

Docente: Dott.ssa Egle Perissinotto

Modulo II

Contenuti:

Misure di tendenza centrale e di posizione:

- Moda, Mediana e quantili, Media aritmetica semplice e ponderata

Misure di variabilità:

- Campo di variazione, differenza interquartile, varianza, deviazione standard, coefficiente di variazione.

MISURE DI CENTRALITÀ E POSIZIONE

Valore tipico che sintetizza adeguatamente i dati di una distribuzione.

- Moda
- Mediana e quantili
- Media aritmetica semplice e ponderata

MISURE DI CENTRALITÀ: **MODA**

Modalità/valore della distribuzione che si presenta più frequentemente.

- *Può sintetizzare sia dati qualitativi che quantitativi*
- *Media di posizione non algebrica*

Cod. Paziente	Sesso	Età	Patologia	Prescrizioni (n)	Glicemia (mg/dl)
1	M	32	MEDICA	2	65
2	M	55	MEDICA	0	143
3	M	63	CHIRURGICA	4	115
4	F	27	PSICHIATRICA	2	110
5	F	83	MEDICA	3	180
6	M	59	CHIRURGICA	3	205
7	M	64	MEDICA	1	237
8	F	47	MEDICA	1	155
9	F	17	ALTRO	2	184
10	M	32	ALTRO	1	68
11	M	68	MEDICA	3	90
12	F	41	MEDICA	2	90
13	M	36	CHIRURGICA	3	145
14	F	47	CHIRURGICA	2	143
15	M	23	MEDICA	0	102
16	F	75	PSICHIATRICA	0	92
17	F	43	MEDICA	1	158
18	F	33	ALTRO	4	60
19	M	16	MEDICA	5	64
20	M	57	MEDICA	0	162

ESEMPIO

Distribuzione di 20 pazienti ricoverati, secondo il tipo di patologia

Tipo di patologia	Frequenze di unità statistiche (f_i)
Medica <i>Moda</i>	11
Chirurgica	4
Psichiatrica	2
Altro	3

ESEMPIO

Distribuzione di 20 pazienti ricoverati, secondo il numero di prescrizioni al momento del ricovero

Numero di prescrizioni	Frequenze di unità statistiche (f_i)
0	4
1	4
2 <i>Moda</i>	5
3	4
4	2
5	1

MISURE DI CENTRALITÀ: **MEDIANA**

Modalità o valore che supera la metà dei valori ed è superato dall'altra metà.

Facendo riferimento ai valori presentati dalle unità statistiche ordinati in senso crescente o decrescente, è il valore dell'unità che occupa il posto centrale.

- *Non risente dei valori estremi*
- *Media di posizione non algebrica*
- *Minimizza la somma delle distanze (in valore assoluto) dei singoli valori rispetto al valore mediano*

Serie non ordinata

Cod. Paziente	Prescrizioni (n)
1	2
2	0
3	4
4	2
5	3
6	3
7	1
8	1
9	2
10	1
11	3
12	2
13	3
14	2
15	0
16	0
17	1
18	4
19	5
20	0

Serie ordinata

Sequenza	Cod. Paziente	Prescrizioni (n)
1	2	0
2	15	0
3	16	0
4	20	0
5	7	1
6	8	1
7	10	1
8	17	1
9	1	2
10	4	2
11	9	2
12	12	2
13	14	2
14	5	3
15	6	3
16	11	3
17	13	3
18	3	4
19	18	4
20	19	5

posto centrale { 10, 11 } *Mediana*

MISURE DI POSIZIONE: **QUANTILI**

I *quantili* sono indicatori di posizione che suddividono la medesima successione ordinata di N unità in un certo numero di parti con lo stesso numero di unità.

In particolare:

I *quartili* dividono la serie ordinata in quattro parti uguali, ciascuna composta da $N/4$ unità statistiche

I *decili* dividono la serie ordinata in dieci parti uguali, ciascuna composta da $N/10$ unità statistiche

I *centili* dividono la serie ordinata in cento parti uguali per numero di unità

MISURE DI POSIZIONE: PERCENTILI

Il **percentile**, suddividendo una serie ordinata di valori continui in cento parti uguali, è il valore della variabile che supera il $p\%$ dei valori ed è superato dal rimanente $(100-p)\%$.

