Estadística Descriptiva

Tema 1

Introducción

CONTENIDO

1. Introducción
2. Definiciones básicas
3. Presentación de datos: tablas y gráficos
4. Medidas de centralización
5. Medidas de dispersión
6. Medidas de posición

Etapas en el análisis estadístico

- Definición del problema de estudio y objetivos del mismo.
- Selección y recogida de la información necesaria para realizar el estudio.
- Ordenación y clasificación de la información en tablas y gráficos.
- Resumen de los datos mediante medidas de centralización, dispersión, asimetría y posición.
- Análisis estadístico formal obteniendo hipótesis y contrastándolas
- Interpretación de resultados y conclusiones.
- Extrapolación y predicción.

Definiciones básicas

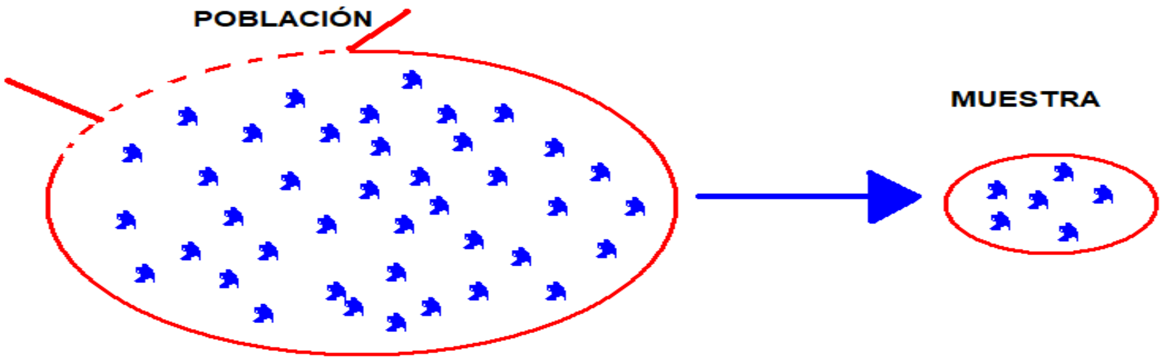
Definiciones básicas

POBLACIÓN

Conjunto de todos los individuos que presentan el carácter objeto de estudio. Según el numero de individuos que la componen puede ser:

Finita: nº finito de individuos. Ejemplo Universitarios de la Comunidad de Madrid.

Infinita: formada por un nº infinito de individuos: Números reales positivos.



Muestra

Subconjunto de la población que sirve para estudiarla y que es representativa de la misma. Para hacer que la **muestra** sea representativa de la **población** se utiliza una rama de la Estadística denominada **muestreo**.

Variable estadística

Cada uno de los rasgos o características de los elementos de una población o muestra objeto de estudio.

Clasificación de variables estadísticas

Cualitativas

Expresan cualidades o atributos. Pueden presentarse en escala nominal u ordinal.

Ej.:

- Estado civil
- Cursos del primer ciclo
- Tipos de envase

Cuantitativas

Expresan caracteres definidos por valores numéricos. Se dividen en:

- Discretas
- Continuas

Clasificación de las variables cuantitativas

Discreta

Cuando el conjunto de valores se puede enumerar (es contable). Se suelen expresar mediante un nº entero

Ej.:

-Número de artículos defectuosos

-Número de hijos de una familia

-Número de terremotos durante un año.

Continua

Es aquella variable que puede tomar cualquier valor en un intervalo de la recta real.

Ej.:

-Peso de una persona

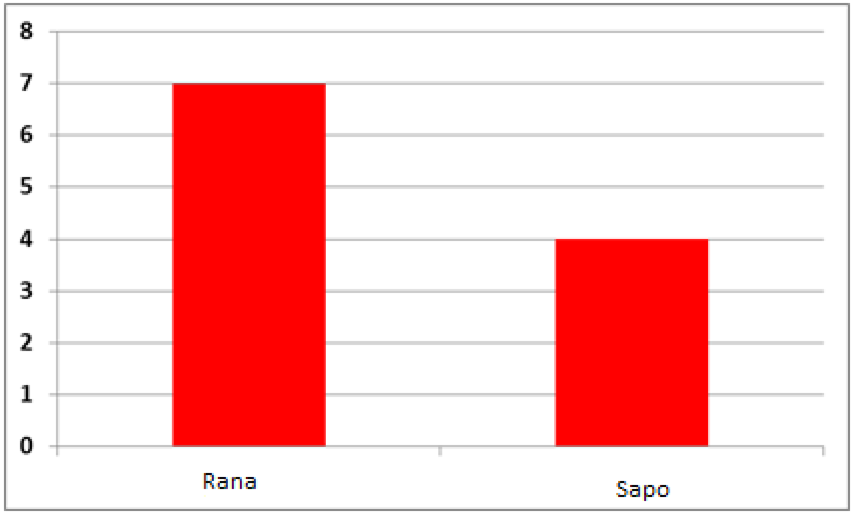
-Temperatura de ignición de un gas

-Resistencia a la tensión de una probeta de cemento

Presentación de datos: Tablas y gráficos

Presentación ordenada de datos: tablas y gráficos

Especie	Frec.
Rana	70
Sapo	40



- Las **tablas de frecuencias** y las representaciones gráficas son dos maneras ***equivalentes*** de presentar la información. Las dos exponen ordenadamente la información contenida en los datos de una muestra.

TABLAS DE FRECUENCIAS

- Una **distribución de frecuencias** o **tabla de frecuencias** es una **ordenación** en forma de **tabla** de los **datos estadísticos**, asignando a cada **dato** su **frecuencia correspondiente**.

Número	Frecuencia
1	3
2	1
3	2
4	1
5	1
6	2

Columnas que aparecen en las tablas de frecuencias

- Variable estadística (dato) x_i
- Frecuencia absoluta n_i
- Frecuencia relativa f_i
- Frecuencia absoluta acumulada N_i
- Frecuencia relativa acumulada F_i

VARIABLE ESTADÍSTICA

Datos sin agrupar (categorías)

Cuando los datos no están agrupados (**categorías**), los valores x_i representan los propios datos. El número total de datos se suele denotar por N.

Datos agrupados (clases)

Cuando el número de datos es muy elevado, estos suelen agruparse en intervalos de la forma $[L_i, L_{i+1})$ denominados **clases.**

En este caso los valores x_i no son los propios datos, sino el punto medio de cada clase (**marca de clase**)

$$x_i = \frac{L_i + L_{i+1}}{2}$$

FRECUENCIAS

Frecuencia absoluta (n_i)

Número de veces que aparece un determinado dato en la muestra.

Frecuencia relativa (fi):

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

Cociente entre la frecuencia absoluta de un determinado dato y el **número total de datos de la muestra (*N*)**.

Frecuencias

Frecuencia absoluta acumulada (Ni)

Suma de la frecuencia absolutas n_i de cada clase o categoría y sus precedentes.

