que toman demasiadas  golosinas?

*Curiosidad*

## Interesantes ejemplos de gráficos

En este enlace tienes una amena e interesante [infografía sobre el gasto medio por familia y persona](#) . Es una encuesta del Instituto nacional de estadística donde puedes comparar en qué gastamos nuestro dinero entre distintas comunidades autónomas. Fíjate en los euros que empleamos en alimentación.

## 4. Calculando números que informan sobre los datos

Antes de continuar, te vamos a preguntar una cosa. Si sacas en dos exámenes un 8 y un 6, ¿a que sabes tu nota media? Lo que haces, quizá sin saberlo, es sumar las dos notas y dividir lo obtenido por dos ¿verdad? Eso que haces se llama media aritmética.

Te vamos a contar más sobre ella y sobre alguna otra cosa... y haremos algunas cuentas

### La media aritmética

#### Importante

Se llama media aritmética de una variable aleatoria a la **suma de todos los valores observados dividida por el total de observaciones**.

Volvamos al ejemplo del desayuno:

Hay **dos formas de calcular la media aritmética**. Una "a lo bruto" y otra "pensando un poco":

- **A lo bruto**: sumamos todas las respuestas que nos han dado los vecinos y dividimos entre los 10 vecinos encuestados:

$$(0+0+9+0+5+5+9+5+15+0):10 = 48$$

$$48:10 = 4,8$$

Esto quiere decir que si todos los vecinos desayunasen el mismo tiempo, **desayunarían todos 4,8 minutos**.

- **Pensando un poco**: ¿y si en vez de sumar los valores como antes primero multiplicamos cada valor por las veces que se repite? Es decir: en la suma anterior hay, por ejemplo, tres cincos... ¿no sería mejor poner 15? Y así con el resto.

En realidad lo que haríamos sería **sumar cada valor multiplicado por su frecuencia absoluta y dividir después por el número total de observaciones**:

$$(0 \cdot 4 + 5 \cdot 3 + 9 \cdot 2 + 15 \cdot 1):10 = 48:10 = 4,8 \text{ minutos.}$$

Podemos usar la tabla de frecuencias para este cálculo.

**Ahora piensa**: aunque hay un vecino que tarda en desayunar 15 minutos... ¿crees que los vecinos pasan bastante tiempo desayunando?

#### Curiosidad

#### ¿Qué ocurriría si...?

¿Qué ocurriría con la media anterior si un vecino tardase dos horas en tomarse su café y sus tostadas?

A la suma anterior tendríamos que añadir 120 minutos y obtendríamos 168, ahora dividimos entre el número de vecinos

encuestados (que serían 11) y... SALE UNA MEDIA DE MÁS DE 15 MINUTOS.

¡Ya quisieran los demás tener tanto tiempo para desayunar! Se ha



Fotografía en Flickr  
de [Olga Berrios](#) .  
Licencia [cc](#)

triplicado la media anterior.

Como ves, los "**valores extraños**" pueden producir **medias extrañas que no reflejen realmente la realidad**.

Para terminar ¿se te ocurre cómo calcular la media aritmética **cuando la variable está agrupada** en intervalos? Es fácil ¿no? Pues en vez de los valores **tomamos las marcas de clase**

## *Ejercicio resuelto*

### **Un caso especial:**

¿Cuánto vale la media si todas las respuestas son iguales? Piensa un poco.

Por ejemplo, si todos los vecinos hacen el mismo número de comidas diarias (4), la media de comidas diarias es justo eso, 4 comidas.

¡Haz la cuenta y verás que no es magia!

---

## **Moda**



¿Qué quiere decir que un color está de moda? Pues que la mayoría de la gente se viste con ese color. ¿Qué significa que está de moda desayunar cereales? Que la mayoría de la gente toma cereales...

Y así podemos seguir.

**¿Qué será la moda de una variable estadística?**

**Efectivamente, acertaste**

Icono en [Pixabay](#).  
Licencia CC0

## *Importante*

La moda de una variable estadística **es el valor más frecuente, el más repetido en las respuestas...** el de mayor frecuencia absoluta (o relativa).

## *Comprueba lo aprendido*

Lea el párrafo que aparece abajo y complete las palabras que faltan.

## Autoevaluación

¿Cuál es la moda en nuestro estudio del tiempo dedicado a desayunar?

De las distintas respuestas una se repitió más que las demás:  minutos dedicados.

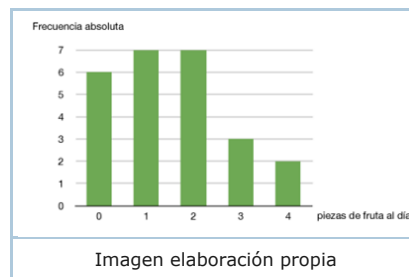
**Enviar**

## Curiosidad

## Piensa un momento

¿Crees que puede haber **más de una moda** en una variable estadística?

Te ayudamos a pensarlo con una pequeña pista: sacude tu memoria... ¿te suena esto?



**Efectivamente, puede haberla.** Observa en la gráfica que hay dos valores con frecuencias iguales y ningún otro valor es más frecuente; ambos serán moda. Así podemos tener variables unimodales (con una moda), bimodales (con dos modas, como en el gráfico)...y hasta polimodales o multimodales (con **varias modas**).

