CAPÍTULO 4:RESUMIENDO DATOS NUMÉRICAMENTE

Estadística - Lic. en Nutrición – UCEL
Marta Ruggieri, Julia Fernández, M. Eugenia Tesser

MEDIDAS RESUMEN

Vamos a aprender a calcular un conjunto de medidas que permiten resumir la información de variables cuantitativas.

Medidas de tendencia central	Medidas de dispersión
• Media aritmética • Mediana • Moda	• Variancia - Desvío estándar • Coeficiente de variación • Rango • Rango intercuartílico

MEDIDAS RESUMEN

También veremos como usar un conjunto de estadísticas de orden para construir el box-plot, o gráfico de caja y bigotes.

Estadísticas de orden

- Cuartiles
- Percentiles

MEDIDAS RESUMEN

Vamos a aprender a calcular un conjunto de medidas que permiten resumir la información de variables cuantitativas.

Medidas de tendencia central

- Media aritmética
- Mediana
- Moda

Medidas de dispersión

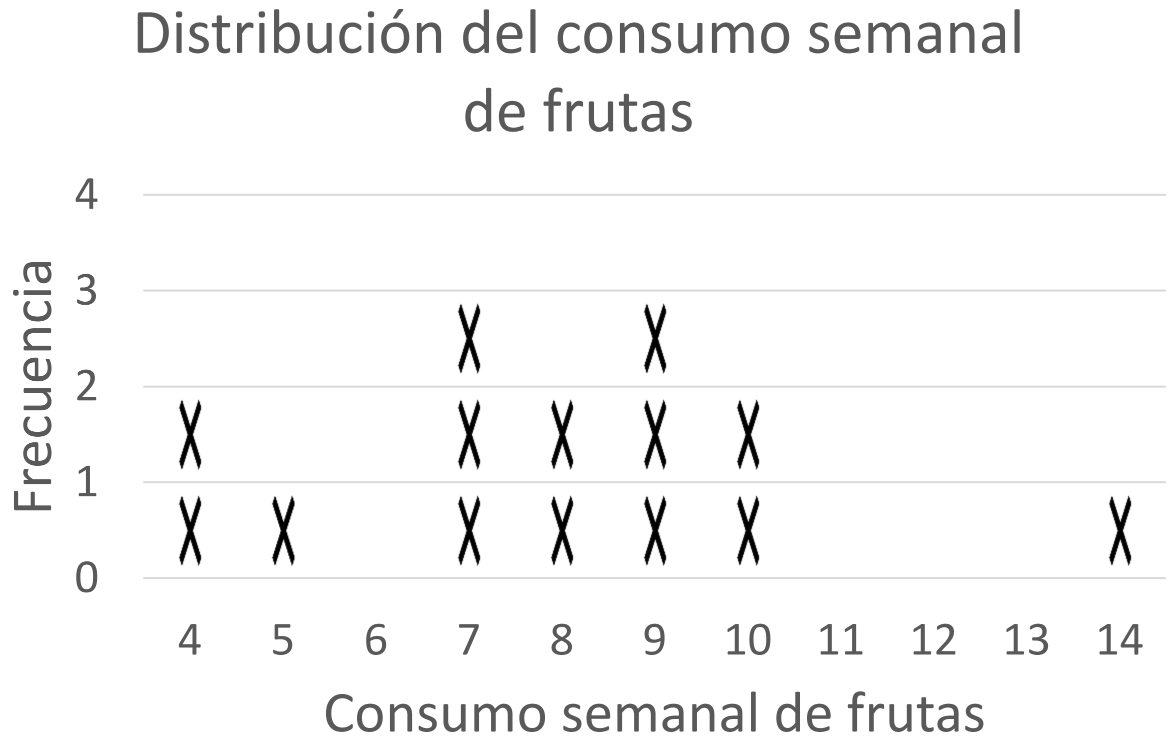
- Variancia - Desvío estándar
- Coeficiente de variación
- Rango
- Rango intercuartílico

MEDIDAS CENTRALES

Volvemos al ejemplo de la Unidad III donde se estudia la cantidad de porciones de fruta que consumen por semana las mujeres de 50 años o más:

10 – 7 – 7 – 5 – 8 – 10 – 7 – 8 – 9 – 14 – 4 – 4 – 9 – 9

Supongamos que queremos usar **un solo número** para describir la cantidad de porciones de fruta que consumen por semana este grupo de mujeres.

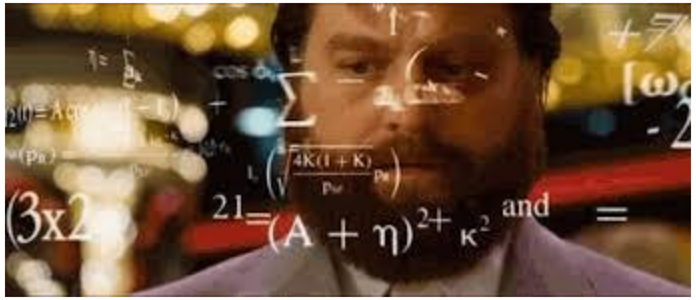


MEDIA ARITMÉTICA

La **media aritmética** de un conjunto de observaciones se obtiene sumando dichas observaciones y dividiendo por el tamaño de muestra (n). Se la simboliza $\bar{x}$.

La media poblacional es un parámetro, al cual se lo indica con la letra griega μ.

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n}$$



Donde $x_1, x_2, \dots, x_n$ representan los valores observados de la variable x en la muestra.

MEDIA ARITMÉTICA

La cantidad de porciones de fruta que consumen por semana un grupo mujeres de 50 años o más:

10 – 7 – 7 – 5 – 8 – 10 – 7 – 8 – 9 – 14 – 4 – 4 – 9 – 9

$$\bar{x} = \frac{10 + 7 + 7 + 5 + 8 + 10 + 7 + 8 + 9 + 14 + 4 + 4 + 9 + 9}{14} = 7{,}93$$

La cantidad promedio de porciones de fruta que consumen por semana las mujeres de 50 años o más es de 7,93.

Este mismo cálculo se puede hacer usando una calculadora científica. A continuación están los tutoriales según el modelo de calculadora:

https://www.youtube.com/watch?v=qguhqq0xvM0

https://www.youtube.com/watch?v=7s77eMDRkdc#action=share

MEDIA ARITMÉTICA

Supongamos que la última observación fue erróneamente registrada y en lugar de 9 es 55. ¿Cuál es el promedio ahora?

$$\bar{x} = \frac{10 + 7 + 7 + 5 + 8 + 10 + 7 + 8 + 9 + 14 + 4 + 4 + 9 + 55}{14} = 11{,}21$$

CUIDADO! La media es sensible a observaciones extremas.

MANOS A LA OBRA N° 1

<u>PORCIONES DE FRUTA-HOMBRES DE 50 AÑOS O MÁS</u>

Encuentre la cantidad promedio de porciones de frutas que consumen por semana hombres de 50 años o más.

4 – 5 – 5 – 6 – 6 – 6 – 7 – 7 – 9

<u>PESO PROMEDIO</u>

Encuentre el peso promedio de las mujeres de 50 años o más.

50,61 – 60,85 – 63,70 – 70,98 – 88,90 – 44,07 – 56,87 – 48,84 – 64,41 – 40,85 – 65,04 – 61,38 – 45,39 – 82,80

<u>PARA PENSAR 4.2. Medias combinadas</u>

La nota promedio de 3 estudiantes es 54 y la nota promedio de otros 4 estudiantes es 76. ¿Cuál es la nota promedio de los 7 estudiantes?

<u>PARA PENSAR 4.1. Una media aritmética no siempre es representativa</u>

Los resultados de los parciales de Gastón son: 7, 98, 25, 19 y 26. Calcule la nota promedio. Explique por qué la media no hace un buen trabajo para resumir las notas de Gastón.

