

Biometria



3 Estadísticos

Estadísticos

■ **Tendencia central**

- Indican valores con respecto a los que los datos parecen agruparse.

- Media, mediana, moda

■ **Dispersión**

- Indican la mayor o menor concentración de los datos con respecto a las medidas de tendencia central.

- Varianza, Desviación estándar, coeficiente de variación, rango intercuartílico

■ **Posición**

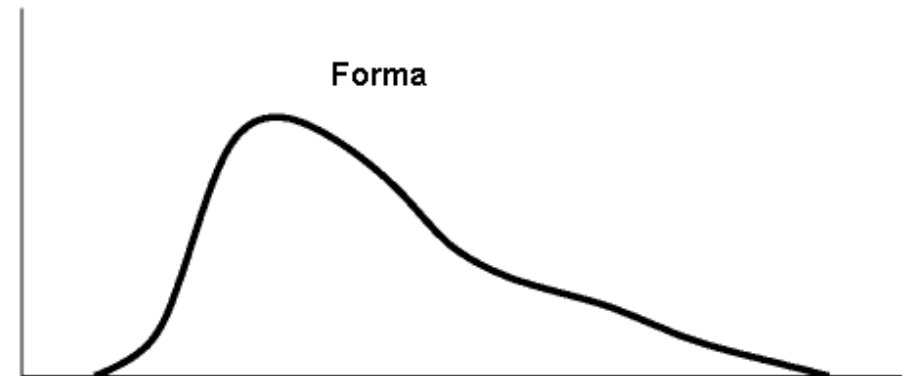
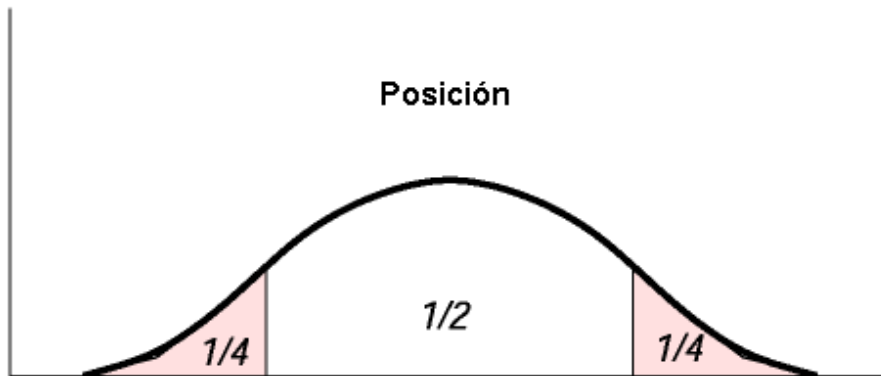
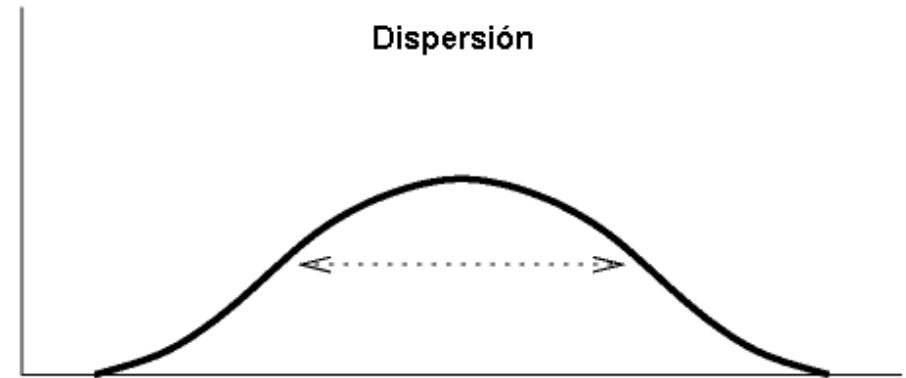
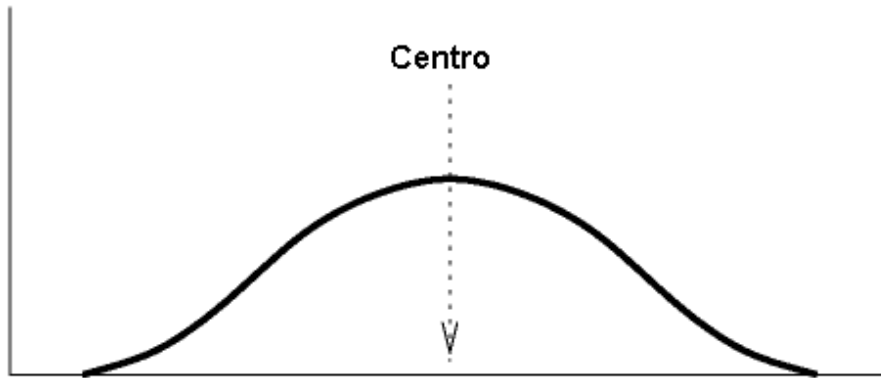
- Dividen un conjunto ordenado de datos en grupos con la misma cantidad de individuos.

- Cuartiles, deciles, percentiles

■ **Forma**

- Asimetría y curtosis

Relacionando gráficos de distribución de frecuencias con estadísticos



Estadísticos de tendencia central

Son medidas que buscan posiciones (valores) con respecto a los cuales los datos muestran tendencia a agruparse.

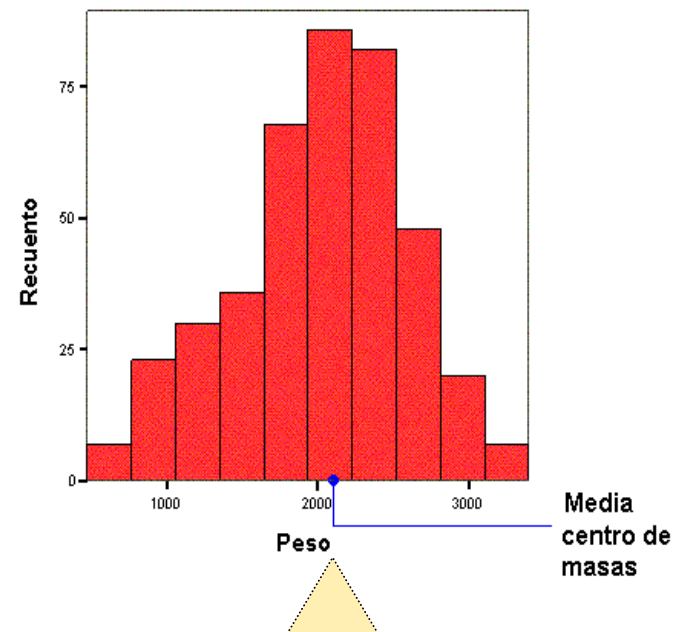
■ Media o promedio aritmético:

Es la suma de los valores de una variable dividido por el total de datos.