Si la variable está agrupada en intervalos, podríamos indicar el **intervalo modal** (el intervalo que más datos "contiene").

## Comprueba lo aprendido

## Autoevaluación

a.

Calcula la media aritmética y la moda para la variable trabajada en las anteriores autoevaluaciones. Te recordamos la tabla de frecuencias:

$x_i$	$n_i$	$f_i$	$N_i$	$F_i$
-------	-------	-------	-------	-------

total	25	1	6	0.24
-------	----	---	---	------



$x_i$	$h_i$	$0,28$	$N_i$	$0,52$
2	7	0,28	20	0,8
3	3	0,12	23	0,92
4	2	0,08	25	1
total	25	1		

Media aritmética:  piezas de fruta.

Moda:  piezas, es distribución  modal.

b. Halla la moda y el intervalo modal de la variable vista anteriormente relativa a la edad y estar apuntado a un gimnasio. La tabla de frecuencias es:

Valores de la variable (edad en años) $I_i$	Marca de clase (edad en años)	Frecuencias absolutas $n_i$	Frecuencias relativas $f_i$	Frecuencias absolutas acumuladas $N_i$	Frecuencias relativas acumuladas $F_i$
15-24	<input type="text"/>	20	0,18	20	0,18
25-34	29,5	41	0,37	61	0,55
35-44	39,5	29	0,26	90	0,81
45-54	49,5	15	0,14	105	0,95
55-64	59,5	4	0,04	109	0,99
65-74	<input type="text"/>	1	0,01	110	1
		110	1		

Media aritmética:  años.

Intervalo modal:  años.

**Enviar**

*Para saber más*

### Para saber más...

La media se llama aritmética por una razón... y es que hay otras medias que no son aritméticas. Puedes verlo en el siguiente enlace:

 [Otras medias y parámetros](#)

## 5. Todos los datos no son iguales: dispersión de los datos

### Ejercicio resuelto

Observa los datos recogidos en estos dos estudios estadísticos:

**Variable 1:**

**NIVEL DE SATISFACCIÓN CON LA IMAGEN CORPORAL EN CIERTA CIUDAD**

- **Población:** habitantes de Megusto con edades entre 11 y 40 años.
- **Muestra:** 230 habitantes de diversas edades, estamentos sociales y profesiones.
- **Pregunta:** ¿Qué nivel de satisfacción con su imagen tiene usted (de 0 a 10)?

Nivel de satisfacción	Frecuencia absoluta
0	56
1	20
2	15
3	2
4	1
5	76
6	39
7	2
8	6
9	3
10	?

**Variable 2:**

**INFLUENCIA DE LOS CÁNONES DE BELLEZA IMPUESTOS POR LA SOCIEDAD COMO CAUSA DE LA ANOREXIA Y/O LA BULIMIA**

- **Población:** habitantes de Nocomono con edades entre 11 y 40 años que han padecido o padecen trastornos alimenticios.
- **Muestra:** 230 habitantes de diversas edades y distintos estamentos sociales y profesiones que padecen o han padecido anorexia o bulimia.
- **Pregunta:** ¿Del cero al diez, cómo califica la influencia de los cánones de belleza impuestos por la sociedad como causa de su trastorno?



Fotografía en [Wikimedia commons](#) de [Lars Aronsson](#). Licencia [cc](#)

Influencia de los cánones de belleza	Frecuencia absoluta
0	1
1	2
2	1

3	2
---	---

Influencia de los cánones de belleza	Frecuencia absoluta
5	5
6	9
7	20
8	45
9	56
10	87

## Trabajemos con ellos...

- ¿Qué valor falta en la tabla primera?
- Para ambos estudios, haz un diagrama de barras y halla la media aritmética y la moda.

### Mostrar retroalimentación

- Para saber la frecuencia que falta sólo hay que recordar que si las sumamos TODAS debe salirnos 230 (el total de observaciones). Como las demás suman 229, sólo una persona ("la número 230") responde 10.
- Vamos allá:

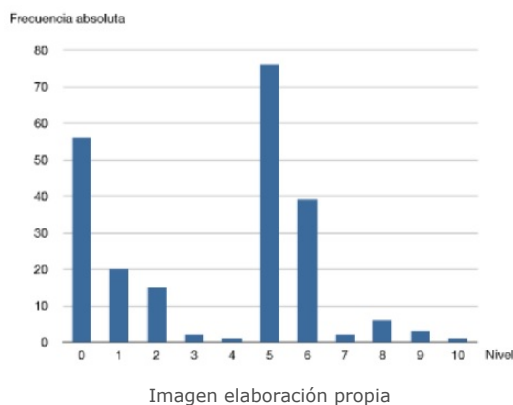
#### ● Variable 1: Nivel de satisfacción con la imagen personal

##### ● La media vale:

$$(0 \cdot 56 + 1 \cdot 20 + 2 \cdot 15 + 3 \cdot 2 + 4 \cdot 10 + 5 \cdot 76 + 6 \cdot 39 + 7 \cdot 2 + 8 \cdot 6 + 9 \cdot 3 + 10 \cdot 1) : 230$$

$$(0 + 20 + 30 + 6 + 40 + 380 + 234 + 14 + 48 + 27 + 10) : 230 = 809 : 230 = \mathbf{3,51}.$$