MEDIANA

¿Qué debemos hacer en situaciones en las cuales la media aritmética no es representativa del centro de un conjunto de datos? Por ejemplo, por la presencia de valores extremos.

La **mediana** de un conjunto de n observaciones ordenadas de menor a mayor es un valor tal que la mitad de las observaciones es menor o igual a ese valor, y la otra mitad de las observaciones es mayor o igual a ese valor. La simbolizamos Q_2 o $\tilde{x}$.

La mediana es una medida de centro **resistente** o **robusta**.

MEDIANA

Ordenamos la cantidad de porciones que consumen por semana este grupo de 14 mujeres. En este caso n=14.

4 – 4 – 5 – 7 – 7 – 7 – 8 – 8 – 9 – 9 – 9 – 10 – 10 – 14

$$\tilde{x} = \frac{8+8}{2} = 8$$

¿Dónde está el centro de este conjunto de observaciones?

El 50% de las mujeres consumen 8 porciones o menos por semana, y el otro 50% de las mujeres consumen 8 porciones o más por semana.

MEDIANA

Observamos los datos ordenados del consumo semanal de frutas de los hombres que tienen 50 años o más. En este caso n=9.

4 – 5 – 5 – 6 – 6 – 6 – 7 – 7 – 9

El 50% de los hombres consumen 6 porciones o menos por semana, y el otro 50% de los hombres consumen 6 porciones o más por semana.

MEDIANA

Si el número de observaciones (n) es impar, tomamos como mediana aquella observación que se encuentre a la mitad de los valores ordenados

En cambio, si el número de observaciones (n) es par, la mediana se obtiene promediando los dos valores centrales del conjunto de datos ordenado.

En el ejemplo de 14 mujeres: $\tilde{x} = \frac{8+8}{2} = 8$

MANOS A LA OBRA N° 2

PARA RESOLVER 4.3 (página 229)

N° de observación	1	2	3	4	5	6	7	8	9	10
N° de chicos	2	3	0	1	4	0	3	0	1	2

1) Encuentre la mediana del número de hijos correspondiente a una muestra de 10 familias.
2) ¿Qué le pasa a la mediana si la observación N° 5 fue incorrectamente registrada como 40?
3) ¿Qué ocurre si la observación N° 3 fue incorrectamente registrada como -20 en lugar de 0?

MUJERES DE 50 AÑOS O MÁS

Encuentre la mediana del peso de las mujeres de 50 años o más de la base de datos de alimentación.
50,61 – 60,85 – 63,70 – 70,98 – 88,90 – 44,07 – 56,87 – 48,84 – 64,41 – 40,85 – 65,04 – 61,38 – 45,39 – 82,80

Encuentre la mediana del consumo semanal de carnes de las mujeres de 50 años o más de la base de datos de alimentación.
3 – 4 – 4 – 4 – 4 – 3 – 4 – 4 – 5 – 4 – 6 – 4 – 4 – 7

MODA

La **moda** de un conjunto de observaciones es el valor de la variable que ocurre con mayor frecuencia. Es aquel valor que tiene la frecuencia más alta entre todas las observaciones. Puede haber más de una moda.

La cantidad de porciones de frutas que consumen por semana las mujeres de 50 años o más:

4 – 4 – 5 – 7 – 7 – 7 – 8 – 8 – 9 – 9 – 9 – 10 – 10 – 14

En este caso hay dos modas y son 7 y 9, es decir, las cantidades de porciones de fruta más frecuentes que consumen por semana las mujeres de 50 años o más son 7 y 9.

MODA

A veces la moda de un conjunto de datos se encuentra alejada del centro de la distribución, y no es usada como medida de centro.

Esta medida es la única que **puede ser utilizada para resumir variables cualitativas**. En ese caso, la moda es la categoría de una variable cualitativa que presenta la mayor frecuencia.

MANOS A LA OBRA N° 3

EJEMPLO 4.3 (página 231). Diferentes medidas pueden dar diferentes impresiones

Consideremos los ingresos anuales de 5 familias de una misma zona:

$12.000 - $12.000 - $30.000 - $90.000 - $100.000

¿Cuál es el ingreso típico o característico de esta zona?
Calcule la media, la mediana y la moda analice cuál de estas medidas es más apropiada en este caso. Explique por qué.

MANOS A LA OBRA N° 3

<u>PARA RESOLVER 4.5 (página 234). La utilidad de la aleatorización</u>

☑ **PARA RESOLVER!!!**

4.5. La utilidad de la aleatorización

Recordemos que aprendimos la importancia de diseñar experi-mentos cuidadosamente. La asignación aleatoria de los sujetos a los distintos tratamientos fue un método usado para reducir el impacto del sesgo.

La aleatorización tiende a balancear los grupos con respecto a otros factores que pueden ser desconocidos por el investigador, pero podrían afectar las conclusiones.

Considere un estudio para comparar dos tipos de antibióticos para el tratamiento de anginas (estreptococos) en los chicos: Amoxicilina y Cefadroxil. En un centro sanitario, 23 chicos fueron aleatoriamente asignados para uno de los dos tratamientos. Se pensó que la edad de los chicos podría influir sobre la efectividad de los antibióticos. Las edades de los chicos para cada tratamiento se presentan a continuación.

Calcule la media, mediana y moda para cada uno de los grupos.

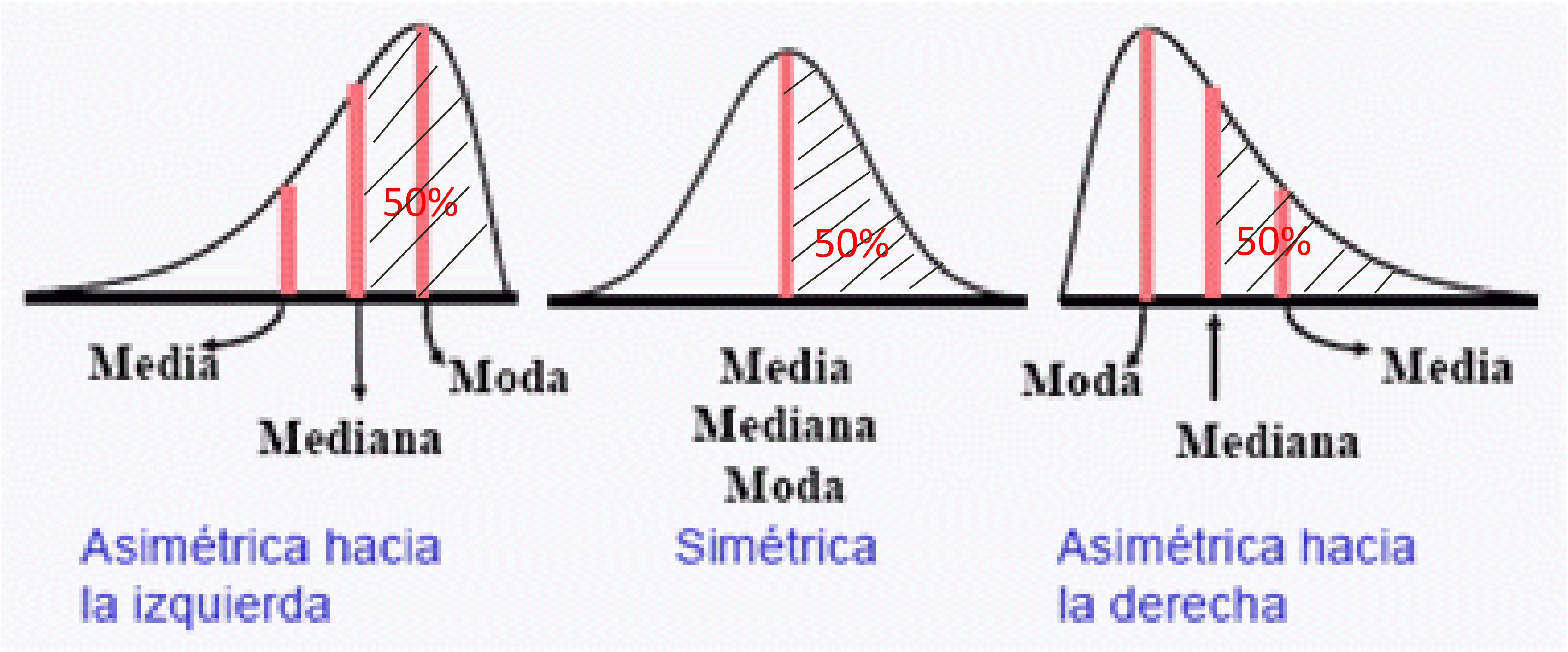
Compare los grupos con respecto a la edad.

