

LEZIONI DI STATISTICA MEDICA

Dott. SIMONE ACCORDINI

Lezione n.8


- Misure di dispersione

- Misure di forma



Sezione di Epidemiologia & Statistica Medica
Università degli Studi di Verona

MISURE DI DISPERSIONE (measures of dispersion)

1. CAMPO DI VARIAZIONE (range)
 2. DISTANZA INTERQUARTILE
 3. DEVIANZA
 4. COEFFICIENTE DI VARIAZIONE
- 
4. VARIANZA
 5. DEVIAZIONE STANDARD

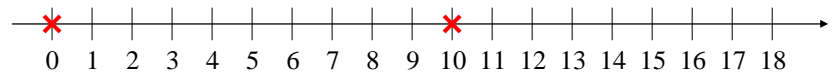


RANGE (CAMPO DI VARIAZIONE)

$$\text{Range} = x_{\max} - x_{\min}$$

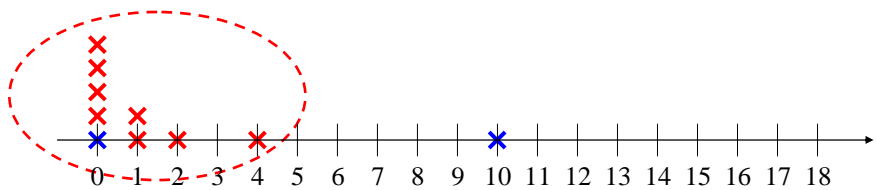
differenza tra il valore massimo e il valore minimo osservato

- Si basa soltanto sui **valori estremi** della distribuzione e non tiene conto dei valori intermedi.
- E' molto influenzato da **osservazioni anomale (outliers)**.
- **Tende ad aumentare al crescere del numero delle osservazioni.**



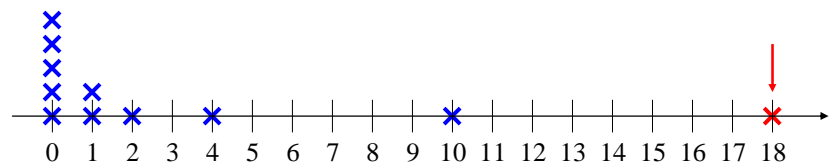
$$n = 2 \rightarrow \text{Range} = x_{\max} - x_{\min} = 10 - 0 = 10$$

num. linfonodi metastatici



$$n = 10 \rightarrow \text{Range} = x_{\max} - x_{\min} = 10 - 0 = 10$$

num. linfonodi metastatici



$$n = 11 \rightarrow \text{Range} = x_{\max} - x_{\min} = 18 - 0 = 18$$

num. linfonodi metastatici



DISTANZA INTERQUARTILE

$$IQR = Q_3 - Q_1$$

differenza tra il III° quartile (Q_3) e il I° quartile (Q_1)

- In questo intervallo ricade la **metà dei valori osservati**, posti esattamente al **centro della distribuzione**.
- Non è influenzata da osservazioni anomale o estreme.



Statura matricole della Facoltà di Medicina (A.A. 95/96)

MASCHI

Statura	Freq.	Cumul.
162	1	1
168	1	2
169	1	3
170	3	6
172	2	8
174	2	10
175	5	15
176	3	18
177	3	21
178	3	24
179	1	25
181	1	26
182	2	28
183	2	30
184	1	31
188	1	32
192	1	33
193	1	34
Totale	34	

$$\text{Range} = x_{\max} - x_{\min} = 193 - 162 = 31 \text{ cm}$$

Calcolo del I° quartile :

(rango percentile = 25)

$$1. \text{ rango} = (34+1) * 25 / 100 \\ = 35 / 4 \approx 9$$

$$2. \text{ I° quartile: } Q_1 = 174 \text{ cm}$$

Calcolo del III° quartile :