##### ● La moda es 5, cuya frecuencia 76 es la mayor.



#### ● Variable 2: Influencia de los cánones de belleza impuestos por la sociedad

##### ● La media vale:

$$(0 \cdot 1 + 1 \cdot 2 + 2 \cdot 1 + 3 \cdot 2 + 4 \cdot 2 + 5 \cdot 5 + 6 \cdot 9 + 7 \cdot 20 + 8 \cdot 45 + 9 \cdot 56 + 10 \cdot 87) : 230$$

$$(0 + 2 + 2 + 6 + 8 + 25 + 54 + 140 + 360 + 504 + 870) : 230 = 1971 : 230 = \mathbf{8,56}.$$

##### ● La moda es 10, cuya frecuencia 87 es la mayor.



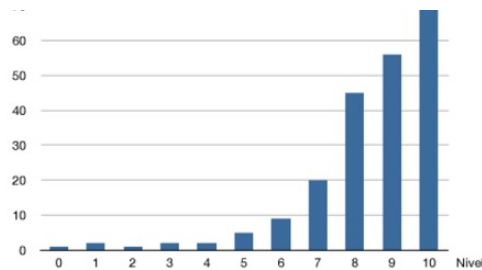


Imagen elaboración propia

### ¿Crees que en ambos casos las medias aritméticas son igual de representativas?

La verdad es que no... Fíjate bien en los diagramas de barras y observa que los datos no están "distribuidos" de la misma manera en los dos ejemplos.

Ahora vas a estudiar que es posible expresar matemáticamente eso de "lo distribuidos que están los datos". Los tres números que vas a aprender a calcular (la varianza, la desviación típica y el coeficiente de variación) **indican lo agrupados o alejados que están los valores respecto de la media**, y por tanto **qué media da información más fiable sobre los datos**.

## Varianza y desviación típica



Imágenes en Wikimedia  
commons de [LGPLY](#) [LGPL](#).  
Licencia [GNU](#)

Podemos estudiar si los datos de nuestras variables 1 y 2 están, en su conjunto, más o menos cerca de las medias respectivas.

Por ejemplo: En la variable1 quien ha respondido 10 se ha **alejado mucho de la media 3.51**. En cambio, los 12 que han respondido 3 o 4 se ha **acercado bastante a la media**. Cuanto más valores cercanos a la media tengamos, **más "fiable" es nuestra media** ¿verdad?

Existen dos números (parámetros estadísticos) que nos ayudan a **MEDIR esta cercanía de los datos a la media, es decir, a medir la dispersión de los datos**.

Uno es la **varianza**, que es la *media de las distancias de los valores a la media, al cuadrado*.

¿Que te da miedo tanta palabrería...? Pues fíjate bien cómo no es para tanto...

Se hace así:

- Se calcula la distancia de cada valor a la media, sencillamente restándole al valor la media (para el valor 10 será  $10 - 3,51 = 6,49$ )
- Se elevan esas distancias al cuadrado (así,  $6,49^2 = 42,1201$ )
- Y se hace la media de los resultados como si fuesen valores (las frecuencias serán las originales)

Y si no te gusta el método anterior, ¡Hay otro mejor!

- Eleva los valores de la variable al cuadrado.
- Haz la media de los resultados obtenidos.
- Eleva la media de la variable al cuadrado y restalo del resultado anterior.

## Ejercicio resuelto

### Por ejemplo:

Para la variable 1...

Los valores al cuadrado son:

$$0^2 = 0, 1^2 = 1, 2^2 = 4, 3^2 = 9, 4^2 = 16, 5^2 = 25, 6^2 = 36, 7^2 = 49, 8^2 = 64, 9^2 = 81, 10^2 = 100.$$

La media de esos resultados es:

$$(0 \times 56 + 1 \times 20 + 4 \times 15 + 9 \times 2 + 16 \times 10 + 25 \times 76 + 36 \times 39 + 49 \times 2 + 64 \times 6 + 81 \times 3 + 100 \times 1) : 230$$

$$(0 + 20 + 60 + 18 + 160 + 1900 + 1296 + 98 + 384 + 243 + 100) : 230 = 4279 : 230 = \mathbf{18,604}.$$

A esa cantidad restamos la media, 3,51, al cuadrado:

$$\mathbf{18,604 - 3,51^2 = 6.283}.$$

Luego nuestra varianza es **6.283**.

## Curiosidad

### Muy MUY importante

Como es una media de "números al cuadrado" y las cosas al cuadrado son siempre positivas:

#### LA VARIANZA ES SIEMPRE POSITIVA

Una varianza negativa se considera un "delito matemático"; si te aparece alguna revisa tus cálculos porque te has equivocado seguro.



Si hemos calculado la varianza, la **desviación típica** es muy fácil de calcular: solo hay que **hacer la raíz cuadrada a la varianza**.

(Por ejemplo: para la variable1 la desviación típica es la raíz cuadrada de 6,283, que es **2.506**).

Y **nos dice la dispersión respecto de la media**. ¿Y eso qué significa? Pues, en nuestro ejemplo, que los valores se alejan un promedio de 2.506 puntos respecto de la media, es decir, que muchos de los valores estarán entre 1 y 6 puntos, lo cual puede verse en la tabla y en el gráfico.

