Metodología de Investigación

Máster Universitario en Ingeniería Matemática y Computación

Nadia Gámez

Presentación de la Asignatura y Tema 6

UNIVERSIDAD

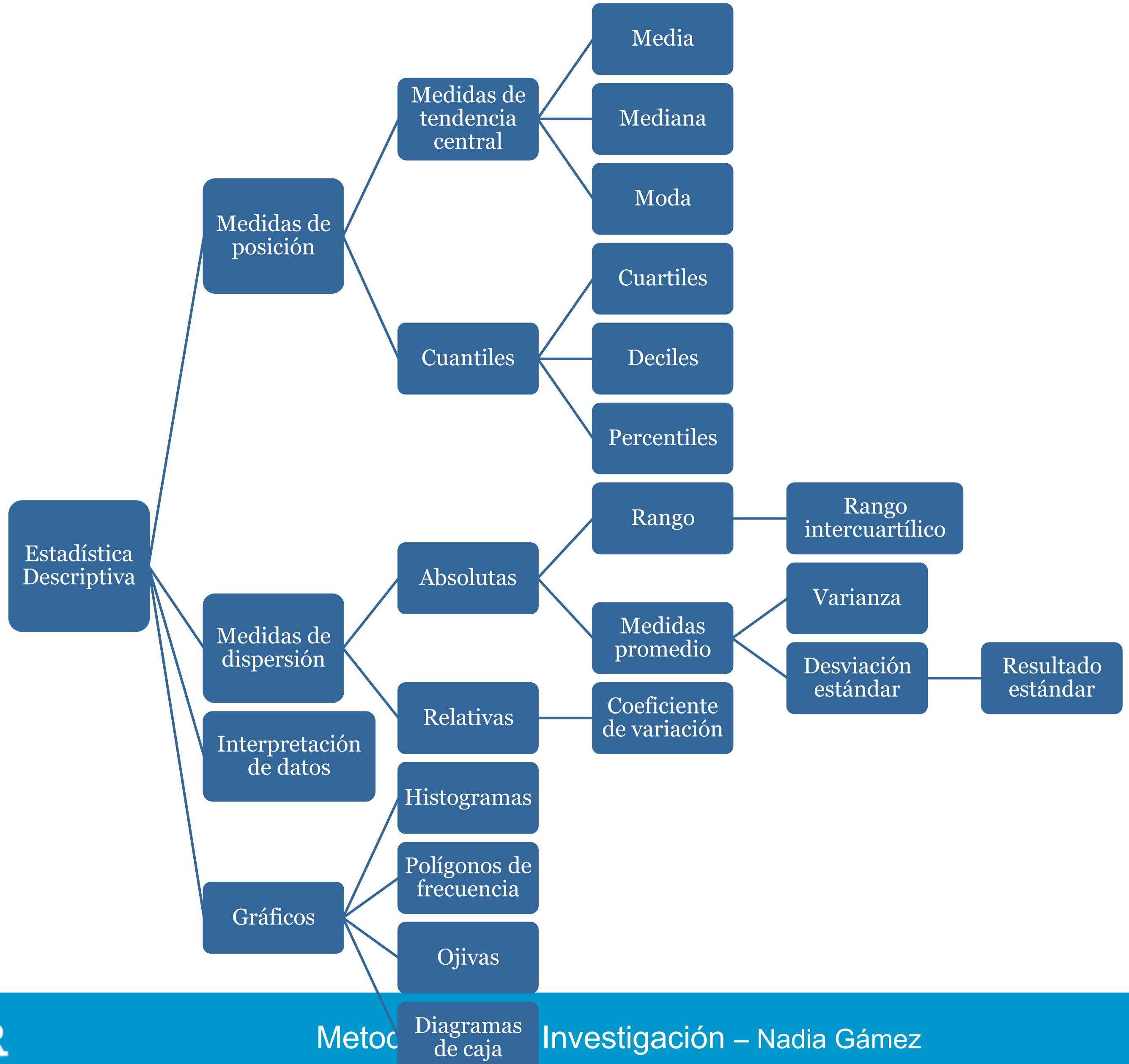
La Universidad en Internet

Tema 1: Estadística descriptiva

6.1 ¿Cómo estudiar este tema?

Leer Ideas Claves	Sesión Virtual	Lección Magistral
Medidas	Posición	Dispersión
Resultados	Interpretación	Gráficos

6.1 ¿Cómo estudiar este tema? → Esquema



6.2. Medidas de posición

¿Para qué sirven?

- La **interpretación de los datos** es crucial en estadística, por ello es preciso saber cómo se pueden organizar los datos para hacer comparaciones y establecer la fiabilidad de los mismos

¿Qué son?

- **Parámetros** que se obtienen de los datos que permiten establecer un **resumen** de los mismos

Tipos

- Tendencia **central**: valores medios a los que tienden un conjunto de datos
  - **media, mediana y moda**
- Tendencia **no central**: dividen el conjunto de datos en partes iguales (**cuantiles**)
  - **cuartiles, deciles y percentiles**

6.2. Medidas de posición

Medidas de Tendencia Central → **Media**

- **Promedio**→ es la medida de tendencia central que puede representar a un conjunto de datos por sí misma

- Tipos:
  - Media **Aritmética**
  - Media **Ponderada**
  - Media **Geométrica**
  - Media **Armónica**
  - Media **Cuadrática**

6.2. Medidas de posición

Medidas de Tendencia Central → **Media Aritmética**

- Se emplea en estadística para saber **cómo es una distribución de datos** y para hacer comparaciones entre distribuciones

- Se calcula como la **suma de los dividida entre el total de datos**

- $\overline{x} = \frac{1}{n}\sum_{i=1}^{k} x_i \cdot f_i$ *donde n* es el número total de datos de la muestra y x_i y f_i son los valores y frecuencias absolutas de cada dato hasta el dato *k,* tomado como el último
- $\overline{x} = \sum_{i=1}^{k} x_i \cdot h_i$ *donde* h_i es la frecuencia relativa de cada dato.

- Cuando dentro del conjunto hay algún **dato anómalo**, la media no proporciona información sobre él, es preciso quitarlo antes de hacer la media

6.2. Medidas de posición

Ejemplo. Cálculo de la media aritmética con datos agrupados en intervalos

Calcular la media aritmética a partir de los siguientes datos, agrupados en intervalos, de faltas de asistencia trimestrales de alumnos de cuatro cursos del mismo año académico. ¿Qué promedio de faltas de asistencia ha habido?

I_i	f_i
4-8	25
9-15	18
16-20	12
21-27	15
28-32	5

6.2. Medidas de posición

Primero hay que calcular las marcas de clase para cada intervalo (nótese que la amplitud de los intervalos no es la misma):

I_i	Marca de clase x_i	f_i
4-8	6	25
9-15	12	18
16-20	18	12
21-27	24	15
28-32	30	5
Total		75

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{k} x_i \cdot f_i = \frac{6 \cdot 25 + 12 \cdot 18 + 18 \cdot 12 + 24 \cdot 15 + 30 \cdot 5}{75} = 14{,}56$$

6.2. Medidas de posición

Medidas de Tendencia Central → **Media Ponderada**

- Se emplea cuando se pretende saber información sobre **un valor en concreto con respecto al total** de valores

- Se calcula **multiplicando los datos por una ponderación**, que es el valor del **peso** de cada dato respecto al total, hacer el sumatorio y dividir entre el total de las ponderaciones

- $\overline{x}_p = \frac{\sum_{i=1}^{k} x_i \cdot w_i}{\sum_{i=1}^{k} w_i}$ *donde w* es el peso para unidad de análisis

6.2. Medidas de posición

Ejemplo. Calcular la media ponderada

A partir de los datos del ejemplo anterior, calcula la media ponderada:

Calculamos las frecuencias relativas, cuya suma ha de ser la unidad.

