

Medidas de Tendencia Central



COMIXPLAIN

Este cómic fue creado en el proyecto de investigación Comixplain, financiado por la Innovation Call 2022 de la Universidad de Ciencias Aplicadas de St. Pölten, Austria.

Equipo:

Victor-Adriel De-Jesus-Oliveira
Hsiang-Yun Wu
Christina Stoiber
Magdalena Boucher
Alena Ertl

Contacto:

victor.oliveira@fhstp.ac.at

Ilustraciones:

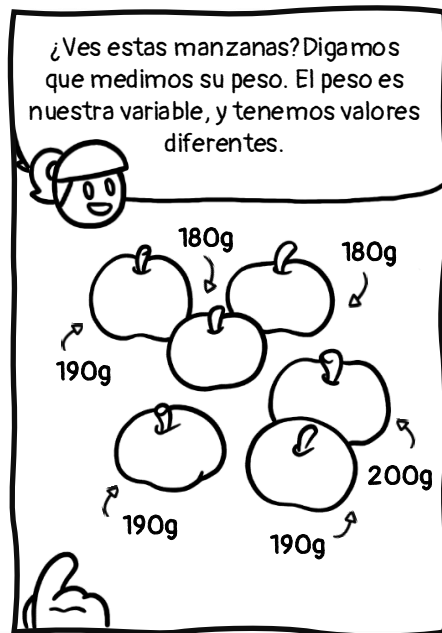
Magdalena Boucher & Alena Ertl

Traducción:


Yhonatan Jesus Iquiapaza



<https://fhstp.github.io/comixplain>



La mejor manera de describir una variable es informar los valores, y la frecuencias en que ellos aparecen. Esto se llama la **DISTRIBUCIÓN** de la variable.



		🍏
🍏	🍏	
🍏	🍏	🍏
180	190	200

Si visualizamos el peso de nuestras manzanas, la distribución sería así, ya que todas las manzanas de la cesta pesan aproximadamente lo mismo.


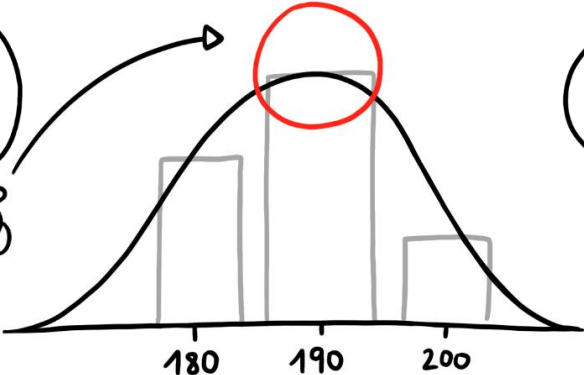
Peso de las manzanas →




Tiene la forma de una campana...




¡Exacto! Si hubiese otras manzanas en la cesta, nuestra mejor estimación para su peso sería un valor cercano a aquel punto.

Oh, vaya. Pero... ¿Cómo calcularía ese valor?



Ese punto medio de nuestra distribución representa bien nuestros datos – nosotros lo llamaremos **TENDENCIA CENTRAL**.

Podemos resumir a la tendencia central de varias maneras. La **MEDIA** es la manera más común, y también la más fácil de calcular.




Estas son nuestras seis manzanas:

🍏	🍏	🍏	🍏	🍏	🍏
200	180	190	190	190	180

Para calcular la media, sumamos todos los valores de peso...

↓ ↓ ↓ ↓ ↓ ↓

$$200 + 180 + 190 + 190 + 190 + 180$$

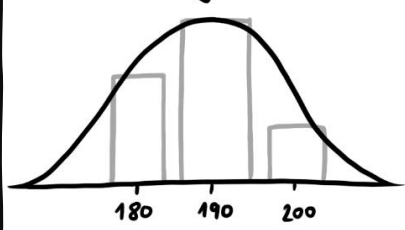
... y después dividimos por el número de manzanas que tenemos

...