## Comprueba lo aprendido

### Autoevaluación

- a. El valor correcto de la varianza de la variable2 (con dos decimales) es:
- b. El valor correcto de la desviación típica de la variable2 es:

**Enviar**

---

## Coeficiente de variación

Con los cálculos anteriores...

**¿Podemos saber qué media de las dos anteriores es más "fiable"?**

Las dos variables tratan cosas muy diferentes, y no podemos establecer la comparación. A priori puede parecer que la segunda variable tiene los datos MENOS DISPERSOS O MÁS AGRUPADOS, ya que su desviación típica es menor... ¿y si una variable estuviese medida en mm y la otra en número de sillas? Está claro que la comparación no es posible... ¿o sí?

Pues con los datos que hemos calculado no, pero sí con otro parámetro estadístico, es decir, otro "numerajo". Pero no te preocupes, **es el coeficiente de variación** y para calcularlo **basta dividir la desviación típica entre la media**.

En el ejemplo que estamos desarrollando...

El coeficiente de variación de la variable1 vale  $2,506:3,51 = 0,7139$ .

*Comprueba lo aprendido*

### Autoevaluación

El valor correcto del coeficiente de variación de la variable2 es (con 2 decimales)

.

Por tanto, ahora sí podemos asegurar que **los datos de la variable  están mucho más cercanos a la media**, la cual es, por tanto, más representativa y fiable.

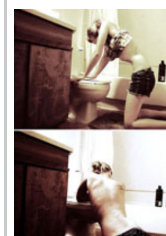
**Enviar**

*Para saber más*

### Para saber más...

Las dos variables vistas están relacionadas con trastornos relativos a la alimentación. Para saber más sobre anorexia y bulimia puedes visitar los enlaces siguientes:

- [Geosalud](#)
- [Tú eres más que una imagen](#)
- [Adaner](#)



Fotografía en  
flickr de [Aimee](#)  
[Heart](#). Licencia [cc](#)

## 6. El estudio estadístico se hace para sacar conclusiones

---



Imagen en [pixabay](#) de [bykst](#). Licencia CC0

Ya sabes hacer muchas cosas con nuestros datos, pero no tiene sentido que nos pongamos a hacer cuentas y gráficos perdiéndonos entre números y tablas simplemente porque sí.

Ten presente que **se trata de dar respuesta** a lo que queríamos saber extrayendo conclusiones:

**Las tablas** de frecuencias y los gráficos tienen por objetivo organizar y facilitar la visualización de los datos.

**La media aritmética** indica un valor representativo de la variable, que resume la información de los datos recogidos y se interpreta en la realidad concreta que estamos estudiando.

**La varianza y desviación típica** nos dicen cómo están de agrupados los datos respecto de la media. Permiten saber hasta qué punto la media aritmética da una buena información de la realidad estudiada.

**El coeficiente de variación**, además de lo anterior, permite comparar la agrupación de los datos respecto de la media en distribuciones que no se parezcan en nada, lo cual puede resultar interesante.

### *Ejercicio resuelto*

#### **Por ejemplo:**

Imaginemos que estudiamos la dieta de 8 personas y extraigamos conclusiones a partir de los datos. Hemos estudiado dos variables: la *cantidad de calorías ingeridas por persona y día* y el *% de grasas ingeridas*.

Supongamos que, después de hacer todos los cálculos de nuestro estudio, estos son los resultados que hemos obtenido:

**La media** de calorías ingeridas por persona y día son 1800 kcal, con una **moda** de 1810 kcal y una **desviación típica** de 50 kcal.

El **porcentaje medio** de grasas que ingieren es un 40%, con una **moda** de 32% y una **desviación típica** de 17 (17%).

¿Qué consecuencias podríamos sacar de estos resultados?

#### **Mostrar retroalimentación**

Respecto a las calorías ingeridas, los valores de la media y la desviación típica nos dicen que **la mayoría ingiere entre 1750 y 1850 kcal diarias, siendo el valor más repetido 1810 kcal**. Podemos interpretar que estas 8 personas tienen un **consumo de calorías diarias razonablemente sano** (acercándose a bajo), siendo su comportamiento muy parecido (puesto que hay poca variación entre las respuestas).

respuestas:

Respecto del % de grasas ingeridas, si bien no ingieren demasiadas calorías (como hemos visto antes) **el porcentaje medio de grasas en su alimentación es demasiado elevado.**

Sin embargo, la desviación típica indica que las respuestas están entre 23% y 57%, lo que refleja que hay grandes diferencias entre unas personas y otras. Algunas cuidan la cantidad de grasa en su dieta pero otras hacen un consumo peligroso para la salud.