I_i	Marca de clase x_i	f_i	h_i
4-8	6	25	0,333
9-15	12	18	0,240
16-20	18	12	0,160
21-27	24	15	0,200
28-32	30	5	0,066
Total		75	1

$$\bar{x}_p = \frac{\sum_{i=1}^{k} x_i \cdot w_i}{\sum_{i=1}^{k} w_i} = \frac{6{\cdot}0{,}333 + 12{\cdot}0{,}240 + 18{\cdot}0{,}160 + 24{\cdot}0{,}200 + 30{\cdot}0{,}066}{0{,}333 + 0{,}240 + 0{,}160 + 0{,}200 + 0{,}066} = 14{,}56$$

6.2. Medidas de posición

Medidas de Tendencia Central → **Media Geométrica**

- Se usa para hacer **promedio de índices**, calcular las tasas de variaciones promedio de ventas, calcular el factor de crecimiento de un capital con el tiempo
- Se calcula con la **raíz enésima del producto de cada dato elevado a la frecuencia absoluta**
- $\overline{x}_g = \sqrt[n]{\prod_{i=1}^{k} {x_i}^{f_i}}$
- $\overline{x}_g = \sqrt[n]{x_1 \cdot x_1 \cdot \cdots \ldots x_2}$ donde los valores de x_i hacen referencia a los factores de crecimiento

6.2. Medidas de posición

Ejemplo. Calcular la media geométrica

Si se invierten 100.000 euros al 2%, 4% y 7% durante tres años, ¿cuál es el valor medio al que se está invirtiendo tal capital?

Se calcula el factor de crecimiento en cada año con la siguiente fórmula:

$$1 + \%\ ingresos$$

Año	% de ingresos	Factor de crecimiento
1	2%	1,02
2	4%	1,04
3	7%	1,07

Tabla. Crecimiento de un capital en 3 años.

Aplicando la fórmula:

$$\bar{x}_g = \sqrt[n]{x_1 \cdot x_2 \cdot \ldots \ldots x_k} = \sqrt[3]{1{,}02 \cdot 1{,}04 \cdot 1{,}07} = 1{,}043$$

Por lo tanto el rédito medio es **4,3%.**

6.2. Medidas de posición

Ejemplo. Cálculo de la media geométrica con valores de porcentaje elevados

A continuación se muestran los datos que pagaron los bancos en cinco años en los que el índice de inflación fue elevado. Calcula el factor de crecimiento promedio.

Año	% (tasa de interés)	Factor de crecimiento
1	75%	1,75
2	100%	2
3	200%	3
4	250%	3,5
5	350%	4,5

Tabla 5. Tasas de interés en 5 años.

Se calcula el factor de crecimiento como 1 + tasa de interés y a continuación se calcula la media geométrica.

$$\bar{x}_g = \sqrt[n]{x_1 \cdot x_2 \cdot \ldots \ldots x_k} = \sqrt[5]{1{,}75 \cdot 2 \cdot 3 \cdot 3{,}5 \cdot 4{,}5} = 2{,}77$$

El resultado obtenido empleando la media aritmética es 2,95. Este resultado difiere más que la media geométrica, por lo que sería menos apropiado su uso.

6.2. Medidas de posición

Medidas de Tendencia Central → **Media Armónica**

- Se usa en situaciones en las que **se promedian magnitudes relativas**, respecto al tiempo, al área,…

- Se calcula como **el inverso de la media aritmética de los valores inversos de los datos multiplicados por las frecuencias**

- $\frac{1}{\bar{x}_a} = \frac{\sum_{i=1}^{k} \frac{f_i}{x_i}}{n} \rightarrow \bar{x}_a = \frac{n}{\sum_{i=1}^{k} \frac{f_i}{x_i}}$

- Se considera **más precisa** que la media aritmética, aunque cuando se trata de valores muy próximos a cero es preferible no emplearla

6.2. Medidas de posición

Ejemplo. Cálculo de la media armónica

Un automóvil realiza 20 km de un trayecto a 90 km/h, 100 km a 120 km/h y 10 km a 70 km/h, ¿Cuál es la velocidad promedio de todo el trayecto?

Como se trata de variaciones con respecto al tiempo se emplea la media armónica:

$$\bar{x}_a = \frac{n}{\sum_{i=1}^{k} \frac{f_i}{x_i}} = \frac{3}{\frac{1}{90} + \frac{1}{120} + \frac{1}{70}} = 88{,}94\ km/h$$

ERROR! Corrección abajo!

$$\bar{x}_a = \frac{n}{\sum_{i=1}^{k} \frac{f_i}{x_i}} = \frac{130}{\frac{20}{90} + \frac{100}{120} + \frac{10}{70}} = \frac{130}{0{,}22 + 0{,}83 + 0{,}14} = 108{,}48\ km/h$$

6.2. Medidas de posición

Medidas de Tendencia Central → **Media Cuadrática**

- Se usa cuando se precisa **calcular el valor medio de valores negativos y positivos** puesto que se basa en elevar al cuadrado tales valores y hacerlos positivos
- $\overline{x}_c = \sqrt{\frac{1}{n}\sum_{i=1}^{k} x_i^2} \cdot f_i$
- Muy útil para calcular el valor medio de **errores en medición**

6.2. Medidas de posición

Ejemplo. Cálculo de la media cuadrática

A continuación se muestran los errores que se han obtenido en la medición de una cantidad de sustancia con un aparato, calcular el error cuadrático medio.

Medida	Error (E)
3,15	-0,15
5,20	0,65
3,65	-0,07
4,90	0,35
4,14	0,12

Tabla 6. Errores en la medición.

$$\bar{x}_c = \sqrt{\frac{1}{n}\sum_{i=1}^{k} x_i^2 \cdot f_i} = \sqrt{\frac{1}{5}\cdot[(-0{,}15)^2 + (0{,}65)^2 + (-0{,}07)^2 + (0{,}35)^2 + (0{,}12)^2]} = 0{,}34$$

6.2. Medidas de posición

Medidas de Tendencia Central → **Mediana**

- Es el **valor central** de un conjunto de datos, los cuales han de estar ordenados
- Su posición se calcula a partir de $\boldsymbol{Posición\ de\ la\ mediana} = \frac{\boldsymbol{n+1}}{\boldsymbol{2}}$
- **Número impar de datos** → la mediana será el valor central que deje el mismo número de datos a la izquierda y a la derecha.
- **Número par de datos** → la mediana será la media aritmética de los valores centrales

6.2. Medidas de posición

Ejemplo. Cálculo de la mediana de un número impar e impar de datos

Calcular la mediana de los siguientes datos:

A. 1,2 2,3 3,6 4,7 5,3 6,9 8,1

B. 100 267 340 456 549 678 750 811

A. En este caso se observa fácilmente que la mediana es el valor central 4,7 ya que deja 3 valores a cada lado. Se puede comprobar mediante la fórmula:

$$Posición\ de\ la\ mediana = \frac{n+1}{2} = \frac{7+1}{2} = 4$$

El dato en posición 4 es 4,7.