(rango percentile = 75)

$$1. \text{ rango} = (34+1) * 75 / 100 \\ = 35 * 3 / 4 \approx 26$$

$$2. \text{ III° quartile: } Q_3 = 181 \text{ cm}$$

$$IQR = Q_3 - Q_1 = 181 - 174 = 7 \text{ cm}$$



DEVIANZA

Nella popolazione

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

dimensione della popolazione

media nella popolazione (parametro)

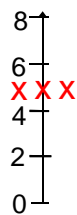
Nel campione

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

dimensione del campione

media nel campione (statistica)

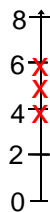
- E' un indice di dispersione definito sulla base del concetto di **SCARTO** rispetto ad un punto centrale della distribuzione (media aritmetica).
- E' la base delle misure di dispersione per variabili quantitative (da essa discendono la **VARIANZA** e la **DEVIAZIONE STANDARD**).



$$\bar{x} = 15 / 3 = 5$$

$$\sum x_i = 5 + 5 + 5 = 15$$

$$\text{devianza} = (5 - 5)^2 + (5 - 5)^2 + (5 - 5)^2 = 0$$



$$\sum x_i = 4 + 5 + 6 = 15$$

$$\text{devianza} = (4 - 5)^2 + (5 - 5)^2 + (6 - 5)^2 = 2$$



$$\sum x_i = 2 + 5 + 8 = 15$$

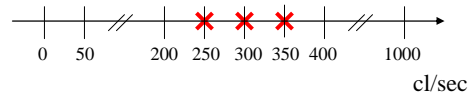
$$\text{devianza} = (2 - 5)^2 + (5 - 5)^2 + (8 - 5)^2 = 18$$



FEV ₁	n _i	p _i	(x _i - \bar{x})	(x _i - \bar{x}) ²
250	1	0.33	-50	2500
300	1	0.33	0	0
350	1	0.33	50	2500
TOT	3	1	0	5000

$$\bar{x} = (250+300+350) / 3 = 300 \text{ cl/sec}$$

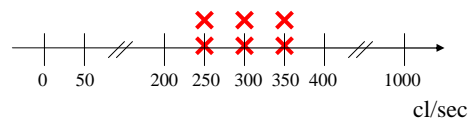
n = 3 → devianza = 5000 cl²/sec²



FEV ₁	n _i	p _i	(x _i - \bar{x})n _i	(x _i - \bar{x}) ² n _i
250	2	0.33	-50*2	2500*2
300	2	0.33	0	0
350	2	0.33	50*2	2500*2
TOT	6	1	0	10000

$$\bar{x} = (250*2+300*2+350*2) / 6 = 300 \text{ cl/sec}$$

n = 6 → devianza = 10000 cl²/sec²



La devianza raddoppia anche se la variabilità è costante!!!



VARIANZA

Nella popolazione

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

parametro

Nel campione (varianza corretta)

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

statistica

Gradi di Libertà

- E' una **devianza media** ossia la devianza rapportata al numero delle osservazioni campionarie (n) o della popolazione (N).
- E' la **media aritmetica del quadrato degli scarti** delle singole osservazioni dalla loro media aritmetica.



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

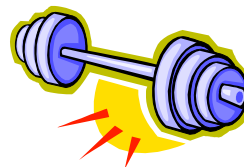
VARIANZA:
formula per
il calcolo

I **GRADI DI LIBERTÀ** rappresentano il **numero di osservazioni indipendenti** del campione, dal momento che sui dati disponibili è già stata calcolata una statistica (la media campionaria).

- Tiene conto di tutte le osservazioni ed è dunque influenzata da eventuali **osservazioni anomale (outliers)**.
- Non è direttamente confrontabile con la media o altri indici di posizione in quanto l'**unità di misura è elevata al quadrato**.