- decile: percentile considerato alla decina
- 1° quartile: = 25° percentile
- 2° quartile: = Mediana nonché 50° percentile
- 3° quartile: = 75° percentile

Sequenza	Cod. Paziente	Prescrizioni (n)
1	2	0
2	15	0
3	16	0
4	20	0
{ 5	7	1
{ 6	8	1
{ 7	10	1
8	17	1
9	1	2
{ 10	4	2
{ 11	9	2
12	12	2
13	14	2
14	5	3
{ 15	6	3
{ 16	11	3
17	13	3
18	3	4
19	18	4
20	19	5

I quartile - 25°percentile

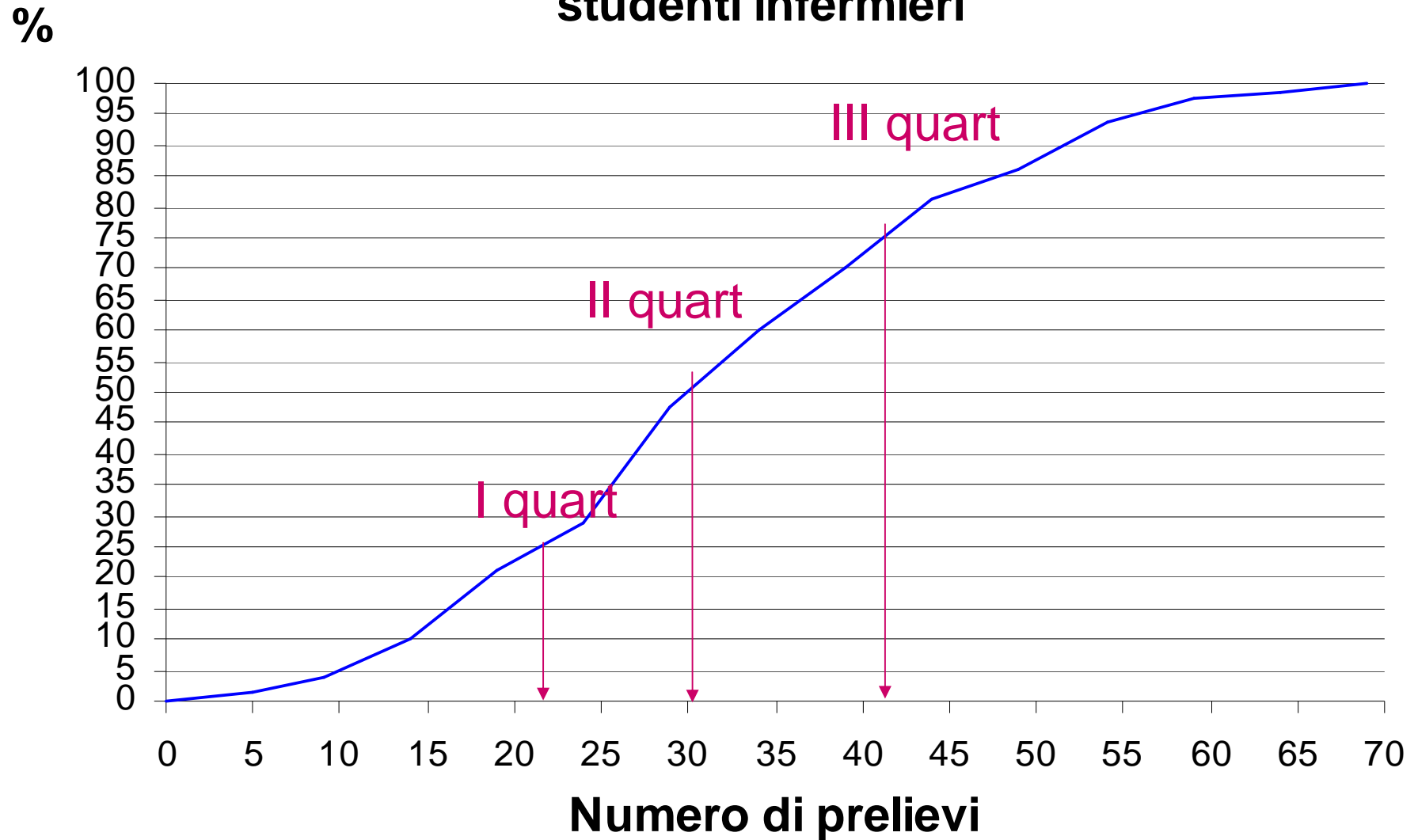
II quartile - 50°percentile-Mediana

III quartile - 75°percentile

Distribuzione del numero di prelievi effettuati mensilmente da 80 studenti infermieri