T1: Amoxicilina (n=11)
Edad: 14 17 11 10 11 14 9 12 8 10 9
Media:
Mediana:
Moda:

T2: Cefadroxil (n=12)
Edad: 9 14 8 10 13 7 9 11 16 10 12 9
Media:
Mediana:
Moda:

- ¿Cómo son los grupos comparados con respecto a la edad?

¿QUÉ MEDIDA DE CENTRO USAR?



¿EL CENTRO ALCANZA?

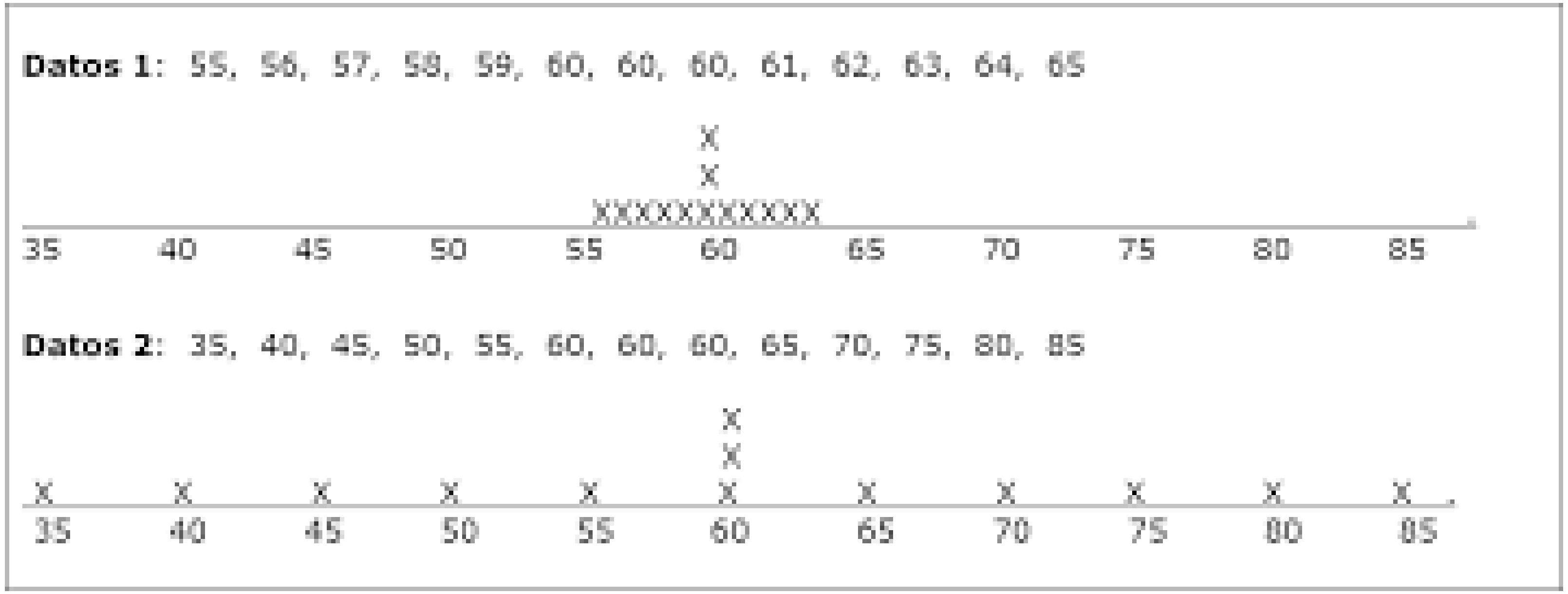
Riesgos de ignorar la variabilidad

Talla única

¿EL CENTRO ALCANZA?

Un único valor nunca es suficiente para describir un conjunto de datos. Además de la medida de centro, vamos a incorporar medidas que describan la variabilidad de los datos.

Puede ocurrir incluso que dos conjuntos de datos tengan la misma medida de centro y variabilidades muy diferentes:



MEDIDAS RESUMEN

Vamos a aprender a calcular un conjunto de medidas que permiten resumir la información de variables cuantitativas.

Medidas de tendencia central

- Media aritmética
- Mediana
- Moda

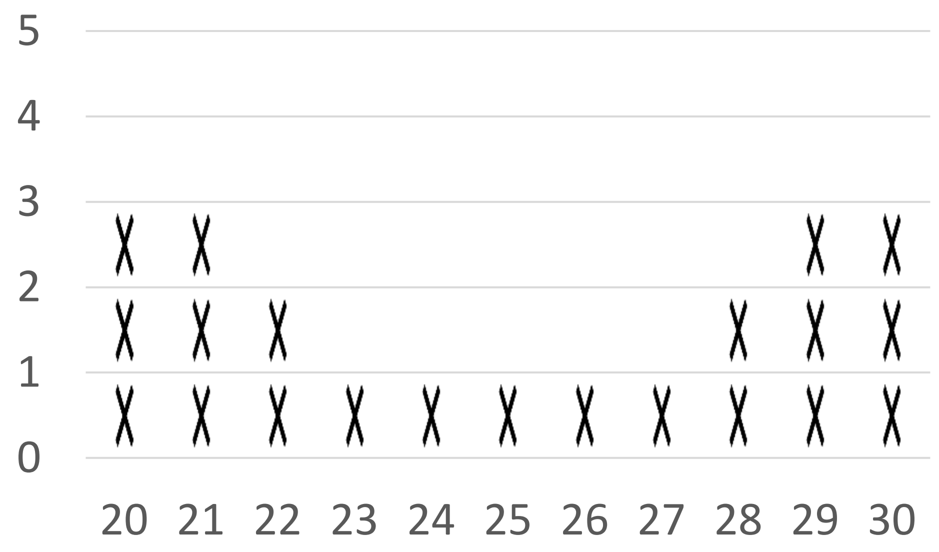
Medidas de dispersión

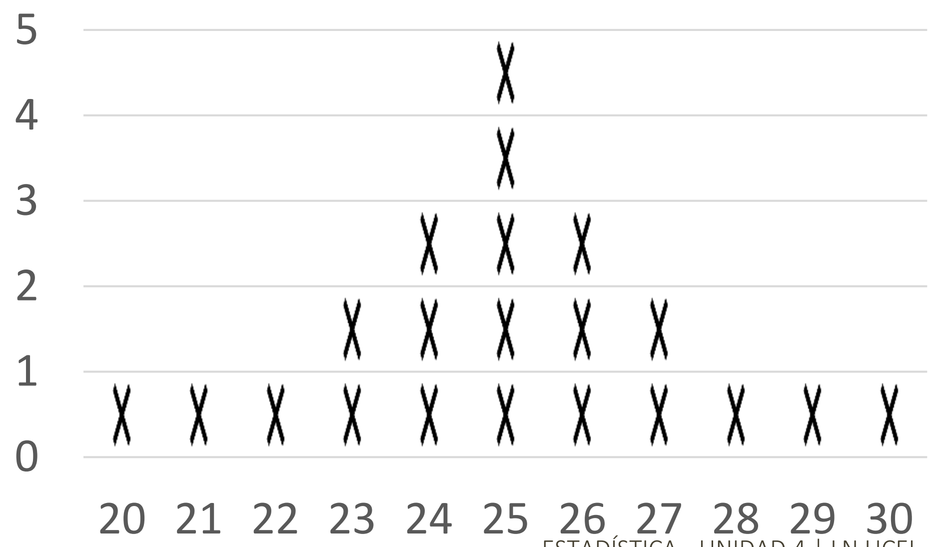
- Variancia - Desvío estándar
- Coeficiente de variación
- Rango
- Rango intercuartílico

RANGO

El **rango** es la diferencia entre la mayor y la menor observación de un conjunto de datos.

