Tema 1. Estadística Unidimensional

1. Conceptos Básicos
2. Gráficas estadísticas
3. Medidas de centralización y dispersión: media, rango, desviación típica, coeficiente de variación
4. Medidas de posición: mediana, cuartiles, percentiles

1. Conceptos generales

Estadística:

Ciencia cuyo objetivo es reunir una información concerniente a individuos, grupos, series de hechos, etc. y deducir de ello gracias al análisis de estos datos unos significados precisos o unas previsiones para el futuro.

Conceptos básicos:

- Población: Conjunto de todos los elementos cuyo conocimiento nos interesa.
- Muestra: Subconjunto de la población, cuyo estudio permite deducir características de toda la población.
- Individuo: Cada uno de los elementos que forman la población o la muestra.
- Variable estadística: Es el aspecto que deseamos estudiar en la población. Puede ser de varios tipos:
  - Cualitativa: No toma valores numéricos.
  - Cuantitativa: Toma valores numéricos.
    - Discreta: Toma valores aislados.
    - Continua: Toma cualquier valor dentro de un intervalo.

EJEMPLO

El conjunto de todos los países representados en la ONU forman una **población**.

Cada país es un **individuo**.

Se pueden analizar múltiples **caracteres**: Por ejemplo: *sistema de gobierno, número de lenguas oficiales y extensión* (km^2).

Las **variables** correspondientes son, respectivamente, **cualitativa**, **cuantitativa discreta** y **cuantitativa continua**.

La variable **número de habitantes** es, evidentemente, discreta. Pero convendrá tratarla como continua, pues las variaciones unidad a unidad son insignificantes respecto del total.

2. Problemas estadísticos

En cada problema de estadística estudiaremos lo siguiente:

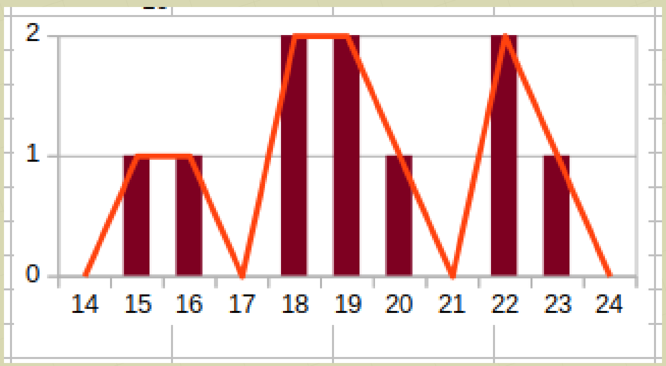
	Cualitativa	Cuantitativa discreta	Cuantitativa continua
Tabla de frecuencias	x	x	x
Gráficas			
Gráfico de barras	x	x	
Histograma			x
Polígono de frecuencias		x	x
Medidas de centralización y dispersión			
Moda	x	x	x
Media		x	x
Desv. Típica		x	x
Coef. Variac.		x	x
Rango		x	x
Medidas de posición			
Mediana		x	x
Cuartiles y percentiles		x	x

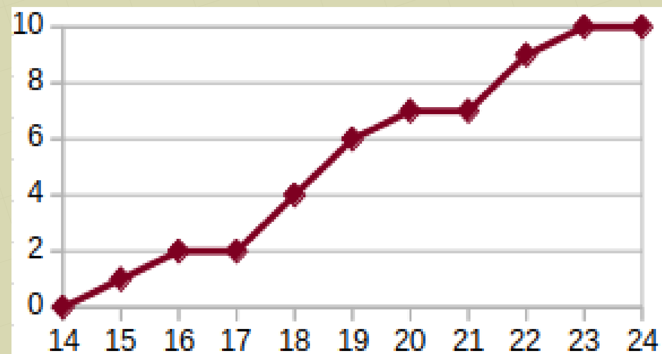
Ejercicio 1: Datos discretos no agrupados.