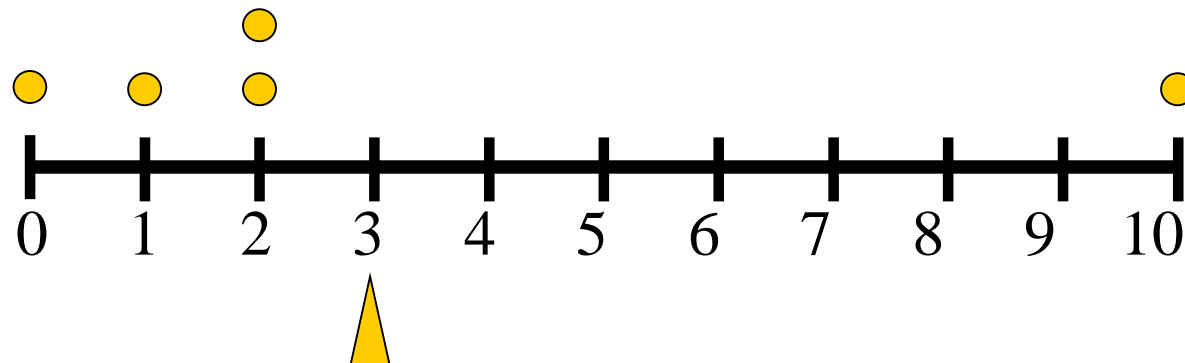
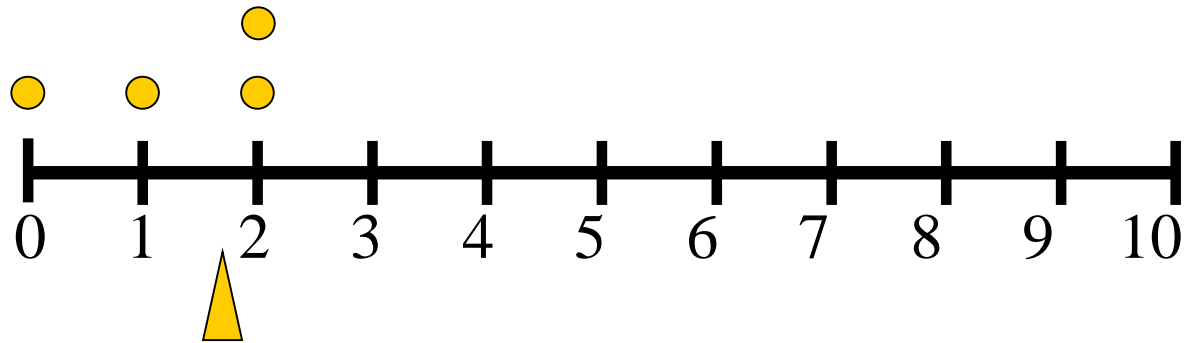
- Media de 2,2,3,7 es

$$\bar{x} = \frac{2+2+3+7}{4} = \frac{14}{4} = 3,5$$

- Centro de gravedad de los datos
- Si se midió a la población, se denomina μ



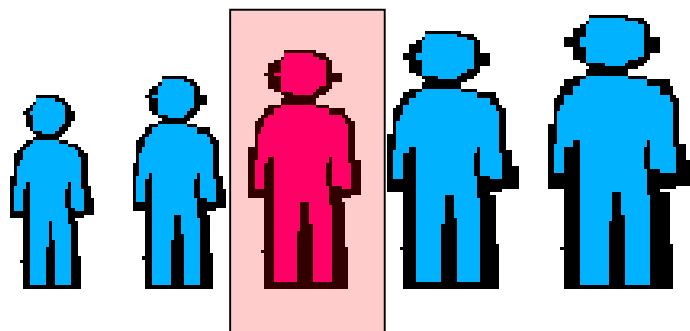
Media



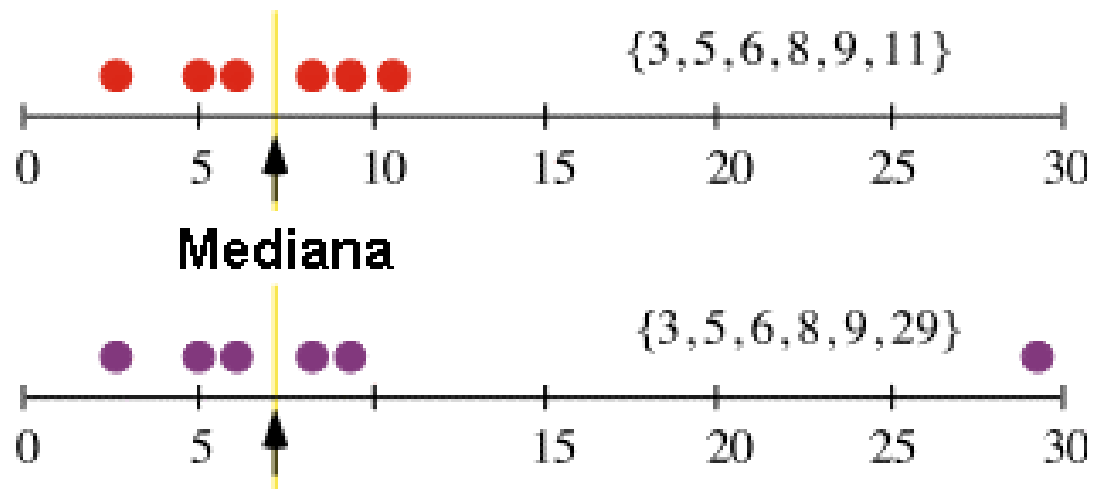
- Conveniente cuando los datos se concentran simétricamente con respecto a ese valor. Muy sensible a valores extremos.