Si bien esta medida de variabilidad es la más fácil de calcular, no es la mejor ya que no tiene en cuenta toda la información de los valores en el centro de la distribución. Veamos el siguiente ejemplo:



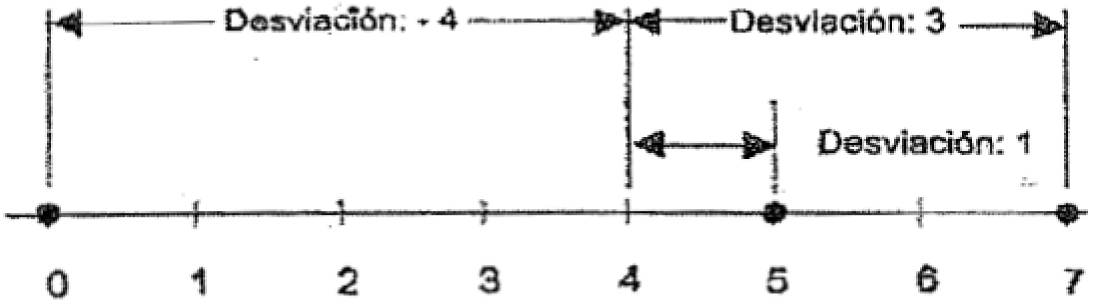


VARIANCIA Y DESVÍO ESTÁNDAR

Otra medida de la <u>dispersión</u> que podemos tomar son las diferencias entre cada valor observado y el valor típico que describe al conjunto de datos: la media.

Tomemos el conjunto de datos: $0 - 5 - 7$, cuya media es $\bar{x} = 4$. ¿Cuán diferentes son los valores observados de la media aritmética?:

- $0 - 4 = -4$
- $5 - 4 = 1$
- $7 - 4 = 3$



VARIANCIA Y DESVÍO ESTÁNDAR

La suma de los desvíos de las observaciones con respecto a la media siempre resulta igual a 0: en el ejemplo $-4+1+3=0$. Para solucionar esta situación, se pueden elevar los desvíos al cuadrado antes de sumarlos:

$$(-4)^2+1^2+3^2=16+1+9=26$$

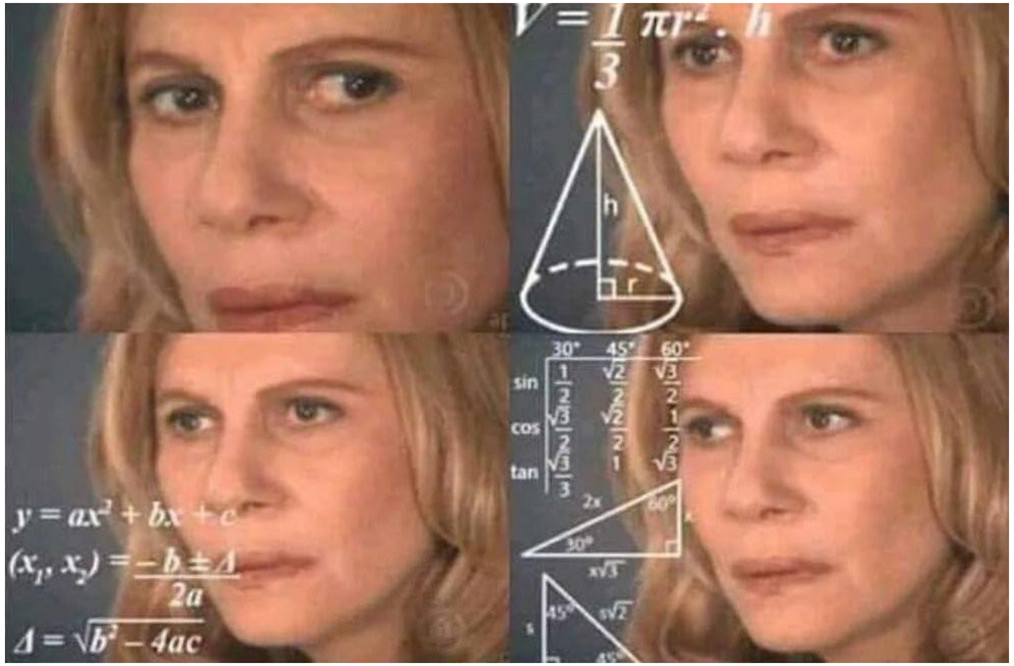
La **variancia muestral**,$\boldsymbol{s^2}$, se define como la suma de los cuadrados de los desvíos de las observaciones con respecto a la media, dividida por la cantidad de observaciones menos uno: $n-1$. Está medida en las unidades de la variable elevadas al cuadrado.

El **desvío estándar muestral**, $\boldsymbol{s}$, es la raíz cuadrada de la variancia muestral. $s=\sqrt{s^2}$. Está medida en la misma unidad que la variable.

VARIANCIA Y DESVÍO ESTÁNDAR

El desvío estándar es aproximadamente la distancia promedio entre las observaciones y la media.

En la población, la **variancia** se simboliza $\boldsymbol{\sigma^2}$ y el **desvío estándar**, $\boldsymbol{\sigma}$. Las fórmulas de cálculo son diferentes a las del caso muestral.



$$s = \sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2}{n-1}}$$

Todos estos cálculos se pueden hacer con una calculadora científica (ver tutoriales).

VARIANCIA Y DESVÍO ESTÁNDAR

En el ejemplo con tres observaciones:

- $s^2 = \frac{(-4)^2+1^2+3^2}{3-1} = \frac{16+1+9}{2} = \frac{26}{2} = 13$
- $s = \sqrt{13} = 3{,}6$

Las observaciones difieren, en promedio, 3,6 unidades de la media aritmética de 4.

CUIDADO! Así como la media es sensible a valores extremos, la variancia y el desvío estándar también lo son, ya que utilizan la media aritmética en su cálculo.

CUIDADO! Una distribución con el desvío estándar más pequeño no es necesariamente la distribución con menor variabilidad.

MANOS A LA OBRA N° 4

PESO Y CONSUMO DE FRUTAS DE LAS MUJERES DE 50 AÑOS O MÁS

Obtenga la media aritmética, la variancia y el desvío estándar de las variables del peso de las mujeres de 50 y más años.

50,61 – 60,85 – 63,70 – 70,98 – 88,90 – 44,07 – 56,87 – 48,84 – 64,41 – 40,85 – 65,04 – 61,38 – 45,39 – 82,80

Obtenga la media aritmética, la variancia y el desvío estándar de la cantidad de porciones de frutas que consumen las mujeres de 50 y más años.

4 – 4 – 5 – 7 – 7 – 7 – 8 – 8 – 9 – 9 – 9 – 10 – 10 – 14

¿Pueden compararse los desvíos estándar de las dos variables?

PARA RESOLVER 4.5

Obtenga la variancia y el desvío estándar de las edades de cada uno de los dos grupos de niños que participaron del estudio para comparar dos tratamientos para la angina. Interprete los valores en términos del problema.