B. En este caso la mediana se halla entre los datos en posiciones 4 y 5 que son los que dejan 3 datos a cada lado. Empelando la fórmula resulta:

$$Posición\ de\ la\ mediana = \frac{n+1}{2} = \frac{8+1}{2} = 4{,}5$$

Como no es número entero hay que hacer el promedio entre los datos de las posiciones 4 y 5 que son los enteros más próximos.

$$\bar{x} = \frac{456+549}{2} = 502{,}5$$

6.2. Medidas de posición

Medidas de Tendencia Central → **Mediana**

- Para **datos agrupados en intervalos** lo que hay que saber es cuál de dichos intervalos contiene a la mediana

- **Posición** igual que antes

- **Cálculo** de la mediana: $\boldsymbol{Mediana} = \left(\frac{\left(\frac{n+1}{2}\right)-(F_{m-1}+1)}{f_m}\right) \cdot a + L_m$

  donde n es el total de datos, F_{m-1} es la frecuencia acumulada hasta el intervalo donde se halla la mediana sin incluir, f_m la frecuencia del intervalo de la mediana, a la amplitud del intervalo y L_m el límite inferior del intervalo de la mediana.

6.2. Medidas de posición

Ejemplo. Cálculo de la mediana

A partir de los datos de la tabla calcular la mediana.

I_i	f_i
4-8	25
9-15	18
16-20	12
21-27	15
28-32	5
Total	75

6.2. Medidas de posición

Se calcula la posición de la mediana:

$$Posición\ de\ la\ mediana = \frac{n+1}{2} = \frac{75+1}{2} = 38$$

Se hace una tabla con las frecuencias acumuladas:

I_i	f_i	F_i
4-8	25	25
9-15	18	43
16-20	12	55
21-27	15	70
28-32	5	75
Total	75	

Comparando con la columna de frecuencias acumuladas, el valor 38 se halla en el intervalo [9-15). Aplicando la fórmula 1.11. se obtiene la mediana:

$$Mediana = \left(\frac{\left(\frac{n+1}{2}\right) - (F_{m-1}+1)}{f_m}\right) \cdot a + L_m = \left(\frac{\left(\frac{75+1}{2}\right) - (25+1)}{18}\right) \cdot 6 + 9 = 13$$

6.2. Medidas de posición

Medidas de Tendencia Central → **Moda**

- Es el valor que más se repite de una serie de datos
- $\boldsymbol{Moda = L_{mo} + \left[\frac{f_{mo} - f_{mo-1}}{(f_{mo} - f_{mo-1}) + (f_{mo} - f_{mo+1})}\right] \cdot a}$ donde L_{mo} es el límite inferior del intervalo modal, f_{mo} es la frecuencia del intervalo modal, f_{mo-1} es la frecuencia del intervalo inferior al intervalo modal, f_{mo+1} es la frecuencia del intervalo por encima del intervalo modal y *a* es la amplitud del intervalo.

6.2. Medidas de posición

Ejemplo. Cálculo de la moda en datos agrupados

Calcular la moda para los datos de la tabla

I_i	f_i
4-8	25
9-15	18
16-20	12
21-27	15
28-32	5
Total	75

El primer intervalo es el que más datos tiene.

$$Moda = L_{mo} + \left[\frac{f_{mo} - f_{mo-1}}{(f_{mo} - f_{mo-1}) + (f_{mo} - f_{mo+1})}\right] \cdot a = 4 + \left[\frac{25 - 0}{(25 - 0) + (25 - 18)}\right] \cdot 4 = 7{,}125$$

6.2. Medidas de posición

Medidas de Tendencia NO Central

- **Cuantiles** que se emplean para calcular otros puntos del conjunto de datos
- Tales valores hacen que el conjunto de datos quede dividido en secciones constantes
- Cuantiles
  - **Cuartiles**
  - **Deciles**
  - **Percentiles**

6.2. Medidas de posición

Medidas de Tendencia NO Central → Cuartiles

- Dividen al conjunto de datos en **4 partes iguales** que albergan el 25% de los datos cada una.

- El primer cuartil, Q_1, corresponde al 25% de los datos; el segundo cuartil, Q_2, al 50% de los datos, por lo que es la mediana; y el tercer cuartil, Q_3, al 75% de los datos.

6.2. Medidas de posición

Ejemplo. Cálculo de los cuartiles de un conjunto de datos

Calcular los cuartiles de este conjunto de datos de faltas de asistencia a un curso:

xi	f_i	F_i
0	3	3
2	12	15
4	21	36
6	34	70
8	22	92
10	3	95
Total	95	

El segundo cuartil, Q_2, es la mediana:

$$Posición\ de\ la\ mediana = \frac{n+1}{2} = \frac{95+1}{2} = 48$$

6.2. Medidas de posición

Comparando con la frecuencia absoluta acumulada, el segundo cuartil, Q_2, corresponde al dato que está en posición 4. El 50% de los asistentes faltaron 6 veces.

La posición del primer cuartil, Q_1, se puede calcular como la mitad de la posición de la mediana, es decir, la que corresponde a un cuarto de los datos:

$$Posición\ de\ Q_1 = \frac{n+1}{4} = \frac{95+1}{4} = 24$$

Comparando con la frecuencia absoluta acumulada, el primer cuartil, Q_1, corresponde al dato que está en posición 3. El 25% de los asistentes faltaron 4 veces.

La posición del tercer cuartil, Q_3, se puede calcular como la que corresponde a las tres cuartas partes de los datos:

$$Posición\ de\ Q_3 = \frac{3(n+1)}{4} = \frac{3(95+1)}{4} = 72$$

Comparando con la frecuencia absoluta acumulada, el tercer cuartil, $Q3$, corresponde al dato en posición 5. El 75% de los asistentes faltaron 8 veces.

6.2. Medidas de posición

Ejemplo. Cálculo de cuartiles de una distribución de datos agrupados en intervalos

A partir de los datos de la tabla 8 calcular los cuartiles.

I_i	f_i	F_i
4-8	25	25
9-15	18	43
16-20	12	55
21-27	15	70
28-32	5	75
Total	75	

Datos de la tabla 7.

El segundo cuartil Q_2 es la mediana que ya se ha calculado en el ejemplo 8 y es 13. El 50% de los datos es menor o igual a 13.

6.2. Medidas de posición

El segundo cuartil Q_2 es la mediana que ya se ha calculado en el ejemplo 8 y es 13.

El 50% de los datos es menor o igual a 13.

El primer cuartil Q1 corresponde a la cuarta parte de los datos:

$$Posición\ de\ Q_1 = \frac{n+1}{4} = \frac{75+1}{4} = 19$$

Comparando con las frecuencias acumuladas, tal posición está en el primer intervalo.

$$Q_1 = \left(\frac{\left(\frac{n+1}{4}\right) - (F_{m-1}+1)}{f_m}\right) \cdot a + L_m = \left(\frac{\left(\frac{75+1}{4}\right) - (0+1)}{25}\right) \cdot 4 + 4 = 6{,}88$$

El 25% de los datos es menor o igual que 6,88.

El tercer cuartil Q3 corresponde a las tres cuartas partes de los datos:

$$Posición\ de\ Q_3 = \frac{3(n+1)}{4} = \frac{3(75+1)}{4} = 57$$

6.2. Medidas de posición

Comparando con las frecuencias acumuladas, tal posición está en el intervalo [21-27)

$$Q_3 = \left(\frac{\frac{3(n+1)}{4} - (F_{m-1}+1)}{f_m} \right) \cdot a + L_m = \left(\frac{\frac{3(75+1)}{4} - (55+1)}{15} \right) \cdot 6 + 21 = 21{,}4$$

El 75% de los datos es menor o igual que 21,4.

