# Statistica

## Lezione 4

Bice Cavallo

# Indice

# Introduzione

- Le medie analitiche e le medie di posizione si utilizzano per sintetizzare con un solo valore o una sola modalità la distribuzione.
  - Le **medie analitiche** vengono calcolate attraverso operazioni matematiche sui valori del carattere, che dovrà essere perciò necessariamente di tipo quantitativo
  - Le **medie di posizione** possono essere determinate anche su caratteri di tipo qualitativo

# Media aritmetica

- La **media aritmetica** di un insieme di $n$ valori osservati $x_1$, $x_2$, …, $x_n$ di un carattere quantitativo $X$ è pari alla somma dei valori divisa per il loro numero:

$$\bar{x}_a = \frac{1}{n}\sum_{i=1}^{n} x_i$$

# Media aritmetica

- Un lavoratore può raggiungere il posto di lavoro utilizzando due mezzi di trasporto: metropolitana, automobile.
- Decide di scegliere il mezzo di trasporto che gli consente di raggiungere il posto di lavoro nel minor tempo possibile.
- Per fare questo decide di andare i primi 12 giorni con l'automobile e i successivi 12 giorni con la metropolitana e di registrare ogni volta il tempo impiegato.

# Media aritmetica

| Tempo impiegato (min.) | | | Tempo impiegato (min.) | | |
|---|---|---|---|---|---|
| Giorno | Auto | Metro | Giorno | Auto | Metro |
| 1 | 23 | 22 | 7 | 28 | 24 |
| 2 | 32 | 24 | 8 | 33 | 28 |
| 3 | 44 | 22 | 9 | 45 | 32 |
| 4 | 21 | 33 | 10 | 34 | 31 |
| 5 | 36 | 26 | 11 | 29 | 37 |
| 6 | 30 | 31 | 12 | 31 | 24 |

# Media aritmetica

| Tempo impiegato (min.) | | | Tempo impiegato (min.) | | |
|---|---|---|---|---|---|
| **Giorno** | **Auto** | **Metro** | **Giorno** | **Auto** | **Metro** |
| 1 | 23 | 22 | 7 | 28 | 24 |
| 2 | 32 | 24 | 8 | 33 | 28 |
| 3 | 44 | 22 | 9 | 45 | 32 |
| 4 | 21 | 33 | 10 | 34 | 31 |
| 5 | 36 | 26 | 11 | 29 | 37 |
| 6 | 30 | 31 | 12 | 31 | 24 |

$$\bar{x}_a(\text{auto}) = 32{,}17$$

$$\bar{x}_a(\text{metropolitana}) = 27{,}83$$

# Media aritmetica

- Se il dato è quantitativo discreto e conosciamo la distribuzione di frequenza:

$$\bar{x}_a = \frac{1}{n}\sum_{j=1}^{k} x_j n_j \quad \text{oppure} \quad \bar{x}_a = \sum_{j=1}^{k} x_j f_j$$

  - $k$=numero di modalità assunte dal carattere
  - $n_j$ = frequenza assoluta della modalità $x_j$
  - $f_j$ = frequenza relativa della modalità $x_j$

# Media aritmetica approssimata

- Se il dato è quantitativo e suddiviso in classi, possiamo **approssimare** la media aritmetica:

$$\bar{x}_a \approx \frac{1}{n}\sum_{j=1}^{k} c_j n_j$$

  - $k$=numero delle classi
  - $n_j$ = frequenza assoluta della classe $j$
  - $c_j$ = valore centrale della classe $j$

- Se c'è una classe non limitata, bisognerà trasformarla scegliendo opportunamente l'estremo mancante della classe

# Media aritmetica approssimata

- Distribuzione di laureati in Economia nel 1986 in Italia

| Classi di età | Frequenza |
|---|---|
| 23-24 | 432 |
| 24-26 | 2836 |
| 26-28 | 1620 |
| 28-30 | 768 |
| oltre 30 | 476 |
| **Totale** | 6132 |