Amoxicilina: 14 – 17 – 11 – 10 – 11 – 14 – 9 – 12 – 8 – 10 – 9

Cefadroxil: 9 – 14 – 8 – 10 – 13 – 7 – 9 – 11 – 16 – 10 – 12 – 9

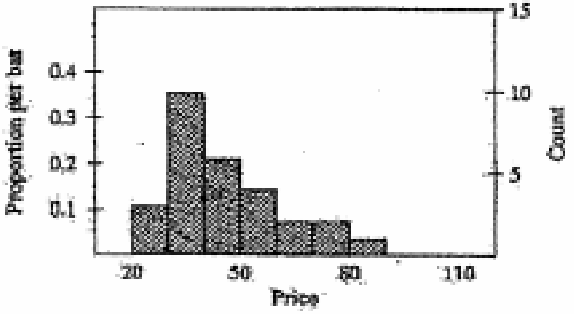
MANOS A LA OBRA N° 4

PARA RESOLVER 4.10 (página 257). Calzado de competición

☑ **PARA RESOLVER !!!**

4.10. Más sobre el calzado de competición

Considere la distribución de los precios de los calzados de competición dada por:



a) ¿Qué valores representarían mejor a la media y a la mediana?

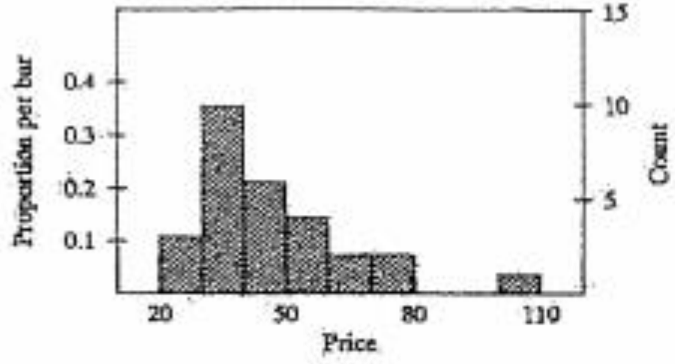
Valores: 40 45

Media:.......... $\boxed{}$ ....

Mediana:...... $\boxed{}$ ....

Explique:...

b) Suponga que una observación más grande fue incorrectamente ingresada, según lo demuestra la siguiente figura:



Marcar con una cruz:

	Incrementará	**Decrecerá**	**Será la misma**
La media			
La mediana			
El desvío estándar			

COEFICIENTE DE VARIACIÓN

En MANOS A LA OBRA N° 4 pregunta si se puede comparar el desvío estándar del peso de las mujeres de 50 años o más y la cantidad de porciones de frutas consumidas por semana. Comparar una desviación estándar medida en kilogramos con otra en porciones no tiene ningún sentido.

El mismo inconveniente se presenta cuando se comparan grupos medidos en la misma unidad pero de magnitudes promedio muy diferentes entre ellos. Por ejemplo el peso de las mujeres de 50 años o más y el peso de bebés recién nacidos. El **COEFICIENTE DE VARIACIÓN** es la solución a estos problemas.

COEFICIENTE DE VARIACIÓN

El **coeficiente de variación** es una medida de variabilidad que elimina las unidades de medida (es un número puro) y cancela el efecto de distorsión que provocan las magnitudes promedios muy diferentes. El coeficiente de variación de Pearson es el cociente entre el desvío estándar y la media aritmética: $CV = \frac{s}{\bar{x}}$ Comúnmente se expresa como porcentaje multiplicando su valor por 100.

COEFICIENTE DE VARIACIÓN

Los siguientes datos corresponden al peso de las mujeres de 50 años o más:

$\bar{x} = 60{,}34\ kg,\ s = 14{,}11\ kg$ $\quad CV = \frac{14{,}11}{60{,}34} = 0{,}234\ Es\ decir\ el\ 23{,}4\%$

Los siguientes datos corresponden a la cantidad de porciones de frutas que se consumen por semana

$\bar{x} = 7{,}93\ porciones,\ s = 2{,}64\ porciones$ $\quad CV = \frac{2{,}64}{7{,}93} = 0{,}333\ Es\ decir\ el\ 33{,}3\%$

La cantidad de porciones de fruta que se consumen por semana presenta mayor variabilidad con respecto a su media que el peso de las mujeres de 50 años o más.

MEDIDAS RESUMEN

También veremos como usar un conjunto de estadísticas de orden para construir el box-plot, o gráfico de caja y bigotes.

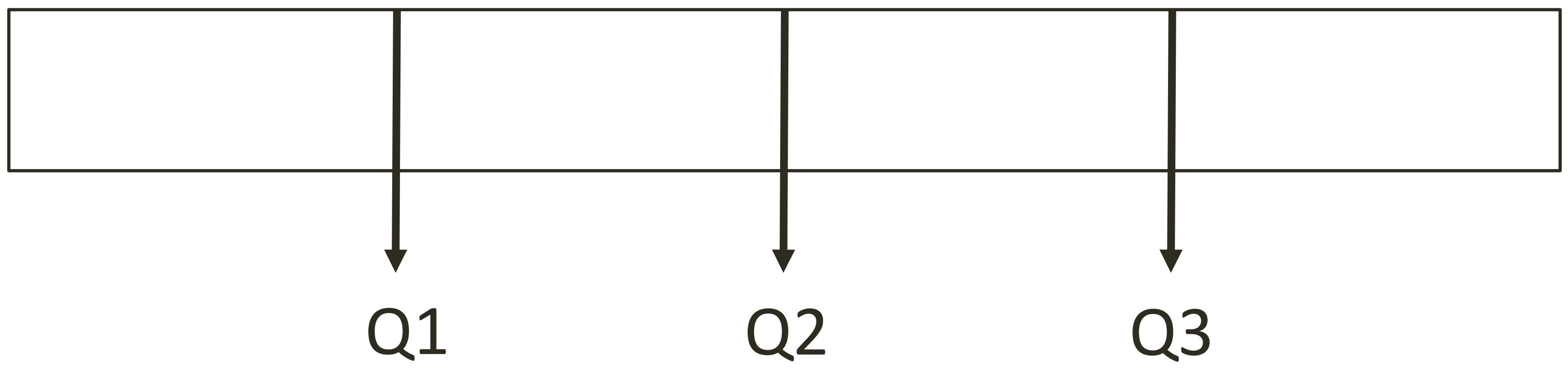
Estadísticas de orden

- Cuartiles
- Percentiles

CUARTILES

Para definir otra medida de dispersión, el rango intercuartílico, primero vamos a encontrar los cuartiles.

Los cuartiles son valores de la variable que dividen al conjunto de datos ordenados en cuatro grupos, cada uno de los cuales contiene aproximadamente el 25% de los valores observados.



CUARTILES

Q_2 es el valor de la variable tal que el 50% de las observaciones es menor o igual que él, mientras que el 50% restante es mayor o igual que él. Este cuartil es la mediana de las observaciones.

- La cantidad de porciones de frutas que consumen por semana las mujeres de 50 años o más (n par):

4 – 4 – 5 – 7 – 7 – 7 – $\boxed{8 – 8}$ – 9 – 9 – 9 – 10 – 10 – 14

$$Q_2 = \frac{8+8}{2} = 8$$

- La cantidad de porciones de frutas que consumen por semana los hombres de 50 años o más (n impar):

4 – 5 – 5 – 6 – 6 – 6 – 7 – 7 – 9

$Q_2 = 6$

CUARTILES

Q_1 es el valor de la variable tal que el 25% de los valores son menores o iguales que él.

- <u>Si el tamaño de muestra es **par**</u>, Q_1 se obtiene calculando la mediana de las observaciones menores o iguales que la mediana:

4 – 4 – 5 – 7 – 7 – 7 – 8 – 8 – 9 – 9 – 9 – 10 – 10 – 14

$Q_1 = 7$

El 25% de las mujeres de 50 años o más consumen 7 o menos frutas por semana.

Q_3 es el valor de la variable tal que el 75% de los valores son menores o iguales que él.

- <u>Si el tamaño de muestra es **par**</u>, Q_3 se obtiene calculando la mediana de las observaciones mayores o iguales que la mediana:

4 – 4 – 5 – 7 – 7 – 7 – 8 – 8 – 9 – 9 – 9 – 10 – 10 – 14

$Q_3 = 9$

El 75% de las mujeres de 50 años o más consumen 9 o menos frutas por semana.