6.2. Medidas de posición

Medidas de Tendencia NO Central → Deciles y Percentiles

- **Deciles**: dividen la distribución en diez partes iguales en las que cada una concentra el 10% de los resultados

- **Percentiles**: dividen el conjunto de datos en 100 partes de manera que cada dato concentra el 1% de los valores

Ejemplo. Calcular los deciles de datos no agrupados

A partir de los datos de la Tabla 10 calcular los deciles 2, 4 y 9.

6.2. Me

xi	f_i	F_i
0	3	3
2	12	15
4	21	36
6	34	70
8	22	92
10	3	95
Total	95	

Posición	Valor
$\frac{2(95+1)}{10} = 19{,}2$	$D_2 = 4$
$\frac{4(95+1)}{10} = 38{,}4$	$D_4 = 6$
$\frac{9(95+1)}{10} = 86{,}4$	$D_9 = 10$

El 20% de los asistentes faltaron 4 veces o menos, El 40% de los asistentes faltaron 6 veces o menos y el 90% de los asistentes faltaron 10 veces o menos.

6.2. Medidas de posición

Ejemplo. Cálculo de deciles en datos agrupados

A partir de los datos de la tabla 8 calcular los deciles 1,5, y 7:

I_i	f_i	F_i
4-8	25	25
9-15	18	43
16-20	12	55
21-27	15	70
28-32	5	75
Total	75	

Posición del decil	Intervalo	Valor del decil
$\frac{(75+1)}{10} = 7{,}6$	[4-8)	$D_1 = \left(\frac{\frac{(75+1)}{10} - (0+1)}{25}\right) \cdot 4 + 4 = 5{,}06$
$\frac{5(75+1)}{10} = 38$	[9-15)	$D_5 = \left(\frac{\frac{5(75+1)}{10} - (25+1)}{18}\right) \cdot 6 + 9 = 13$
$\frac{7(75+1)}{10} = 53{,}2$	[16-20)	$D_7 = \left(\frac{\frac{7(n+1)}{10} - (F+1)}{f_m}\right) \cdot w + L_m = \left(\frac{\frac{7(75+1)}{10} - (43+1)}{12}\right) \cdot 4 + 16 = 19{,}06$

El 10% de los datos incluyen hasta el dato 5.06, el 50% de los datos incluyen hasta el dato 13, la mediana, y el 70% de los datos incluyen hasta el dato 19,06.

Ejemplo. Cálculo de percentiles en datos no agrupados

Calcular el percentil 50 con los datos de la tabla

xi	f_i	F_i
0	3	3
2	12	15
4	21	36
6	34	70
8	22	92
10	3	95
Total	95	

La posición del percentil se calcula de la siguiente manera:

$$Posición\ del\ percentil = \frac{50(n+1)}{100} = \frac{50(95+1)}{100} = 48$$

Este valor pertenece al cuarto dato. El 50% de los asistentes faltaron 6 veces o menos.

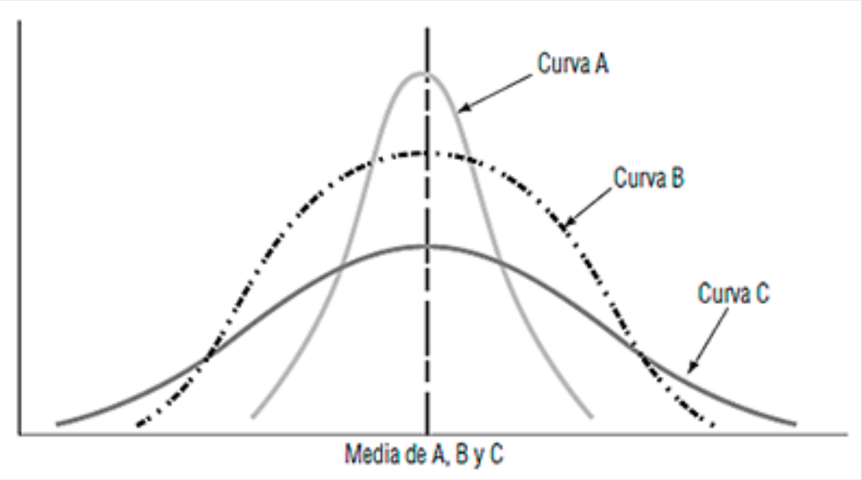
Si se calculase el percentil 70 con los datos agrupados de la tabla 7, el valor sería el mismo que el decil 7 calculado en el ejemplo 13.

6.3. Medidas de dispersión

Medidas de Dispersión

- Saber cuánto **difieren los datos entre sí**
- Distribuciones de valores con la misma media pero diferente dispersión

.



- Medidas de Dispersión:
  - **Absolutas**
  - **Relativas**

Medidas de Dispersión Absolutas

- Nos dan información sobre la dispersión de una distribución sin permitir la comparación entre distribuciones

- Medidas:
  - **Rango**
  - **Medidas de desviación promedio**

6.3. Medidas de dispersión

Medidas de Dispersión Absolutas → **Rango**

- Se calcula como la **diferencia entre los valores más grande y más pequeño** de una serie de datos
- $\boldsymbol{Rango = valor\ más\ grande - valor\ más\ pequeño}$

- No da información de los datos interiores, de cómo se dispersan

- **Rango interfractil**: se calcula entre dos medidas de cuantiles de una distribución y da información de la dispersión entre dichos cuantiles.
- **Rango intercuartil**: es un tipo de rango interfractil definido como la diferencia entre el primer cuartil y el tercer cuartil. La mitad del rango intercuartil se llama desviación intercuartil.

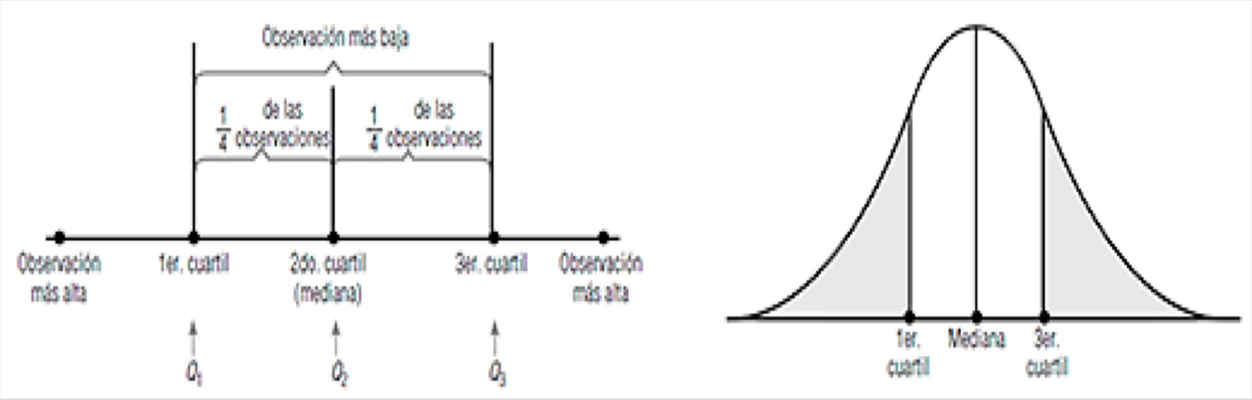
6.3. Medidas de dispersión

Medidas de Dispersión Absolutas → **Rango Intercuartil**

- Medir cuánto hay que **desplazarse de la mediana** hacia valores mayores y menores antes de recorrer la mitad de los mismos

- Se emplea para hacer **diagramas de caja**

.