En una fábrica de tornillos se mide la longitud (en mm) de algunos de ellos y se obtiene:

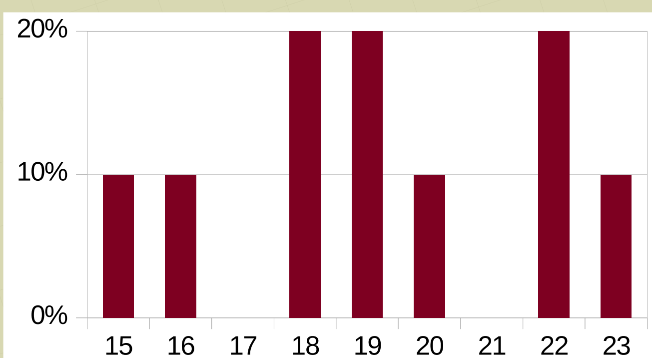
22, 20, 18, 15, 19 22, 16, 19, 23, 18

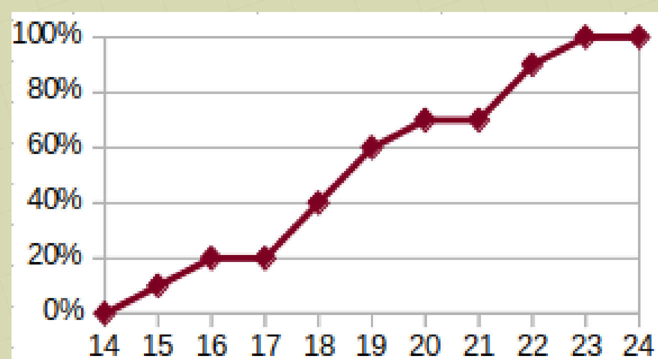
- Tabla de frecuencias no tiene mucho sentido puesto que son muy pocos datos.
- Diagrama de barras, polígono de frecuencias y polígono de frecuencias acumuladas absolutas:





- Diagrama de barras porcentuales y polígono de frecuencias acumuladas porcentuales:





- Media: es una medida representativa de toda la población.

$$\bar{X}=\frac{22+20+18+...+23+18}{10}=19{,}2\ mm$$

$$\boxed{\bar{x}=\frac{\sum x_i}{n}}$$

- Desviación típica: mide lo dispersos que están los datos

$$\text{Varianza: } Var=\frac{22^2+20^2+18^2+...+23^2+18^2}{10}-19{,}2^2=6{,}16\ mm^2$$

$$\text{Desv. Típ.: } s=\sqrt{\frac{22^2+20^2+18^2+...+23^2+18^2}{10}-19{,}2^2}=2{,}5\ mm$$

$$\boxed{s=\sqrt{\frac{\sum x_i^2}{n}-\bar{x}^2}}$$

- Coeficiente de variación: sirve para comparar la dispersión entre dos poblaciones

$$C.V.=\frac{s}{\bar{x}}=0{,}13$$

- Moda: es el valor que más se repite.

  En este caso hay tres, por lo que no tiene mucho significado

  Mo = 18 , 19 y 22 mm.

- Rango: intervalo en que están los datos.

  Rango = [15 , 23] mm.

- Mediana: Es el valor que tiene a la mitad de la población por debajo y a la otra mitad por encima.

  15 16 18 18 19 | 19 20 22 22 23

  Me = 19 mm.

Ejercicio 2: Datos discretos agrupados.

