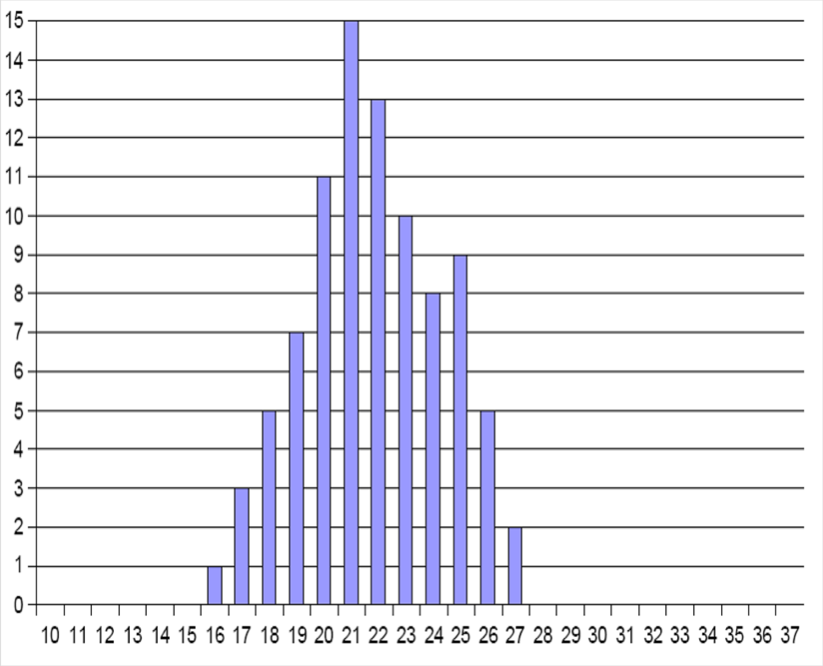
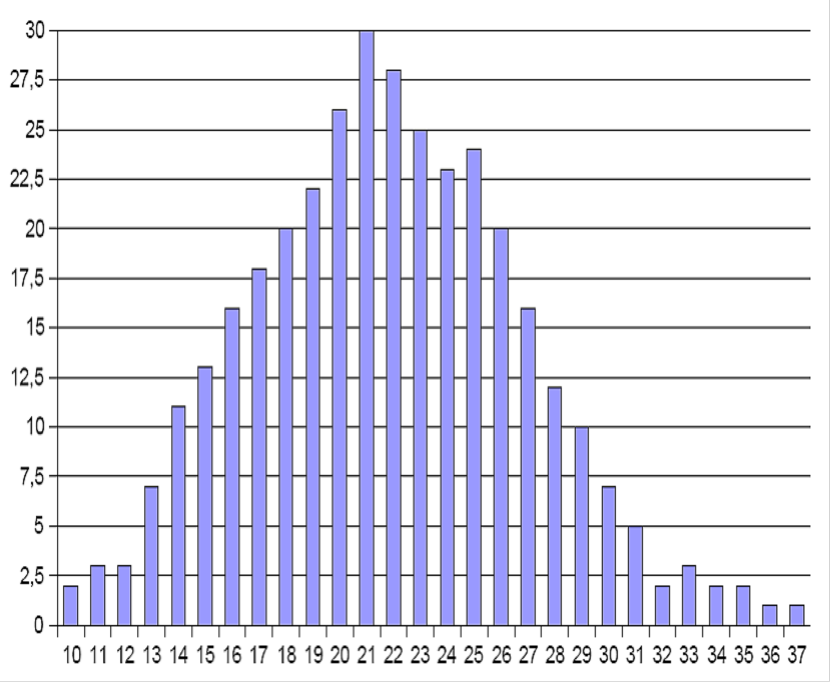
Mesures de dispersió

– Les mesures de tendència central no són suficients per a descriure una distribució de valors

– Exemple:



Moda= 21
Mediana= 22
Mitjana= 21,82



Moda= 21
Mediana= 22
Mitjana= 21,82

Mesures de dispersió

– Els **estadístics de dispersió** més habituals són:

- el ratio de variació
- el rang
- el rang interquartílic
- la variància (desviació típica)

Mesures de dispersió

• ***Ratio* de variació**: medeix la concentració dels valors respecte a la moda.
(Única mesura de dispersió possible per a variables nominals).

$$RV = 1 - \frac{n_{moda}}{N}$$

(concentració) 0 ≤ *RV* ≤ 1 (dispersió)

Exemple:

Immigració per nacionalitats

Nacionalitat	Freq absoluta
Colòmbia	350
Ecuador	250
Perú	120
Argentina	100
Rumania	80
Marruecos	70
Senegal	30
	1000 =N

moda: Colòmbia

$$RV = 1 - \frac{350}{1000} = 0,65$$

Mesures de dispersió

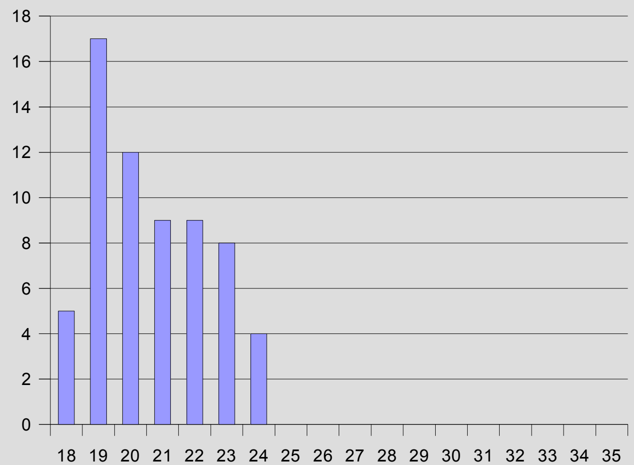
- **Rang**: diferència entre els valors màxim i mínim

$$Rang = max - min$$

- **Rang interquartílic (RIQ)**: diferència entre el tercer i el primer quartils. Menos sensible a valors extrems que el rang.

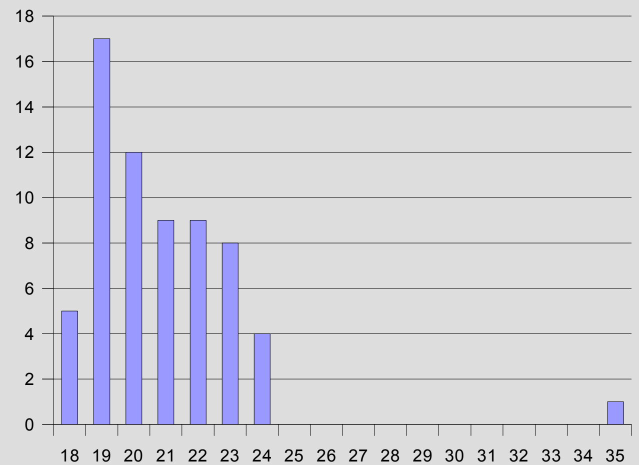
$$RIQ = Q_3 - Q_1$$

Exemple:



Rang=6

RIQ=3



Rang=17

RIQ=3

Mesures de dispersió

- **Desviació típica**: mesura la diferència entre les dades i el seu valor mitjà

$$s = \sqrt{Var}$$

Var és la **Variància** de les dades

Es compleix que:

- al menys un 75% dels valors estan entre $\bar{x} - 2 \cdot s$ i $\bar{x} + 2 \cdot s$

- al menys un 89% dels valors estan entre $\bar{x} - 3 \cdot s$ i $\bar{x} + 3 \cdot s$

- al menys un 93% dels valors estan entre $\bar{x} - 4 \cdot s$ i $\bar{x} + 4 \cdot s$

Mesures de dispersió

- Càlcul de la variància:

Dades brutes:

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

$$Var_X=\frac{x_1^2+x_2^2+\cdots+x_n^2}{n}-\bar{x}^2$$

(**variància poblacional** ($VarP_X$))

Taula de freqüències:

$$\bar{x}=\frac{x_1\cdot n_1+x_2\cdot n_2+\cdots+x_k\cdot n_k}{n}$$

$$Var_X=\frac{x_1^2\cdot n_1+x_2^2\cdot n_2+\cdots+x_k^2\cdot n_k}{n}-\bar{x}^2$$

(**variància poblacional** ($VarP_X$))

Observacions:

- Per a dades agrupades en **intervals** els valors de xi es es substitueixen per les **marques de classe**
- Per a dades procedents d'una **mostra** es calcula la **variància mostral**:

$$VarM_X=s_X^2=\frac{n}{n-1}\cdot VarP_X$$

Mesures de dispersió

– Exemple desviació típica i variància: nota d'estadística de 8 persones

Dades brutes

7
5
9
7
5
6
7
6
4

9 valors

$$\bar{x} = \frac{7+5+\cdots+4}{9} = 6{,}22$$

$$Var\,x = \frac{7^2+5^2+\cdots+4^2}{9} - 6{,}22^2 = 1{,}98$$

$$s = \sqrt{1{,}98} = 1{,}4$$

Taula de freqüències

x_i	n_i
4	*1*
5	*2*
6	*2*
7	*3*
9	*1*

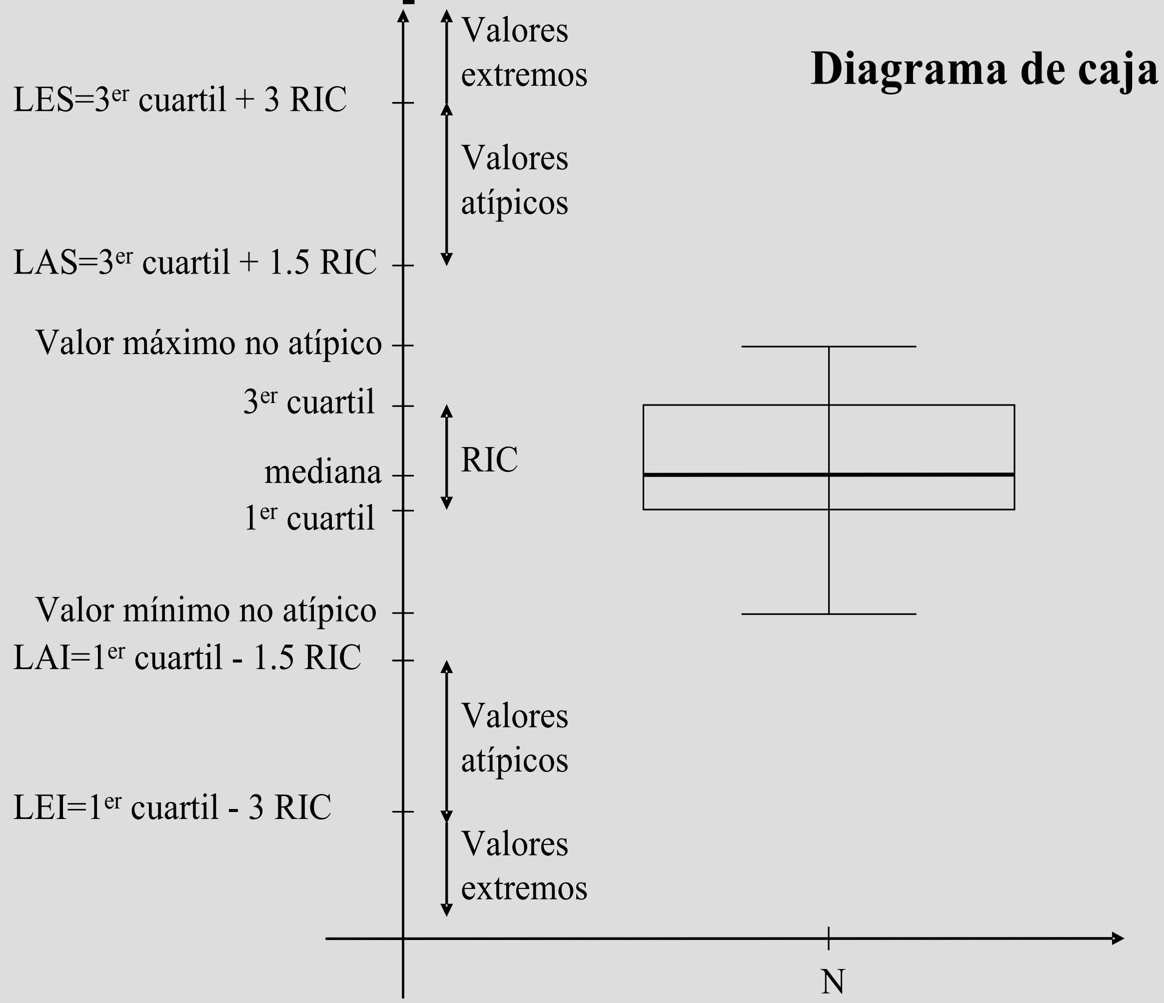
$n=9$

$$\bar{x} = \frac{4\cdot 1+5\cdot 2+\cdots+9\cdot 1}{9} = 6{,}22$$

$$Var\,x = \frac{4^2\cdot 1+5^2\cdot 2+\cdots+9^2\cdot 1}{9} - 6{,}22^2 = 1{,}98$$