6.3. Medidas de dispersión

Medidas de Dispersión Absolutas → **Medidas de Desviación Promedio**

- Permiten calcular la dispersión de los datos de manera más precisa
- Estudian la **dispersión de los datos respecto a la media**
- Medidas:
  - **varianza**
  - **desviación estándar**

6.3. Medidas de dispersión

Medidas de Dispersión Absolutas → **Medidas de Desviación Promedio: Varianza**

- **Permite identificar la media de las desviaciones** de entre los datos de una variable.

- Es el sumatorio de los cuadrados de las diferencias entre los valores y la media dividido entre el total de datos

- $\sigma^2 = \frac{\sum (x-\mu)^2}{N} = \frac{\sum x^2}{N} - \mu^2$ *donde x* es cada dato, *μ* es la media poblacional y *N* es el total de observaciones poblacionales

- $\sigma^2 = \frac{\sum f(x-\mu)^2}{N} = \frac{\sum f x^2}{N} - \mu^2$ *donde f* es la frecuencia de cada clase

- $S^2 = \frac{\sum (x-\overline{x})^2}{n-1} = \frac{\sum x^2}{n-1} - \frac{n\overline{x}}{n-1}$ $\boldsymbol{donde\ S^2}$ es la varianza de la muestra, $\overline{x}$ es la media muestral y *n* el tamaño muestral. Al tomar varias muestras y comparar la suma de las varianzas con la varianza de la población total los datos se acercan más cuando se toma para las muestran un valor de datos de *n-6.*

6.3. Medidas de dispersión

Medidas de Dispersión Absolutas → **Medidas de Desviación Promedio: Desviación Estándar**

- La **raíz cuadrada de la varianza**

- Se emplea mucho más para hacer los cálculos, ya que permite obtener la dispersión en las **mismas unidades que los datos**

- Es la raíz cuadrada del sumatorio de los cuadrados de las distancias entre cada una de las observaciones y la media dividido entre el total de observaciones

- $\boldsymbol{\sigma = \sqrt{\sigma^2} = \sqrt{\frac{\sum(x-\mu)^2}{N}} = \sqrt{\frac{\sum x^2}{N} - \mu^2}}$

- $\boldsymbol{\sigma = \sqrt{\sigma^2} = \sqrt{\frac{\sum f(x-\mu)^2}{N}} = \sqrt{\frac{\sum fx^2}{N} - \mu^2}}$ *donde f* es la frecuencia de cada clase

- $\boldsymbol{S = \sqrt{S^2} = \sqrt{\frac{\sum f(x-\bar{x})^2}{n-1}}}$

6.3. Medidas de dispersión

Ejemplo. Calculo de la desviación estándar para una muestra

A partir de los datos de la siguiente tabla de salarios anuales de personas de una empresa, calcular la desviación estándar. Un individuo que cobre 22.000 euros anuales, ¿cuánto se aleja de la media?

Salarios	Empleados
8.000-10.000	25
10.000-15.000	32
15.000-20.000	12
20.000-30.000	16
30.000-50.000	7
50.000-100.000	2

6.3. Medidas de dispersión

Calculamos la media con las marcas de clase de cada intervalo:

$$\bar{x} = \frac{9.000 \cdot 25 + 12.500 \cdot 32 + 17.500 \cdot 12 + 25.000 \cdot 16 + 40.000 \cdot 7 + 75.000 \cdot 2}{94} = 17.712{,}8$$

Intervalo	x	f_i	$x - \bar{x}$	$(x - \bar{x})^2$	$f_i \cdot (x - \bar{x})^2$
8.000-10.000	9.000	25	-8.713	75912361,1	1897809027
10.000-15.000	12.500	32	-5.213	27172971,1	869535074
15.000-20.000	17.500	12	-213	45271,07	543252,875
20.000-30.000	25.000	16	7.287	53103721,1	849659537
30.000-50.000	40.000	7	22.287	496720621	3477044348
50.000-100.000	75.000	2	57.287	3281826721	6563653442
Total		94			13658244681

$$S = \sqrt{S^2} = \sqrt{\frac{\sum f(x-\bar{x})^2}{n-1}} = \sqrt{\frac{13658244681}{94-1}} = 12.118{,}7$$

6.3. Medidas de dispersión

Para calcular lo que se aleja de la media alguien que cobre 22.000 euros anuales, se calcula el resultado estándar:

$$Resultado\ estándar = \frac{x - \bar{x}}{S} = \frac{22.000 - 17.712,8}{12.118,7} = 1,8$$

Ese individuo está alejado en 1,8 desviaciones estándar de la media.

6.3. Medidas de dispersión

Medidas de Dispersión Relativas

- Situaciones en las que se toman **muestran iguales pero en diferentes** intervalos de tiempo, o llevadas a cabo por diferentes personas o instrumentos para obtener algún tipo de información sobre un parámetro

- Las desviaciones estándar de cada muestra no nos permiten hacer una comparación entre todos los ensayos ya que cada ensayo tiene una media

- Es preciso saber cuánto vale cada desviación con respecto a su correspondiente media

- El **coeficiente de variación** permite relacionar la **desviación estándar con la media** de un conjunto de datos en tanto por ciento
  - $C.V = \frac{\sigma}{\mu} \cdot 100$
  - $C.V = \frac{S}{\bar{x}} \cdot 100$

6.4. Interpretación de los datos obtenidos

Aprender a interpretar los datos

Saber qué parámetros debemos calcular en determinadas situaciones

6 ejemplos

6.4. Interpretación de los datos obtenidos

Ejemplo

A continuación se muestra la producción de los trabajadores en cinco líneas de una empresa de embotado de encurtidos por hora, ¿Cuál es la productividad media? ¿Cuál es el tiempo medio para embotar un bote? En una jornada laboral de 8 horas, ¿cuántos botes se hacen?

Línea	Botes/hora
1	150
2	120
3	167
4	135
5	113

6.4. Interpretación de los datos obtenidos

Hay que calcular la media armónica:

$$\bar{x}_a = \frac{n}{\sum_{i=1}^{k} \frac{f_i}{x_i}} = \frac{5}{\frac{1}{150} + \frac{1}{120} + \frac{1}{167} + \frac{1}{135} + \frac{1}{113}} = 134\ botes/hora$$

El tiempo empleado para cada bote es:

$$\bar{t} = \frac{1}{\bar{x}_a} = \frac{1}{134} = 0{,}0075\,\frac{horas}{bote} = 0{,}45\ \frac{minutos}{bote}$$

En 8 horas se hacen:

$$134 \cdot 8 = 1072\ botes$$

6.4. Interpretación de los datos obtenidos

Ejemplo

Para saber cuántas personas ven al día un canal televisivo se han tomados dos muestras estratificadas de una población y se ha calculado la media y la desviación estándar. ¿Cuál es la media y la varianza en la población?