- Consideriamo 40 come estremo superiore

# Media aritmetica approssimata

- Distribuzione di laureati in Economia nel 1986 in Italia

| Classi di età | Frequenza |
|---|---|
| 23-24 | 432 |
| 24-26 | 2836 |
| 26-28 | 1620 |
| 28-30 | 768 |
| oltre 30 | 476 |
| **Totale** | 6132 |

→

| **Centrali** |
|---|
| 23,5 |
| 25 |
| 27 |
| 29 |
| 35 |

$$\bar{x}_a \approx 26{,}7$$

- Consideriamo 40 come estremo superiore

# Media aritmetica esatta per un carattere suddiviso in classi

- Comuni per ampiezza demografica in Italia, anno 2011 (ISTAT)

| Classi di ampiezza demografica | Comuni |
|---|---|
| Fino a 1.000 | 1.942 |
| 1.001-3.000 | 2.608 |
| 3.001-5.000 | 1.149 |
| 5.001-10.000 | 1.187 |
| 10.001-15.000 | 479 |
| 15.001-20.000 | 221 |
| 20.001-50.000 | 364 |
| 50.001-80.000 | 77 |
| 80.001-100.000 | 20 |
| 100.001-250.000 | 33 |
| 250.001-500.000 | 6 |
| oltre 500.000 | 6 |
| **Totale** | 8.092 |

- Supponiamo 2.500.000 ultimo estremo

| Centrali | Comuni |
|---|---|
| 500 | 1942 |
| 2000 | 2608 |
| 4000 | 1149 |
| 7500 | 1187 |
| 12500 | 479 |
| 17500 | 221 |
| 35000 | 364 |
| 65000 | 77 |
| 90000 | 20 |
| 175000 | 33 |
| 375000 | 6 |
| 1500000 | 6 |
| **Totale** | 8092 |

- Popolazione media per comune (approssimazione)

$$\bar{x}_a \approx 8.170$$

# Media aritmetica esatta per un carattere suddiviso in classi

**TABELLA 3.2.1** Comuni e popolazione per ampiezza demografica. Italia, anno 2011 (dati provvisori, fonte ISTAT)

| Classi di ampiezza demografica | Comuni | Popolazione |
|---|---|---|
| Fino a 1.000 | 1.942 | 1.056.740 |
| 1.001-3.000 | 2.608 | 4.800.689 |
| 3.001-5.000 | 1.149 | 4.472.328 |
| 5.001-10.000 | 1.187 | 8.397.010 |
| 10.001-15.000 | 479 | 5.836.597 |
| 15.001-20.000 | 221 | 3.800.346 |
| 20.001-50.000 | 364 | 11.081.116 |
| 50.001-80.000 | 77 | 4.686.699 |
| 80.001-100.000 | 20 | 1.786.011 |
| 100.001-250.000 | 33 | 4.772.982 |
| 250.001-500.000 | 6 | 1.859.262 |
| oltre 500.000 | 6 | 7.020.801 |
| **Totale** | 8.092 | 59.570.581 |

$$\bar{x}_a = \frac{59.570.581}{8.092} = 7.362$$

# Media aritmetica ponderata

- In alcuni casi si vuole dare diversa importanza alle diverse osservazioni del carattere attribuendo ad ognuna di esse un peso specifico

- **Media aritmetica ponderata** di $n$ valori osservati $x_1, x_2, \ldots, x_n$ di un carattere quantitativo $X$ con pesi $p_1, p_2, \ldots, p_n$ non negativi è data da:

$$\bar{x}_a = \frac{\sum_{i=1}^{n} x_i p_i}{\sum_{i=1}^{n} p_i}$$

  - Pesi uguali → media aritmetica
  - $$\sum_{i=1}^{n} p_i = 1 \Rightarrow \bar{x}_a = \sum_{i=1}^{n} x_i p_i$$

# Media aritmetica ponderata

- Esami sostenuti da uno studente

| Corso | Voto | CFU |
|---|---|---|
| Matematica e Statistica | 30 | 9 |
| Fondamenti di Informatica | 24 | 6 |
| La città che cambia (corso introduttivo) | 24 | 4 |

- Media aritmetica=
- Media aritmetica ponderata=

# Media aritmetica ponderata

- Esami sostenuti da uno studente

| Corso | Voto | CFU |
|---|---|---|
| Matematica e Statistica | 30 | 9 |
| Fondamenti di Informatica | 24 | 6 |
| La città che cambia (corso introduttivo) | 24 | 4 |