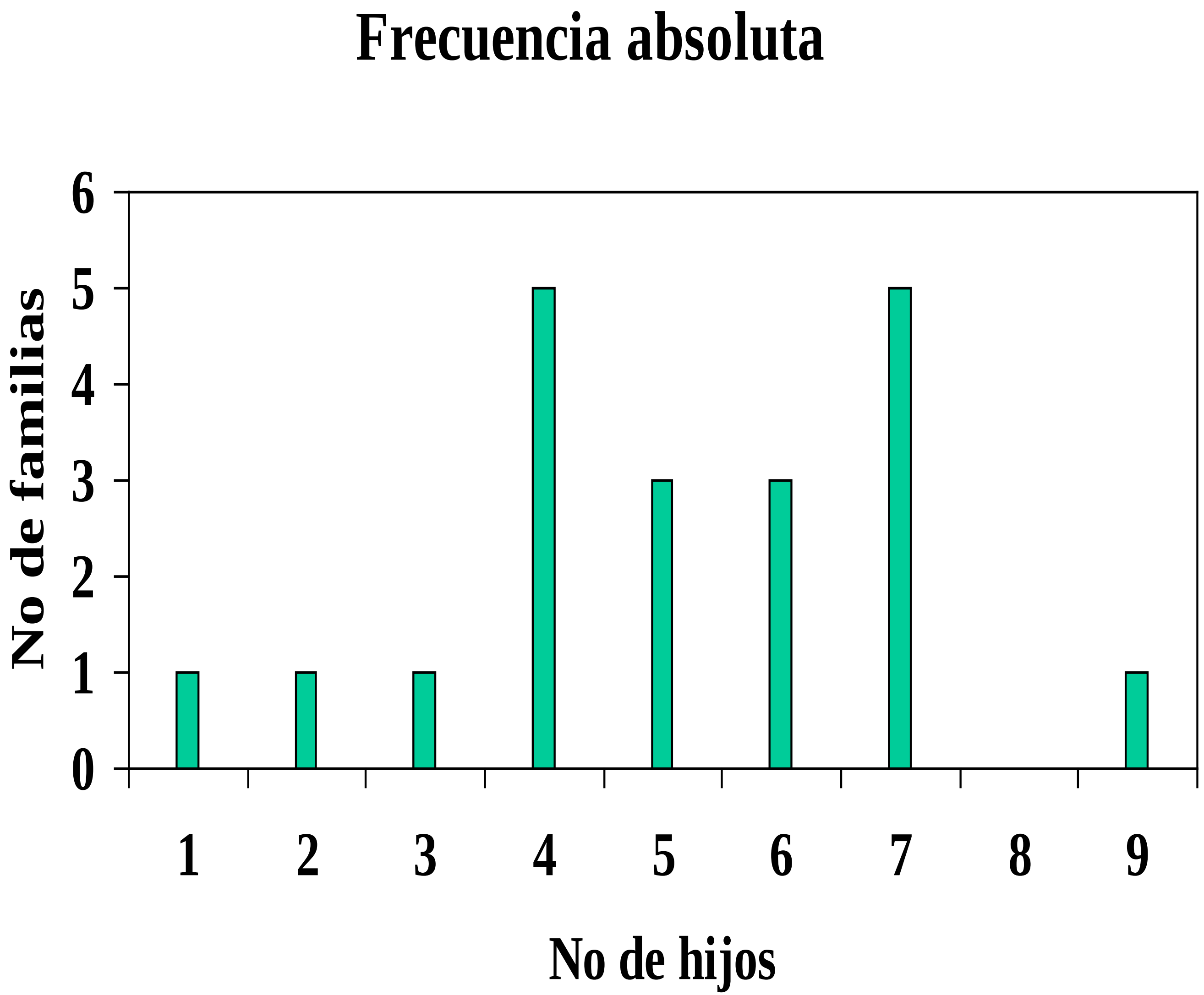
Frecuencia relativa acumulada (Fi)

Suma de la frecuencia relativa f_i de cada clase o categoría y sus precedentes.

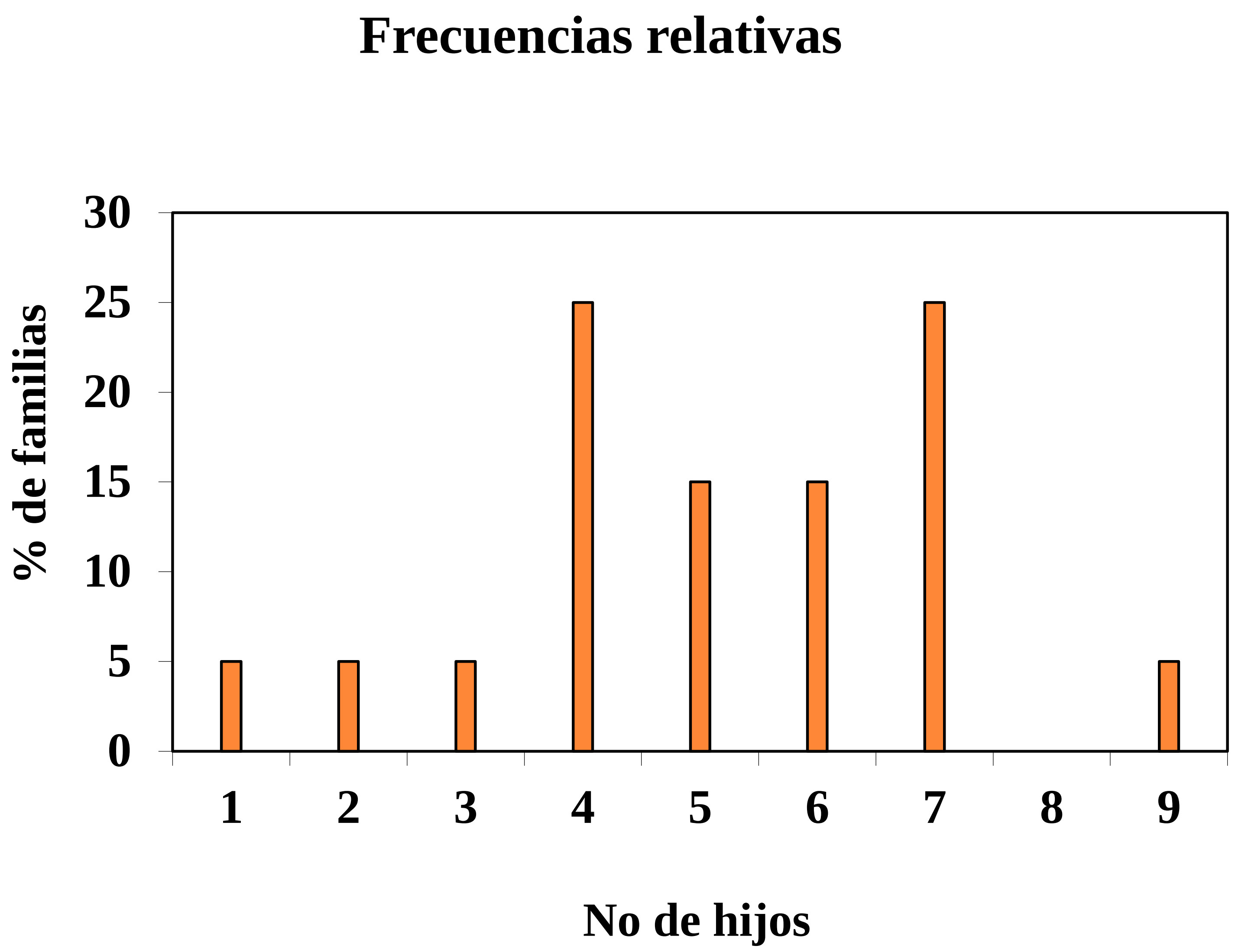
Ejemplo de Tabla con datos no agrupados

Nota	Frecuencia Absoluta	Frecuencia Relativa	Frecuencia Relativa Porcentual (%)
2,8	1	0,041	4,166
3,2	4	0,166	16,666
3,9	3	0,125	12,500
4,2	5	0,208	20,833
5,0	4	0,166	16,666
5,6	3	0,125	12,500
6,0	4	0,166	16,666

Presentación de datos mediante gráficos: Diagrama de barras o bastones



Presentación de datos mediante gráficos: Diagrama de barras o bastones



TABLAS DE DATOS AGRUPADOS

"VARIABLE"