En una fábrica de tornillos se mide la longitud (en mm) de algunos de ellos y se obtiene:

22, 20, 18, 15, 19   22, 16, 19, 23, 18
17, 23, 23, 21, 18   20, 22, 18, 25, 23
22, 22, 19, 19, 20   21, 18, 24, 17, 20

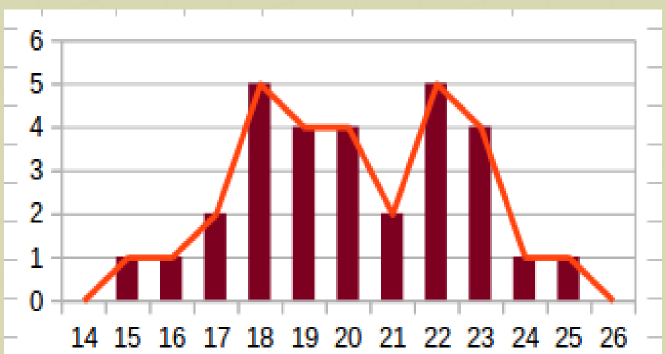
frecuencias acumuladas

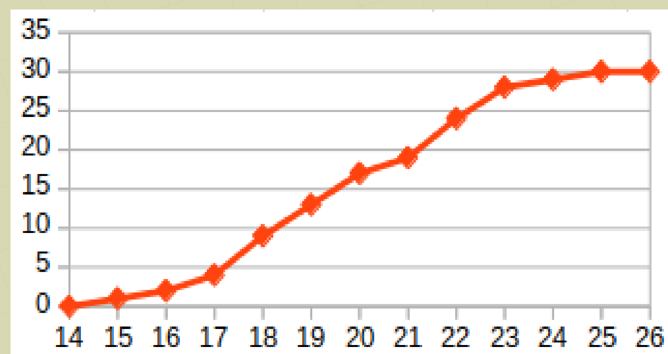
frecuencias

variable

x_i	f_i	F_i
15	1	1
16	1	2
17	2	4
18	5	9
19	4	13
20	4	17
21	2	19
22	5	24
23	4	28
24	1	29
25	1	30
Total	30	

- Tabla de frecuencias. Como ahora son más datos los vamos a agrupar en una tabla.
- Diagrama de barras, polígono de frecuencias y polígono de frecuencias acumuladas absolutas:





- Diagrama de barras y polígono de frecuencias acumuladas porcentuales: se haría de forma similar

- Media:

$$\bar{X}=\frac{15\cdot 1+16\cdot 1+17\cdot 2+18\cdot 5+\ldots+25\cdot 1}{30}=20{,}13\ mm$$

$$\boxed{\bar{x}=\frac{\sum x_i\cdot f_i}{n}}$$

- Desviación típica:

$$\text{Varianza: } Var=\frac{15^2\cdot 1+16^2\cdot 1+17^2\cdot 2+18^2\cdot 5+\ldots+25^2\cdot 1}{30}-20{,}13^2=6\ mm^2$$

$$\text{Desv. Típ.: } s=\sqrt{\frac{15^2\cdot 1+16^2\cdot 1+17^2\cdot 2+18^2\cdot 5+\ldots+25^2\cdot 1}{30}-20{,}13^2}=2{,}5\ mm$$

$$\boxed{s=\sqrt{\frac{\sum x_i^2\cdot f_i}{n}-\bar{x}^2}}$$

x_i	f_i	F_i	$x_i\cdot f_i$	$x_i^2\cdot f_i$
15	1	1	15	225
16	1	2	16	256
17	2	4	34	578
18	5	9	90	1620
19	4	13	76	1444
20	4	17	80	1600
21	2	19	42	882
22	5	24	110	2420
23	4	28	92	2116
24	1	29	24	576
25	1	30	25	625
Total	30		604	12342

- Coeficiente de variación:

$$C.V.=\frac{s}{\bar{x}}=0{,}12$$

- Moda:

  Bimodal: 18 y 22 mm

- Rango:

  Rango = [15 , 25] mm

- Mediana: Hacemos 30/2 = 15. Buscamos 15 o más en la columna F_i. La mediana es su x_i correspondiente. En este caso:

Me = 20 mm

- Cuartiles:
  - Q_1: Deja por debajo al 25% de la población y al 75% por encima

    Hacemos 30/4 = 7,5. Buscamos 7,5 o más en la columna F_i. El primer cuartil es su x_i correspondiente. En este caso:

    $$Q_1 = 18 \text{ mm}$$

  - Q_3: Deja por debajo al 75% de la población y al 25% por encima

    Hacemos (30/4)·3 = 22,5. Buscamos 22,5 o más en la columna F_i. El tercer cuartil es su x_i correspondiente. En este caso:

    $$Q_3 = 22 \text{ mm}$$

- Rango intercuartílico: El 50% central de la población está comprendido entre 18 y 22 mm.
- Percentiles:
  - P_{90}: Deja por debajo al 90% de la población y al 10% por encima

    Hacemos (30/100)·90 = 27. Buscamos 27 o más en la columna F_i. El percentil 90 es su x_i correspondiente. En este caso:

    $$P_{90} = 23 \text{ mm}$$

Ejercicio 3: Datos continuos. Esta vez hay muchos datos por lo que los vamos a agrupar en intervalos

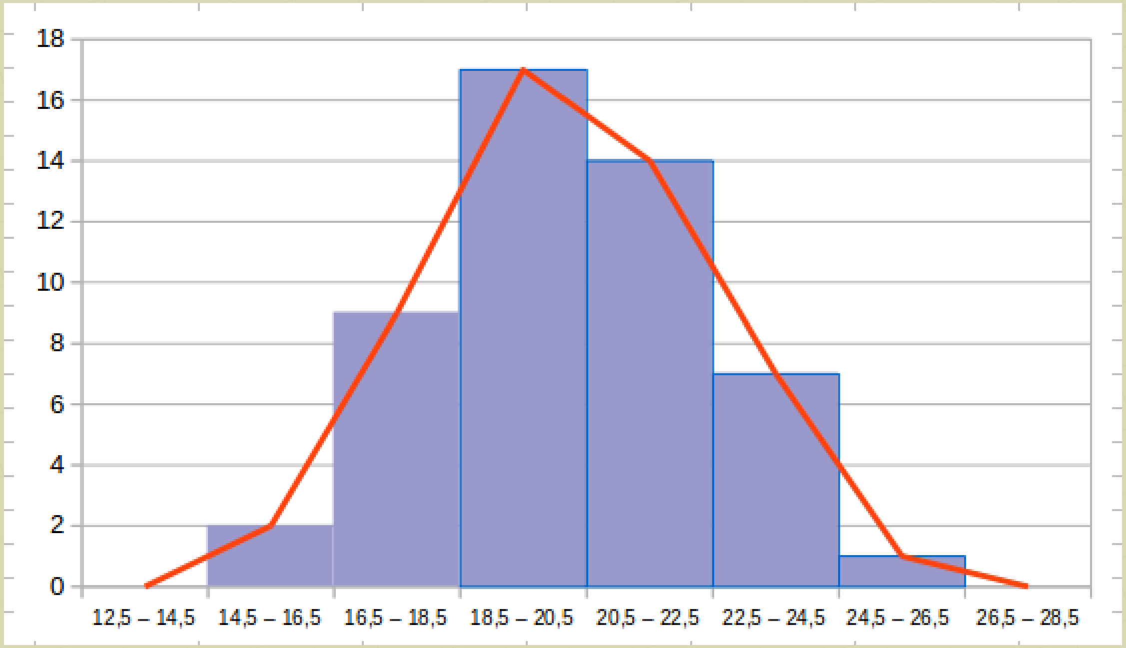
En una fábrica de tornillos se mide la longitud (en mm) de algunos de ellos y se obtiene:

22,	20,	18,	15,	19	22,	16,	19,	23,	18
17,	23,	23,	21,	18	20,	22,	18,	25,	23
22,	22,	19,	19,	20	21,	18,	24,	17,	20
19,	23,	21,	23,	21	20,	19,	21,	20,	22
19,	20,	18,	21,	19	18,	20,	22,	21,	19

- Tabla de frecuencias. Ponemos los intervalos y calculamos la **marca de clase**: valor medio de cada intervalo