Intervallo di classe	Valore centrale della classe	Frequenza assoluta	Frequenza relativa	Frequenza cumulata	
				assoluta	relativa %
0-4	2	1	0,0125	1	1,25
5-9	7	2	0,0250	3	3,75
10-14	12	5	0,0625	8	10,00
15-19	17	9	0,1125	17	21,25
20-24	22 I quart	6	0,0750	23	28,75
25-29	27	15	0,1875	38	47,50
30-34	32 II quart	10	0,1250	48	60,00
35-39	37	8	0,1000	56	70,00
40-44	42 III quart	9	0,1125	65	81,25
45-49	47	4	0,0500	69	86,25
50-54	52	6	0,0750	75	93,75
55-59	57	3	0,0375	78	97,50
60-64	62	1	0,0125	79	98,75
65-69	67	1	0,0125	80	100,00
Totale	-	80	1,0000		

Frequenze cumulative relative percentuali del numero di prelievi mensilmente eseguiti da 80 studenti infermieri



MISURE DI CENTRALITÀ : **MEDIA ARITMETICA (M o \bar{x})**

Per una serie di n valori, la media aritmetica è data da:

$$\text{Media aritmetica} = \bar{x} = \frac{\sum x_i}{n}$$

Se gli n valori sono espressi in una distribuzione di frequenza (seriazione) di k classi, la media è data da:

$$\text{Media aritmetica} = \bar{x} = \frac{\sum x_i f_i}{\sum f_i} = \frac{\sum x_i f_i}{n}$$

Caratteristiche della media aritmetica

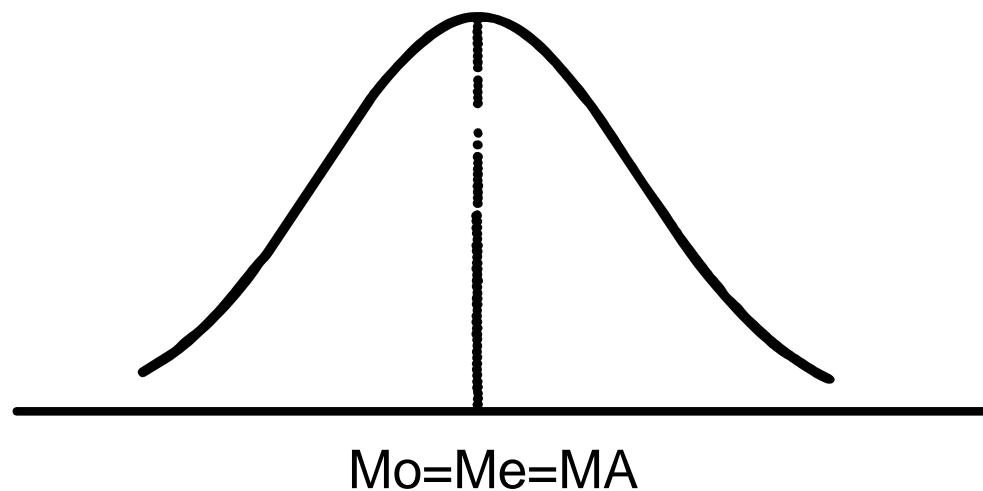
- *Media di posizione algebrica*
- *Risente dei valori estremi*
- *Proprietà di equilibrio: è nulla la somma degli scarti dalla media $\sum (x_i - \bar{x})$*
- *Proprietà di minimo: è minima la somma dei quadrati degli scarti dalla media $\sum (x_i - \bar{x})^2 = \text{minimo}$
La somma del quadrato degli scarti dalla media è detta
DEVIANZA*