$[L_i , L_{i+1})$	x_i	n_i	f_i	N_i	F_i
$[L_{i+1} , L_{i+2})$					
…					
Total					

Donde: $[L_i , L_{i+1})$ **: Intervalos o Clases**

x_i **: Marca de clase**

$\mathbf{n_i}$ **: frecuencia absoluta**

$\mathbf{f_i}$ **: frecuencia relativa**

$\mathbf{N_i}$ **: Frecuencia absoluta acumulada**

$\mathbf{F_i}$ **: Frecuencia relativa acumulada**

$\mathbf{L_i}$: Límite Inferior

$\mathbf{L_{i+1}}$: Límite Superior

COMO AGRUPAR LOS DATOS

1. Calcular el Rango de la distribución

$$R = x_{max} - x_{min}$$

2. Obtener el nº de clases K:

   a. Regla de Sturges: $\mathrm{K} = 1 + 3{,}33 \log N$

   b. $\mathrm{K} = \sqrt{N}$ ó $\mathrm{K} = 2\sqrt{N}$

3. Amplitud de clase: $A = \frac{R}{K}$ (para clases de igual amplitud)

4. Calcular la **marca de clase**: $x_i = \frac{\mathrm{L_i} + \mathrm{L_{i+1}}}{2}$

TABLAS PARA VARIABLES AGRUPADAS

Intervalos	x_i	n_i	f_i	N_i	F_i
[42-58)	50	2	0.050	2	0.050
[58-74)	66	5	0.125	7	0.175
[74-90)	82	15	0.375	22	0.550
[90-106)	98	9	0.225	31	0.775
[106-122)	114	6	0.150	37	0.925
[122-138)	130	3	0.075	40	1.000
TOTAL	**N**	**N=40**	**1.000**		

EJEMPLO : CONSTRUIR TABLA DE FRECUENCIAS AGRUPADAS

Se ha realizado una encuesta en **40** centros de salud con el fin de observar el número de pacientes atendidos en ellos, obteniéndose:

42	58	79	86	98	120	134	120	59	62
85	89	76	110	104	78	84	96	90	75
120	130	122	95	82	94	108	79	105	115
102	80	56	78	84	66	69	78	84	98

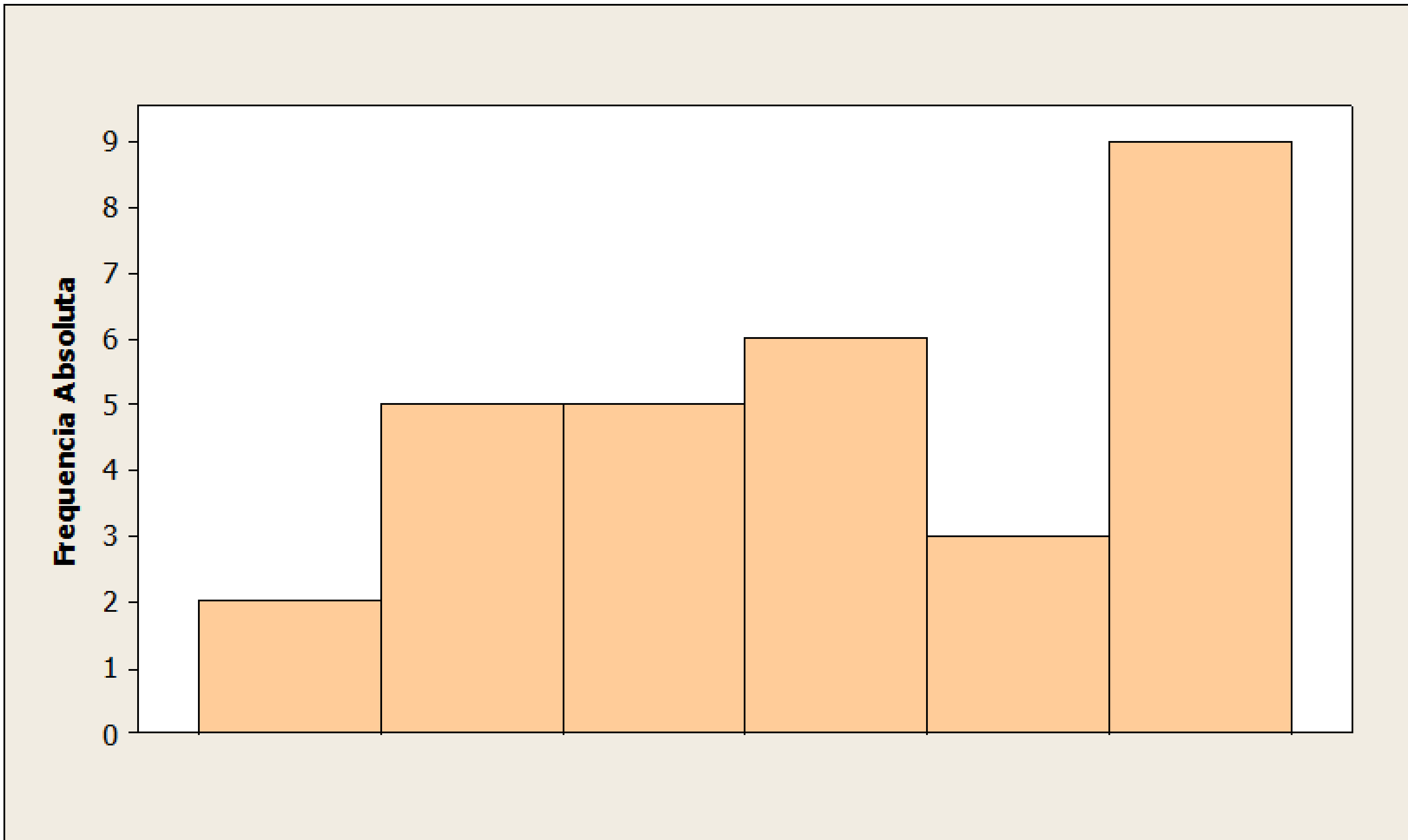
Procedimiento

- Rango: $R = 134 - 42 = 92$
- Número de intervalos: $K = 1 + 3.3 \log 40 = 6.29$ (fórmula de Sturges) redondeo a 6 intervalos
- Amplitud de clase: $A = 92 / 6 = 15.33$ redondeamos por exceso a 16