	Individuos	Media	Varianza
Estrato 1	500	2	4
Estrato 2	160	5	9

6.4. Interpretación de los datos obtenidos

El tamaño de cada muestra estratificada es:

Estrato 1:

$$f_1 = \frac{n_1}{n} = \frac{500}{660} = 0{,}76$$

Estrato 2:

$$f_2 = \frac{n_2}{n} = \frac{160}{660} = 0{,}24$$

La media total es:

$$\bar{x} = \bar{x}_1 \cdot f_1 + \bar{x}_2 \cdot f_2 = 2 \cdot 0{,}76 + 5 \cdot 0{,}24 = 2{,}72$$

La varianza total se calcula como:

$$S^2 = f_1 \cdot S_1^2 + f_2 \cdot S_2^2 + f_1(\bar{x}_1 - \bar{x})^2 + f_2(\bar{x}_2 - \bar{x})^2$$
$$= 0{,}76 \cdot 4 + 0{,}24 \cdot 9 + 0{,}76(2 - 2{,}72)^2 + 0{,}24(5 - 2{,}72)^2 = 6{,}84$$

6.4. Interpretación de los datos obtenidos

Ejemplo

A continuación se muestran las edades de los miembros de un curso:

22	18	17	21	21	29	36	16	23	16
25	25	29	23	16	23	52	33	17	23
35	16	32	31	40	23	27	26	22	17

A. Calcular la media, la mediana y la moda.

B. ¿Cuál es la edad máxima del 75% de los miembros?

Aplicando las fórmulas necesarias, el valor de la media es 25,13 y la mediana y la moda salen 23.

Para calcular la edad máxima hay que calcular el percentil 75 o el tercer cuartil, Q_3. Aplicando la fórmula necesaria se obtiene que el el 75% de los miembros tiene 29 o menos años.

6.4. Interpretación de los datos obtenidos

Ejemplo

Una compañía telefónica A saca un anuncio que dice: «Si nuestros precios promedio no son iguales o menores que los de cualquier otra compañía, le sale la primera factura grati». Un cliente fue una tienda y dijo que había visto 4 tarifas, en función de los megas, con precios menores que los que esta en cuestión ofrecía y que iba a denunciar. Tales tarifas eran:

11,9	16,9	21,9	26,9

Los precios de la compañía A para esas mismas tarifas eran:

9,9	16,9	22,9	28,9

El trabajador le dice al cliente que el anuncio hace referencia a un promedio ponderado de las tarifas y que el promedio es menor porque las ventas de esas tarifas han sido:

25	20	17	12

¿Puede denunciar el cliente o queda resulto el asunto con la justificacion de las medias ponderadas?

6.4. Interpretación de los datos obtenidos

Sin hacer medias ponderadas resulta que:

$$Media\ de\ la\ compañía\ A = \frac{9{,}9 + 16{,}9 + 22{,}9 + 28{,}9}{4} = 19{,}65$$

$$Media\ de\ la\ otra\ compañía = \frac{11{,}9 + 16{,}9 + 19{,}9 + 26{,}9}{4} = 19{,}4$$

Haciendo medias ponderadas resulta que:

$$Medida\ ponderada\ de\ la\ compañía\ A = \frac{9{,}9 \cdot 25 + 16{,}9 \cdot 20 + 22{,}9 \cdot 17 + 28{,}9 \cdot 12}{25 + 20 + 17 + 12}$$
$$= 17{,}86$$

$$Medida\ ponderada\ de\ la\ compañía\ A = \frac{11{,}9 \cdot 25 + 16{,}9 \cdot 20 + 19{,}9 \cdot 17 + 26{,}9 \cdot 12}{25 + 20 + 17 + 12}$$
$$= 17{,}98$$

Tal y como se puede observar el cliente tiene razón ya que el promedio sale mayor para la compañía A, pero con las medias ponderadas sale menor por lo que la empresa se cura en salud.

6.5. Gráficos y su interpretación

Tipos de diagramas (variables cuantitativas)

- Histogramas
- Ojivas
- Polígonos de Frecuencia
- Diagramas de Caja

6.5. Gráficos y su interpretación

Histogramas

- **Diagramas de barras** para variables cuantitativas
- **Histograma** → de datos agrupados en intervalos
- **Diagramas de barras** → de datos no agrupados
- Se basan en una **representación de las frecuencias**, ya sean absolutas o relativas, **frente a los datos**
- En el caso de los histogramas, si se considera igual amplitud de intervalos, la anchura de los rectángulos ha de ser la misma y las alturas coincidirán con las frecuencias de cada clase

6.5. Gráficos y su interpretación

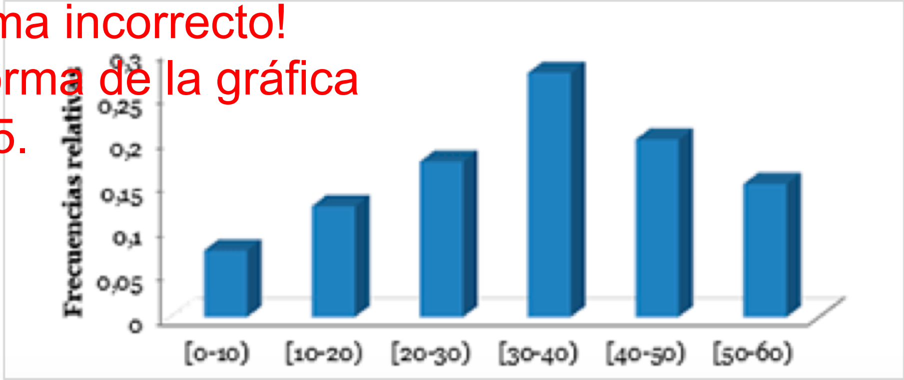
Ejemplo - Construcción de un histograma

Los datos de la siguiente tabla muestran los aciertos que han tenido un grupo de personas en un test. Construir un histograma.

I_i	f_i	h_i
0-10	33	0,236
10-20	25	0,179
20-30	45	0,321
30-40	17	0,121
40-50	5	0,036
50-60	15	0,107
Total	140	1

Tabla 12. Datos de aciertos en un test.

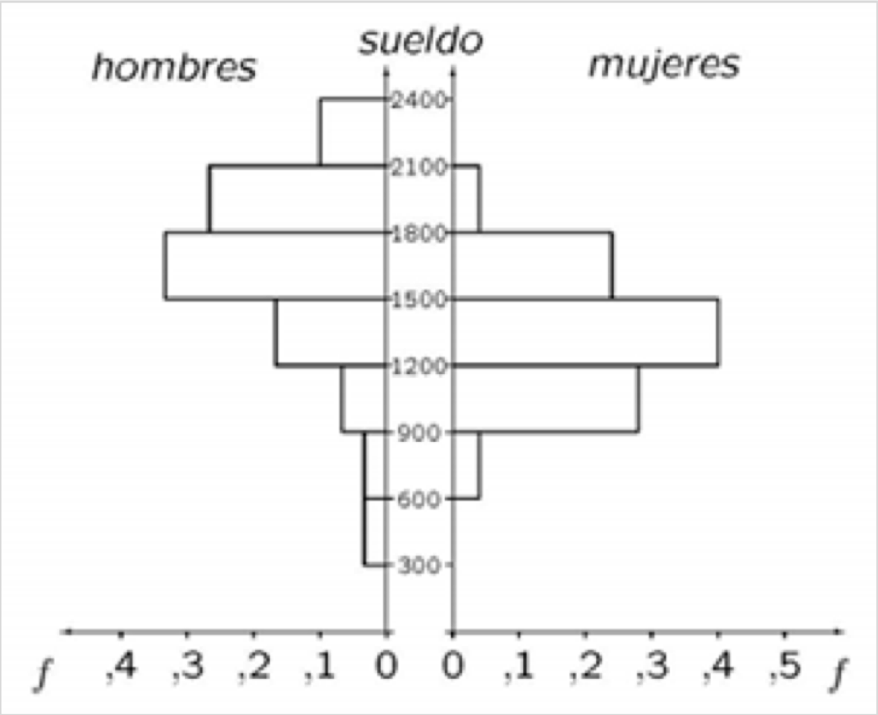
ERROR! Histograma incorrecto!
Debería tener la forma de la gráfica de la diapositiva 75.