Mediana

- Es un valor que divide a las observaciones **ordenadas** en dos grupos con el mismo número de individuos
 - Mediana de 1,2,4, **5**, 6,6,8 es
 - Mediana de 1,2,4, **5**, 6,6,8,9 es
 - Es conveniente cuando los datos son asimétricos. No es sensible a valores extremos.

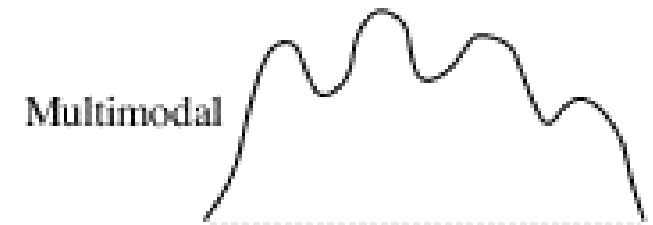
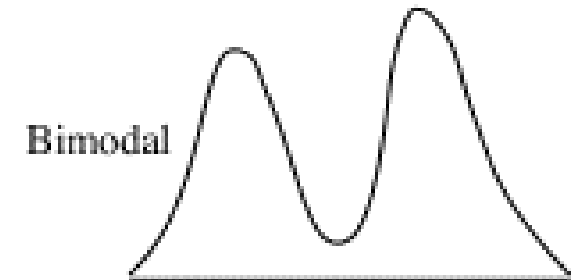
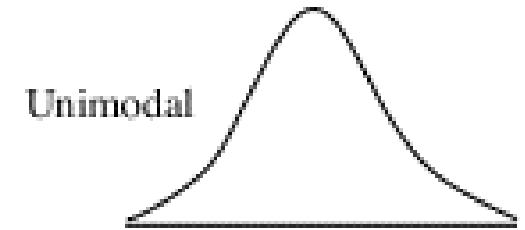


Altura mediana



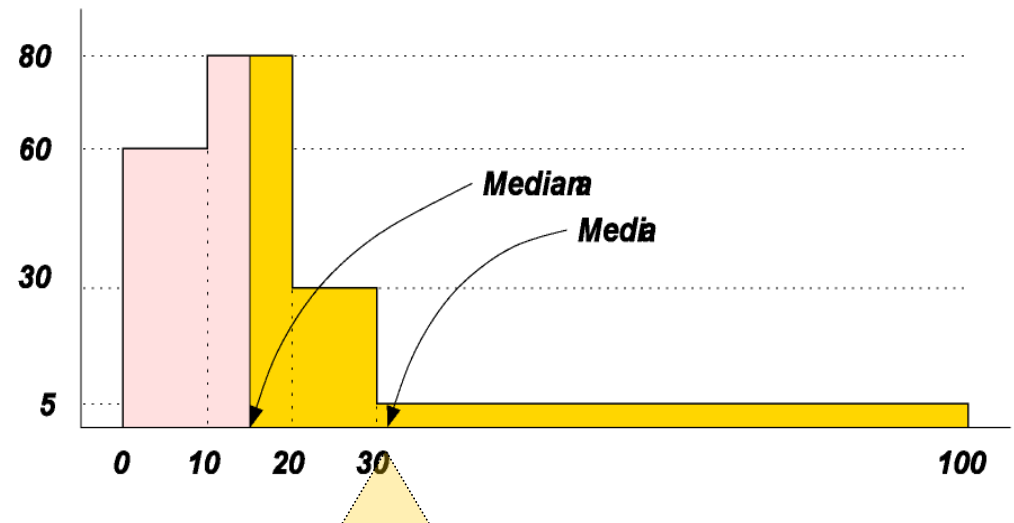
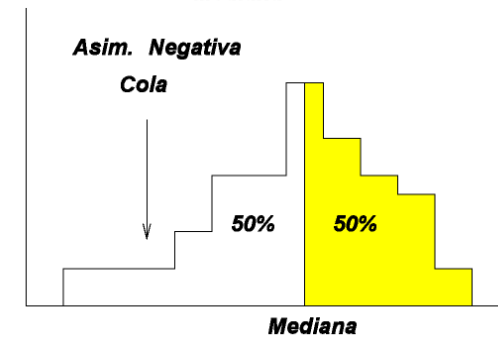
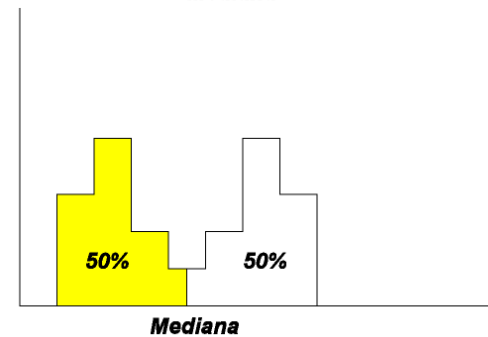
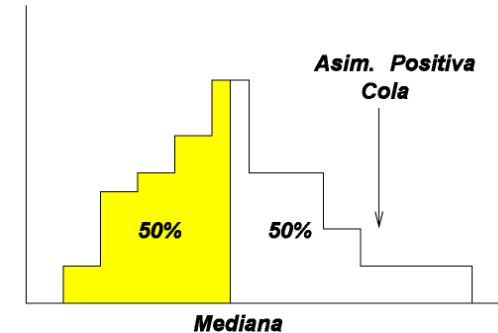
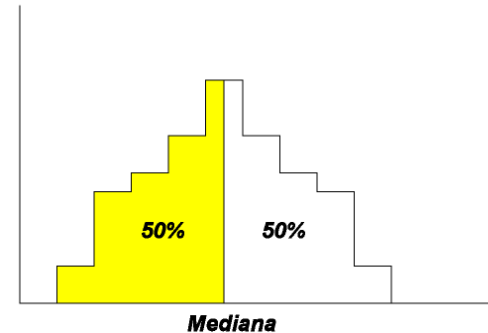
Moda

- Es el valor de la variable que más se repite
 - Moda de 1,2,4,5,6,6,8 es
 - Moda de 1,2,2,5,6,6,8 es
 - Moda de 1,2,4,5,6,7,8,9 es
- Es el único estadístico calculable en variables cualitativas
- En variables continuas, es el/los valor/es donde la distribución de frecuencia alcanza un máximo, relativo o absoluto.