- Media aritmetica=26
- Media aritmetica ponderata= 26,84

# Media geometrica

- La **media geometrica** di un insieme di $n$ valori osservati $x_1$, $x_2$, …, $x_n$ di un carattere quantitativo $X$ è:

$$\bar{x}_g = \sqrt[n]{x_1 \cdot x_2 \cdot \ldots x_n}$$

# Media geometrica

- Se conosciamo la distribuzione di frequenza assoluta:

$$\bar{x}_g = \sqrt[n]{x_1^{n_1} \cdot x_2^{n_2} \cdot \ldots x_k^{n_k}}$$

  - $n_j$ = frequenza assoluta della modalità $x_j$
  - $k$=numero di modalità assunte dal carattere

- Se conosciamo la distribuzione di frequenza relativa:

$$\bar{x}_g = x_1^{f_1} \cdot x_2^{f_2} \cdot \ldots x_k^{f_k}$$

  - $f_j$ = frequenza relativa della modalità $x_j$

# Media geometrica

Un risparmiatore ha investito € 10.000 nell'acquisto di obbligazioni a tasso variabile. I tassi di interesse percentuali ottenuti alla fine di ogni anno sono riportati nella tabella seguente:

| Anno | Tasso di interesse | Tasso di interesse (%) | Capitale a fine anno |
|---|---|---|---|
| I | 0,015 | 1,5 | 10.150,0 |
| II | 0,020 | 2,0 | 10.353,0 |
| III | 0,072 | 7,2 | 11.098,4 |
| IV | 0,090 | 9,0 | 12.097,3 |
| V | 0,074 | 7,4 | 12.992,5 |
| VI | 0,045 | 4,5 | 13.577,1 |

- Calcolare il tasso di interesse percentuale medio annuo.

# Media geometrica

Un risparmiatore ha investito € 10.000 nell'acquisto di obbligazioni a tasso variabile. I tassi di interesse percentuali ottenuti alla fine di ogni anno sono riportati nella tabella seguente:

| Anno | Tasso di interesse | Tasso di interesse (%) | Capitale a fine anno |
|---|---|---|---|
| I | 0,015 | 1,5 | 10.150,0 |
| II | 0,020 | 2,0 | 10.353,0 |
| III | 0,072 | 7,2 | 11.098,4 |
| IV | 0,090 | 9,0 | 12.097,3 |
| V | 0,074 | 7,4 | 12.992,5 |
| VI | 0,045 | 4,5 | 13.577,1 |

- Capitale al sesto anno:

$$10.000 \cdot (1+0{,}015) \cdot (1+0{,}02) \cdot (1+0{,}072) \cdot (1+0{,}09) \cdot (1+0{,}074) \cdot (1+0{,}045) = 13.577{,}13$$

# Media geometrica

Il coefficiente medio di incremento è quel valore che applicato per tutti i 6 anni permette di ottenere il medesimo ammontare finale; si tratta quindi di determinare la media geometrica degli incrementi osservati:

$$\sqrt[6]{(1{,}015)(1{,}02)(1{,}072)(1{,}09)(1{,}074)(1{,}045)} = 1{,}05229$$

Quindi il tasso di interesse percentuale medio annuo è del 5,23 %. Infatti si ha:

$$10.000 \cdot (1 + 0{,}05229)^6 = 13.577{,}28$$

# Trimmed mean

- Maggior difetto della media aritmetica (e media geometrica) è che risente fortemente dei valori estremi
- Es. 3, 5, 5, 6, 8, 8, 9, 150
  - Media aritmetica = 24,25 (primi 7 valori più piccoli della media)
  - Media geometrica= 8,89

# Trimmed mean

- La trimmed mean al 50% di un carattere quantitativo è la media aritmetica del 50% dei valori più centrali di un insieme di osservazioni. Ovvero nel calcolo della media non vengono considerati il 25% dei valori più piccoli ed il 25% dei valori più grandi.