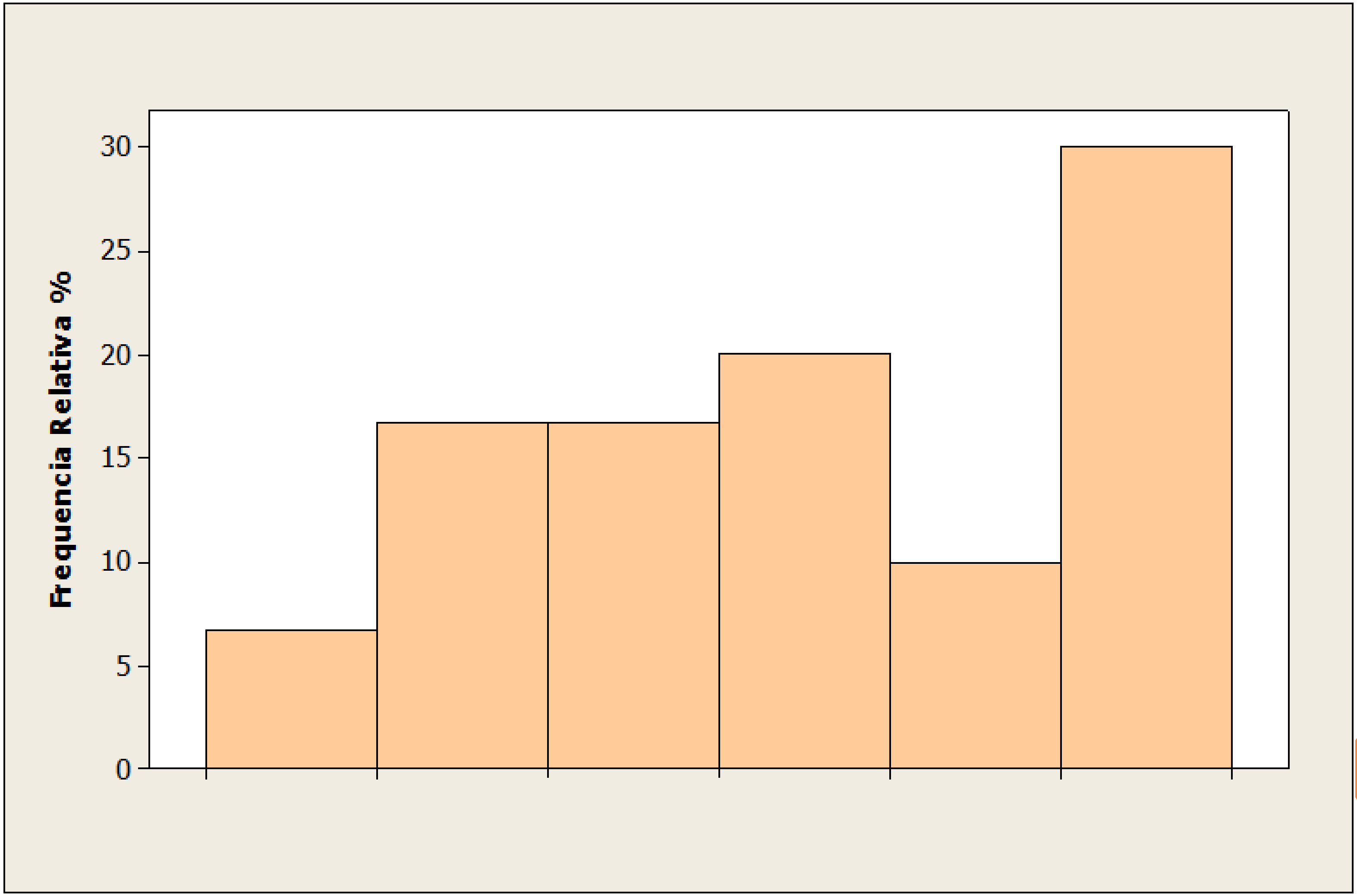
Número de pacientes (variable estadística agrupada en clases)

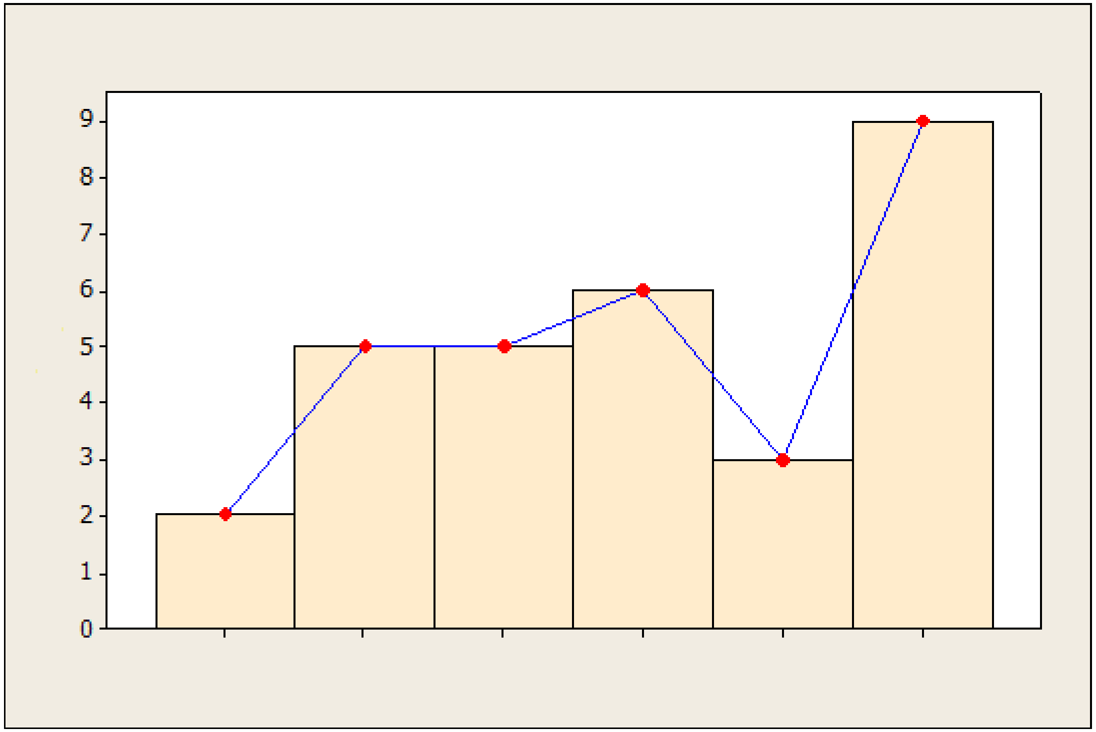
Intervalos	x_i	n_i	f_i	N_i	F_i
[42-58)	50	2	0.050	2	0.050
[58-74)	66	5	0.125	7	0.175
[74-90)	82	15	0.375	22	0.550
[90-106)	98	9	0.225	31	0.775
[106-122)	114	6	0.150	37	0.925
[122-138)	130	3	0.075	40	1.000
TOTAL	**N**	**N=40**	**1.000**		

Presentación de datos agrupados en clases mediante gráficos: Histograma de frecuencias absolutas

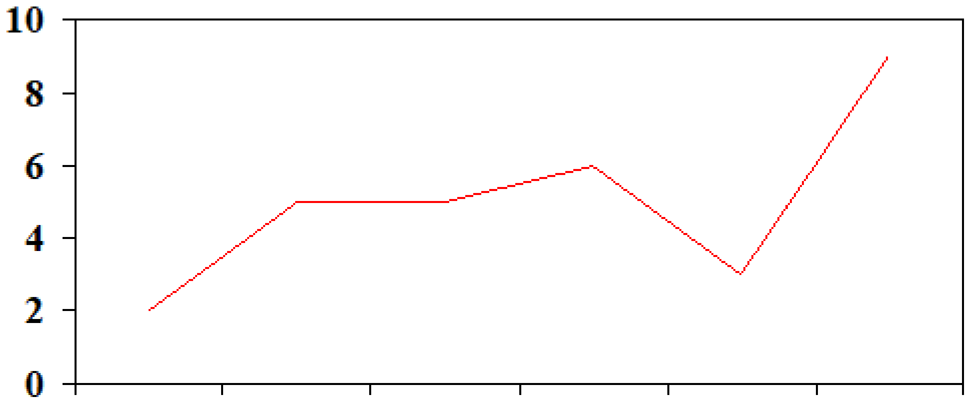


Presentación de datos agrupados en clases mediante gráficos: Histograma de frecuencias relativas