$$\frac{(200 + 180 + 190 + 190 + 190 + 180)}{6}$$

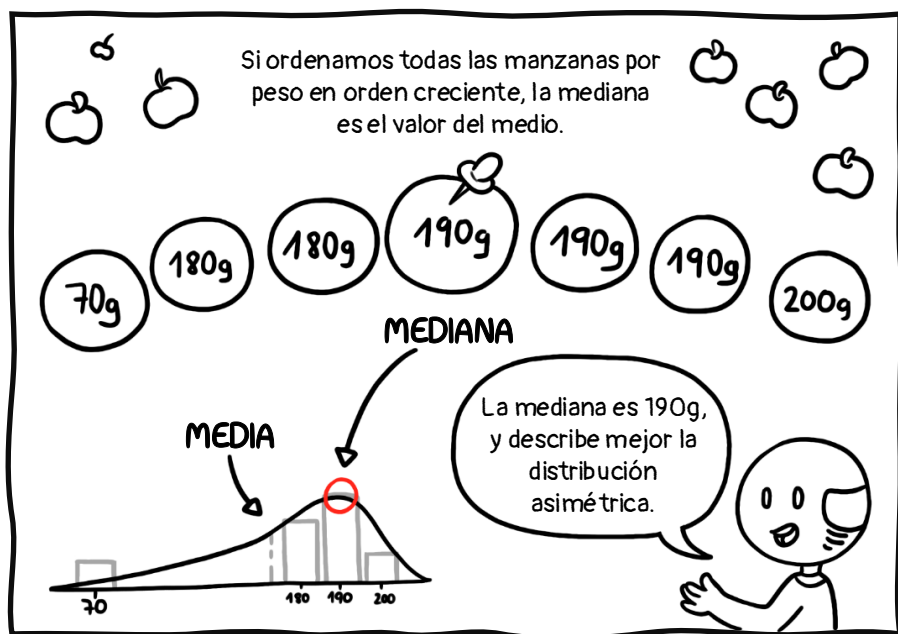
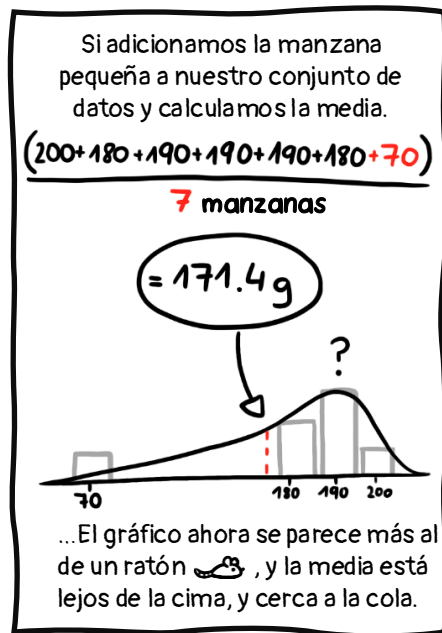
Eso nos da un valor medio de

188.3g



Oh, eso es casi el valor de la parte superior de nuestro gráfico en forma de campana





En este caso, nuestro gráfico de repente tiene dos picos.

Como ratones enamorados

Seguro...

Peso de las manzanas grandes y pequeñas

De cualquier forma, podemos usar algo llamado **MODA** para describir la tendencia central de nuestra distribución si esta tuviera varios picos.

La moda define el o los valores que ocurren con más frecuencias en un conjunto de datos.

En este caso, tenemos varias modas. Pero algunas veces existirá solamente una o ninguna moda.

Puedes aplicar la media, mediana y moda a diferentes muestras de manzanas. Pero a menudo, algunos representarán los datos mejor que otros.

180, 180, 190, 190, 190, 200

$M = 188.3$ ↪ Buenos
 $MD = 190$ ↪ parámetros
 $Moda = 190$ ↪

70, 180, 181, 190, 191, 191, 200

$M = 171.8$ ↪ Buenos
 $MD = 190$ ↪ parámetros
 $Moda = 191$ ↪

60, 70, 70, 70, 80, 180, 180, 190, 190, 190, 200

$M = 134.5$ ↪ Buen
 $MD = 180$ ↪ parámetro
 $Moda = 70 \text{ \& } 190$ ↪

Ok, gracias... Aprendi mucho. Ahora solo tengo que aplicar a los datos que tengo que presentar.

Son de un aplicativo que rastrea mediciones de frecuencia cardiaca.

User ID	Heart Rate (bpm)	Time of Use	User Rating
1	45	13:00	1
2	50	9:00	5
3	55	10:00	3
4	57	9:00	4
5	63	14:00	5
6	70	15:00	5
7	65	16:00	4
8	75	15:00	2

↪ Echa un vistazo a tus datos y sigue los mismo pasos que acabamos de hacer con las manzanas. Puedes usar las siguientes páginas para tus cálculos.



Antes de pasar de página, trata de calcular la media, mediana y moda para cada variable, y verifique cuál parámetro es más adecuado para describir la tendencia central.

Puedes hacer anotaciones en esta página.

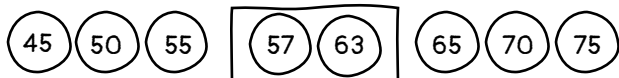
Siéntete libre de verificar tus cálculos.
Puedes hacer más anotaciones en esta página.

FRECUENCIA CARDIACA

Calculando la Media:

$$\frac{45+50+55+57+63+70+65+75}{8 \text{ Usuarios}} = \frac{480}{8} = 60 \text{ bpm}$$

Calculando la MEDIANA:



Si hubiera dos valores centrales, la media de los valores es la mediana:
 $(57+63)/2 = 60 \text{ bpm}$

Calculando la MODA:

45, 50, 55, 57, 63, 70, 65, 75

Cada valor solo existe una vez –
 la moda no existe!

Si la distribución de los valores es simétrica, sin distorsiones, la media es generalmente igual a la mediana.