6.5. Gráficos y su interpretación

Ejemplo - Interpretación de un histograma

¿Qué se puede decir sobre el salario de los hombres en relación al de las mujeres en función de los siguientes histogramas?

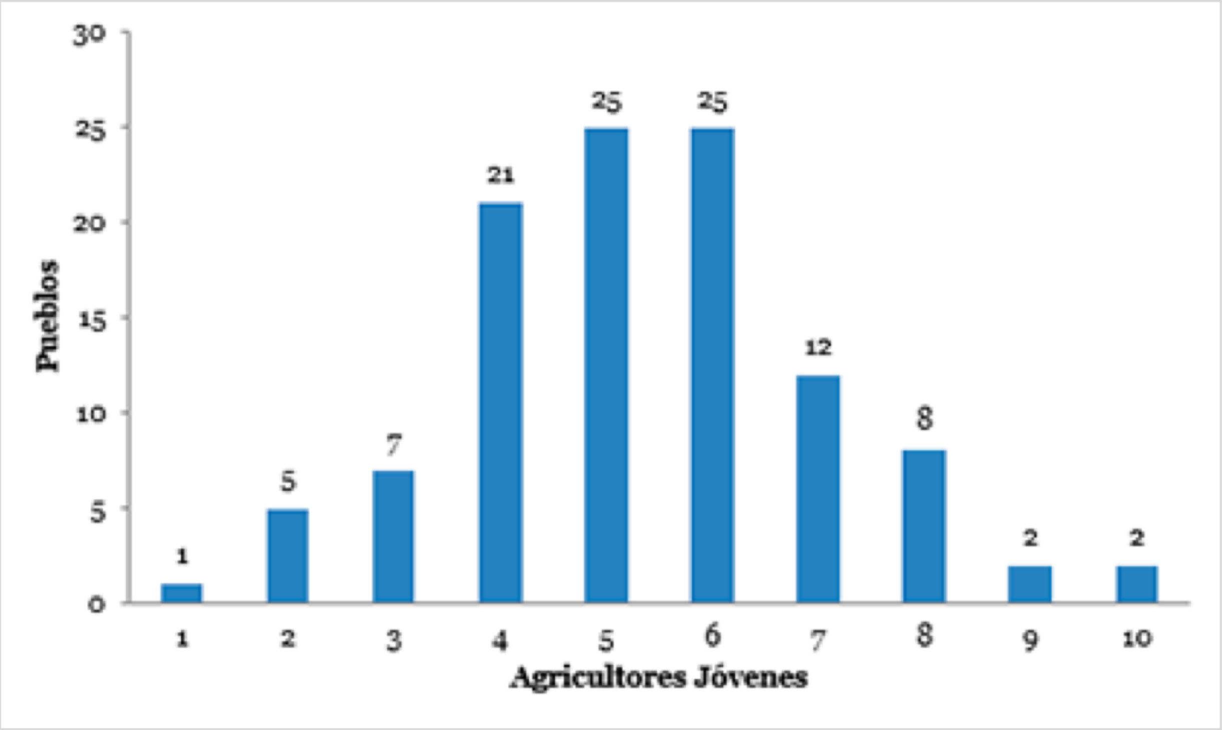


El sueldo medio de los hombres es más alto y se distribuye más asimétricamente.

6.5. Gráficos y su interpretación

Ejemplo - Interpretación de un diagrama de barras con datos discretos

El siguiente diagrama de barras muestra el número de agricultores jóvenes en unos pueblos de una comunidad autónoma. Responder a la siguientes preguntas:



6.5. Gráficos y su interpretación

A. ¿Cuántos pueblos se han tomado para el estudio?

108 pueblos.

B. ¿Cuántos pueblos tienen como máximo 7 agricultores jóvenes?

92 pueblos.

C. ¿Cuál es el porcentaje de pueblos con 3 agricultores jóvenes?

6,5% (7/108)

D. ¿Cuál es el número más frecuente de agricultores?

5 y 6 agricultores en 25 empresas.

6.5. Gráficos y su interpretación

E. ¿Cuál es el número máximo de agricultores jóvenes en los 25 pueblos más pequeños, es decir, con menos trabajadores?

Se hace la tabla de frecuencias acumuladas:

Agricultores jóvenes	Pueblos	Frecuencia acumulada
1	1	1
2	5	6
3	7	13
4	21	34
5	25	59
6	25	84
7	12	96
8	8	104
9	2	106
10	2	108

En los 25 pueblos con menos agricultores el máximo número de los mismos es 4.

6.5. Gráficos y su interpretación

F. Si una empresa quiere suministrar fertilizantes y solo quiere visitar a aquellos pueblos con más de 7 agricultores jóvenes, ¿a qué porcentaje de pueblos irá?

A los que tienen más de 7 agricultores, 12/108=11,1%

G. Si el 50% de pueblos con mayor número de agricultores puede beneficiarse de una subvención, ¿qué mínimo de agricultores ha de haber en los pueblos para conseguir esa subvencion?

El 50% del 108 pueblos es 54. Comparando con las frecuencias acumuladas, el máximo número de agricultores es 5.

6.5. Gráficos y su interpretación

Polígonos de Frecuencia

- Son diagramas de **puntos conectados** por una línea
- El área que dejan debajo coincide con el área de las barras del histograma y tiene forma poligonal, o curva cuando hay muchos datos
- Proporcionan la misma información que los histogramas
- En estos gráficos, se dan dos puntos en los extremos en los que las frecuencias sean cero para cerrar el polígono con el eje horizontal

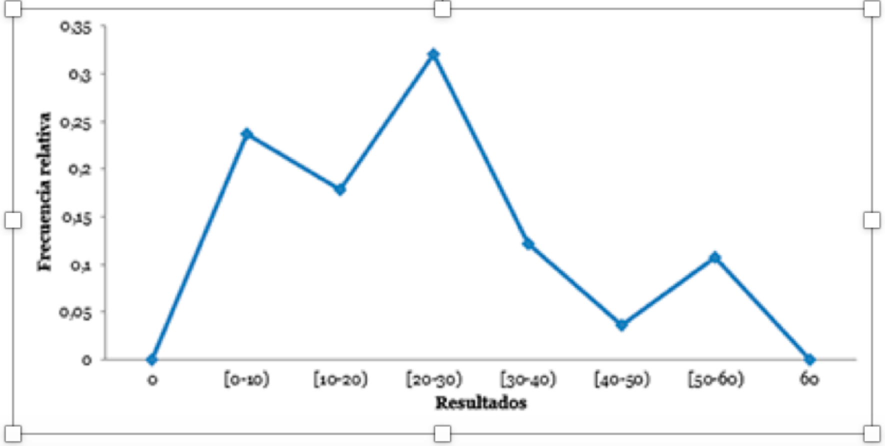
6.5. Gráficos y su interpretación

Ejemplo - Construir un polígono de frecuencias

Construir un polígono de frecuencias con los datos del Ejemplo anterior.

I_i	f_i	h_i
0-10	33	0,236
10-20	25	0,179
20-30	45	0,321
30-40	17	0,121
40-50	5	0,036
50-60	15	0,107
Total	140	1

Datos del ejemplo.