CUARTILES

- <u>Si el tamaño de muestra es **impar**</u>, se excluye el Q_2 y se calcula la mediana de las observaciones menores que la mediana:

$$4 - \boxed{5 - 5} - 6 - \cancel{6} - 6 - 7 - 7 - 9$$

$$Q_1 = \frac{5 + 5}{2} = 5$$

El 25% de los hombres de 50 años o más consumen 5 o menos frutas por semana.

- <u>Si el tamaño de muestra es **impar**</u>, se excluye el Q_2 y se calcula la mediana de las observaciones mayores que la mediana:

$$4 - 5 - 5 - 6 - \cancel{6} - 6 - \boxed{7 - 7} - 9$$

$$Q_3 = \frac{7 + 7}{2} = 7$$

- El 75% de los hombres de 50 años o más consumen 7 o menos frutas por semana.

MANOS A LA OBRA N° 5

<u>PESO DE LAS MUJERES DE 50 AÑOS O MÁS</u>

Obtenga los cuartiles del peso de las mujeres de 50 años o más.
50,61 – 60,85 – 63,70 – 70,98 – 88,90 – 44,07 – 56,87 – 48,84 – 64,41 – 40,85 – 65,04 – 61,38 – 45,39 – 82,80

<u>EJEMPLO 4.4. (página 241)</u>

Encuentre los cuartiles de la edad de los individuos que participaron del estudio médico del Conjunto de datos 3. Interprete los valores en términos del problema.

45 – 41 – 51 – 46 – 47 – 42 – 43 – 50 – 39 – 32 – 41 - 44 – 47 – 49 – 45 – 42 – 41 – 40 – 45 - 37

<u>PARA RESOLVER 4.5</u>

Encuentre los cuartiles de las edades de cada uno de los dos grupos de niños que participaron del estudio para comparar dos tratamientos para la angina. Interprete los valores en términos del problema.
<u>Amoxicilina</u>: 14 – 17 – 11 – 10 – 11 – 14 – 9 – 12 – 8 – 10 – 9
<u>Cefadroxil</u>: 9 – 14 – 8 – 10 – 13 – 7 – 9 – 11 – 16 – 10 – 12 – 9

RANGO INTERCUARTÍLICO

El **rango intercuartílico** es una medida de dispersión que se obtiene haciendo la diferencia entre el tercer cuartil y el primer cuartil:

$$RI = Q_3 - Q_1$$

El RI considera la dispersión del 50% central de los datos.

Esta medida de variabilidad acompaña a la mediana, y como ella, no se ve afectada por valores extremos. Considerando La cantidad de porciones de frutas que consumen por semana las mujeres de 50 años o más

RI = 9-7 = 2. En el 50% central de la distribución de la cantidad de porciones de frutas consumidas por semana, el rango de variación es de 2 porciones.

DIAGRAMA DE CAJA (BOX-PLOT)

El diagrama de caja y bigotes se construye utilizando 5 medidas resumen: mínimo, Q_1, Q_2, Q_3 y máximo.

- Se construye una caja cuyos límites son Q_1 y Q_3.
- Dentro de la caja, se traza una línea a la altura de la mediana.
- Los “bigotes” o “patillas” se extienden desde los extremos de la caja hasta el máximo y el mínimo, respectivamente.

DIAGRAMA DE CAJA MODIFICADO

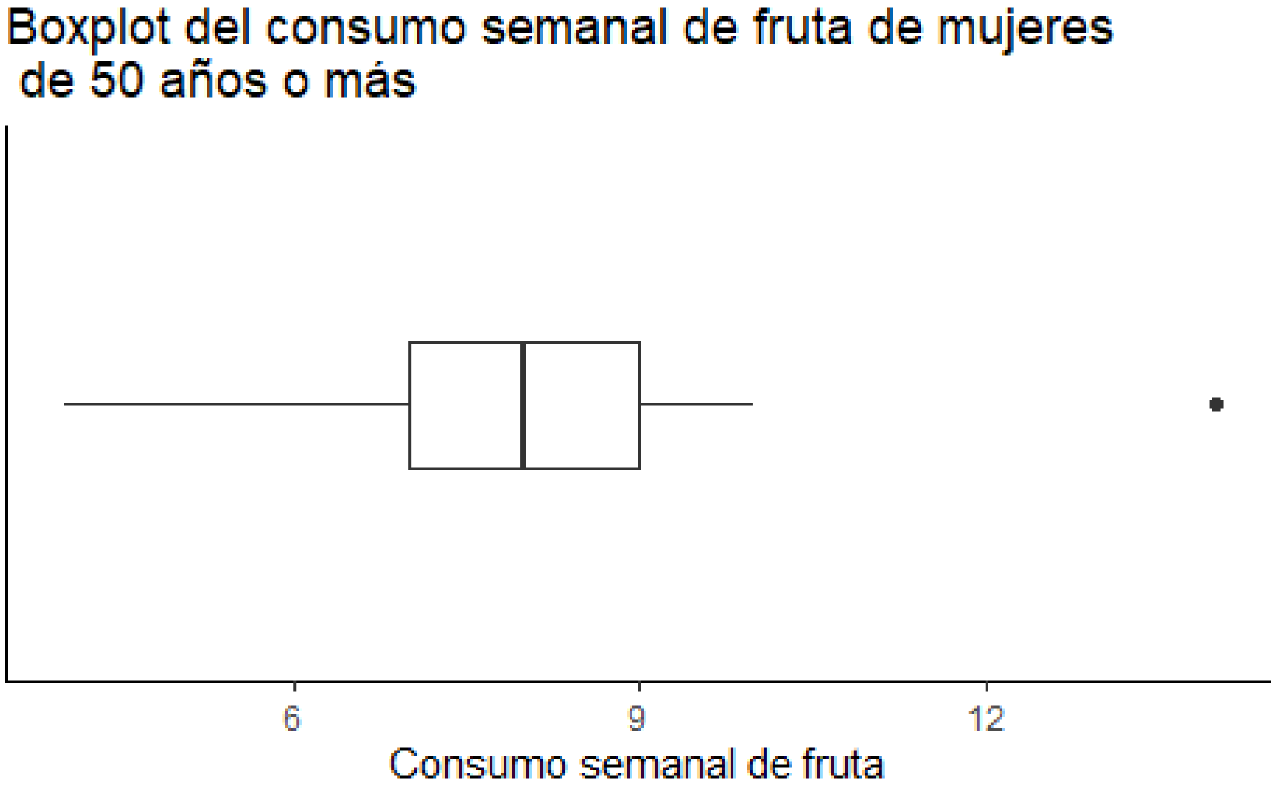
Los diagramas de caja también se utilizan para detectar potenciales outliers o valores extremos, es decir, valores "muy diferentes" de los demás. Para esto, definimos las **barreras internas**:

- $Q_1 - 1{,}5 \times RI$
- $Q_3 + 1{,}5 \times RI$

Los bigotes se grafican hasta el último valor observado mayor a $Q_1 - 1{,}5 \times RI$ y hasta el último valor observado menor a $Q_3 + 1{,}5 \times RI$.

Estos límites también se conocen como **regla del pulgar**.