Distribuzione del numero di prelievi effettuati mensilmente da 80 studenti

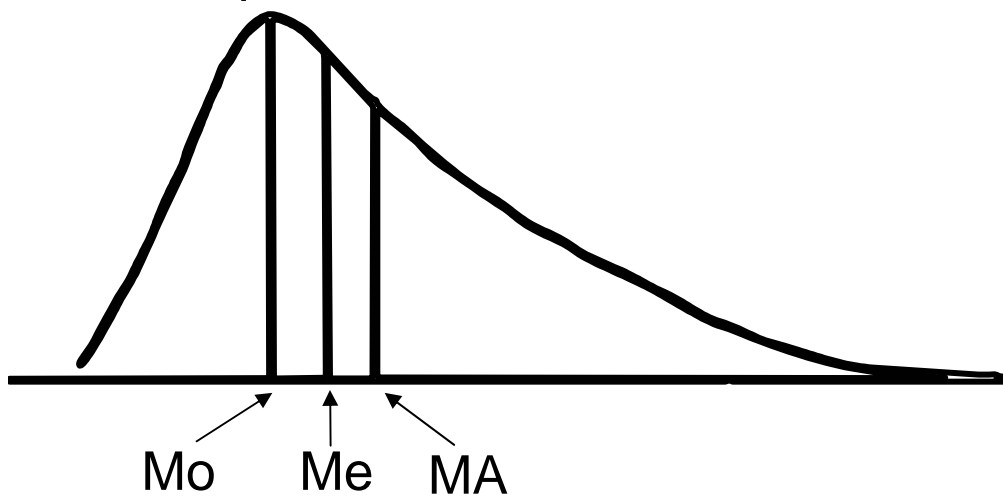
Intervallo di classe	Valore centrale x_i	Frequenza assoluta f_i	$x_i f_i$
0-4	2	1	2
5-9	7	2	14
10-14	12	5	60
15-19	17	9	153
20-24	22	6	132
25-29	27	15	405
30-34	32	10	320
35-39	37	8	296
40-44	42	9	378
45-49	47	4	188
50-54	52	6	312
55-59	57	3	171
60-64	62	1	62
65-69	67	1	67
Totale		80	2560

$$\bar{x} = \frac{\sum x_i f_i}{\sum f_i} = \frac{\sum x_i f_i}{n} = \frac{2560}{80} = 32$$