Hay mucha disparidad en las respuestas:



*Unos comen un número adecuado de calorías y una cantidad saludable de grasa. Seguramente hacen una dieta mediterránea, que es una de las mejores del mundo según los expertos... ¡Y la tenemos tan cerca!*  
Fotografía en [freeimages](#) de [iker.freeimages](#) licencia



*Otros no comen demasiadas calorías, pero de lo que comen un porcentaje demasiado elevado es grasa. Pueden ser personas con poco tiempo para comer que abusan de la comida rápida, poco saludable.*  
Fotografía en [freeimages](#) de [iker](#) . [freeimages](#) licencia

## Ejercicio resuelto

Otro ejemplo:



Fotografía en [pixabay](#) de PublicDomainPictures. Licencia CC0

**Variable:** Índice de masa corporal (IMC) (para saber qué es el índice de masa corporal o IMC mira [éste enlace](#))

**Población:** niños de nuestro país.

**Muestra:** 200 niños seleccionados al azar.

IMC	Frecuencia absoluta
-----	---------------------

menor	c
-------	---



que 18,5 IMC	Frecuencia absoluta
18,5 - 24,9	110
25-29,9	66
30-34,9	16
35-39,9	1
40 o mayor	1

**¿Qué podemos decir a la vista de los datos?** Seguro que se te ocurren muchas cosas, y eso que aún no hemos hecho ni una cuenta. Casi a simple vista, vemos que...

#### Mostrar retroalimentación

- 110 niños (el **55%**) tienen un IMC **saludable**, entre 18,5 y 24,9
- un **3%** tiene un peso demasiado **bajo**,
- y un **9% (18 de los 200)** tienen **obesidad**.

Además, el **IMC medio** es **24,51**, es decir, **normal cercano al sobrepeso**. (Hemos tomando como marcas de clase en los intervalos extremos la media de las 6 respuestas para el intervalo "menor que 18,5", que es 18 y la respuesta en el intervalo "40 o mayor", que ha sido 40).

Parece claro que todos tenemos que poner de nuestra parte, y las autoridades sanitarias, para cuidar la alimentación de los niños, ya que un niño obeso es un joven enfermo. Las prisas, el trabajo, la comodidad no pueden ganar la batalla a la dedicación y el interés por lo que nuestros niños comen.

Ya ves. Se pueden sacar muchas conclusiones, a veces de gran importancia social, con un estudio estadístico... ¡y eso que sólo hemos metido un pie en el mundo de la Estadística!

## Ejercicio resuelto

### Autoevaluación

Ahora extrae tú tus propias **conclusiones**, sin hacer nuevos cálculos, sobre:

1. El estudio sobre el número de piezas de fruta.
2. Las variables 1 y 2 que estudiaste en el apartado anterior.

#### Mostrar retroalimentación

1. Posibles conclusiones son:

- El número **medio** de piezas de fruta consumidas al día es **1,52**.
- Las respuestas más frecuente son **1 y 2** piezas de fruta.
- Más de la mitad (52%) toma **una o ninguna** pieza de fruta.
- Sólo un 20% toma **3 o más** piezas.
- Se debe promover un mayor consumo de frutas como parte de una alimentación completa y sana. Debemos sustituir por fruta la bollería industrial, perjudicial para nuestra salud y calidad de vida.

2. Veamos:

- **Variable 1:**

- Por término medio la puntuación de la imagen personal es **3,51**, lo que indica bajo nivel de autoestima respecto del aspecto físico.
- La puntuación **más frecuente es 5**, indicada por 103 de los 230 encuestados.
- Casi el **45% "suspende"** su imagen.
- Sólo **el 0.52% está satisfecho** con su imagen (puntuándola con 7 o más).
- Se deben promover patrones de belleza saludables y realistas, ya que los actuales producen frustración y baja autoestima respecto del aspecto físico.
- **Variable 2:**
  - El nivel medio es **8,56**.
  - La respuesta **más frecuente es 10**.
  - Un elevado porcentaje (superior al **90%**, 208 de los 230 encuestados) atribuye una gran importancia como causa de su enfermedad (respondiendo 7 o más)
  - Sólo un **0.34%** atribuye un nivel de importancia **inferior a 5**.
  - Parece claro que la influencia de los cánones de belleza promovidos socialmente es un factor importante en la aparición de trastornos alimentarios, por tanto todos los componentes de la sociedad deberían tomar medidas al respecto, para que disminuyan los casos de éstas enfermedades.

*Para saber más*

Para hacer tantas operaciones, es mejor que uses una calculadora. Aquí tienes **4 vídeos** que te enseñan a manejarla, dependiendo de la calculadora que uses.

Pero si quieres ahorrarte hacer todas las cuentas, y además representar los datos en una bonita gráfica, lo mejor es usar una hoja de cálculo. En estos **vídeos** te enseñan su funcionamiento.

¿Te parece difícil? Ya sabes que, al principio, casi todo lo es. Pero si practicas un poco, verás que no es para tanto y, sobre todo, te ahorrarás tanto cálculos que, al final, estarás encantado de haber aprendido a manejarla.

Para ayudarte a dar tus primeros pasos, ¿te acuerdas del estudio del consumo semanal de golosinas de los niños de un colegio? Esta es su **hoja de cálculo en excel** y así se vería en **versión libre en formato ods**.



Imagen en Wikimedia  
commons de  
[Michael180](#). Licencia  
GNU

## 7. ¿Y si la información no es fiable?

Constantemente estás recibiendo información en gráficos pero muchos están manipulados voluntariamente. Fíjate en el siguiente caso extraído del periódico [eldiario.es](http://eldiario.es)



En un programa de televisión se usó el gráfico de la izquierda con la evolución del paro registrado en el año 2013. El eje estaba tan cortado que parecía que el paro se había reducido a casi cero durante esos meses. En este caso puedes comprobar lo importantísimo que es usar la escala adecuada en cada eje.