VARIANZA PONDERATA



Quando la distribuzione di frequenza è stata discretizzata in **k classi**

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



Distribuzione di frequenza della statura delle matricole di Medicina dell'Università di Verona nell'A.A. 95/96

CLASSE	PUNTO CENTRALE (x_i)	FREQUENZA ASSOLUTA	$x_i * n_i$	$x_i^2 * n_i$
[150-155)	152.5	1	152.5 * 1 = 152.5	(152.5) ² * 1 = 23256.25
[155-160)	157.5	8	157.5 * 8 = 1260.0	(157.5) ² * 8 = 198450.00
[160-165)	162.5	24	162.5 * 24 = 3900.0	(162.5) ² * 24 = 633750.00
[165-170)	167.5	34	5695.0	953912.50
[170-175)	172.5	27	4657.5	803418.75
[175-180)	177.5	19	3372.5	598618.75
[180-185)	182.5	9	1642.5	299756.25
[185-190)	187.5	1	187.5	35156.25
[190-195]	192.5	2	385.0	74112.50
TOTALE		125	21252.5	3620431.25

$$S^2 = \frac{\sum_{i=1}^k x_i^2 n_i - (\sum_{i=1}^k x_i n_i)^2 / n}{n - 1} = \frac{3620431.25 - (21252.5)^2 / 125}{124} = 57.1 \text{ cm}^2$$



DEVIAZIONE STANDARD

Nella popolazione

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

Nel campione (dev. st. corretta)

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

- Ha sempre un **valore positivo**.
- E' una misura della dispersione della variabile intorno alla media.
- E' direttamente **confrontabile con le misure di posizione**, essendo calcolata con la stessa unità di misura.



Distribuzione di frequenza della statura delle matricole di Medicina dell'Università di Verona nell'A.A. 95/96

CLASSE	PUNTO CENTRALE (x_i)	FREQUENZA ASSOLUTA	$x_i * n_i$	$x_i^2 * n_i$
[150-155)	152.5	1	152.5 * 1 = 152.5	(152.5) ² * 1 = 23256.25
[155-160)	157.5	8	157.5 * 8 = 1260.0	(157.5) ² * 8 = 198450.00
[160-165)	162.5	24	162.5 * 24 = 3900.0	(162.5) ² * 24 = 633750.00
[165-170)	167.5	34	5695.0	953912.50
[170-175)	172.5	27	4657.5	803418.75
[175-180)	177.5	19	3372.5	598618.75
[180-185)	182.5	9	1642.5	299756.25
[185-190)	187.5	1	187.5	35156.25
[190-195]	192.5	2	385.0	74112.50
TOTALE		125	21252.5	3620431.25

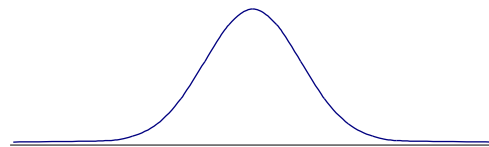
$$S = \sqrt{\frac{\sum_{i=1}^k x_i^2 n_i - (\sum_{i=1}^k x_i n_i)^2 / n}{n-1}} = \sqrt{\frac{3620431.25 - (21252.5)^2 / 125}{124}} = \sqrt{57.1} = 7.6 \text{ cm}$$



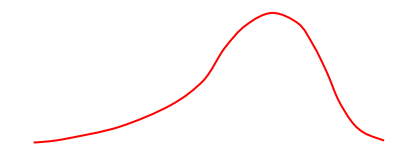
MISURE DI FORMA

• Misure di Simmetria (skewness)

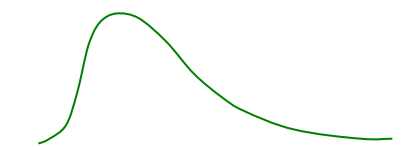
Indice di asimmetria: $[\sum(x_i - \bar{x})^3] / [\sum(x_i - \bar{x})^2]^{3/2}$



Simmetria (indice = 0)



Asimmetria negativa (indice < 0)



Asimmetria positiva (indice > 0)



• Misure di Appiattimento o di Curtosi (kurtosis)

Indice di curtosi:

$$\frac{[\sum(x_i - \bar{x})^4]}{[\sum(x_i - \bar{x})^2]^2}$$

