LEZIONI DI STATISTICA MEDICA

Prof. SIMONE ACCORDINI

Lezione n.8

- Misure di dispersione

Sezione di Epidemiologia & Statistica Medica
Università degli Studi di Verona

MISURE DI DISPERSIONE

(measures of dispersion)

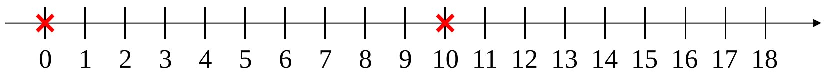
- CAMPO DI VARIAZIONE (range)
- DISTANZA INTERQUARTILE
- DEVIANZA → { VARIANZA; DEVIAZIONE STANDARD }
- COEFFICIENTE DI VARIAZIONE

RANGE (CAMPO DI VARIAZIONE)

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

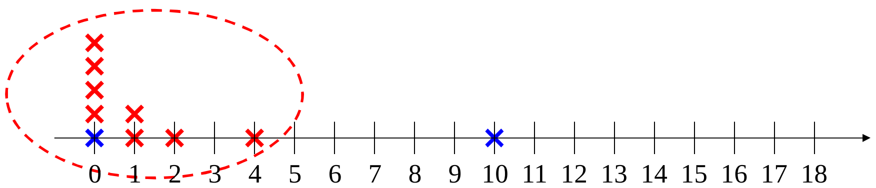
differenza tra il valore massimo e il valore minimo osservato

- 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.



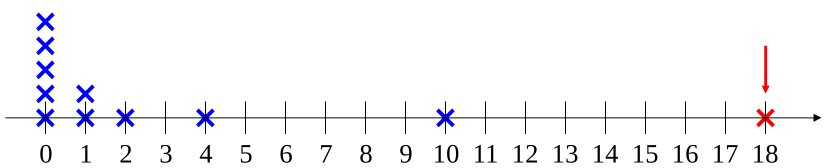
n = 2 → Range = x_{max} - x_{min} = 10 - 0 = 10

num. linfonodi metastatici



n = 10 → Range = x_{max} - x_{min} = 10 - 0 = 10

num. linfonodi metastatici



n = 11 → Range = x_{max} - x_{min} = 18 - 0 = 18

num. linfonodi metastatici

DISTANZA INTERQUARTILE

$$\boxed{IQR = Q_3 - Q_1}$$

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	

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

Calcolo del I° quartile:

(rango percentile = 25)

1. **rango = (34+1) * 25 / 100 = 35 / 4 ≈ 9**
2. **I° quartile: Q_1 = 174 cm**

Calcolo del III° quartile:

(rango percentile = 75)

1. **rango = (34+1) * 75 / 100 = 35 * 3 / 4 ≈ 26**
2. **III° quartile: Q_3 = 181 cm**

$IQR = Q_3 - Q_1$ = 181 - 174 = 7 cm

DEVIANZA

Nella popolazione

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

dimensione della popolazione

media nella popolazione (parametro)

Nel campione

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

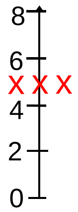
dimensione del campione

media nel campione (statistica)

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

SESM

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

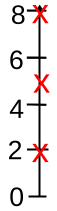


$\Sigma x_i = 5 + 5 + 5 = 15$

devianza = $(5 - 5)^2 + (5 - 5)^2 + (5 - 5)^2 =$ **0**

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

devianza = $(4 - 5)^2 + (5 - 5)^2 + (6 - 5)^2 =$ **2**



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

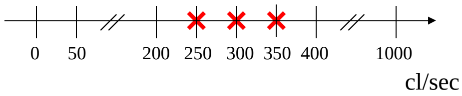
devianza = $(2 - 5)^2 + (5 - 5)^2 + (8 - 5)^2 =$ **18**

SESM

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 cl/sec

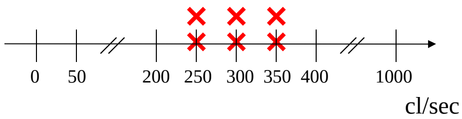
n = 3 → devianza = 5000 cl²/sec²



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 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.**