- Es. 3, 5, 5, 6, 8, 8, 9, 150

# Trimmed mean

- Es. 3, 5,5,6,8,8,9,150

$$\frac{5+6+8+8}{4} = 6{,}75$$

# Trimmed mean

- Es. 3, 5,5,6,8,8,9,150 $\frac{5+6+8+8}{4} = 6{,}75$

- Osservare che sono stata esclusi anche 3,5,9 che non sono "valori estremi"

- Si può anche utilizzare altra percentuale diversa dal 50%

# Mediana

- Media aritmetica e media geometrica sono molto sensibili a valori estremi
- Solo per caratteri quantitativi

- La **mediana** è meno sensibile a valori estremi e può essere utilizzata anche per dati qualitativi ordinabili
- La mediana è una media di posizione

# Mediana

- La **mediana** ($M_e$) di un insieme di unità ordinate (secondo un carattere ordinabile) è la modalità presentata dall'unità centrale
  - **Unità centrale**= unità che divide il collettivo in due parti di uguale numerosità:
    - una parte formata dalle unità che presentano una modalità **precedente o uguale** a quella dell'unità centrale
    - una parte formata dalle unità che presentano una modalità **successiva o uguale** a quella dell'unità centrale

# Mediana

- A 7 individui viene chiesto di esprimere un giudizio su una marca di yogurt:

| Individuo | Giudizio |
|---|---|
| 1 | insufficiente |
| 2 | sufficiente |
| 3 | buono |
| 4 | sufficiente |
| 5 | discreto |
| 6 | ottimo |
| 7 | ottimo |

- Ordinare e determinare mediana

# Mediana

| Ordine crescente | |
|---|---|
| **Individuo** | **Giudizio** |
| 1 | insufficiente |
| 4 | sufficiente |
| 2 | sufficiente |
| 5 | discreto |
| 3 | buono |
| 6 | ottimo |
| 7 | ottimo |

# Mediana

| Ordine crescente | |
|---|---|
| **Individuo** | **Giudizio** |
| 1 | insufficiente |
| 4 | sufficiente |
| 2 | sufficiente |
| 5 | discreto |
| 3 | buono |
| 6 | ottimo |
| 7 | ottimo |

- Possiamo ordinare anche in ordine decrescente

| Ordine decrescente | |
|---|---|
| **Individuo** | **Giudizio** |
| 7 | ottimo |
| 6 | ottimo |
| 3 | buono |
| 5 | discreto |
| 2 | sufficiente |
| 4 | sufficiente |
| 1 | insufficiente |

# Mediana

- Si ordinano le unità in maniera crescente rispetto alle modalità del carattere
  - se n è dispari, la posizione dell'unità centrale è $\frac{n+1}{2}$, pertanto:

$$M_e = x_{\frac{n+1}{2}}$$

  - se n è pari: ci sono due unità centrali $x_{\frac{n}{2}}, \quad x_{\frac{n}{2}+1}$
    - se il carattere è quantitativo:

$$M_e = \frac{1}{2}\left(x_{\frac{n}{2}} + x_{\frac{n}{2}+1}\right)$$

# Mediana

- Il proprietario di 5 negozi di abbigliamento conteggia il numero di capi venduti in un giorno da ognuno dei negozi:
  - 18, 27, 15, 11, 6
- Calcolare la mediana

# Mediana

- Il proprietario di 5 negozi di abbigliamento conteggia il numero di capi venduti in un giorno da ognuno dei negozi:
  - 18, 27, 15, 11, 6
- Calcolare la mediana

| Capi venduti |
| ---: |
| 6 |
| 11 |
| 15 |
| 18 |
| 27 |

# Mediana

- Il proprietario di 6 negozi di abbigliamento conteggia il numero di capi venduti in un giorno da ognuno dei negozi:
  - 18, 27, 15, 11, 6, 20
- Calcolare la mediana