Asimetría

- 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 los valores extremos (colas).
- Las discrepancias entre las medidas de centralización son indicación de asimetría.



¿Cuánto ganan los españoles?

EL PAÍS

PORTADA INTERNACIONAL POLÍTICA

ECONOMÍA

ECONOMÍA EMPRESAS MERCADOS BOLSA FINANZAS PERSONALES VIVIENDA TECNOLOGÍA

▶ ESTÁ PASANDO NCG vende Evo Banco Seat Expatriados Revolución en las 'telecos'

ENTREVISTA DIGITAL

 Charla con José Sámano. El redactor jefe de Deportes de EL PAÍS en un especial sobre la derrota de la candidatura olímpica de Madrid 2020 »

El sueldo más común en España no llega a 16.000 euros

lne.es NOTICIAS **Economía** [HEMEROTECA »](#) [EL TIEMPO »](#)

PORTADA NOTICIAS DEPORTES **ECONOMÍA** OPINIÓN/BLOGS OCIO

[Noticias](#) [Bolsa y Mercados](#) [Empresas en Asturias](#) [Constituciones](#) [Nombramientos](#) [Licitaciones](#) [Subvenciones](#) [Finanzas](#)

 **EN DIRECTO** [Sigue la decimosexta etapa de la Vuelta: Graus - Aramón Formigal](#)

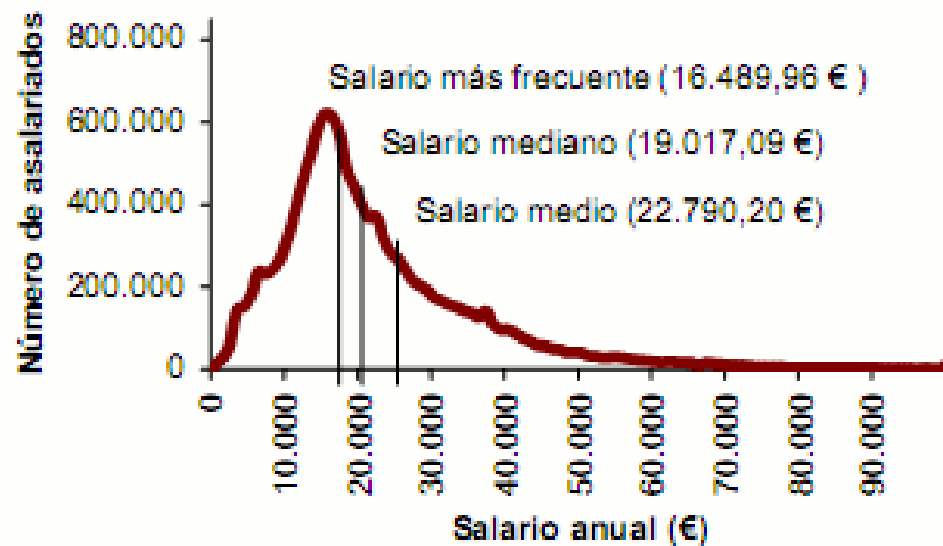
[Lne.es » Economía](#)  

Datos del INE de 2009

El sueldo medio en España es de 22.511 euros, según el INE

¿Cuánto ganan los españoles?

Distribución del salario bruto anual



Estadísticos de dispersión

Miden el grado de dispersión (variabilidad) de los datos, independientemente de su causa.

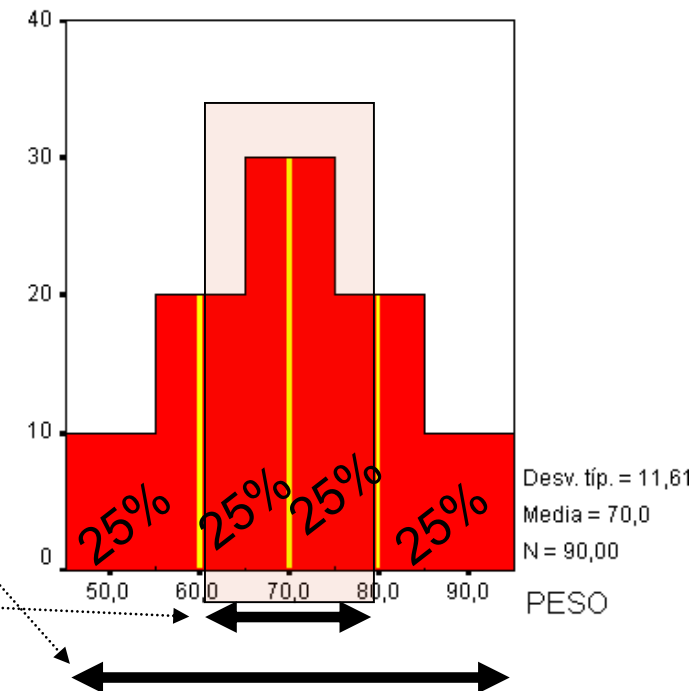
■ Amplitud o Rango:

La diferencia entre las observaciones extremas.

- 2, 1, 4, 3, 8, 4. El rango es $8 - 1 = 7$
- Es muy sensible a los valores extremos.

