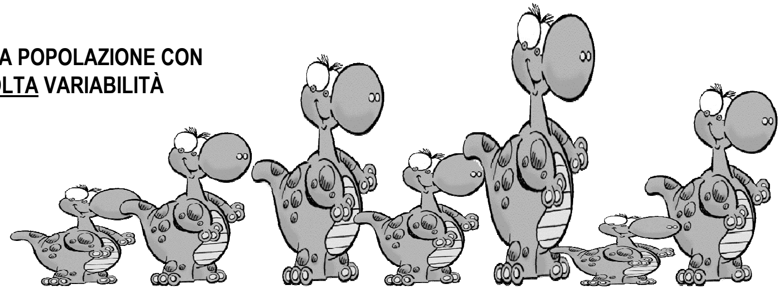
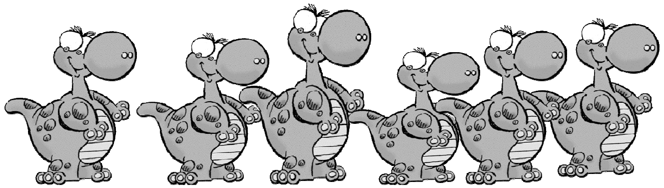
la variabile d'interesse è l'ALTEZZA

**UNA POPOLAZIONE CON <u>MOLTA</u> VARIABILITÀ**



**UNA POPOLAZIONE CON <u>POCA</u> VARIABILITÀ**



# Misure di dispersione

- 27 28 29 30 31
- 9 18 23 45 50

$\bar{x} = 29$ $\bar{x} = 29$

- range=31-27=4
- range=50-9=41

- Ma anche il range ha dei limiti

# RANGE (CAMPO DI VARIAZIONE)

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

***differenza tra il valore massimo e il valore minimo osservati***

✓ Si basa soltanto sui valori estremi della distribuzione e non tiene conto dei valori intermedi

✓ E' molto influenzato da osservazioni anomale (***outliers***)

# DISTANZA INTERQUARTILE

$$IQR = Q_3 - Q_1$$

***differenza tra il III°quartile (Q3) ed il I°quartile (Q1)***

✓ In questo intervallo ricade la metà dei valori osservati, posta esattamente al centro della distribuzione.

✓ Non è influenzata da osservazioni anomale o estreme.

*Calcolo della deviazione standard del volume plasmatico di 8 adulti maschi sani. Media, $\bar{x}$=3.00*

| *Volume plasmatico* $x$ | *Deviazione Dalla media* $x-\bar{x}$ | *Deviazione al quadrato* $(x-\bar{x})^2$ | *Osservazione elevata al quadrato* $x^2$ |
|---|---|---|---|
| 2.75 | -0.25 | 0.0625 | 7.5625 |
| 2.86 | -0.14 | 0.0196 | 8.1796 |
| 3.37 | 0.37 | 0.1369 | 11.3569 |
| 2.76 | -0.24 | 0.0576 | 7.6176 |
| 2.62 | -0.38 | 0.1444 | 6.8644 |
| 3.49 | 0.49 | 0.2401 | 12.1801 |
| 3.05 | 0.05 | 0.0025 | 9.3025 |
| 3.12 | 0.12 | 0.0144 | 9.7344 |
| Totali 24.02 | Totali 0.00 | Totali 0.6780 | Totali 72.7980 |

## ESEMPIO VOLUME PLASMATICO DI 8 MASCHI ADULTI SANI

**DEVIANZA**

$$\sum(x-\bar{x})^2 = 0{,}6780$$

**VARIANZA** $s^2$

$$s^2 = \frac{devianza}{N-1} = \frac{0{,}6780}{7} = 0{,}0969 l^2$$

**DEVIAZIONE STANDARD** $s$

$$s = \sqrt{0{,}0969} = 0{,}311 l$$

**DEVIANZA** $\sum\left(x-\bar{x}\right)^2$

**VARIANZA** $s^2$

$$s^2 = \frac{\sum\left(x-\bar{x}\right)^2}{N-1}$$

**DEVIAZIONE STANDARD** $s$

$$s = \sqrt{VARIANZA} = \sqrt{\frac{\sum\left(x-\bar{x}\right)^2}{N-1}}$$

**ESEMPIO: La seguente tabella mostra le ore di sollievo dovute a due differenti farmaci somministrati consecutivamente a pazienti che soffrono di artrite**