# Mediana

- Il proprietario di 6 negozi di abbigliamento conteggia il numero di capi venduti in un giorno da ognuno dei negozi:
  - 18, 27, 15, 11, 6, 20
- Calcolare la mediana

| Capi venduti |
| ---: |
| 6 |
| 11 |
| 15 |
| 18 |
| 20 |
| 27 |

$$M_e = \frac{15+18}{2} = 16{,}5$$

# Mediana

## Sensibilità delle medie ai valori anomali

Consideriamo ora una distribuzione in cui è presente un valore anomalo e vediamo come la media aritmetica e la mediana vengono influenzate da questo valore. Da 10 cartelle mediche di un pediatra rileviamo l'età dei corrispondenti pazienti: 5, 6, 3, 7, 3, 6, 8, 2, 5, 60. È probabile che l'età dell'ultimo paziente sia frutto di un errore di registrazione. L'età media, pari a 10,5, è chiaramente poco rappresentativa, giacché nove pazienti su dieci risultano avere un'età inferiore a questa. Per calcolare la mediana bisogna prima ordinare i dati in ordine crescente:

# Mediana

**Sensibilità delle medie ai valori anomali**

- Consideriamo ora una distribuzione in cui è presente un valore anomalo e vediamo come la media aritmetica e la mediana vengono influenzate da questo valore. Da 10 cartelle mediche di un pediatra rileviamo l'età dei corrispondenti pazienti: 5, 6, 3, 7, 3, 6, 8, 2, 5, 60. È probabile che l'età dell'ultimo paziente sia frutto di un errore di registrazione. L'età media, pari a 10,5, è chiaramente poco rappresentativa, giacché nove pazienti su dieci risultano avere un'età inferiore a questa. Per calcolare la mediana bisogna prima ordinare i dati in ordine crescente:

| Età | 2 | 3 | 3 | 5 | 5 | 6 | 6 | 7 | 8 | 60 |
|---|---|---|---|---|---|---|---|---|---|---|

$$M_e = \frac{5+6}{2} = 5{,}5$$

- La mediana è più rappresentativa della media e non è influenzata dalla presenza del valore anomalo.

# Moda

- La moda è una media di posizione
- Può essere calcolata per qualunque tipo di carattere
  - anche per i qualitativi sconnessi
- La moda è la modalità della distribuzione che si presenta con la **massima frequenza** (assoluta, relativa o percentuale)

# Moda

- Popolazione italiana residente in età per grado di istruzione al censimento del 1981

| **Titolo di studio** | **Freq. %** |
|---|---:|
| Analfabeti | 3,1 |
| Alfabeti | 18,2 |
| Licenza elementare | 40,6 |
| Licenza media | 23,8 |
| Diploma | 11,5 |
| Laurea | 2,8 |

# Moda

- Popolazione italiana residente in età per grado di istruzione al censimento del 1981

| **Titolo di studio** | **Freq. %** |
|---|---|
| Analfabeti | 3,1 |
| Alfabeti | 18,2 |
| Licenza elementare | 40,6 |
| Licenza media | 23,8 |
| Diploma | 11,5 |
| Laurea | 2,8 |

# Moda

## Difetti della moda

Supponiamo di osservare la seguente distribuzione:

| *X* | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | **Totale** |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Freq.** | 4 | 2 | 2 | 3 | 2 | 3 | 3 | 3 | 3 | 2 | 27 |

La moda di questa distribuzione è 1 a cui corrisponde una frequenza del 15%. Se ora calcoliamo sulla medesima distribuzione anche la media aritmetica e la mediana otteniamo per la prima un valore di 5,41 e per la seconda un valore di 6; questi valori sono molto distanti da quello relativo alla moda, ma certamente riescono a sintetizzare molto meglio i valori della distribuzione.

# Sintesi delle medie

**TABELLA 3.6.1** Uso delle medie in base al tipo di carattere

| | Caratteri | | |
|---|---|---|---|
| | **Quantitativi** | **Qualitativi ordinati** | **Qualitativi sconnessi** |
| Media aritmetica | • | – | – |
| Media geometrica | • | – | – |
| Trimmed mean | • | – | – |
| Mediana | • | • | – |
| Moda | • | • | • |

# Variabilità

- Una distribuzione è ben rappresentata da una media quando la gran parte delle unità presentano una modalità vicino alla media
- Un **indice di variabilità** deve
  - assumere il suo valore minimo se e solo se le unità della distribuzione presentano uguale modalità del carattere
  - aumentare all'aumentare della diversità tra le modalità assunte dalle varie unità