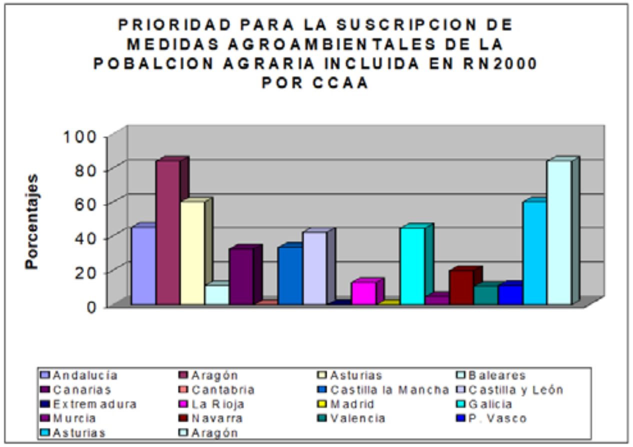


Polígono de frecuencias absolutas

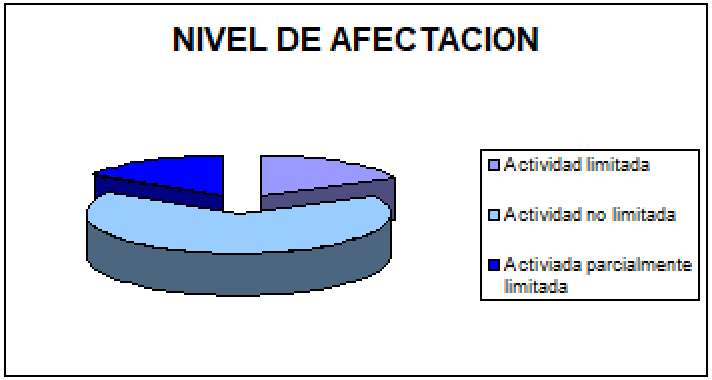


Presentación de datos para variables cualitativas

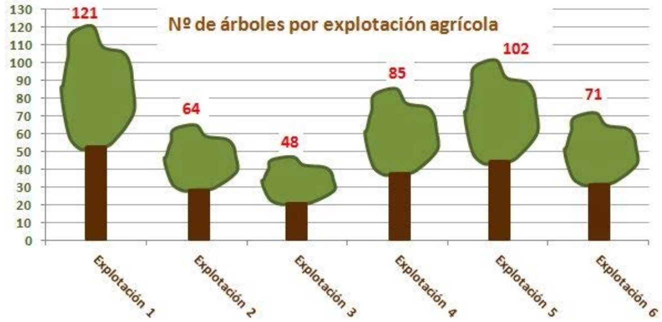
- Diagramas de barras
  - Alturas proporcionales a las frecuencias (absolutas o relativas)
  - Se pueden aplicar también a variables discretas



- Diagramas de sectores
  - El área de cada sector es proporcional a su frecuencia (absolutas o relativas)

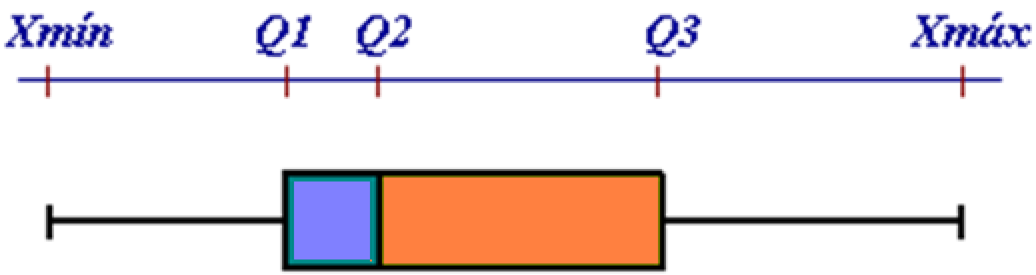


- Pictogramas
  - Fáciles de entender.
  - El área de cada modalidad debe ser proporcional a la frecuencia.



Diagramas de cajas y bigotes (box and whisker-plot)

- Presentación visual que describe varias características al mismo tiempo, tales como la dispersión y la simetría y la existencia de **datos atípicos**.
- En la caja aparecen **los tres cuartiles** y los bigotes alcanzan los valores menor/mayor de los datos en un rango de 1.5 veces el ancho de la caja (rango intercuartílico o iqr) por debajo/encima de los cuartiles primero/tercero.



Medidas de centralización

CONTENIDO

- Introducción
- Media aritmética
- Mediana
- Moda

INTRODUCCION

- **Parámetro (poblacional):** Es una cantidad numérica calculada sobre una población

  - Ejemplo: La altura media de los individuos de un país

  - La idea es resumir toda la información que hay en la población en unos pocos números (parámetros).

- **Estadístico:** Es una cantidad numérica calculada sobre la muestra.

  - La altura media de los que estamos en este aula.
    - Somos una muestra (¿representativa?) de la población.
    - Si un estadístico se usa para aproximar un parámetro también se le suele llamar estimador.

  - **Normalmente nos interesa conocer un parámetro, pero por la dificultad que conlleva estudiar a *TODA* la población, calculamos un estimador sobre una muestra y "confiamos" en que sean próximos. Más adelante veremos como elegir muestras para que el error sea "confiablemente" pequeño.**

Introducción

- **Centralización**
  - **Indican valores con respecto a los que los datos parecen agruparse.**
    - **Media, mediana y moda**
- **Posición**
  - **Dividen un conjunto ordenado de datos en grupos con la misma cantidad de individuos.**
    - **Cuantiles: percentiles, cuartiles, deciles,...**
- **Dispersión**
  - **Indican la mayor o menor concentración de los datos con respecto a las medidas de centralización.**
    - **Desviación típica, coeficiente de variación, rango, varianza**
- **Forma**
  - **Asimetría o sesgo**
  - **Apuntamiento o curtosis**

Media aritmética

- Media Aritmética para una muestra: Consiste en sumar todos los datos y dividirlos por N.

$$\bar{X} = \frac{x_1 + x_2 + \cdots + x_N}{N} = \frac{\sum_{i=1}^{N} x_i}{N}$$

Ejemplo

- El numero de hijos por familia queda representado en la siguiente tabla. Calcular el numero medio de hijos por familia

Xi	ni	xi.ni
0	5	0
1	6	6
2	8	16
3	4	12
4	2	8
	25	**42**

$$\bar{X} = \frac{42}{25} = 1.68$$

MEDIANA

- **La mediana** (Me) de un conjunto de observaciones se define como el valor que queda en la parte central de este conjunto ordenados de menor a mayor.

Mediana en datos no agrupados

- Se ordenan los datos de menor a mayor y se busca el dato que ocupa la posición central.
  - Si la distribución tiene un nº impar de datos, habrá un solo dato, que represente el valor central
  - Si la distribución tiene un nº par de datos, habrá dos valores, por tanto la mediana será la media aritmética de los dos datos que ocupan la posición central.

