

Aplicacions Estadístiques

Enginyeria Edificació 2009/10.

Antonio E. Teruel

Temari

- ▶ **Estadística Descriptiva**

 - Tema 1. **Anàlisi exploratori de dades.**

 - Tema 2. Distribucions estadístiques bidimensionals.

- ▶ Probabilitat.

 - Tema 3. Teoria de la probabilitat.

- ▶ Estadística Inferencial.

 - Tema 4. Variables aleatòries discretes.

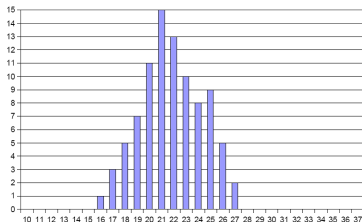
 - Tema 5. Variables aleatòries contínues.

 - Tema 6. Estimació de paràmetres.

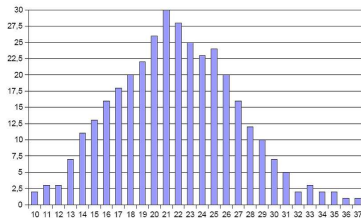
 - Tema 7. Contrast d'hipotesis.

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).

$$RV = 1 - f_{moda} = 1 - \frac{n_{moda}}{n}$$

- ▶ $0 \leq RV \leq 1$
- ▶ Exemple:

x_i	n_i	f_i
Colòmbia	350	0.35
Equador	250	0.25
Perú	120	0.12
Argentina	100	0.1
Romania	80	0.08
Marroc	70	0.07
Senegal	30	0.03

$$n = 1000$$

Mesures de dispersió

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

$$RV = 1 - f_{moda} = 1 - \frac{n_{moda}}{n}$$

- ▶ $0 \leq RV \leq 1$
- ▶ Exemple:

x_i	n_i	f_i
Colòmbia	350	0.35
Equador	250	0.25
Perú	120	0.12
Argentina	100	0.1
Romania	80	0.08
Marroc	70	0.07
Senegal	30	0.03

$$n = 1000$$

$$RV = 1 - 0.35 = 1 - \frac{350}{1000} = 0.65$$

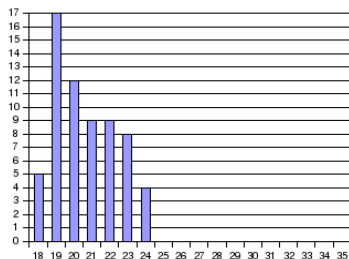
Mesures de dispersió

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

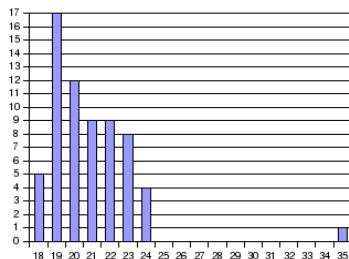
$$Rang = \max x_k - \min x_k$$

- **Rang interquartíl:** diferència entre el tercer i primer quartíl

$$RIQ = Q_3 - Q_1$$



◀────────▶ Rang=6
◀──▶ RIQ=3



◀────────────────────────▶ Rang=17
◀──▶ RIQ=3

Mesures de dispersió

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

$$\sigma_X = \sqrt{Var_X}$$

- ▶ Var_X és la variància de les dades
- ▶ Es compleix que:
 - ▶ Menys del 75% dels valors estan continguts en $(\bar{x} - 2\sigma_X, \bar{x} + 2\sigma_X)$.
 - ▶ Menys del 89% dels valors estan continguts en $(\bar{x} - 3\sigma_X, \bar{x} + 3\sigma_X)$.
 - ▶ Menys del 93% dels valors estan continguts en $(\bar{x} - 3\sigma_X, \bar{x} + 3\sigma_X)$.

Mesures de dispersió

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

$$\sigma_X = \sqrt{Var_X}$$

- ▶ Var_X és la variància de les dades
- ▶ Càlcul de la variància amb dades brutes

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

Mesures de dispersió

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

$$\sigma_X = \sqrt{\text{Var}_X}$$

- ▶ Var_X és la variància de les dades
- ▶ Càlcul de la variància amb la taula de freqüències

$$\bar{x} = \frac{x_1 n_1 + x_2 n_2 + \dots + x_k n_k}{n}, \quad \text{Var}_X = \frac{x_1^2 n_1 + x_2^2 n_2 + \dots + x_k^2 n_k}{n} - \bar{x}^2$$

Mesures de dispersió

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

$$\sigma_X = \sqrt{\text{Var}_X}$$

- ▶ Var_X és la variància de les dades
- ▶ Càlcul de la variància amb la taula de freqüències

$$\bar{x} = \frac{x_1 n_1 + x_2 n_2 + \dots + x_k n_k}{n}, \quad \text{Var}_X = \frac{x_1^2 n_1 + x_2^2 n_2 + \dots + x_k^2 n_k}{n} - \bar{x}^2$$