# Deviazione standard

- Per misurare la variabilità di una distribuzione rispetto alla medi aritmetica
- La **deviazione standard** di un insieme di *n* valori osservati $x_1, x_2, \ldots, x_n$ di un carattere quantitativo *X* con media aritmetica $\bar{x}$ è pari a:

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

# Deviazione standard

- Se conosciamo le distribuzione di frequenza di una variabile $X$ con $k$ modalità:

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

  - $k$=numero di modalità assunte dal carattere
  - $n_j$ = frequenza assoluta della modalità $x_j$
  - $f_j$ = frequenza relativa della modalità $x_j$

# Scostamento semplice dalla media aritmetica

- Per misurare la variabilità di una distribuzione rispetto alla medi aritmetica
- **Lo scostamento semplice medio dalla media aritmetica** è:

$$S_{\bar{x}} = \frac{1}{n} \sum_{i=1}^{n} | x_i - \bar{x} |$$

# Scostamento semplice dalla media aritmetica

- Se conosciamo le distribuzione di frequenza di una variabile *X* con *k* modalità:

$$S_{\bar{x}} = \frac{1}{n}\sum_{i=1}^{k} |x_i - \bar{x}| n_i = \sum_{i=1}^{k} |x_i - \bar{x}| f_i$$

  - $k$=numero di modalità assunte dal carattere
  - $n_j$ = frequenza assoluta della modalità $x_j$
  - $f_j$ = frequenza relativa della modalità $x_j$

# Scostamento semplice dalla mediana

- Per misurare la variabilità di una distribuzione rispetto alla mediana
- **Lo scostamento semplice medio dalla mediana** è:

$$S_{M_e} = \frac{1}{n}\sum_{i=1}^{n} | x_i - M_e |$$

- Può essere vantaggioso, rispetto alla deviazione standard, quando si sospetta la presenza di valori anomali
  - La mediana è più robusta
  - Gli scarti dei valori anomali non vengono esaltati dall'elevamento al quadrato

# Scostamento semplice dalla mediana

- Se conosciamo le distribuzione di frequenza di una variabile $X$ con $k$ modalità:

$$S_{M_e} = \frac{1}{n}\sum_{i=1}^{k} |x_i - M_e| n_i = \sum_{i=1}^{k} |x_i - M_e| f_i$$

  - $k$=numero di modalità assunte dal carattere
  - $n_j$ = frequenza assoluta della modalità $x_j$
  - $f_j$ = frequenza relativa della modalità $x_j$

# Variabilità

- La seguente tabella riporta il tempo (in giorni) impiegato da sei individui per il consumo di una confezione di pasta da 250 grammi:

| Individuo | N. giorni |
|---|---|
| 1 | 1 |
| 2 | 6 |
| 3 | 5 |
| 4 | 3 |
| 5 | 30 |
| 6 | 15 |

- Calcolare deviazione standard, scostamento semplice dalla media aritmetica, scostamento semplice dalla mediana

# Variabilità

- Media aritmetica

$$\bar{x} = \frac{1+6+5+3+30+15}{6} = 10$$

- Mediana

| Individuo | N. giorni |
|---|---|
| 1 | 1 |
| 4 | 3 |
| 3 | 5 |
| 2 | 6 |
| 6 | 15 |
| 5 | 30 |

$$M_e = \frac{5+6}{2} = 5{,}5$$

# Variabilità

| $x_i$ | $\lvert x_i - Me \rvert$ | $\lvert x_i - \bar{x} \rvert$ | $(x_i - \bar{x})^2$ |
|---|---|---|---|
| 1 | 4.5 | 9 | 81 |
| 3 | 2.5 | 7 | 49 |
| 5 | 0.5 | 5 | 25 |
| 6 | 0.5 | 4 | 16 |
| 15 | 9.5 | 5 | 25 |
| 30 | 24.5 | 20 | 400 |
| *Totale* | *42* | *50* | *596* |