Ejemplo 1

- El numero medio de hijos por familia en un país africano es el siguiente:

2, 5, 6, 4, 5, 6, 4, 5, 3

- Para calcular la mediana:
- 1º ordenamos los datos de menos a mayor:

2, 3, 4, 4, 5, 5, 5, 6, 6

- La mediana es el valor que ocupa la posición central es decir:

Me= 5

Ejemplo 2

- Los sueldos de 8 trabajadores de una fábrica son los siguientes:

650, 556, 722, 478, 570, 660, 814, 670

Para calcular la mediana:

1. Ordenamos los datos de menos a mayor:

478, 556, 570, 650, 660, 670, 722, 814

En este caso tenemos dos valores que ocupan posiciones centrales: 650 y 660

Luego la mediana será:

$$\frac{650+660}{2} = 655$$

Mediana en datos agrupados

- Para calcular la mediana de *n* datos tabulados por intervalos, primero se determina el *intervalo mediano* que es aquél cuya frecuencia absoluta acumulada N_i excede de N/2.
- Luego se utiliza la fórmula:

$$M_e = L_i + \left(\frac{\frac{N}{2} - N_{i-1}}{n_i} \right) a_i$$

- donde:
  - L_i es el límite inferior del intervalo que contiene a la mediana.
  - $N/2$ semisuma de las frecuencias absolutas
  - N_{i-1} = frecuencia acumulada de la clase anterior a la que contiene a la mediana
  - a_i= amplitud del intervalo que contiene a la mediana
  - n_i= frecuencia absoluta del intervalo mediano

Ejemplo 3

	ni	Ni
[10, 15)	3	3
[15, 20)	5	8
[20, 25)	7	15
[25, 30)	4	19
[30, 35)	2	21
	21	

El intervalo mediano es [20,25) es decir, i=3

$$M_e = L_i + \left(\frac{\frac{N}{2} - N_{i-1}}{n_i} \right) a_i$$

$$\frac{N}{2} = 10.5$$

$$Me = 20 + \frac{10.5 - 8}{7} \cdot 5 = 21.786$$

Moda para datos no agrupados

- Valor que más se repite, o que presenta mayor frecuencia absoluta (mayor n_i).

- Ejemplo: 2,3,3,4,4,4,5,5

Mo= 4

- Puede haber distribuciones multimodales:
- Ejemplo: 2,3,3,3,4,4,4,5,5
- En este caso, hay dos modas:

Mo=3 y Mo=4

Moda para datos agrupados en intervalos

- Se aplica la siguiente formula:

$$Mo = L_i + \frac{n_i - n_{i-1}}{(n_i - n_{i-1}) + (n_i - n_{i+1})} \cdot a_i$$

- Donde:
- L_i: limite inferior del intervalo que contiene a la moda
- n_i: frecuencia absoluta de la clase que contiene a la moda
- n_{i-1}: frecuencia absoluta de la clase inmediatamente anterior a la que contiene a la moda
- n_{i+1}: frecuencia absoluta de la clase posterior a la que contiene a la moda
- a_i: amplitud de la clase que contiene a la moda

Ejemplo: calculo de la moda para valores agrupados

	n_i
[60-63)	5
[63-66)	18
[66-69)	42
[69-72)	27
[72-75]	8
	Σ100

$$M_o = L_i + \frac{n_i - n_{i-1}}{(n_i - n_{i-1}) + (n_i - n_{i+1})} a_i$$

$$M_o = 66 + \frac{42 - 18}{(42 - 18) + (42 - 27)} 3 = 67{,}8$$

Medidas de posición

Medidas de posición

- Dividen la muestra en varias partes.
- Para calcular las **medidas de posición** es necesario que los **datos** estén ordenados de **menor a mayor**.
- La **medidas de posición** se llaman en general cuantiles.

Principales medidas de posición

- Percentil o centil
- Decil
- Cuartil
- Quintil
- Tercil

PERCENTILES

- El percentil k, al que simbolizaremos como P_k, es el valor numérico de la variable tal que el k por ciento de los datos ordenados está por debajo de ese valor. En consecuencia, el $(100 - k)$ por ciento de los datos está por encima de P_k.

$$\boldsymbol{k = 1, 2, 3 \dots 99}$$

- Para hallar los percentiles primero detectamos a qué intervalo pertenece el percentil que buscamos y después procedemos a su cálculo.

CALCULO DEL PERCENTIL PARA DATOS NO AGRUPADOS

Se siguen los siguientes pasos:

1) Se construye la tabla de frecuencias con los datos ordenados de menos a mayor
2) Se calcula la columna de frecuencias absolutas acumuladas
3) Es el primer valor cuya frecuencia absoluta acumulada excede la cantidad:

$$\frac{k \cdot N}{100}$$

Siendo:

$$k\text{: } 1, 2, 3, \ldots, 99$$

N: Número de datos de la distribución (tamaño muestral)

Ejemplo

- Calcular los percentiles 14 y 62

X_i	n_i
2	6
3	15
4	10
5	9
	40

N_i
6
21
31
40

$$P_{14} \rightarrow K = 14; \ \frac{14 \times 40}{100} = 5{,}6 \rightarrow P_{14} = 2$$

$$P_{62} \rightarrow K = 62; \ \frac{62 \times 40}{100} = 24{,}8 \rightarrow P_{62} = 4$$

Calculo del percentil para datos agrupados

Localización

$$\frac{k \cdot N}{100}$$

Identificación

$$P_k = L_i + \frac{\frac{k \cdot N}{100} - N_{i-1}}{n_i} a_i$$

EJEMPLO

Calcular el percentil 35 de la siguiente distribución:

	ni
[50-60)	8
[60-70)	10
[70-80)	16
[80-90)	14
[90-100)	10
[100-110)	5
[110-120)	2
	65

Ni
8
18
34
48
58
63
65

$$P_k = L_i + \frac{\frac{k \cdot N}{100} - N_{i-1}}{n_i} a_i$$

$$\text{K} = 35 \rightarrow \frac{35 \times 65}{100} = 22{,}27$$

$$P_{35} = 70 + \frac{22{,}27 - 18}{16} 10 = 72{,}95$$

DECILES

- El decil k, al que simbolizaremos como D_k, es el valor numérico de la variable tal que el k por ciento de los datos ordenados está por debajo de ese valor. En consecuencia, el $(100-k)$ por ciento de los datos está por encima de D_k.

$$\boldsymbol{k = 1, 2, 3, \ldots 9}$$

- Para hallar los deciles primero detectamos a que intervalo pertenece el decil que buscamos y después procedemos a su cálculo.

Calculo de deciles para datos no agrupados

Se siguen los siguientes pasos:

1) Se construye la tabla de frecuencias con los datos ordenados de menos a mayor
2) Se calcula la columna de frecuencias absolutas acumuladas
3) Es el primer valor cuya frecuencia absoluta acumulada excede la cantidad:

$$\frac{k \cdot N}{10}$$

Siendo:

k: 1, 2,……. 9

N: Número de datos de la distribución

EJEMPLO DECILES PARA DATOS NO AGRUPADOS

- Calcular los Deciles 3 y 8

Xi	ni
2	6
3	15
4	10
5	9
	40

Ni
6
21
31
40

$$D_3 \rightarrow k = 3; \ \frac{3\times 40}{10} = 12 \rightarrow D_3 = 3$$

$$D_8 \rightarrow k = 8; \ \frac{8\times 40}{10} = 32 \rightarrow D_8 = 5$$

Calculo de deciles para datos agrupados

Localización

$$D_k = \frac{k \cdot N}{10}$$

Identificación

$$D_k = L_i + \frac{\frac{k \cdot N}{10} - N_{i-1}}{n_i} a_i$$

EJEMPLO

Calcular el decil 6 de la siguiente distribución:

$$D_k = L_i + \frac{\frac{k \cdot N}{10} - N_{i-1}}{n_i} a_i$$

	ni	Ni
[50-60)	8	8
[60-70)	10	18
[70-80)	16	34
[80-90)	14	48
[90-100)	10	58
[100-110)	5	63
[110-120)	2	65
	65	

$$\text{k} = 6 \rightarrow \frac{6 \times 65}{10} = 39$$

$$D_6 = 80 + \frac{39-34}{14}\, 10 = 83{,}57$$

CUARTILES

- Son tres valores que dividen a un conjunto de datos ordenados en cuatro partes del mismo tamaño.
- Q_1, Q_2 y Q_3 determinan los valores correspondientes al 25%, 50% y 75% de los datos. Q_2 coincide con la mediana, el percentil 50 y el decil 5
- En este caso:

$$\boldsymbol{k = 1, 2, 3}$$

- Para hallar los percentiles primero detectamos a qué intervalo pertenece el percentil que buscamos y después procedemos a su calculo.