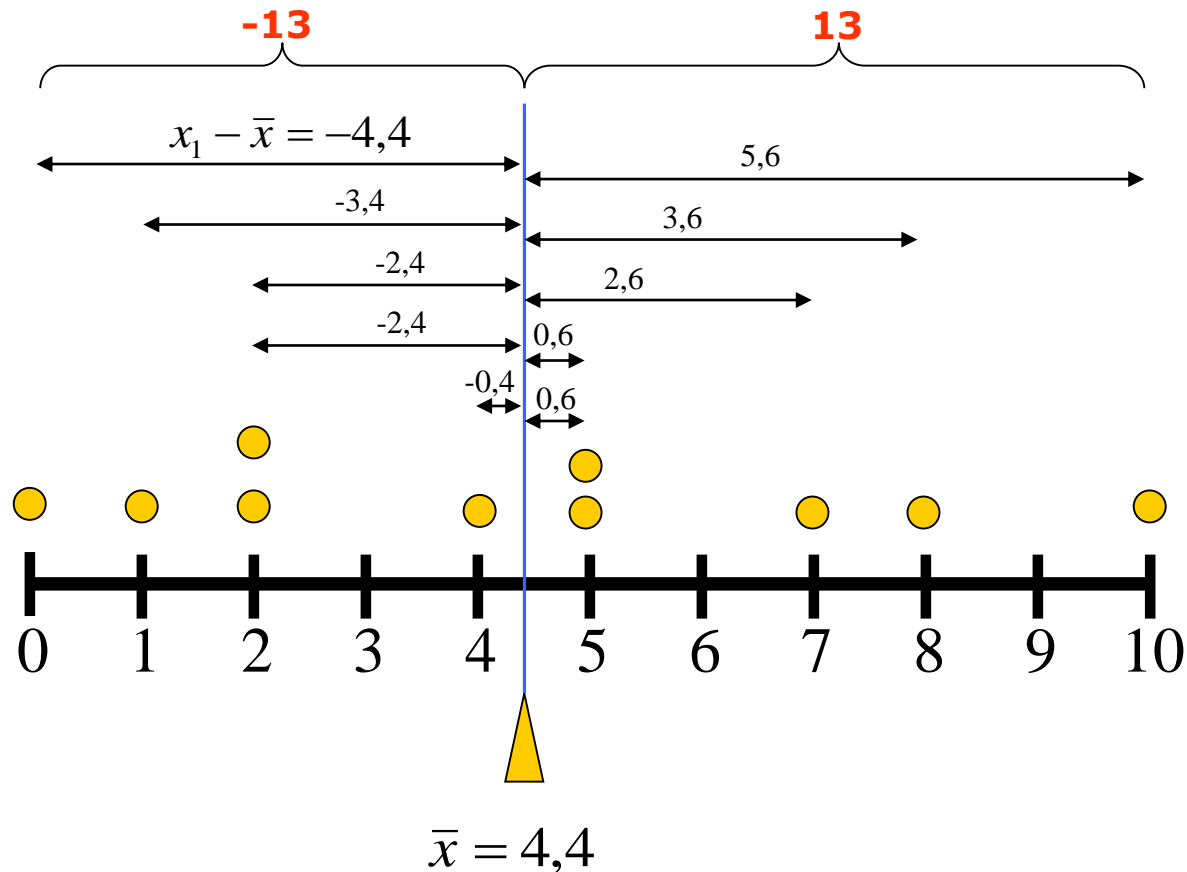
■ Rango intercuartílico (RIQ):

- Es la distancia entre el primer y tercer cuartil.
 - Rango intercuartílico = $P_{75} - P_{25}$
- Parecida al rango, pero eliminando las observaciones más extremas inferiores y superiores.
- No es tan sensible a valores extremos.



Estadísticos de dispersión

$\{0, 1, 2, 2, 4, 5, 5, 7, 8, 10\}$ $n=10$



Varianza

$$\sigma^2 = \frac{\sum_{i=1}^N (x_i - \mu)^2}{N}$$

poblacional

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

muestral

- ✓ Es el promedio de las desviaciones (al cuadrado) de cada dato con respecto a la media
- ✓ A mayor variabilidad, mayor varianza
- ✓ Es sensible a valores extremos
- ✓ Sus unidades son las de la variable pero al cuadrado!

Desvío estándar o típico

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{\sum_{i=1}^N (x_i - \mu)^2}{N}}$$

poblacional

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

muestral

- ✓ Tiene las mismas unidades que la variable
- ✓ A mayor variabilidad, mayor desvío estándar
- ✓ No confundir con error estándar o típico

Coeficiente de variación

$$CV = \frac{\sigma}{\mu} \times 100$$

poblacional

$$CV = \frac{s}{x} \times 100$$

muestral

- ✓ No tiene unidades
- ✓ Es una medida de variabilidad relativa: Mide la desviación típica en forma de “qué tamaño tiene con respecto a la media”
- ✓ Se utiliza para comparar variabilidad

Prevalencia de la diabetes mellitus no dependiente de la insulina en Lejona (Vizcaya)

TABLA 1
Características generales de los distintos grupos estudiados

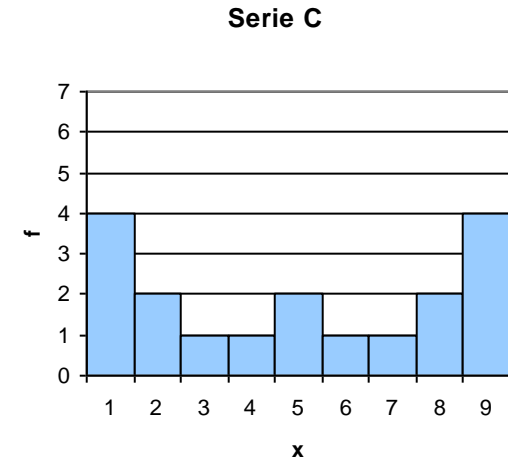
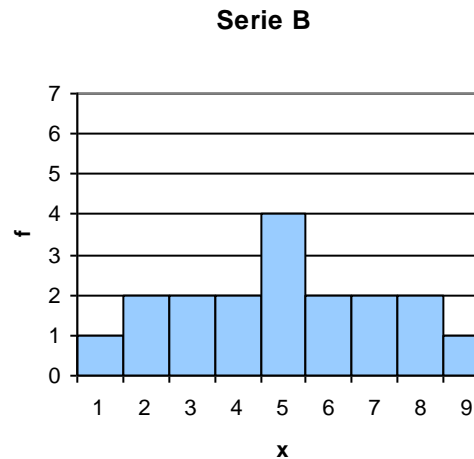
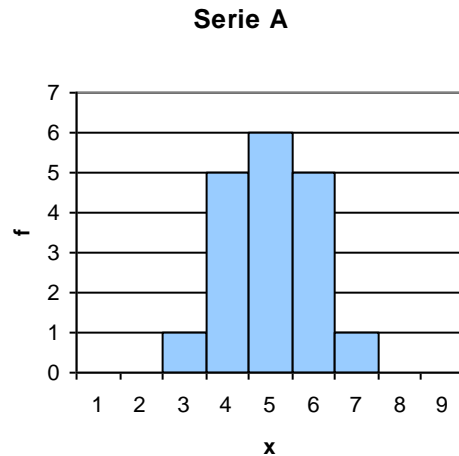
	Normales (grupo N)		Diabéticos (grupo DM)	
	Varones	Mujeres	Varones	Mujeres
Número de individuos	370	347	15	16
Edad (años)	44,7 ± 11,5	45,4 ± 13,8 ^a	61,0 ± 11,8	61,7 ± 13,1 ^a
Talla (cm)	170,6 ± 5,8	158,9 ± 6,2 ^b	164,5 ± 6,3	151,6 ± 7,1 ^b
Peso (kg)	73,4 ± 9	62,1 ± 8,8 ^b	75,8 ± 14,8	67,2 ± 9,9 ^b
IMC (kg/m ²)	25,2 ± 2,6	24,6 ± 3,4 ^b	27,9 ± 4,5	29,1 ± 3,7 ^b
PAS (mmHg)	125,7 ± 16	125,4 ± 21,1 ^a	148,6 ± 24	158,7 ± 29 ^a
PAD (mmHg)	71,3 ± 9,2	71,5 ± 11,8 ^a	82 ± 10,6	87,5 ± 14,4 ^a
Antecedentes familiares (%)	18,1	16,7 ^a	18,8	23,5 ^a