Mesures de dispersió

– Representació gràfica de la dispersió: **Diagrames de capsa**



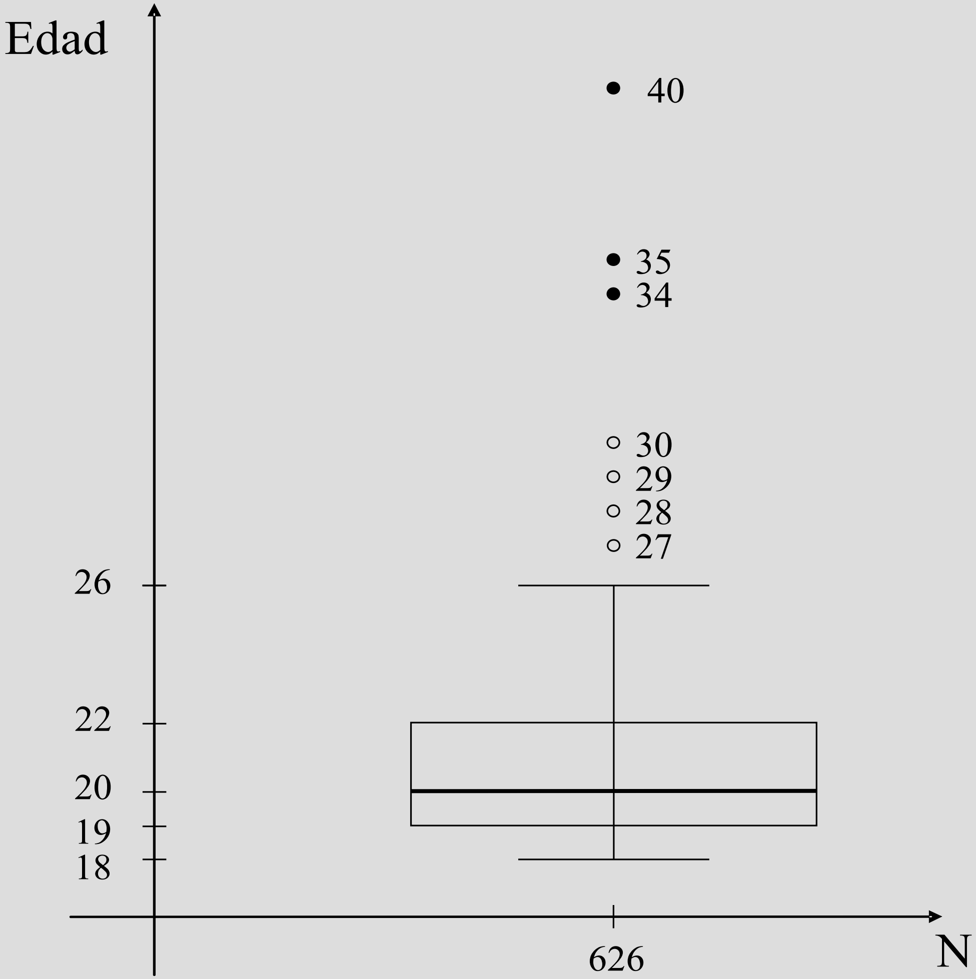
Mesures de dispersió

– Diagrames de capsa (Exemple)

Edat	Freq. Absoluta
18	120
19	150
20	90
21	70
22	65
23	50
24	30
25	20
26	10
27	7
28	8
29	2
30	1
34	1
35	1
40	1
	626

1er quartil → 19
mediana → 20
3er quartil → 22

RIC=22-19=3
1er quartil - 1,5 RIC=19-4,5=14,5
3er quartil + 1,5 RIC=22+4,5=26,5



Mesures de dispersió

– Tipificació de dades estadístiques: z-scores

- La tipificació és un càlcul que permet comparar dades procedents de diferents estudis estadístics

- Per a cada valor x d'una variable estadística (quantitativa) es calcula l'anomenat z-score

$$z = \frac{x - \bar{x}}{s}$$

$\bar{x}$: mitjana
s: desviació típica

- Dues variables diferents es comparen comparant els seus respectius z-scores

Mesures de dispersió

– Tipificació de dades estadístiques: z-scores

Exemple:

Dues persones opten a una beca esportiva. La primera té una marca de 7.60m en salt de longitud i la segona una de 65.4s en 100 metres lliures de natació. Quina persona mereix més la beca?

Per saber quin dels candidats destaca més en el seu respectiu esport hem d'analitzar les marques de la resta d'esportistes. Si tenim que:

Salt de longitud: mitjana 7.40m, desviació típica 0.4m

100m lliures natació: mitjana 68.30s, desviació típica 1.6s

Llavors:

$$z_1 = \frac{7.6-7.4}{0.4} = 0.5 \qquad z_2 = \frac{65.4-68.3}{1.6} = -1 \longrightarrow 1$$

Donariem la beca al nadador perquè destaca més en la seva especialitat.