INTERVALOS	x_i	f_i	$f_i \cdot x_i$	$f_i \cdot x_i^2$	F_i	EN %
14,5 - 16,5	15,5	2	31	480,5	2	4
16,5 - 18,5	17,5	9	157,5	2756,25	11	22
18,5 - 20,5	19,5	17	331,5	6464,25	28	56
20,5 - 22,5	21,5	14	301	6471,5	42	84
22,5 - 24,5	23,5	7	164,5	3865,75	49	98
24,5 - 26,5	25,5	1	25,5	650,25	50	100
		50	1011	20688,5		

- Histograma y polígono de frecuencias absolutas:



- Media, desviación típica y coeficiente de variación: se hacen igual que en el ejercicio 2, usando las marcas de clase. Se obtiene:

$$\bar{x} = \frac{1\,011}{50} = 20{,}22 \text{ mm}$$

$$\sigma = \sqrt{\frac{20\,688{,}5}{50} - 20{,}22^2} = 2{,}22 \text{ mm}$$

- Moda: Intervalo modal: 18,5 – 20,5 mm
- Rango: [14,5 ; 26,5] mm

- Mediana: Hacemos 50/2 = 25. Buscamos 25 o más en la columna F_i.

  Así encontramos el intervalo correspondiente: [18,5 – 20,5]. Ahora hay que hacer una regla de tres:

INTERVALOS	f_i	F_i
14,5 - 16,5	2	2
16,5 - 18,5	9	11
18,5 - 20,5	17	28
20,5 - 22,5	14	42
22,5 - 24,5	7	49
24,5 - 26,5	1	50

Intervalo	Anchura	Individuos
18,5 – 20,5	2	17
	x	14

Hasta 18,5 hay 11.
Hasta 25 nos faltan 14 *(25 – 11)*

$$x=\frac{2\cdot 14}{17}=1{,}6 \quad ; \quad Me=18{,}5+1{,}6=20{,}1 \text{ mm}$$

El 50% de los tornillos mide menos de 20,1mm, el otro 50% mide más.

- Cuartiles y Percentiles: Igual que la mediana, pero con los porcentajes correspondientes.
  - Q3: 75% de 50 = 37,5 → Intervalo: (20,5 – 22,5)

Regla de tres:

Intervalo	Anchura	Individuos
20,5 – 22,5	2	14
	x	9,5

$$x=\frac{2\cdot 9{,}5}{14}=1{,}4$$

$$Q_3=20{,}5+1{,}4=21{,}9 \text{ mm}$$

El 75% de los tornillos mide menos de 21,9 mm, el otro 25% mide más.

Individuos en el intervalo $(\bar{x} - s\ ,\ \bar{x} + s)$. Si una población es “normal”, en este intervalo debe estar alrededor del 65 – 70% de la población.

Lo calculamos para el ejercicio 3.

INTERVALOS	f_i	F_i
14,5 - 16,5	2	2
16,5 - 18,5	9	11
18,5 - 20,5	17	28
20,5 - 22,5	14	42
22,5 - 24,5	7	49
24,5 - 26,5	1	50

Intervalo: *(20,22 – 2,22 ; 20,22 + 2,22) = (18 – 22,44).*

Hay que hacer igual que para la mediana. Necesitamos un poco del 2º intervalo, el 3º completo y casi todo el 4º.

Hacemos reglas de tres.

Intervalo 2º:

Intervalo	Anchura	Individuos
16,5 – 18,5	2	9
	0,5	x

$$x = \frac{9 \cdot 0{,}5}{2} = 2{,}25$$

Intervalo 3º: completo: 17

Intervalo 4º:

Intervalo	Anchura	Individuos
20,5 – 22,5	2	14
	1,96	x

$$x = \frac{14 \cdot 1{,}96}{2} = 13{,}7$$

Por tanto, en el intervalo $(\bar{x} - s\ ,\ \bar{x} + s)$ = *(18 – 22,44)* hay

2,25 + 17 + 13,7 = 32,95 tornillos.

Esto supone un 66%. La población es “normal”