IMC: índice de masa corporal; PAS: presión arterial sistólica; PAD: presión arterial diastólica. Resultados expresados como $\bar{x} \pm DE$.

Actividad

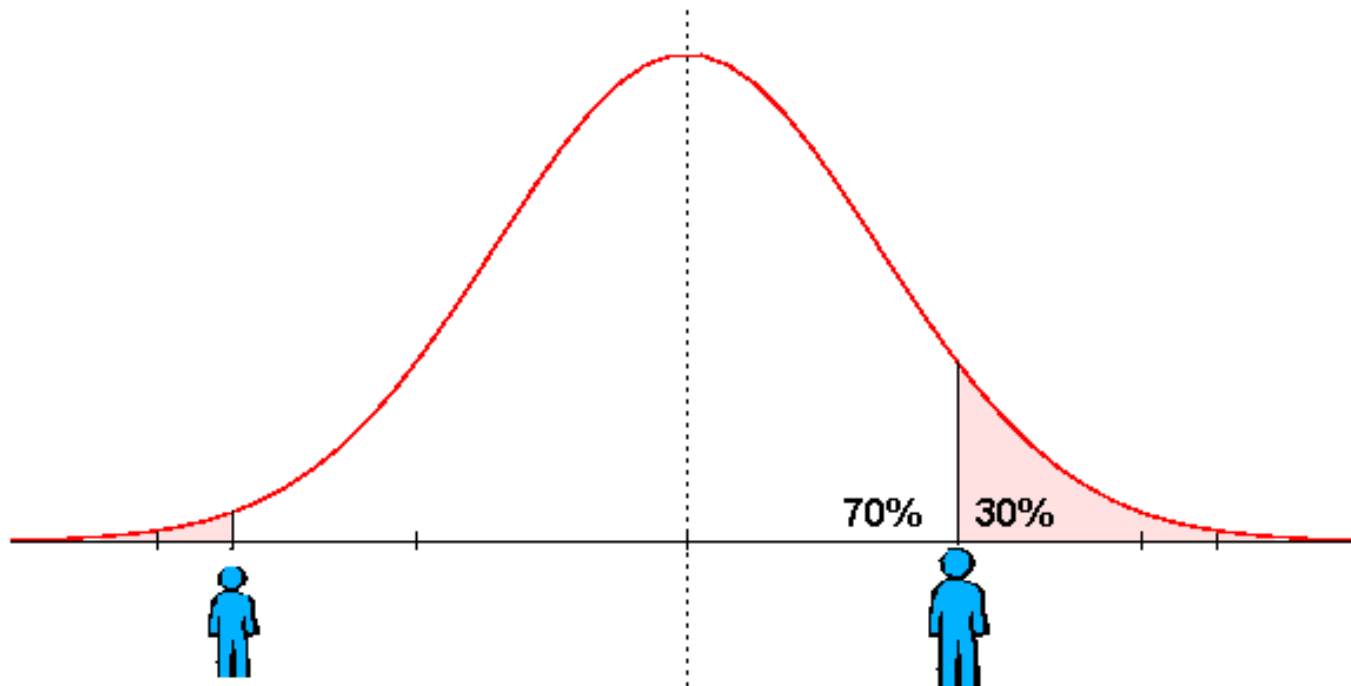
Las siguientes distribuciones están basadas en la misma cantidad de datos

Determinar el promedio y ordenar según el desvío estándar en orden creciente



Estadísticos de posición

- Se define el **fractil** de orden k como un valor de la variable por debajo del cual se encuentra una frecuencia acumulada k .
- Casos particulares son los percentiles, cuartiles, deciles, percentiles,...