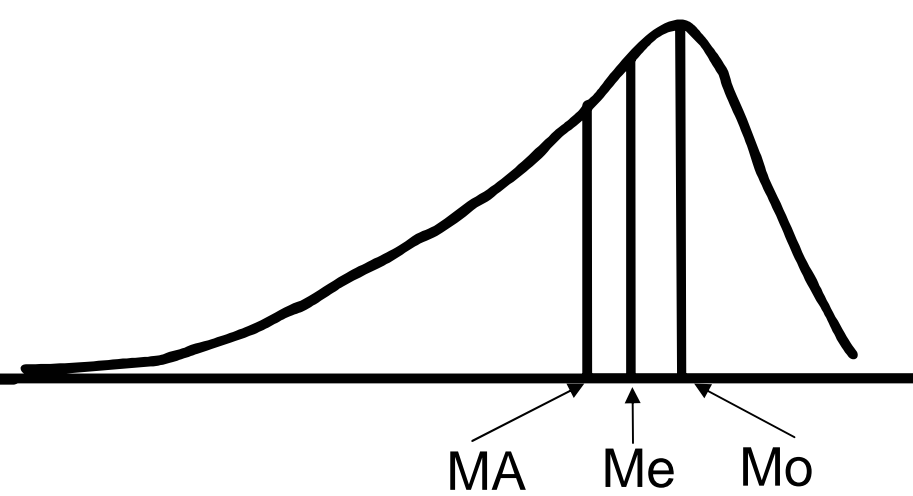
Distribuzione simmetrica a campana



Asimmetria positiva,
asimmetrica destra,
obliqua sinistra



Asimmetria negativa,
asimmetrica sinistra,
obliqua destra



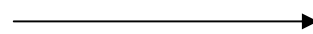
Criteria di scelta tra le medie

- Tipo di variabile
- Forma della distribuzione

CARATTERE

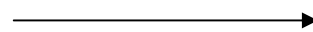
MISURE DI CENTRALITÀ E POSIZIONE

Qualitativo nominale



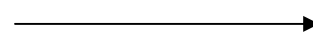
Moda

Qualitativo ordinale



Moda, Mediana, Quantili

Quantitativo discreto



Moda, Mediana, Quantili,
Media aritmetica

Quantitativo continuo



Moda, Mediana, Quantili,
Media aritmetica

MISURE DI VARIABILITÀ

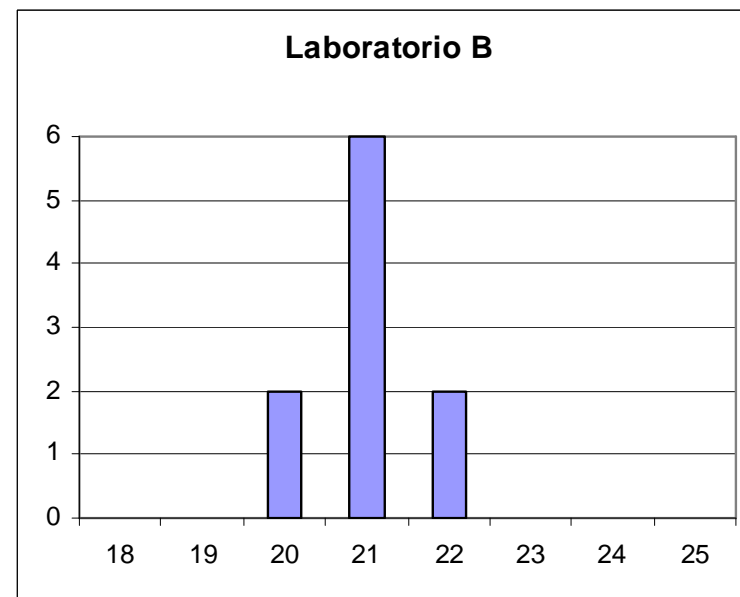
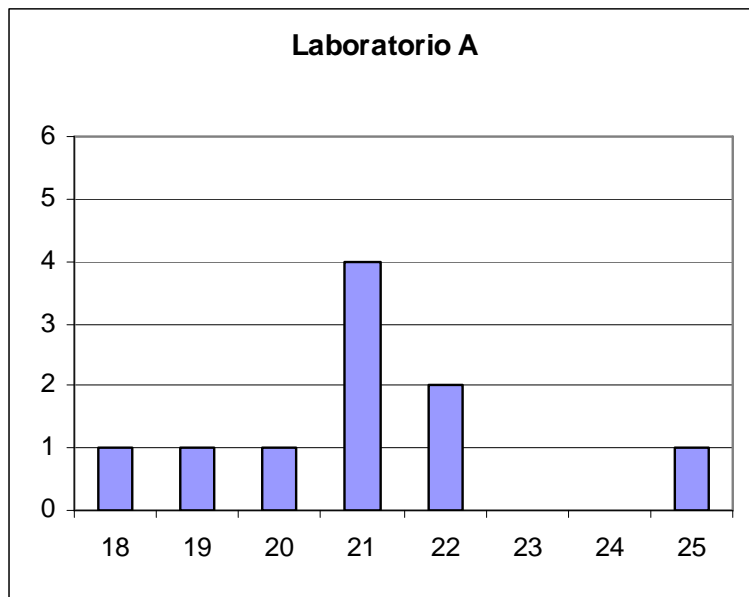
La variabilità è l'attitudine dei fenomeni ad assumere livelli o modalità differenti al variare dell'unità statistica

Gli indici di variabilità valgono zero nella situazione di nessuna variabilità (unità statistiche uguali)

Si distinguono:

- indici di variabilità assoluta (dipendenti dall'unità di misura della variabile)
- indici di variabilità relativa (adimensionali)

LABORATORIO A	LABORATORIO B
19	21
21	22
22	21
20	21
21	21
21	20
25	20
22	22
18	21
21	21
Moda = 21 Mediana = 21 Media = 21	Moda = 21 Mediana = 21 Media = 21



MISURE DI VARIABILITÀ: **Numero di modalità e Frequenza relativa della moda**

Il **numero di modalità** è il numero di categorie, modi, secondo cui la variabile qualitativa può presentarsi

La **frequenza relativa** della moda è data dal rapporto tra la frequenza assoluta della categoria modale e il totale delle frequenze

A differenza degli altri indici di variabilità, più è elevato il suo valore e meno variabile è la distribuzione

MISURE DI VARIABILITÀ : **Campo di variazione**

Il **campo di variazione** (range): è la differenza tra il valore più grande e quello più piccolo assunti dalla variabile nella serie di dati

- *Dipende dal numero di unità statistiche esaminate*
- *Dipende dai valori estremi*

Esempio: Numero di prestazioni ricevute in un anno da 13 soggetti dell'ULSS

Serie statistica	
Soggetto i	Numero x_i di prestazioni
1	0
2	2
3	6
4	2
5	6
6	5
7	22
8	3
9	1
10	10
11	2
12	5
13	1

Serie ordinata di valori			
Rango	Soggetto i	Numero x_i di prestazioni	Indice posizione
1°	1	0	
2°	9	1	
3°	13	1	
4°	2	2	1° Quartile
5°	4	2	
6°	11	2	
7°	8	3	2° Quartile
8°	6	5	
9°	12	5	
10°	3	6	3° Quartile
11°	5	6	
12°	10	10	
13°	7	22	