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

$$\text{Var}M_X = \frac{n}{n-1} \text{Var}P_X$$

Mesures de dispersió

- ▶ Exemple de desviació típica i variància

7
5
9
7
5
6
7
6
4

4	1
5	2
6	2
7	3
9	1

Mesures de dispersió

► Exemple de desviació típica i variància

7
5
9
7
5
6
7
6
4

$$\bar{x} = \frac{7 + 5 + \dots + 4}{9} = 6.22$$

$$\begin{aligned} Var_X &= \frac{7^2 + 5^2 + \dots + 4^2}{9} - (6.22)^2 \\ &= 1.98 \end{aligned}$$

$$\sigma_X = \sqrt{Var_X} = 1.4$$

4	1
5	2
6	2
7	3
9	1

Mesures de dispersió

► Exemple de desviació típica i variància

7
5
9
7
5
6
7
6
4

$$\bar{x} = \frac{7 + 5 + \dots + 4}{9} = 6.22$$

$$\begin{aligned} Var_X &= \frac{7^2 + 5^2 + \dots + 4^2}{9} - (6.22)^2 \\ &= 1.98 \end{aligned}$$

$$\sigma_X = \sqrt{Var_X} = 1.4$$

4	1
5	2
6	2
7	3
9	1

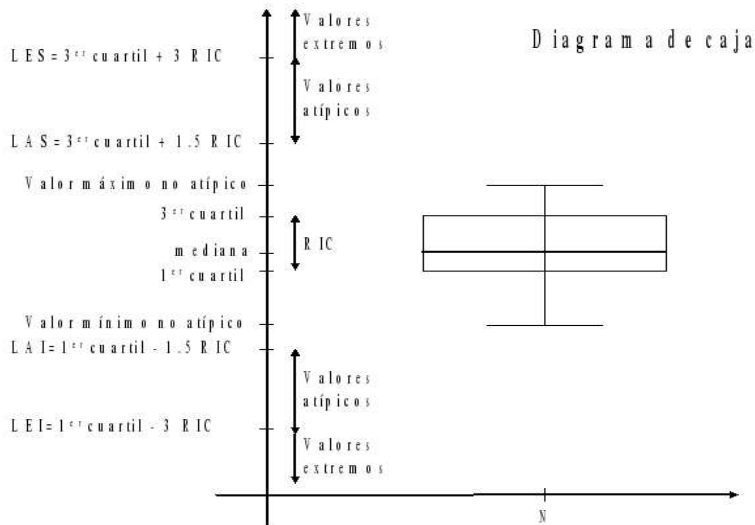
$$\bar{x} = \frac{4 * 1 + \dots + 9 * 1}{9} = 6.22$$

$$\begin{aligned} Var_X &= \frac{4^2 * 1 + \dots + 9^2 * 1}{9} \\ &\quad - (6.22)^2 = 1.98 \end{aligned}$$

$$\sigma_X = \sqrt{Var_X} = 1.4$$

Mesures de dispersió

- Representació gràfica de la dispersió: **Diagrames de caps**



Mesures de dispersió

► Exemple de diagrames de capsa

x_i	n_i	P_i
18	120	19,1693
19	150	43,1310
20	90	57,5080
21	70	68,6901
22	65	79,0735
23	50	87,0607
24	30	91,8530
25	20	95,0479
26	10	96,6454
27	7	97,7636
28	8	99,0415
29	2	99,3610
30	1	99,5208
34	1	99,6805
35	1	99,8403
40	1	100,0000

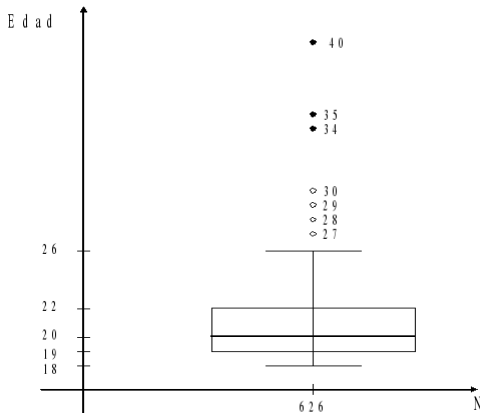
$n = 626$

$Q_1 = 19$

$Q_3 = 22$

mediana=20

$RIC = 3$



$Q_1 - 1.5RIC = 14.5$

$Q_3 + 1.5RIC = 26.5$

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

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

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

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

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

- ▶ Dues variables diferents es comparen comparant els seus respectius z-scores
- ▶ 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. Si tenim que en salt de longitud la mitjana és 7.40m i desviació típica és de 0.4m i en 100m lliures natació la mitjana és de 68.30s i la desviació típica és de 1.6s, quina persona mereix ms la beca?

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

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

- ▶ Dues variables diferents es comparen comparant els seus respectius z-scores
- ▶ 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. Si tenim que en salt de longitud la mitjana és 7.40m i desviació típica és de 0.4m i en 100m lliures natació la mitjana és de 68.30s i la desviació típica és de 1.6s, quina persona mereix ms la beca?

$$z_1 = \frac{7.60 - 7.40}{0.4} = 0.5 \quad z_2 = \frac{65.4 - 68.3}{1.6} = -1$$