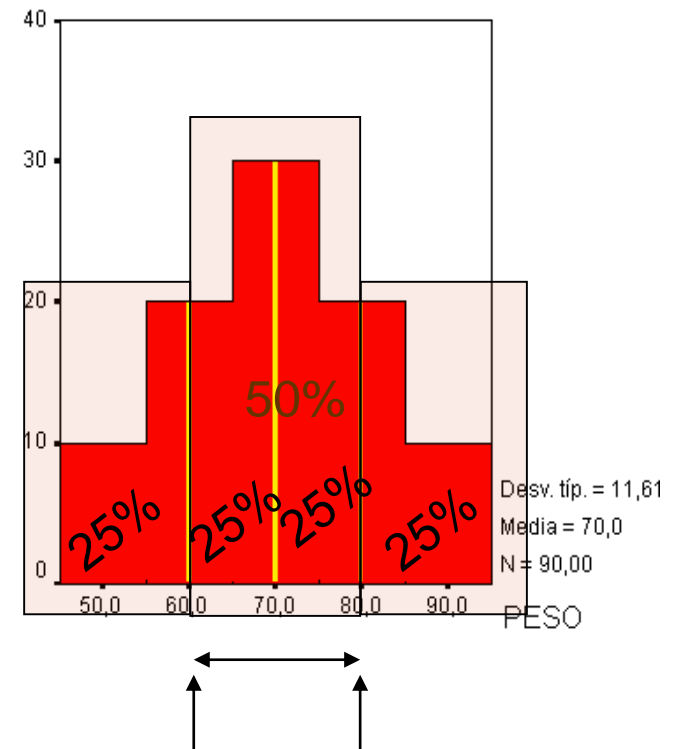


Estadísticos de posición

- **Cuartiles:** Dividen a la muestra en 4 grupos con la misma cantidad de datos.
 - Q1 = Deja por debajo el 25% de los datos
 - Q2 = Deja por debajo el 50% de los datos = mediana
 - Q3 = Deja por debajo el 75% de los datos
- **Deciles:** Dividen a la muestra en 10 grupos con la misma cantidad de datos
 - El D1 deja por debajo al 10% de las observaciones. Por encima queda el 90%
- **Percentiles:** Dividen a la muestra en 100 grupos con la misma cantidad de datos
 - La mediana es el percentil 50
 - El percentil de orden 15 deja por debajo al 15% de las observaciones. Por encima queda el 85%

Ejemplo

- ¿Qué peso no llega a alcanzar el 25% de los individuos?
 - Primer cuartil = percentil 25 = 60 Kg.
- ¿Qué peso es superado por el 25% de los individuos?
 - Tercer cuartil = percentil 75 = 80 kg.
- ¿Entre qué valores se encuentra el 50% de los individuos con un peso “más normal”?
 - Entre el primer y tercer cuartil = entre 60 y 80 kg.



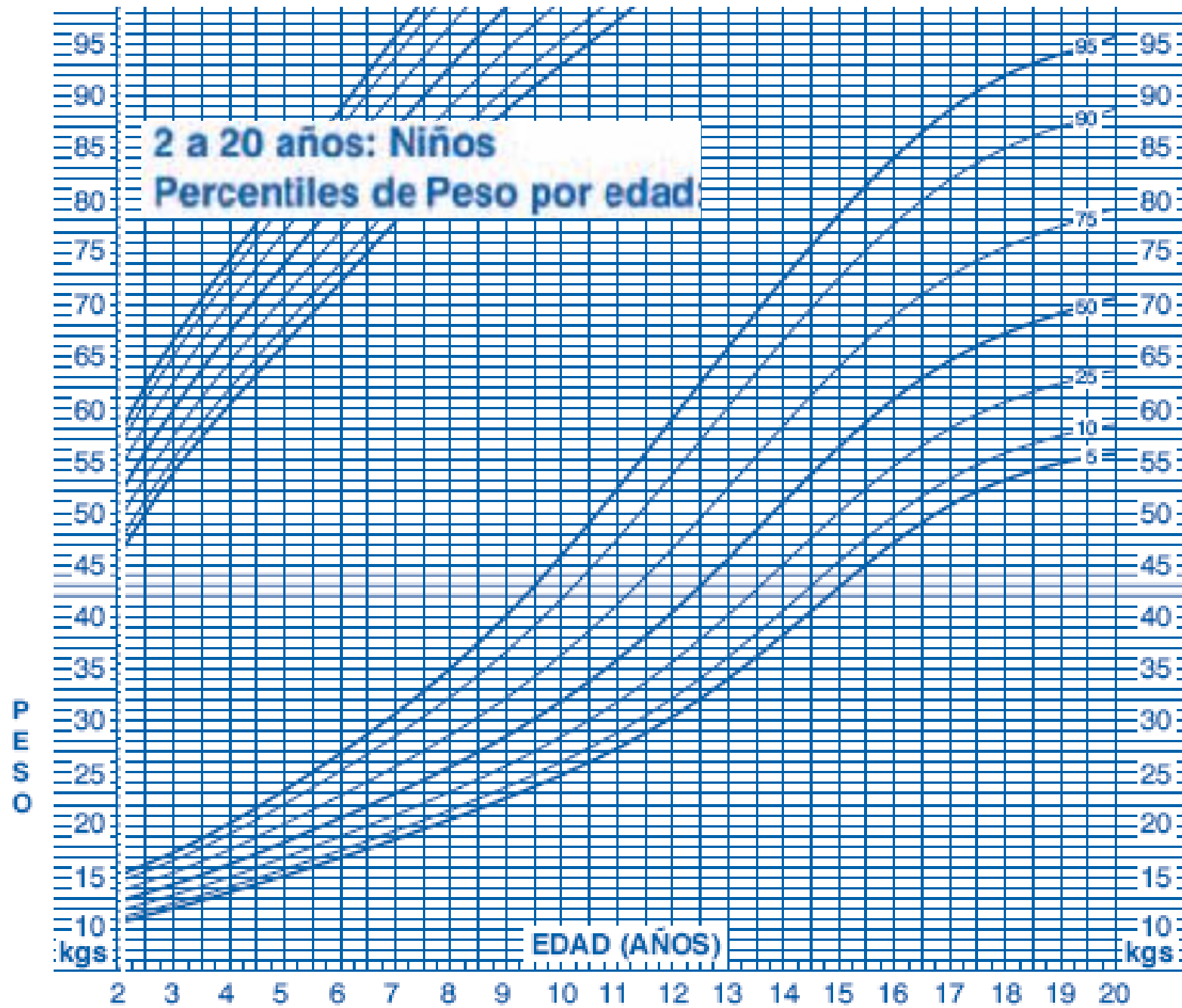
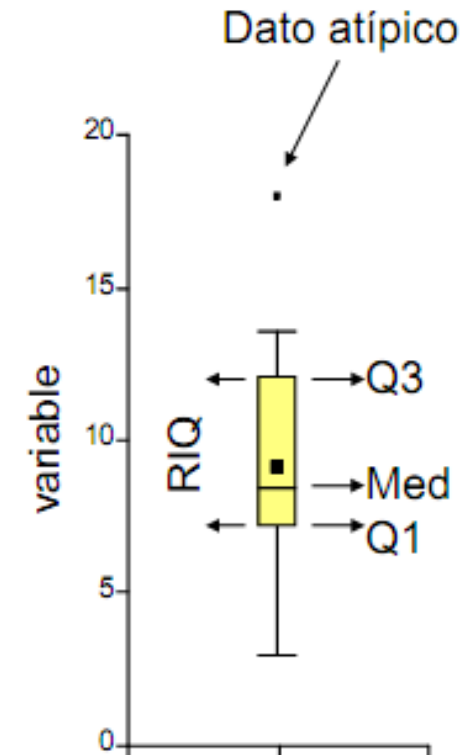