DIAGRAMA DE CAJA MODIFICADO



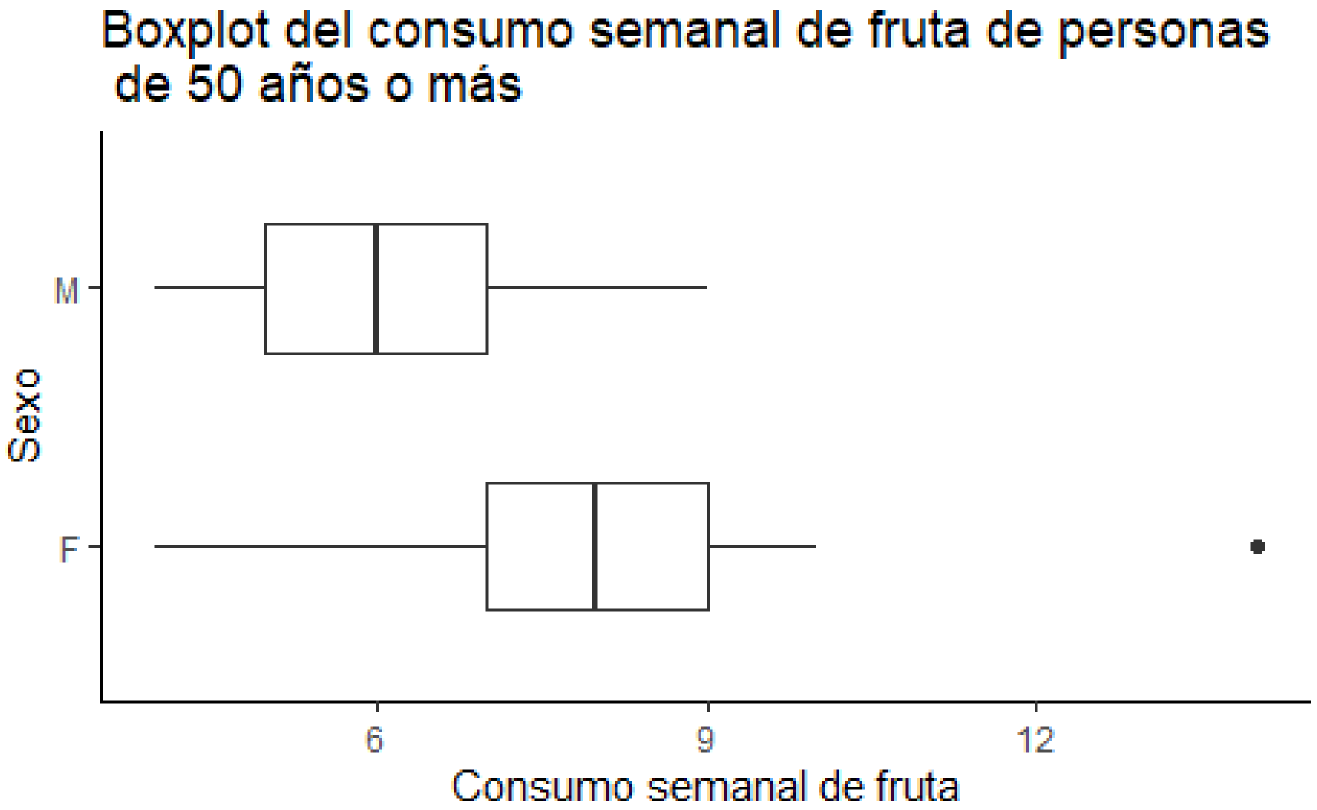
- $Q_1 - 1{,}5 \times RI = 7 - 1{,}5 \times 2 = 4$
- $Q_3 + 1{,}5 \times RI = 9 + 1{,}5 \times 2 = 12$

Recordemos que ninguna mujer de 50 o más consume menos de 4 frutas por semana.

Una mujer de 50 o más consume 14 frutas semanales (más que 12) y las mujeres que consumen menos que 12 frutas consumen 10 frutas por semana (límite del bigote).

DIAGRAMAS DE CAJA LADO A LADO

Estos diagramas se utilizan mucho para comparar la distribución de la misma variable en distintos grupos:



MANOS A LA OBRA N° 6

PARA RESOLVER! 4.7: COMPARANDO EDADES. ESTUDIO DE LOS ANTIBIÓTICOS (página 246)

Se estudia la edad de los niños que participaron de un estudio sobre dos antibióticos: Amoxicilina y Cefadroxil. En este tipo de estudios es importante comparar si las condiciones generales de los grupos son homogéneas, en este caso si los niños que recibieron los distintos antibióticos tienen edades similares, ya que niños más pequeños o más grandes podrían tener reacciones diferentes a los tratamientos.

1) Con las 5 medidas resumen de las edades de los niños que participaron en cada grupo que ya se calcularon en el Manos a la Obra N° 5, construya los box-plots lado a lado. Utilice la regla del pulgar para cada uno de ellos.
2) Interprete el gráfico obtenido. ¿Cómo es la distribución de la edad en cada uno de los dos grupos de niños?
3) ¿Puede identificarse algún outlier en alguno de los dos grupos?

MANOS A LA OBRA N° 6

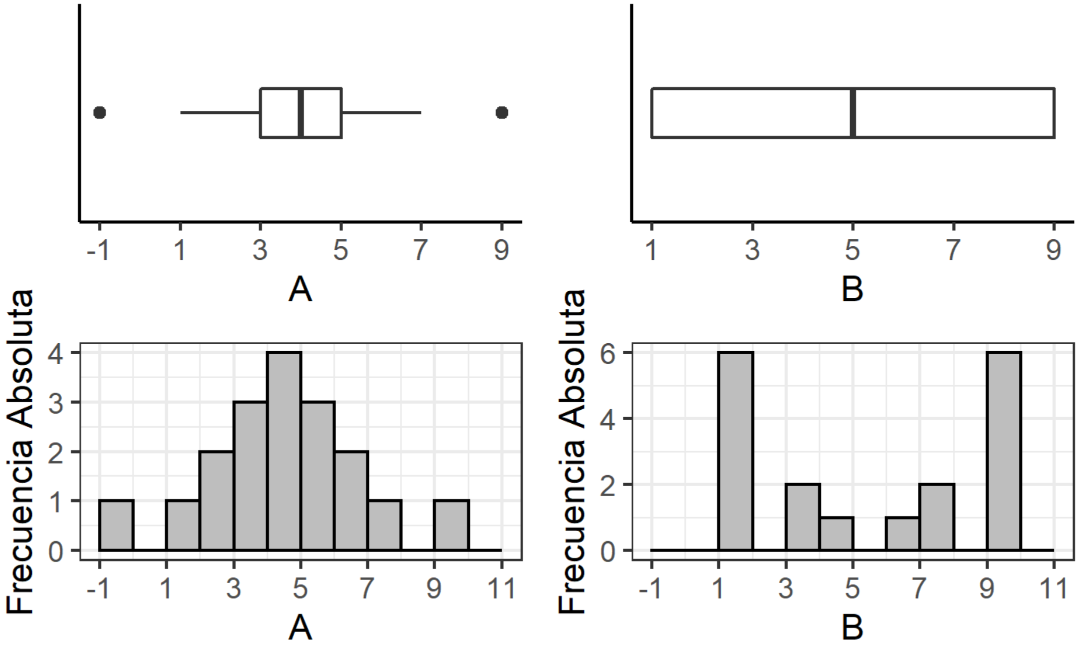
EJEMPLO 4.7: DISTRIBUCIÓN SIMÉTRICA IMPLICA BOX-PLOT SIMÉTRICO (página 248)

Consideremos los siguientes dos conjuntos de observaciones:

A: -1, 1, 2, 2, 3, 3, 3, 4, 4, 4, 4, 5, 5, 5, 6, 6, 7, 9

B: 1, 1, 1, 1, 1, 1, 3, 3, 4, 6, 7, 7, 9, 9, 9, 9, 9, 9

Medida	A	B
Mínimo	-1	1
Q1	3	1
Q2	4	5
Q3	5	9
Máximo	9	9



MANOS A LA OBRA N° 6

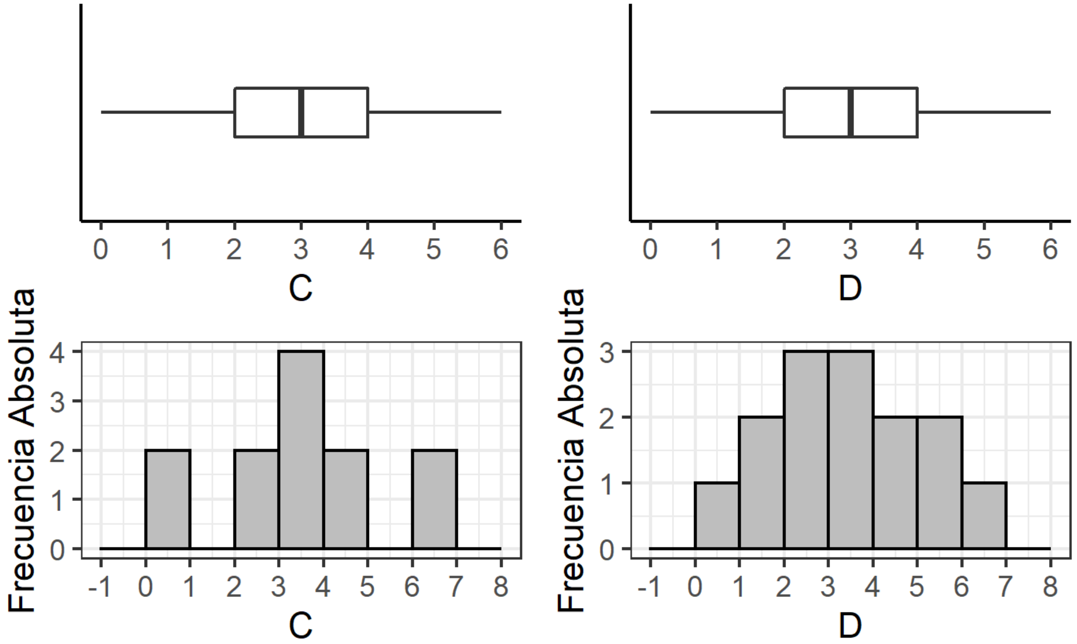
EJEMPLO 4.7: BOX-PLOT SIMÉTRICO NO IMPLICA DISTRIBUCIÓN SIMÉTRICA (página 249)

Consideremos los siguientes dos conjuntos de observaciones:

C: 0, 0, 2, 2, 3, 3, 3, 3, 4, 4, 6, 6

D: 0, 1, 1, 2, 2, 2, 3, 3, 3, 4, 4, 5, 5, 6

Medida	C	D
Mínimo	0	0
Q1	2	2
Q2	3	3
Q3	4	4
Máximo	6	6



CÁLCULO DE MEDIDAS RESUMEN A PARTIR DE UNA TABLA DE FRECUENCIAS

¿Qué pasa si contamos con la información de la edad de un grupo de 50 personas en una tabla de frecuencias?

Para ver tutoriales sobre el uso de calculadora para obtener la media y el desvío estándar:

https://www.youtube.com/watch?v=xdYPXk__ZeY

https://www.youtube.com/watch?v=39MjYi7RQkk

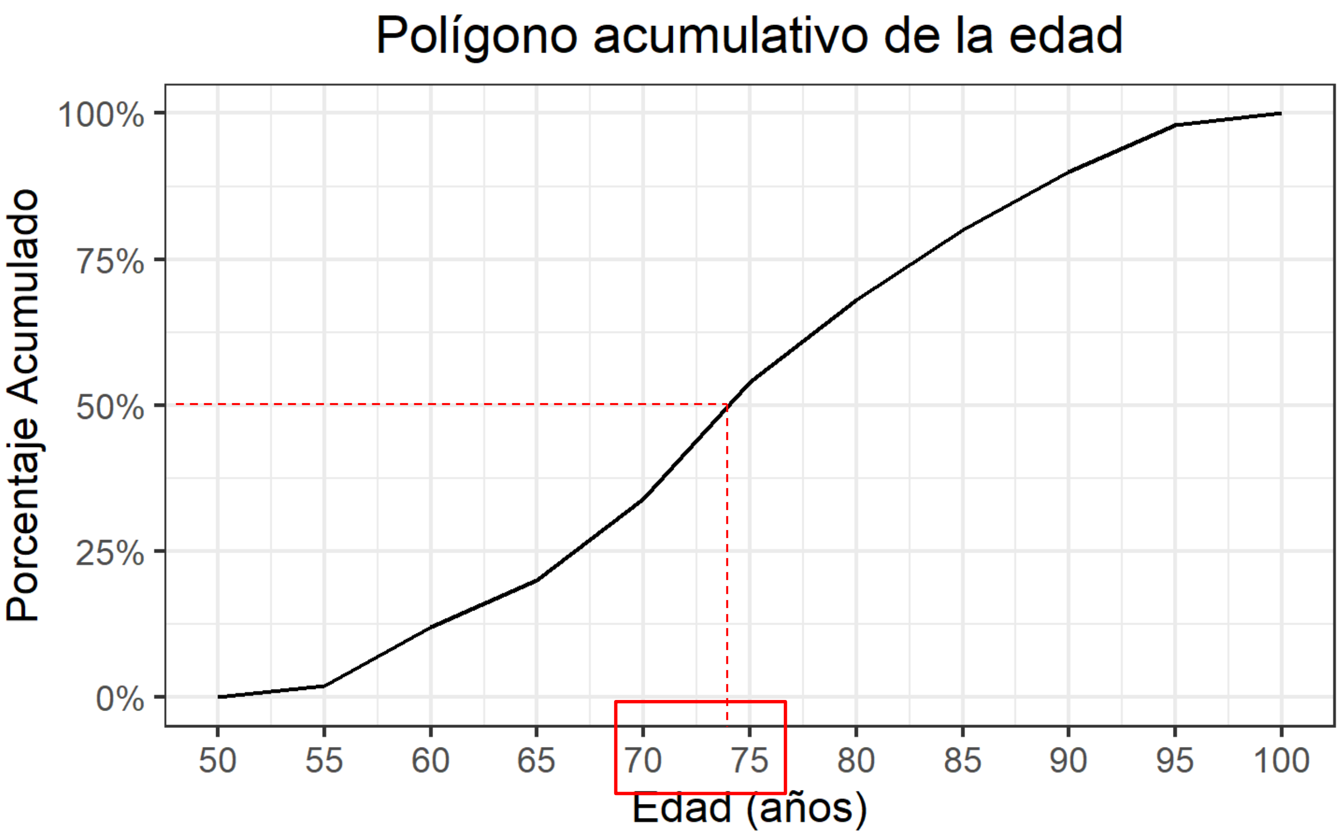
Clase	Punto medio	f_j	H_j
[50-55)	52,5	1	0,02
[55-60)	57,5	5	0,12
[60-65)	62,5	4	0,20
[65-70)	67,5	7	0,34
[70-75)	72,5	10	0,54
[75-80)	77,5	7	0,68
[80-85)	82,5	6	0,80
[85-90)	87,5	5	0,90
[90-95)	92,5	4	0,98
[95-100)	97,5	1	1,00
Total		50	

CÁLCULO DE MEDIDAS RESUMEN A PARTIR DE UNA TABLA DE FRECUENCIAS

Encontrar el intervalo que contiene la mediana:

Podemos usar la información de las frecuencias relativas acumuladas, con ayuda del polígono de frecuencias acumuladas.

La mediana de la edad de las personas está en el intervalo de 70 a 75 años.



MANOS A LA OBRA N° 7

Describir numéricamente y gráficamente la edad de los docentes de la Universidad UCEL.

Tabla de distribución de freucencia de la edad de los docentes de Facultad de Ciencias Empresariales 2012

Lim. Inferior (incluye)	Lim. Superior	Punto medio	Frecuencia Abolsuta (fj)	Frec. Relativa (hj)	Frec. Abs. Acumulada (Fj)	Frec. Relativa Acumulada (Hj)
24	31	27,5	4	0,08	4	0,08
31	38	34,5	13	0,26	17	0,34
38	45	41,5	12	0,24	29	0,58
45	52	48,5	9	0,18	38	0,76
52	59	55,5	7	0,14	45	0,9
59	66	62,5	4	0,08	49	0,98
66	73	69,5	1	0,02	50	1