| Paziente | Farmaco A | Farmaco B |
|---|---|---|
| 1 | 2.0 | 3.5 |
| 2 | 3.6 | 5.7 |
| 3 | 2.6 | 2.9 |
| 4 | 2.6 | 2.4 |
| 5 | 7.3 | 9.9 |
| 6 | 3.4 | 3.3 |
| 7 | 14.9 | 16.7 |
| 8 | 6.6 | 6.0 |
| 9 | 2.3 | 3.8 |
| 10 | 2.0 | 4.0 |
| 11 | 6.8 | 9.1 |
| 12 | 8.5 | 20.9 |

1) CALCOLARE LA MEDIA NELLE ORE DI SOLLIEVO DATE DAI DUE FARMACI

A: X = (2.0+3.6+2.6…+8.5) /12= 5.22

B: X = 7.35

2) CALCOLARE LA DEVIANZA:

$A: \sum(X - X)^2 (2.0-5.22)^2 + (3.6-5.22)^2 + (2.6-5.22)^2 + \ldots + (8.5-5.22)^2 = 162.12$

$$B : \sum \left( X - X \right)^2 = 384.49$$

**3) CALCOLARE LA VARIANZA**

$$A : \sum \left( X - X \right)^2 / (N - 1) = 132.12 / 11 = 14{,}74 = S_A^2$$

$$B : \sum \left( X - X \right)^2 / (N - 1) = 384.49 / 11 = 34.95 = S_B^2$$

**4) CALCOLARE LA DEVIAZIONE STANDARD**

$$A : s = radq \left( \sum \left( x - x \right)^2 / (N - 1) \right) = radq \left( 14.74 \right) = 3.84 = S_A$$

$$B : s = radq \left( \sum \left( x - x \right)^2 / (N - 1) \right) = radq \left( 34.95 \right) = 5.91 = S_B$$

**5) CALCOLARE L' INTERVALLO DI VARIAZIONE E RANGE DELLA VARIABILE ORE DI SOLLIEVO PER I DUE FARMACI**

| | FARMACO A | FARMACO B |
|---|---|---|
| RANGE | 14.9-2=12.9 | 20.9-2.4=18.5 |

In alcune situazioni il confronto della variabilità all'interno di due gruppi di osservazioni utilizzando la deviazione standard è fuorviante

**Due variabili diverse:**

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)*

**Due gruppi con valori medi molto distanti**

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)*

# COEFFICIENTE DI VARIAZIONE PERCENTUALE

CV% = (deviazione standard / media) * 100%

Ci permette di misurare la variabilità **indipendentemente** dalla grandezza e dalla scala di misura delle osservazioni

| | 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.**

## Esempio di confronto tra variabilità

Nell' ipotesi di aver riscontrato in un gruppo di soggetti i seguenti valori

glicemia: $\bar{x}$ = 85 mg / 100 ml

s = 11 mg / 100 ml,

calcemia: $\bar{x}$ = 9 mg / 100 ml

s = 1,5 mg / 100 ml,

non è corretto considerare più dispersa la variabile glicemia che presenta una deviazione standard circa dieci volte maggiore.

Il confronto deve essere effettuato tramite i CV % :

glicemia: $CV\,\% = \dfrac{11\,mg\ /100\ ml}{85\,mg\ /100\ ml} \cdot 100\ = 12{,}9\%$

calcemia: $CV\,\% = \dfrac{1{,}5\,mg\ /100\ ml}{9\,mg\ /100\ ml} \cdot 100\ = 16{,}7\%$

In realtà, in base alle misure ottenute risulta più dispersa, anche se di poco, la calcemia

**ESEMPIO**

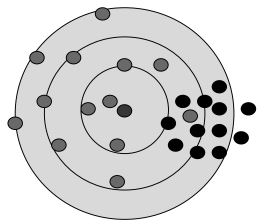
Per valutare l' attendibilità di due metodi (A e B) per la determinazione del colesterolo sierico, un siero standard a concentrazione nota di colesterolo, pari a 180 mg/dl, viene misurato 100 volte con ognuno dei due metodi.