Observa estos dos diagramas de barras. En ambas tablas los porcentajes son los mismos (27 % prefiere Sociales, algo menos del doble que el 15 % que prefiere Mates). Sin embargo en la primera gráfica (i precisamente la que la hoja de cálculo de Google ofrece por defecto! ) el porcentaje de preferencias de las Sociales sobre las Mates parece ser siete u ocho veces mayor. (Información extraída de la [página de Francisco Durán Ceacero](#))

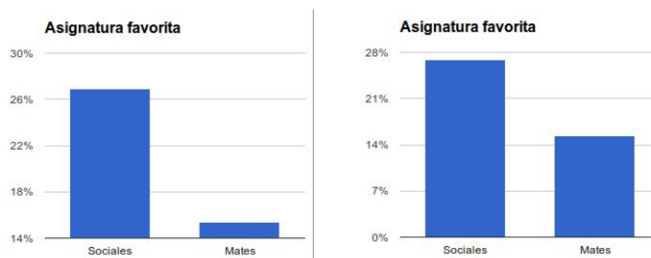


Imagen en site de [Francisco Durán Ceacero](#). Licencia CC 3.0



Sin embargo, los gráficos han evolucionado muchísimo con la tecnología y su utilidad es enorme. En el siguiente [enlace](#) del periódico [El Mundo](#) puedes ver una comparativa de las leches de España realizada por la OCU.

## 8. Resumen

### Importante

En este tema se trabajan los aspectos más importantes relacionados con la Estadística, una parte de las Matemáticas que se ocupa de estudiar grandes cantidades de datos sobre un tema concreto. Verás desde cómo se hace una encuesta hasta cómo hacer cuentas con los datos y sacar conclusiones. Aplicamos las técnicas de la estadística a ejemplos sobre alimentación saludable y al análisis crítico de la información presentada con gráficos.

### Importante

#### **A lo que queremos estudiar lo llamamos variable estadística**

Las variables que podemos expresar con números se llaman cuantitativas.

Las variables cuantitativas pueden ser discretas (número de libros leídos el último mes) o continuas (la estatura).

Las variables que no se expresan numéricamente se llaman cualitativas (tipo de café que se consume en el desayuno).

#### **¿Qué o quiénes son los que queremos analizar?**

La muestra se debe escoger con cuidado para que sea representativa.

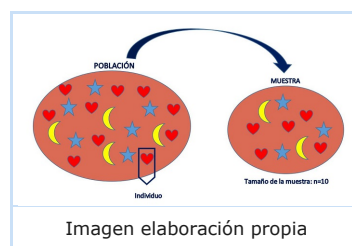


Imagen elaboración propia

### Importante

#### **¿Cómo obtenemos los datos?**

Los datos pueden obtenerse de manera directa o indirecta.

La obtención directa puede hacerse por medición o cuestionario.

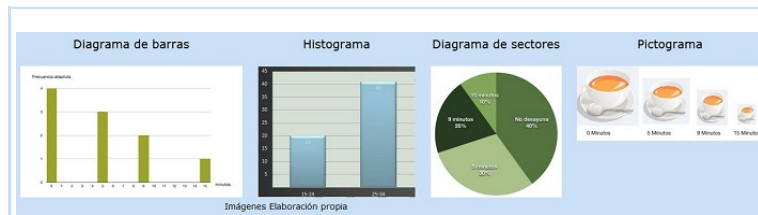
#### **¿Qué hacemos con los datos obtenidos?**

Los datos obtenidos se organizan en una tabla de frecuencias (a veces los datos se agrupan en intervalos)

Tendremos frecuencias absolutas ( $n_i$ =número de veces que se ha observado el valor  $x_i$ ) y frecuencias relativas ( $f_i = n_i/n$  donde  $n$  es el número total de observaciones)

## Importante

### Representamos los datos con gráficos



## Importante

### Números que nos resumen la información de todos los datos

Con los datos podemos obtener valores que representan información sobre los datos recogidos. Los más importantes son:

<b>Media</b> $\bar{x} = \frac{\text{suma de todos los } x_i \cdot f_i}{n}$ o bien $\bar{x} = \frac{\text{suma de todos los } x_i \cdot n_i}{n}$	<b>Moda</b> = el valor que más se repite (el que tiene mayor frecuencia absoluta)
<b>Varianza</b> $\sigma^2 = \frac{\text{suma de todos los } x_i^2 \cdot f_i}{n} - \bar{x}^2$ o bien $\sigma^2 = \frac{\text{suma de todos los } x_i^2 \cdot n_i}{n} - \bar{x}^2$	<b>Desviación típica</b> $\sigma = \sqrt{\text{varianza}}$
<b>El coeficiente de variación</b> = CV (desviación típica / la media) = $\frac{\sigma}{\bar{x}}$	

## Importante

### Sacamos información a partir del estudio estadístico.

Por último, no olvidemos que el objetivo es interpretar los datos y extraer información sobre la variable estudiada.

Los cálculos realizados nos llevan a valores que debemos interpretar.