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

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.**

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 (n_i)	$(x_i - x)^2 * n_i$
[150-155)	152.5	1	$(152.5-170.0)^2 * 1 =$ 307.0
[155-160)	157.5	8	$(157.5-170.0)^2 * 8 =$ 1254.0
[160-165)	162.5	24	$(162.5-170.0)^2 * 24 =$ 1357.2
[165-170)	167.5	34	215.9
[170-175)	172.5	27	166.1
[175-180)	177.5	19	1063.1
[180-185)	182.5	9	1401.8
[185-190)	187.5	1	305.6
[190-195]	192.5	2	1010.7
TOTALE		125	7081.2

$$S^2 = \frac{\sum_{i=1}^{k}(x_i - \bar{x})^2 n_i}{n-1} = \frac{7081.2}{124} = 57.1 \text{ cm}^2$$

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.**

SESM

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 (n_i)	$(x_i - \bar{x})^2 * n_i$
[150-155)	152.5	1	$(152.5-170.0)^2 * 1 =$ 307.0
[155-160)	157.5	8	$(157.5-170.0)^2 * 8 =$ 1254.0
[160-165)	162.5	24	$(162.5-170.0)^2 * 24 =$ 1357.2
[165-170)	167.5	34	215.9
[170-175)	172.5	27	166.1
[175-180)	177.5	19	1063.1
[180-185)	182.5	9	1401.8
[185-190)	187.5	1	305.6
[190-195]	192.5	2	1010.7
TOTALE		125	7081.2

$$S = \sqrt{\frac{\sum_{i=1}^{k}(x_i - \bar{x})^2 n_i}{n-1}} = \sqrt{\frac{7081.2}{124}} = \sqrt{57.1} = 7.6 \text{ cm}$$

SESM

In alcune situazioni il confronto della variabilità all'interno di due gruppi di osservazioni è fuorviante se si utilizza la deviazione standard.

<u>Due gruppi con valori medi molto distanti</u>:

Tre neonati pesano rispettivamente **3**, **4** e **5 Kg** (media = **4 Kg**; dev.st. = **1 Kg**).
Tre bambini di 1 anno pesano **10**, **11** e **12 Kg** (media = **11 Kg**; dev.st. = **1 Kg**).

La deviazione standard è uguale nei due gruppi, ma il buon senso suggerisce che la variabilità del peso sia maggiore nei neonati.

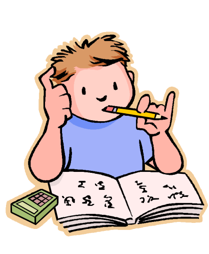
1. La variabile misurata è la stessa ma i valori medi delle osservazioni nei due gruppi sono molto distanti *(le osservazioni nei due gruppi sono su diversi ordini di grandezza)*

SESM

<u>Due variabili diverse</u>:

In 91 ragazze matricole di Medicina a Verona nell'A.A. 95/96,
la media del **peso** era pari a **55.1 Kg** e la deviazione standard era pari a **5.7 Kg**,
la media della **statura** era pari a **166.1 cm** e la deviazione standard era pari a **6.1 cm**.

E' maggiore la variabilità del peso o la variabilità della statura?



1. Le variabili misurate nei due gruppi sono diverse *(le osservazioni nei due gruppi sono espresse con diverse unità di misura)*

SESM

COEFFICIENTE DI VARIAZIONE (CV)

La deviazione standard viene espressa in percentuale della media.

CV = (deviazione standard / media) * 100

	Media	Dev. standard	CV
Neonati	4 Kg	1 Kg	**25.0 %**
Bambini 1 anno	11 Kg	1 Kg	**9.1 %**

La variabilità del peso è maggiore nei neonati.

	Media	Dev. standard	CV
Peso	55.1 Kg	5.7 Kg	**10.3 %**
Statura	166.1 cm	6.1 cm	**3.7 %**

La variabilità del peso è maggiore della variabilità della statura.

SESM