EJEMPLO

- Calcular los cuartiles 1 y 3 de la siguiente distribución:

Xi	ni
2	6
3	15
4	10
5	9
	40

Ni
6
21
31
40

- $Q_k = \frac{K \cdot N}{4}$

$$Q_1 \rightarrow K = 1 \rightarrow \frac{1 \cdot 40}{4} = 10 \rightarrow Q_1 = 3$$

$$Q_3 \rightarrow K = 3 \rightarrow \frac{3 \cdot 40}{4} = 30 \rightarrow Q_3 = 4$$

Calculo de los cuartiles para datos agrupados

Localización

$$Q_k = \frac{k \cdot N}{4}$$

Identificación

$$Q_k = L_i + \frac{\frac{k \cdot N}{4} - N_{i-1}}{n_i} a_i$$

EJEMPLO

Calcular el segundo cuartil de la siguiente distribución:

	ni	Ni
[50-60)	8	8
[60-70)	10	18
[70-80)	16	34
[80-90)	14	48
[90-100)	10	58
[100-110)	5	63
[110-120)	2	65
	65	

$$Q_2 = L_i + \frac{\frac{kN}{4} - N_{i-1}}{n_i} a_i$$

$$\text{K}= 2 \rightarrow \frac{2 \times 65}{4} = 32{,}5$$

$$Q_2 = 70 + \frac{32{,}5 - 18}{16} 10 = 79{,}062$$

EQUIVALENCIAS

- Equivalencias entre cuartiles, deciles y percentiles:

Cuartil 1 : $Q_1 = P_{25}$

Cuartil 2 : $Q_2 = P_{50} = Me$

Cuartil 3 : $Q_3 = P_{75}$

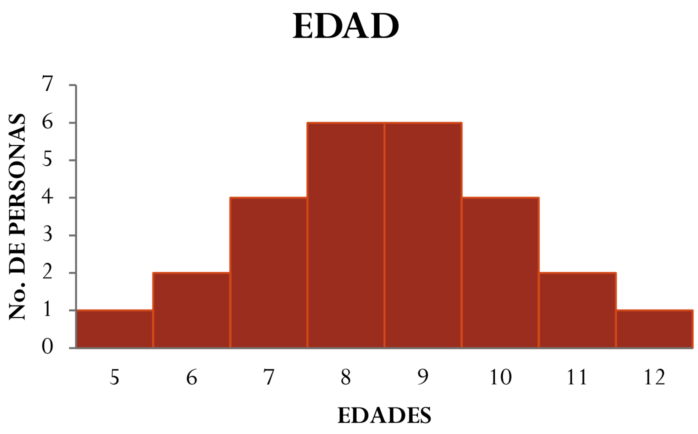
Medidas de dispersión

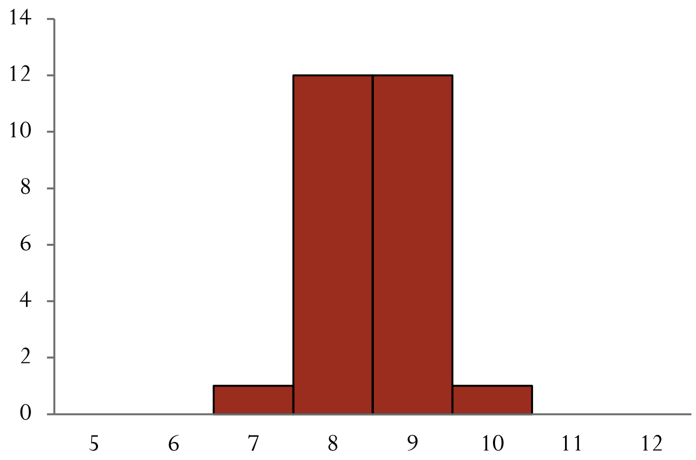
INDICE

- Rango
- Rango intercuartílico
- Varianza
- Cuasivarianza
- Desviación típica
- Cuasidesviación típica
- Coeficiente de variación (de Pearson)

MEDIDAS DE DISPERSIÓN

- La *Dispersión* hace referencia a la manera en que se dispersan o alejan los datos de la media





MEDIA: 8.5 años

MEDIDAS DE DISPERSIÓN

- La *DISPERSIÓN* (o variabilidad) de un conjunto de datos es una medida de la distancia entre los datos, y su media. Las medidas de dispersión miden la variabilidad de una nuestra independientemente de su causa.
- Poca dispersión = Datos homogéneos = Media muy representativa
- Mucha dispersión = Datos heterogéneos = Media poco representativa

Rango

- *Amplitud o Rango*:
  Diferencia entre observaciones extremas.
  - 2,1,4,3,8,4. El rango es 8-1=7
  - Es muy sensible a los valores atípicos.

- *Rango intercuartílico*:
  - Es la distancia entre el primer y el tercer cuartil:
    Rango intercuartílico = (P_{75} - P_{25}) ó (Q_3 - Q_1)
  - Parecida al rango, pero eliminando las observaciones más extremas inferiores y superiores.
  - No es tan sensible a valores atípicos.

EJEMPLO

La siguiente tabla representa las edades de un conjunto de personas. Calcular su rango

10
13
22
26
16
23
35
53
17
32
41
35
24
23
27
16
20
60
48

Valor Máximo: 60

Valor Mínimo: 10

$$\text{Rango} = X_{Max} - X_{Min}$$
$$= 60 - 10$$
$$= 50$$

VARIANZA S^2

Se llama varianza de una serie de datos x_1, x_2, x_3,…… x_n

Y se representa como S^2 al siguiente cociente:

$$S^2 = \frac{\sum_{i=1}^{N}(x_i - \bar{x})^2}{N}$$

Si desarrollamos el cociente anterior y operamos queda:

$$S^2 = \frac{\sum_{i=1}^{N} x_i^2}{N} - \bar{x}^2$$

CUASIVARIANZA

$$\hat{s}^2 = \frac{\sum_{i=1}^{N}(x_i - \bar{x})^2}{N-1}$$

Ejemplo cálculo de la varianza

Tras encuestar a 25 familias sobre el número de hijos que tenían, se obtuvieron los de la siguiente tabla. Calcular la varianza.

Xi	ni	fi	Ni	Fi
0	5	0,20	5	0,20
1	6	0,24	11	0,44
2	8	0,32	19	0,76
3	4	0,16	23	0,92
4	2	0,08	25	1
	25	1		

Ejemplo de calculo de la varianza

Xi	ni	xi.ni	xi2	xi2.ni
0	5	0	0	0
1	6	6	1	6
2	8	16	4	32
3	4	12	9	36
4	2	8	16	32
	25	**42**	**30**	**106**

$$S^2 = \frac{\sum_{i=1} n_i x_i^2}{N} - \bar{x}^2$$

$$S^2 = \frac{106}{25} - 1{,}68^2 = 1'4176$$

$$\bar{x} = \frac{1}{N} \sum_{i=1} x_i n_i$$

$$\bar{x} = \frac{1}{25} 42 = 1.68$$

Desviación típica

Es la raíz cuadrada de la varianza

$$S = \sqrt{S^2}$$

- Tiene la misma dimensionalidad (unidades) que la variable.
- La distribución de Gauss, que veremos más adelante, quedará completamente determinada por la media y la desviación típica.