Campo di variazione: $x_{13} - x_1 = 22 - 0 = 22$ prestazioni

MISURE DI VARIABILITÀ: **Differenza interquartile**

La **differenza interquartile** è la differenza tra il 3° e il 1° quartile

- *Include il 50% dei valori della serie*
- *Non dipende dai valori estremi*

Serie ordinata di valori			
Rango	Soggetto i	Numero x_i di prestazioni	Indice posizione
1°	1	0	
2°	9	1	
3°	13	1	
4°	2	2	1° Quartile
5°	4	2	
6°	11	2	
7°	8	3	2° Quartile
8°	6	5	
9°	12	5	
10°	3	6	3° Quartile
11°	5	6	
12°	10	10	
13°	7	22	

Differenza interquartile= 3°quartile- 1°quartile= 6 - 2= 4 prestazioni

MISURE DI VARIABILITÀ: **Varianza e Deviazione standard** **(Scarto Quadratico Medio)**

$$\sigma^2 = \frac{\sum (x_i - M)^2}{n} = \text{Varianza}$$

Il numeratore $\sum (x_i - M)^2$ si chiama Devianza

$$\sigma = \sqrt{\frac{\sum (x_i - M)^2}{n}} = \text{Deviazione standard (scarto quadratico medio)}$$

Per dati in seriazione:

$$\sigma^2 = \frac{\sum (x_i - M)^2 f_i}{n}$$

Risultati analisi Laboratorio A

Campione	Concentrazione (mg/dl) x_i	$(x_i - \bar{X})$	$(x_i - \bar{X})^2$
1	19	-2	4
2	21	0	0
3	22	1	1
4	20	-1	1
5	21	0	0
6	21	0	0
7	25	4	16
8	22	1	1
9	18	-3	9
10	21	0	0
Totale		0	32

Media = 21 mg/dl

$$\text{Varianza} = \frac{\sum (x_i - M)^2}{n} = \frac{32}{10} = 3,2 \text{ (mg/dl)}^2$$

$$\text{Deviazione standard} = \sqrt{3,2 \text{ (mg/dl)}^2} = 1,8 \text{ mg/dl}$$

Risultati analisi Laboratorio B

Campione	Concentrazione (mg/dl) x_i	$(x_i - \bar{X})$	$(x_i - \bar{X})^2$
1	21	0	0
2	22	1	1
3	21	0	0
4	21	0	0
5	21	0	0
6	20	-1	1
7	20	-1	1
8	22	1	1
9	21	0	0
10	21	0	0
Totale		0	4

Media= 21 mg/dl

$$\text{Varianza} = \frac{\sum (x_i - M)^2}{n} = \frac{4}{10} = 0,4 \text{ (mg/dl)}^2$$

$$\text{Deviazione standard} = \sqrt{0,4 \text{ (mg/dl)}^2} = 0,6 \text{ mg/dl}$$

MISURE DI VARIABILITÀ RELATIVA: **Coefficiente di variazione**

Il **coefficiente di variazione** è il rapporto tra lo scarto quadratico medio e la media aritmetica, moltiplicato per 100

$$CV\% = \frac{\sigma}{M} \times 100$$

Nell'esempio dei due laboratori si ha:

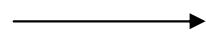
Laboratorio A: $CV\% = (1,8/21) \times 100 = 8,6\%$

Laboratorio B: $CV\% = (0,6/21) \times 100 = 2,9\%$

TIPO DI VARIABILE

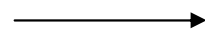
INDICE DI VARIABILITÀ

Variabili
qualitative



- Numero di modalità
- Frequenza relativa della moda

Variabili
quantitative



- Campo di variazione
- Differenza interquartile
- Deviazione standard (Scarto Quadratico Medio), Varianza
- Coefficiente di variazione

PRECISIONE (REPLICABILITÀ)

Attitudine di un procedimento di misura a fornire sempre lo stesso risultato quando applicato ripetutamente sullo stesso oggetto.

Misurata mediante lo Scarto Quadratico Medio.

ACCURATEZZA (ESATTEZZA)

Differenza tra la misura effettuata e la misura reale (nota per altra via). Espressa in forma percentuale:

$$\left[\frac{(X - \mu)}{\mu} \right] \times 100.$$