Gráfico de caja (Box Plot)

- ❑ Ordenar los datos de menor a mayor
- ❑ Hallar los 5 números resumen: mín, Q1, mediana, Q3 y máx
- ❑ Los dos cuartiles determinan los bordes de la caja y la mediana, la línea que la atraviesa
- ❑ Calcular el RIQ = $Q3 - Q1$
- ❑ Calcular los límites inferior y superior como $Q1 - 1.5 \text{ RIQ}$ y $Q3 + 1.5 \text{ RIQ}$ respectivamente
- ❑ Las observaciones que caen por fuera se consideran **datos atípicos**
- ❑ Prolongar la línea hasta el valor menor y mayor, respectivamente, dentro de los límites



Datos atípicos o outliers

- ▣ Son datos muy distintos al resto
- ▣ Pueden aparecer por:
 - error en el procedimiento (toma de datos, registro)
 - Como consecuencia de un evento extraordinario
 - Indicativos de un segmento menor de la población o de un fenómeno novedoso
- ▣ Se detectan por métodos gráficos (gráfico de caja) o analíticos (estandarización, pruebas estadísticas)

Volviendo al ejemplo

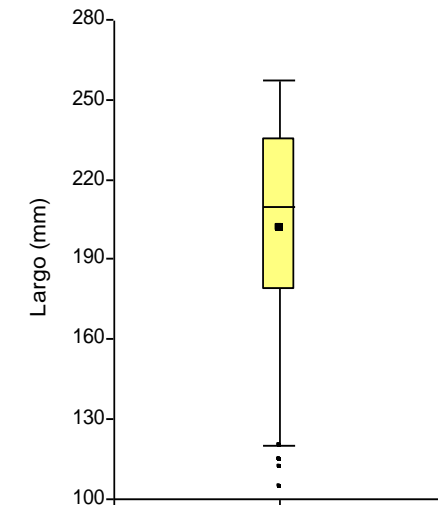
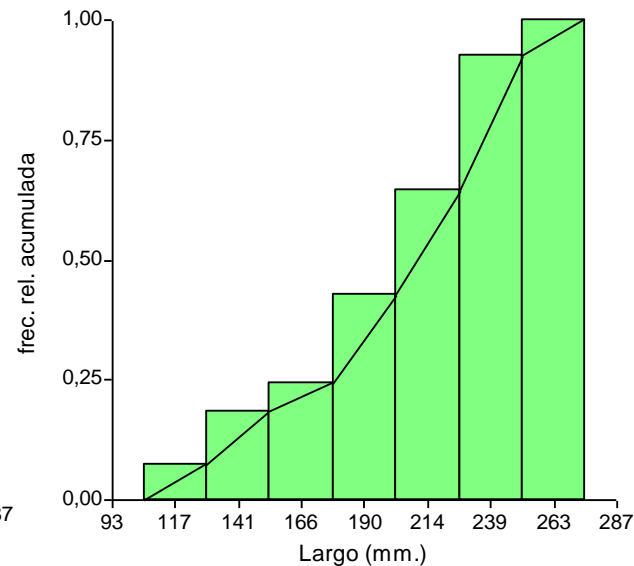
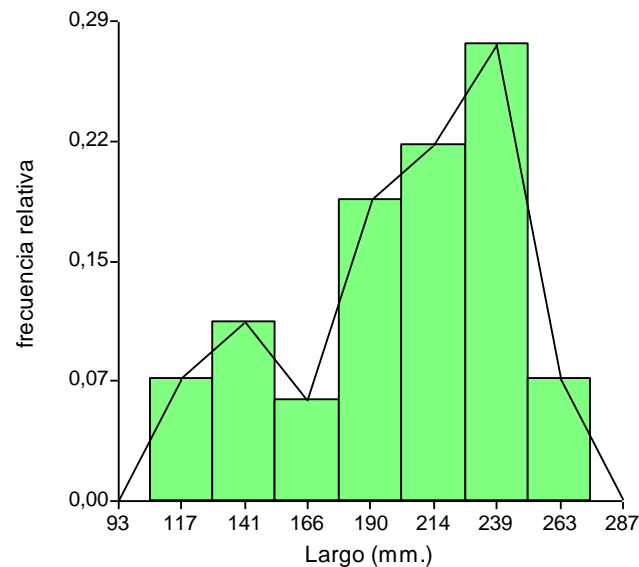


Tablas de frecuencias

Variable	Clase	LI	LS	MC	FA	FR	FAA	FRA
Largo	1	105,00	129,29	117,14	11	0,07	11	0,07
Largo	2	129,29	153,57	141,43	16	0,11	27	0,18
Largo	3	153,57	177,86	165,71	9	0,06	36	0,24
Largo	4	177,86	202,14	190,00	27	0,18	63	0,43
Largo	5	202,14	226,43	214,29	32	0,22	95	0,65
Largo	6	226,43	250,71	238,57	41	0,28	136	0,93
Largo	7	250,71	275,00	262,86	11	0,07	147	1,00
					147	1,00		

Estadística descriptiva

Resumen	Largo
n	147
Media	201,79
D.E.	42,49
Var (n-1)	1805,00
Var (n)	1792,72
CV	21,05
Mín	105,00
Máx	275,00
Mediana	210,00
Q1	178,00
Q3	235,00
Asimetría	-0,63
Kurtosis	-0,60



Estadísticos y gráficos recomendados según el tipo de variable

Tipo de variable	Tipo de gráfico	Medida de tend. central	Medida de dispersión
Cualitativa Nominal	<ul style="list-style-type: none"> ➤ Circular ➤ De barras 	modo	-
Cualitativa ordinal	<ul style="list-style-type: none"> ➤ De barras 	modo	-
Cuantitativa Discreta	<ul style="list-style-type: none"> ➤ De barras ➤ Box Plot 	modo	-
		media	desvío std
		mediana	RIQ
Cuantitativa Continua	<ul style="list-style-type: none"> ➤ Histograma ➤ Box Plot 	modo	-
		media	desvío std
		mediana	RIQ