Cuidado con los engaños intencionados de la información que recibimos a través de los medios de comunicación

## 9. Para aprender... hazlo tú

En este tema has aprendido a hacer estudios estadísticos. Ahora vas a poner en práctica lo aprendido y... lo que aún no hayas entendido, te quedará más claro después de estudiar las siguientes actividades.

### Actividad de lectura

Se ha realizado el siguiente estudio estadístico:

Se han seleccionado al azar 25 personas que hacen algún tipo de dieta estricta, tipo "milagro", (de distinto sexo, edad, condición social).

Sobre esa muestra se ha estudiado el % de reducción de vitaminas que sufren como consecuencia de su dieta, siendo los resultados:

10%, 50%, 72%, 30%, 40%, 65%, 24%, 8%, 45%, 30%, 45%, 67%, 76%, 81%, 23%, 43%, 12%, 58%, 32%, 66%, 34%, 23%, 34%, 71%, 15%.



Fotografía en morguefile de  
[keyseeker](#). Licencia  
[morguefile](#)

- Realiza la **tabla completa de frecuencias** (agrupa los datos en intervalos de amplitud razonable para tu trabajo).
- Realiza el correspondiente **histograma**.
- ¿Cuál es el **intervalo modal**?
- ¿Qué **porcentaje medio** de vitaminas se pierde?
- Halla la **varianza**, **desviación típica** y el **coeficiente de variación**.
- Extrae conclusiones** a partir de los datos sobre las dietas "milagro" y las consecuencias para la salud.

### Mostrar retroalimentación

Vamos allá...

#### a. Tabla de frecuencias:

Aunque no es la única opción, vamos a agrupar los datos en intervalos de amplitud 20 aproximadamente:

Porcentaje de reducción de vitaminas	Frecuencia absoluta	Frecuencia relativa	Frecuencia absoluta acumulada	Frecuencia relativa acumulada
0-19	4	0.16	4	0.16
20-39	8	0.32	12	0.48
40-59	6	0.24	18	0.72
60-79	6	0.24	24	0.96
80-100	1	0.04	25	1
Suma total:	25	1		

#### b. Histograma:

Frecuencia  
absoluta

8



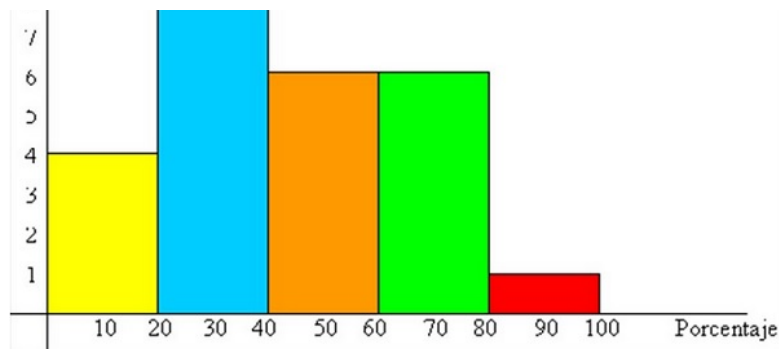


Imagen elaboración propia

#### b. Intervalo modal:

El intervalo modal es **20-39**, cuya frecuencia absoluta vale 8 y es superior a las demás.

#### c. Porcentaje medio de vitaminas que se pierde:

Lo que nos están pidiendo aquí es, sencillamente, que calculemos la media de la variable.

Para hacer la media, necesitamos un representante de cada intervalo: tomaremos, como es habitual, el punto medio. Así, **los representantes son: 9,5 - 29,5 - 49,5 - 69,9 - 89,5** (que hemos calculado sumando los valores de los extremos de cada intervalo y dividiendo entre 2 los resultados)

Usando estas marcas de clase y las frecuencias absolutas calculamos la media:

$$\begin{aligned}
 & (9,5 \times 4 + 29,5 \times 8 + 49,5 \times 6 + 69,9 \times 6 + 89,5 \times 1) \div 25 = \\
 & = (38 + 236 + 297 + 419 + 89,5) \div 25 = \\
 & = 1079,9 \div 25 = 43,19
 \end{aligned}$$

**Así que la reducción media de vitaminas es del 43,1%**

#### d. Cálculo de la varianza, desviación típica y coeficiente de variación:

Lo primero que tenemos que calcular es la varianza. Paso a paso, tomando como valores los representantes (marcas de clase) anteriores:

1.

Valores al cuadrado: 90,25 - 870,25 - 2450,25 - 4886 - 8010,25

2.

Media de los valores al cuadrado:

$$(90,25 \times 4 + 870,25 \times 8 + 2450,25 \times 6 + 4886 \times 6 + 8010,25) : 25 = 59350,75 : 25 = 2374,03$$

3.

A lo anterior le restamos la media al cuadrado y tendremos la varianza:  
 $2374,03 - 43,19^2 = 2374,03 - 1865,37 = 508,66$

De modo que **la varianza es: 508,66**

Luego calculamos la desviación típica; solo hay que hacer la raíz cuadrada al resultado anterior:

$$\sqrt{508,66} = 22,55$$

e.

Por tanto, **la desviación típica es 22,55 %**

Por último, dividiendo la desviación típica entre la media obtenemos el coeficiente de variación:

$$22,55 \div 43,19 = 0,52$$

f.