# Scostamento semplice dalla mediana

| $x_i$ | $\|x_i - Me\|$ | $\|x_i - \overline{x}\|$ | $(x_i - \overline{x})^2$ |
|---|---|---|---|
| 1 | 4.5 | 9 | 81 |
| 3 | 2.5 | 7 | 49 |
| 5 | 0.5 | 5 | 25 |
| 6 | 0.5 | 4 | 16 |
| 15 | 9.5 | 5 | 25 |
| 30 | 24.5 | 20 | 400 |
| *Totale* | *42* | *50* | *596* |

$$S_{M_e} = \frac{1}{6}\sum_{i=1}^{6}|x_i - M_e| = \frac{42}{6} = 7$$

mediamente le durate del pacchetto di pasta differiscono (si discostano) dalla durata mediana di 7 giorni.

# Scostamento semplice dalla media aritmetica

| $x_i$ | $\|x_i - Me\|$ | $\|x_i - \bar{x}\|$ | $(x_i - \bar{x})^2$ |
|---|---|---|---|
| 1 | 4.5 | 9 | 81 |
| 3 | 2.5 | 7 | 49 |
| 5 | 0.5 | 5 | 25 |
| 6 | 0.5 | 4 | 16 |
| 15 | 9.5 | 5 | 25 |
| 30 | 24.5 | 20 | 400 |
| *Totale* | *42* | *50* | *596* |

$$S_{\bar{x}} = \frac{1}{6}\sum_{i=1}^{6} |x_i - \bar{x}| = \frac{50}{6} = 8,333$$

mediamente le durate del pacchetto di pasta differiscono (si discostano) dalla durata media di 8,333 giorni.

# Deviazione standard

| $x_i$ | $\lvert x_i - Me \rvert$ | $\lvert x_i - \bar{x} \rvert$ | $(x_i - \bar{x})^2$ |
|---|---|---|---|
| 1 | 4.5 | 9 | 81 |
| 3 | 2.5 | 7 | 49 |
| 5 | 0.5 | 5 | 25 |
| 6 | 0.5 | 4 | 16 |
| 15 | 9.5 | 5 | 25 |
| 30 | 24.5 | 20 | 400 |
| *Totale* | *42* | *50* | *596* |

$$\sigma = \sqrt{\frac{1}{6}\sum_{i=1}^{6}(x_i - \bar{x})^2} = \sqrt{\frac{596}{6}} = 9,967$$

mediamente le durate del pacchetto di pasta differiscono (si discostano) dalla durata media di 9,967 giorni.

# Confronti di variabilità

- Deviazione standard e scostamenti semplici (sia dalla media aritmetica che dalla mediana) dipendono
  - dall'unità di misura
  - dall'ordine di grandezza dei dati (anche se con stessa unità di misura)

- Non consentono di eseguire confronti tra la variabilità di fenomeni con diverse unità di misura e ordini di grandezza diversi

# Coefficienti di variazione

- Per deviazione standard

$$\frac{\sigma}{\bar{x}} \cdot 100$$

- Per scostamento semplice dalla media aritmetica

$$\frac{S_{\bar{x}}}{\bar{x}} \cdot 100$$

- Per scostamento semplice dalla mediana

$$\frac{S_{M_e}}{M_e} \cdot 100$$

# Coefficienti di variazione

Consideriamo la variabile "quantità di cenere inquinante" che fuoriesce dalla ciminiera di un certo tipo di industria. In una regione si hanno nove industrie che hanno installato un dispositivo anti-inquinante di tipo A e altre nove che hanno installato un dispositivo di tipo B.

Quantità di pulviscolo eliminate dai dispositivi A e B

| Tipo | Quantità di pulviscolo (g/min) | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| A | 69 | 80 | 44 | 52 | 54 | 54 | 86 | 77 | 66 |
| B | 35 | 62 | 43 | 23 | 30 | 28 | 22 | 40 | 25 |

$$\bar{x}_A = 64{,}67 \qquad \bar{x}_B = 34{,}22$$

$$\sigma_A = 13{,}65 \qquad \sigma_B = 12{,}02$$

$$\frac{\sigma_A}{\bar{x}_A} \cdot 100 = 21{,}11\%$$

$$\frac{\sigma_B}{\bar{x}_B} \cdot 100 = 35{,}11\%$$