Le medie e le deviazioni standard delle determinazioni sono rispettivamente:

per A: media 178 mg/dl, ds=2 mg/dl

per B: media 179.2 mg/dl, ds=4 mg/dl

**Quale dei due metodi è più preciso?**



< ds ⇨ > precisione

**A + precisa**

**B** + accurata, si avvicina di più al valorevero

# Misure di posizione e di dispersione: simbologia

| **popolazione** | | **campione** |
|---|---|---|
| $\mu$ = media | *stimata da* | x |
| $\sigma^2$ = varianza | *stimata da* | $s^2$ |
| $\sigma$ = deviazione standard | *stimata da* | s |

**1.** Si hanno i pesi (kg) alla nascita di 10 bambini:

3,5 4,2 3,2 2,9 4,0 3,5 3,1 3,0 3,9 4,1

a) Calcolare media, moda, mediana del peso dei 10 neonati.
b) Calcolare range e deviazione standard del peso dei 10 neonati.

**2.** 50 neonati presentavano un peso medio di 3,8 kg e 10 neonate un peso medio di 3,0 kg. Qual è il peso medio dei 60 neonati?

FREQUENZE ASSOLUTE E RELATIVE DEI LIVELLI DI COLESTEROLO SIERICO IN 2.294 SOGGETTI DELLA POPOLAZIONE MASCHILE DEGLI STATI UNITI, 1976 – 1980
RAPPRESENTAZIONE GRAFICA TRAMITE POLIGONI DI FREQUENZA
ENTRAMBI SULLO STESSO GRAFICO.

| **X Livello di colesterolo (mg / 100 ml)** | Età 25 - 34 | | Età 55 - 64 | |
|---|---|---|---|---|
| | **Numero di soggetti** | **Frequenza relativa (%)** | **Numero di soggetti** | **Frequenza relativa (%)** |
| 80-119 | 13 | 1,2 | 5 | 0,4 |
| 120-159 | 150 | 14,1 | 48 | 3,9 |
| 160-199 | 442 | 41,4 | 265 | 21,6 |
| 200-239 | 299 | 28,0 | 458 | 37,3 |
| 240-279 | 115 | 10,8 | 281 | 22,9 |
| 280-319 | 34 | 3,2 | 128 | 10,4 |
| 320-359 | 9 | 0,8 | 35 | 2,9 |
| 360-399 | 5 | 0,5 | 7 | 0,6 |
| **TOTALE** | **1.067** | **100,0** | **1.227** | **100,0** |

FREQUENZE RELATIVE E FREQUENZE RELATIVE CUMULATIVE DEI LIVELLI DI COLESTEROLO SIERICO IN 2.294 SOGGETTI DELLA POPOLAZIONE MASCHILE DEGLI STATI UNITI, 1976 – 1980

| **Livello di colesterolo (mg / 100 ml)** | Età 25 - 34 | | Età 55 - 64 | |
|---|---|---|---|---|
| | **Frequenza relativa (%)** | **Frequenza relativa cumulativa (%)** | **Frequenza relativa (%)** | **Frequenza relativa cumulativa (%)** |
| 80-100 | 1,2 | 1,2 | 0,4 | 0,4 |
| 120-159 | 14,1 | 15,3 | 3,9 | 4,3 |
| 160-199 | 41,4 | 56,7 | 21,6 | 25,9 |
| 200-239 | 28,0 | 84,7 | 37,3 | 63,2 |
| 240-279 | 10,8 | 95,5 | 22,9 | 86,1 |
| 280-319 | 3,2 | 98,7 | 10,4 | 96,5 |
| 320-359 | 0,8 | 99,5 | 2,9 | 99,4 |
| 360-399 | 0,5 | 100,0 | 0,6 | 100,0 |

**Misure di sintesi del livello di colesterolo sierico in 2294 soggetti della popolazione maschile statunitense**

*Età 25 – 34 anni*

*Classe modale (160 – 199) mg/100 ml*

*Classe mediana (160 – 199) mg/100 ml*

$$\bar{x} = 199,3$$

$$s = 43,9\, mg / 100\, ml$$

$$CV\% = 22,0\%$$

*Età 55 – 64 anni*

*Classe modale (200 – 239) mg/100 ml*

*Classe mediana (200 – 239) mg/100 ml*

$$\bar{X} = 229,7$$

$$s = 46,4\, mg / 100\, ml$$

$$CV\% = 20,2\%$$