Cuasi-desviación típica

$$\hat{s} = \sqrt{\hat{s}^2}$$

Coeficiente de variación (Pearson)

Es la razón entre la desviación típica y la media (cuando la media es distinta de cero)

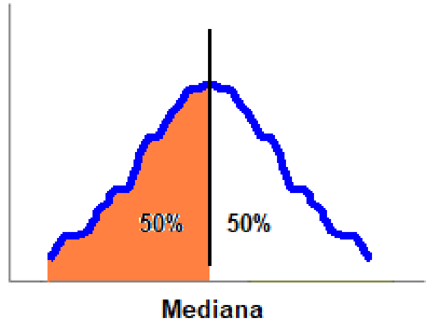
$$CV = \frac{S}{\bar{x}}$$

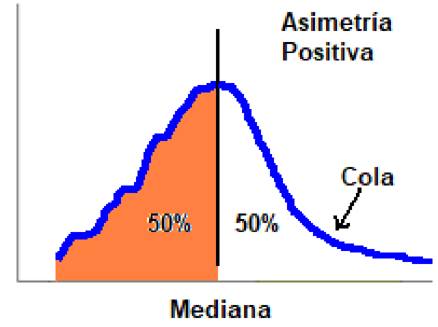
- Se la denomina también variabilidad relativa.
- Es frecuente mostrarla en porcentajes
  - Si la media es 80 y la desviación típica 20 entonces CV=20/80=0,25=25% (variabilidad relativa)

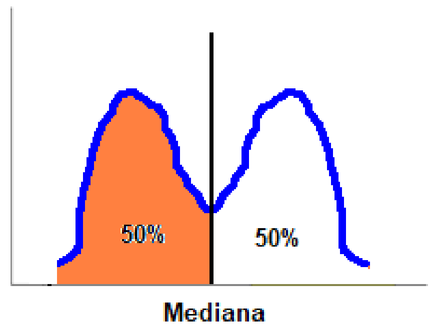
- Es una cantidad **adimensional**. Interesante para comparar la variabilidad de diferentes variables.
  - Si el peso tiene CV=30% y la altura tiene CV=10%, los individuos presentan más dispersión en peso que en altura.

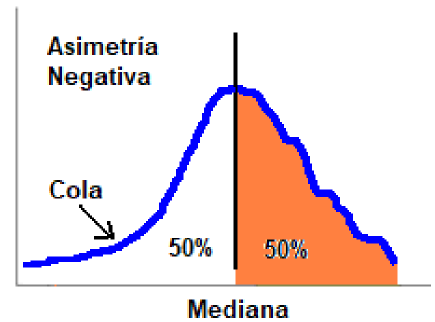
Asimetría o Sesgo

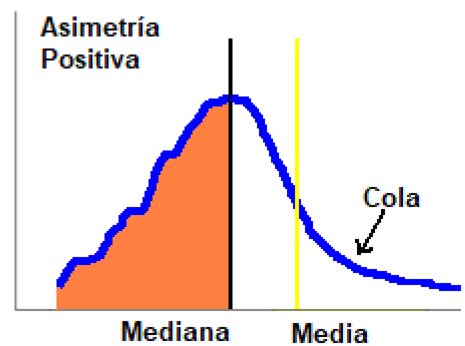
- Una distribución es simétrica si la mitad izquierda de su distribución es la imagen especular de su mitad derecha.
- En las distribuciones simétricas media y mediana coinciden. Si sólo hay una moda también coincide
- La asimetría es positiva o negativa en función de a qué lado se encuentra la cola de la distribución.
- La media tiende a desplazarse hacia las valores extremos (colas).
- Las discrepancias entre las medidas de centralización son indicación de asimetría.











Apuntamiento o curtosis

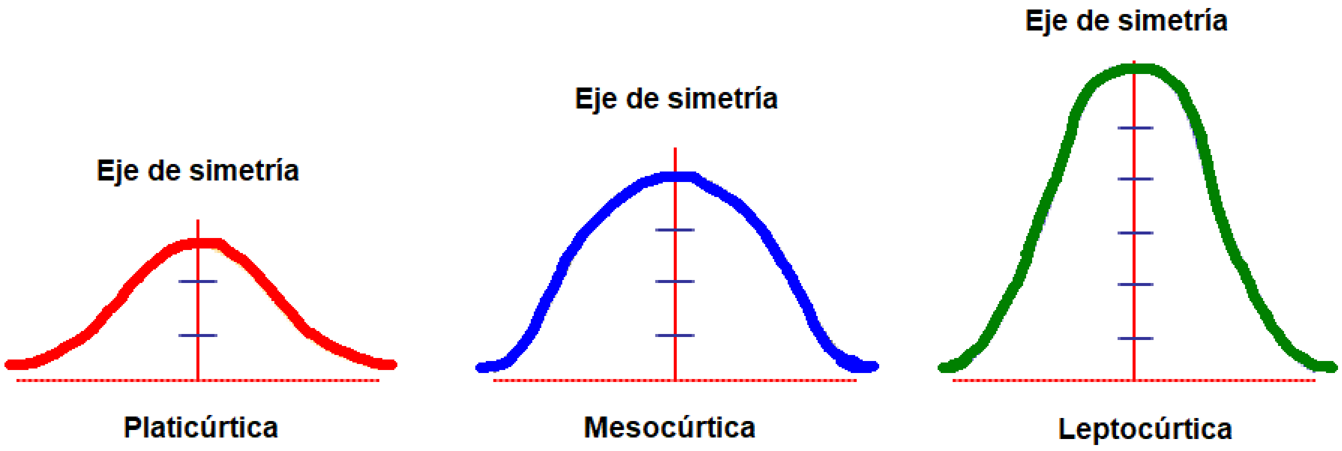
La curtosis nos indica el grado de apuntamiento (aplastamiento) de una distribución con respecto a la distribución normal o gaussiana. Es adimensional.

Platicúrtica (aplanada): curtosis < 0

Mesocúrtica (como la normal): curtosis = 0

Leptocúrtica (apuntada): curtosis > 0

En el curso serán de especial interés las mesocúrticas y simétricas (parecidas a la normal).



EJEMPLO

- Para los datos reflejados en esta tabla calcular:
  - Media
  - Mediana
  - Moda
  - Varianza
  - Desviación típica

	ni
25-35	45
35-45	23
45-55	15
55-65	3
	86

Ejemplo

	ni
25-35	45
35-45	23
45-55	15
55-65	3
	86

xi	xi.ni	Ni
30	1350	45
40	920	68
50	750	83
60	180	86
	3200	

$$\bar{x} = \frac{1}{n}\sum_{i=1} x_i n_i = \frac{3200}{86} = 37{,}21$$

$$M_e = L_i + \left(\frac{\frac{n}{2} - N_{i-1}}{n_i} \right) a_i$$

$$25 + \frac{\frac{1}{2}86 - 0}{45} 10 = 34{,}55$$

$$\text{Mo} = l_i + \frac{n_i - n_{i-1}}{(n_i - n_{i-1}) + (n_i - n_{i+1})} a_i$$

$$\text{Mo} = 25 + \frac{45-0}{(45-0)+(45-23)} 10 = 31{,}716$$

Ejemplo

	x_i	n_i
25-35	30	45
35-45	40	23
45-55	50	15
55-65	60	3
		86

${x_i}^2$	${n_i}^2$
900	40500
1600	36800
2500	37500
3600	10800
8600	125600

- $S^2 = \frac{\sum_{i=1} n_i x_i^2}{N} - \bar{x}^2 = \frac{125600}{86} - 37{,}21^2 = 75{,}88$
- $S = \sqrt{S^2} = \sqrt{75{,}88} = 8{,}71$