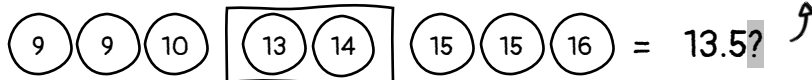
TIEMPO DE USO MÁS FRECUENTE

Calculando la Media:

$$\frac{9+9+10+13+14+15+15+16}{8 \text{ Usuarios}} = \frac{101}{8} = 12.6?$$

El tiempo de uso no es un valor cuantitativo – por lo tanto, calcular la media y mediana no tiene sentido alguno!

Calculando la MEDIANA:



Calculando la MODA:

9:00, 10:00, 13:00, 14:00, 15:00, 16:00
 2x 1x 1x 1x 2x 1x = 2 modas:
 9:00 & 15:00

Moda no solo es adecuado para distribuciones multimodales, sino también cuando se trabaja con datos ordinales y categóricos.



CLASIFICACIÓN POR ESTRELLAS

Calculando la Media:

$$\frac{1+2+3+4+4+5+5+5}{8 \text{ Usuarios}} = \frac{29}{8} = 3.6 \text{ estrellas}$$

Calculando la MEDIANA:



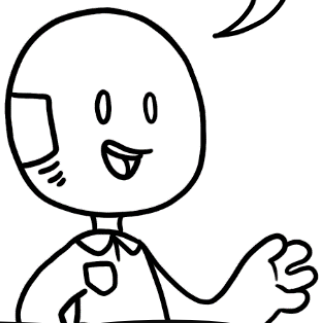
Calculando la MODA:

1 2 3 4 5
 1x 1x 1x 2x 3x = 5 estrellas

Para conjuntos de datos con distribución asimétrica, la mediana es la mejor manera para describir la tendencia central.



Lenguajes de programación como **R**, te ayudan a calcular la tendencia central de atributos en grandes conjuntos de datos. Con bibliotecas de R, como **tidyverse**, podrás visualizar rápidamente la distribución de datos.



	model	year	hwy
1	jetta	1999	44
2	corolla	2008	37
3	civic	2008	36
4	civic	2008	36
5	corolla	1999	35
6	altima	2008	32
7	sonata	2008	31
+ otros 227 items			

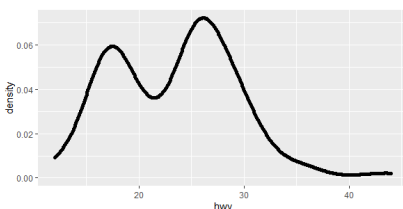
Dentro de tidyverse, tendrás acceso a los conjuntos de datos, como **mpg** con datos de economía de combustible. Este incluye 11 atributos, como modelo del carro (**model**), año de fabricación (**year**) y millas por galón en carretera (**hwy**).



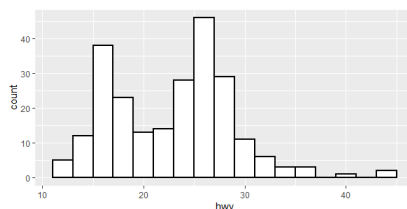
Puedes usar **ggplot**, que está incluido en tidyverse, para visualizar la distribución de datos de millas por galón en carretera (**hwy**) usando un histograma, una curva de densidad, o ambos.

`install.packages("tidyverse")` # Instalelo solo la primera vez que va usar la biblioteca

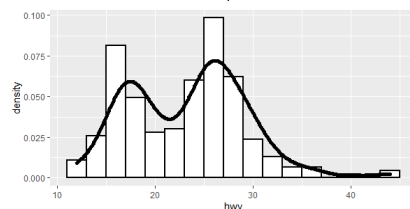
```
library(tidyverse)
plot <- ggplot(mpg, aes(x=hwy))
plot +
  geom_density()
```



```
plot +
  geom_histogram(
    colour="black",
    fill="white")
```



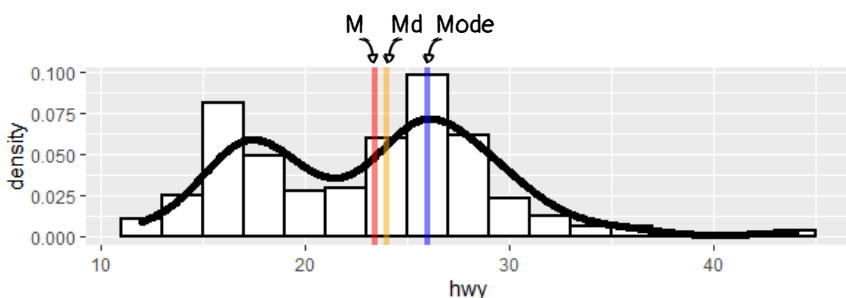
```
plot +
  geom_histogram(aes(y=..density..),
    colour="black",
    fill="white") +
  geom_density()
```



`mean(mpg$hwy)` = 23.4

`median(mpg$hwy)` = 24

`library(modeest)`
`mlv(mpg$hwy)` = 26



R incluye funciones nativas para calcular la media y mediana. Para la moda, puedes crear tu propia función o usar Most Likely Values (**mlv**) de la biblioteca **modeest**.



Fuentes:

Downey, A. (2014). Think stats: exploratory data analysis. O'Reilly Media, Inc.

Field, A. (2022). An adventure in statistics: The reality enigma. Sage.