Ojivas

- Tipos de diagramas para datos agrupados que dan información de las **frecuencias acumuladas**
- Se muestran los datos que están por debajo o por encima de unos determinados valores en vez de los datos de los intervalos o las marcas de clase.
- Hay dos tipos de ojivas: ojivas ‘menor que’ y ojivas ‘mayor que’

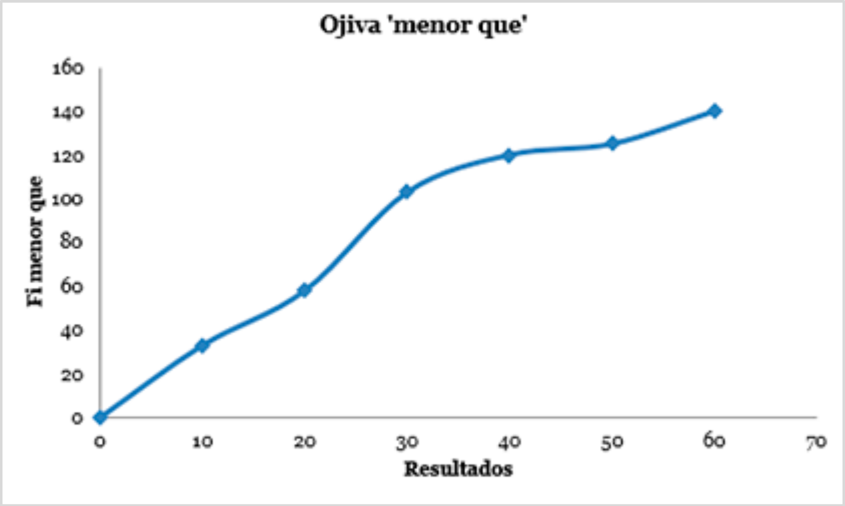
6.5. Gráficos y su interpretación

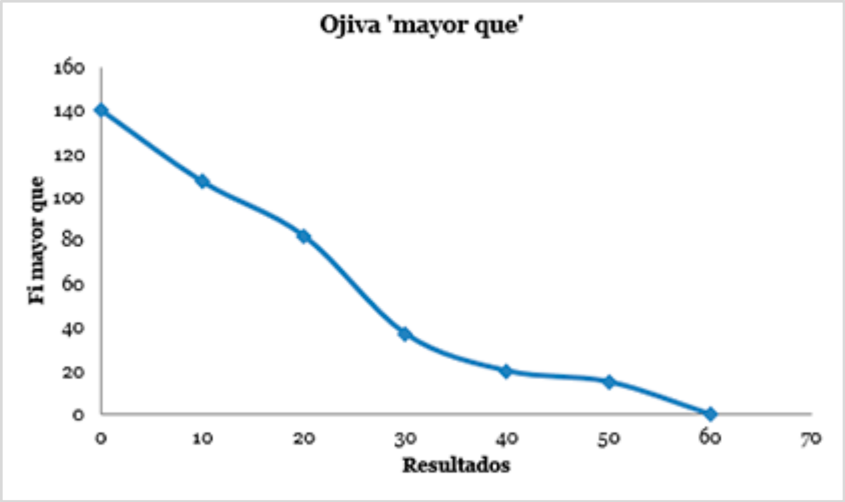
Ejemplo . Construir una ojiva 'menor que' y otra 'mayor que'

A partir de los datos del ejemplo 22 hacer una ojiva 'menor que' y una ojiva 'mayor que'.

Ojiva 'menor que'		Ojiva 'mayor que'	
Clase	F_i (menor que)	Clase	F_i (mayor que)
Menor que 0	0	Mayor que 0	140
Menor que 10	33	Mayor que 10	107
Menor que 20	58	Mayor que 20	82
Menor que 30	103	Mayor que 30	37
Menor que 40	120	Mayor que 40	20
Menor que 50	125	Mayor que 50	15
Menor que 60	140	Mayor que 60	0

6.5. Gráficos y su interpretación





6.5. Gráficos y su interpretación

Diagramas de Caja

- Tipo de representación basada en los **cuartiles** de una distribución
- Su representación es una caja, donde quedan los valores de los cuartiles y unos brazos llamados bigotes que van hasta los límites de la distribución
- Son ampliamente empleados para detectar la presencia de algún dato anómalo dentro de un conjunto de datos

6.5. Gráficos y su interpretación

Ejemplo - Elaborar un diagrama de cajas

A partir de los datos presentados a continuación elaborar un diagrama de cajas e interpretar los datos.

34	48	49	51	52	54	55	71	76

Calculamos la mediana y los cuartiles 1 y 3.

$$Posición\ de\ la\ Mediana = \frac{n+1}{2} = \frac{9+1}{2} = 5$$

6.5. Gráficos y su interpretación

La mediana es el dato 52.

$$Posición\ de\ Q_1 = \frac{n+1}{4} = \frac{9+1}{4} = 2{,}5$$

Q_1 es 48,5.

$$Posición\ de\ Q_3 = \frac{3(n+1)}{4} = \frac{3(9+1)}{4} = 7{,}5$$

Q_3 es 63.

Calculamos el rango intercuartílico:

$$Rango\ intercuartílico = Q_3 - Q_1 = 63 - 48{,}5 = 14{,}5$$

Calculamos los límites inferior y superior:

$$L_i = Q_1 - 1{,}5 \cdot RIC = 48{,}5 - 1{,}5 \cdot 14{,}5 = 26{,}75$$

$$L_s = Q_3 + 1{,}5 \cdot RIC = 63 + 1{,}5 \cdot 14{,}5 = 84{,}75$$

6.5. Gráficos y su interpretación

A continuación, se calcula la anchura de cada rectángulo del diagrama:

	Valores	Anchos
Mínimo	34	34
Q_1	48,5	14,5
Mediana	52	3,5
Q_3	63	11
Máximo	76	13

6.5. Gráficos y su interpretación

Se dibuja el diagrama de cajas:

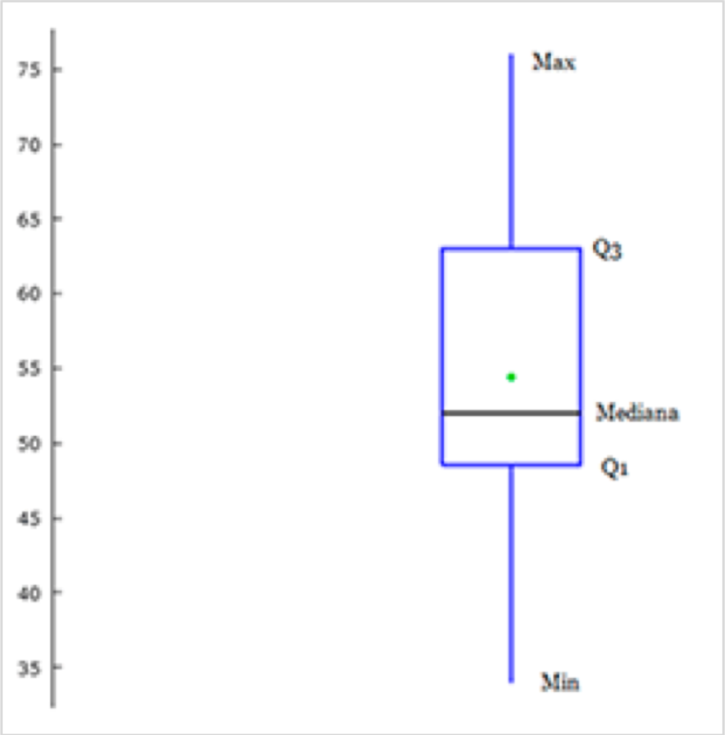


Gráfico 7. Diagrama de cajas.

Usa el Campus Virtual, Pregunta, Participa!

UNIVERSIDAD INTERNACIONAL DE LA RIOJA unir

www.unir.net