Luego **el coeficiente de variación es 0,52**

g. **Algunas de las conclusiones que podemos extraer son las siguientes:**

- El 84% de los encuestados sufre una reducción de vitaminas del 20% o más.
- Sólo el 16% tiene una reducción de sus vitaminas de menos del 20%.
- Lo más frecuente es que se reduzcan las vitaminas entre el 20 y el 40%, siendo el porcentaje medio del 43.1% (de media se reducen casi a la mitad las vitaminas ingeridas).
- La varianza nos dice que es muy probable experimentar una reducción de vitaminas entre 20,68% y 65,52%
- Vemos bastante dispersión respecto del porcentaje medio, lo que indica que no todas las personas experimentan el mismo efecto ante una dieta milagro, o que no todas las dietas milagro provocan el mismo efecto.
- Las dietas milagro conllevan la falta de sustancias esenciales para el organismo, como las vitaminas estudiadas en nuestro caso, las sales minerales, etc. Las grasas tomadas adecuadamente son fundamentales, entre otras cosas permiten asimilar las llamadas vitaminas liposolubles (disueltas en la grasa), por ello una drástica disminución de la grasa en la dieta provoca un serio déficit vitamínico.

**Recuerda**, las dietas milagro de poco ayudan en realidad a perder peso... y ayudan a perder salud, así que no dudes en ponerte en manos de tu médico si realmente quieres perder peso pero... sentirte bien.

## Actividad de lectura

**Queremos saber si el consumo habitual de "comida basura" influye en la obesidad, en la población Andaluza.**

Para ello se ha estudiado el IMC en 10 personas adultas consumidoras habituales de comida rápida de alto contenido en grasas saturadas industriales, siendo éstos los resultados:

1, 28, 39, 27, 36, 39, 29, 38, 41, 35.



Fotografía en [pixabay](#) de  
Meditations. Licencia CC0

a. ¿Cómo habría que haber seleccionado a las 10 personas? ¿Es una muestra suficiente? Especifica la información sobre la población y la muestra que consideres necesaria para que los datos que se obtengan sean útiles



necesaria para que los datos que se obtengan sean útiles.

- b. Describe el rango de valores que toma la variable y calcula el IMC medio.
- c. Extrae conclusiones, para ello deberás volver a visitar la página sobre el IMC, prestando atención a la tabla.
- d. ¿Qué

### Mostrar retroalimentación

Veamos lo bien que lo has hecho...

a. Para saber la influencia de la comida basura en los casos de obesidad en Andalucía, es obvio que **no bastará con preguntar a sólo 10 personas, es una muestra insuficiente**. Si preguntamos a 10 personas, la población a estudiar debe ser más reducida. Una población adecuada puede ser, por ejemplo, las personas de nuestro vecindario que comen un considerable número de veces "comida basura". Además, **debemos seleccionar a los 10 entrevistados al azar**.

b. Los valores obtenidos oscilan entre 28 y 41, siendo el **IMC medio de 34,3**.

$$(1+28+39+27+36+39+28+39+41+35):10=34,3$$

c. **7 de las 10 las personas** encuestadas tienen un IMC igual o mayor que 30, lo que indica que **son obesas**. Las otras tres padecen sobrepeso no muy lejano a la obesidad. Además, el IMC medio revela claramente las malas consecuencias de la comida basura frecuentemente ingerida.

d. Por tanto, y aunque se podrían (y deberían) tener más factores y cálculos en cuenta, **con la información vista parece bastante claro que las personas asiduas a la "comida basura" sufre obesidad**, con los graves problemas que ello conlleva.

**Recuerda**, comer alguna vez en tu restaurante favorito de comida rápida no "es delito" alimenticio, **pero abusar de ello puede provocarte serios problemas de salud**.

**La dieta mediterránea es mucho mejor, está hasta más rica y puedes abusar todo lo que quieras y tu salud lo agradecerá.**

# Imprimible

---

Descargar [PDF](#)

## Aviso Legal

---

El presente texto (en adelante, el "**Aviso Legal**") regula el acceso y el uso de los contenidos desde los que se enlaza. La utilización de estos contenidos atribuye la condición de usuario del mismo (en adelante, el "**Usuario**") e implica la aceptación plena y sin reservas de todas y cada una de las disposiciones incluidas en este Aviso Legal publicado en el momento de acceso al sitio web. Tal y como se explica más adelante, la autoría de estos materiales corresponde a un trabajo de la **Comunidad Autónoma Andaluza, Consejería de Educación y Deporte (en adelante Consejería de Educación y Deporte)**.

Con el fin de mejorar las prestaciones de los contenidos ofrecidos, la Consejería de Educación y Deporte se reserva el derecho, en cualquier momento, de forma unilateral y sin previa notificación al usuario, a modificar, ampliar o suspender temporalmente la presentación, configuración, especificaciones técnicas y servicios del sitio web que da soporte a los contenidos educativos objeto del presente Aviso Legal. En consecuencia, se recomienda al Usuario que lea atentamente el presente Aviso Legal en el momento que acceda al referido sitio web, ya que dicho Aviso puede ser modificado en cualquier momento, de conformidad con lo expuesto anteriormente.

**Régimen de Propiedad Intelectual e Industrial sobre los contenidos del**

