

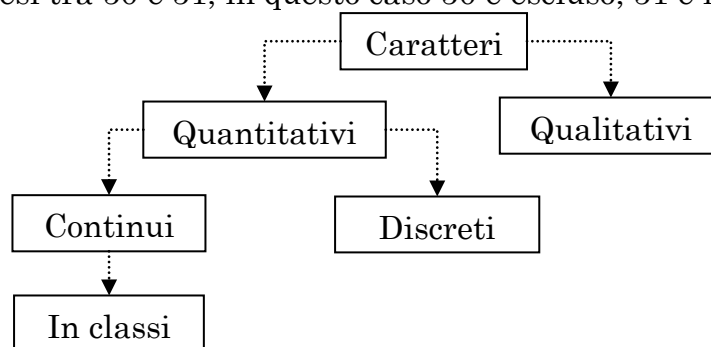
MAPPA 1: STATISTICA DESCRITTIVA

SETTEMBRE 2008

FRANCA GRESSINI

APPUNTI DI STATISTICA

- La **statistica** è la scienza che si occupa della raccolta e della elaborazione di dati.
- La **popolazione** è l'insieme degli individui o degli elementi sui quali si vuole indagare (insieme delle persone che vivono in Italia, le aziende agrarie del Lazio, le lampadine prodotte da una certa azienda, le precipitazioni atmosferiche giornaliere in una certa località, le autovetture in circolazione in Italia, le estrazioni di numeri nel gioco del Lotto).
- L'**unità statistica** è ciascun elemento della popolazione statistica.
- Il **carattere** di una indagine statistica è il tipo di dati sul quale l'indagine si effettua: altezza, numeri di figli, titolo di studio, stato civile, fatturato.
- Un **carattere qualitativo** esprime una qualità, cioè valori non numerici come il mezzo di trasporto usato, il titolo di studio, il colore degli occhi. Può essere ordinabile (titolo di studio) o non ordinabile (colore degli occhi).
- Un **carattere quantitativo** esprime una quantità, cioè un valore numerico: numero di figli, età, altezza, peso, fatturato di un'azienda.
- I dati quantitativi si distinguono in **discreti**, se il valore numerico può essere espresso solo da un numero intero (numero di scarpa, numero di figli, anni di età) o **continui**, se il valore numerico può variare anche di piccole quantità (peso, altezza, superficie di un appartamento).
- Se i dati sono continui è opportuno organizzarli in **classi** cioè intervalli di valori tra i quali oscilla un gruppo (peso compreso tra 50 e 51 Kg, si può indicare come $50 < p \leq 51$, raccoglie tutti i valori compresi tra 50 e 51, in questo caso 50 è escluso, 51 è incluso).



- La **rilevazione** dei dati statistici consiste nel raccogliere le informazioni relative a dei caratteri oggetto di indagine e nel raggruppare i dati ottenuti.
- Il **censimento** è una indagine statistica che riguarda tutti gli elementi della popolazione
- Il **rilevamento per campione** riguarda un sottoinsieme della popolazione scelto in maniera opportuna in modo che sia rappresentativo dell'intera popolazione.
- La **frequenza assoluta** è il numero di volte con cui si rileva lo stesso valore; la **frequenza relativa** si ottiene facendo il rapporto tra la frequenza assoluta e il numero totale delle rilevazioni, può variare da 0 a 1; la **frequenza percentuale** si ottiene moltiplicando per 100 la frequenza relativa e aggiungendo il simbolo di percentuale %; le **frequenze cumulate** si ottengono sommando la frequenza percentuale di ciascun valore a quelle dei precedenti.
- La **moda** di un insieme di dati è il valore, o classe di valori, che si presenta con maggiore frequenza.
- La **media aritmetica** si ottiene sommando tutti i dati e dividendo il risultato per il numero di dati. La **media aritmetica ponderata** si ottiene moltiplicando i dati per il numero di volte in cui compaiono, si sommano tutti i prodotti ottenuti e si divide il risultato per il numero totale di dati.
- La **mediana** si ottiene ordinando i dati in modo crescente (o decrescente) e prendendo il valore che occupa il posto centrale, nel caso in cui i dati siano in numero pari avremmo due posizioni al centro e per trovare la mediana sommiamo i due valori centrali e dividiamo per due il risultato.

- Il **campo di variazione** di un insieme di dati statistici è la differenza tra il valore più grande e quello più piccolo.

Esempio. Uno studente ha preso i seguenti voti in pagella: 6, 6, 5, 7, 5, 6, 8, 4. La media

aritmetica è data da $\frac{6+6+5+7+5+6+8+4}{8} = 5,875$. La stessa media si può calcolare come

media aritmetica ponderata $\frac{3 \times 6 + 2 \times 5 + 7 + 8 + 4}{8} = \frac{18 + 10 + 7 + 8 + 4}{8} = 5,875$.

I grafici ci aiutano a interpretare i dati più rapidamente rispetto a una tabella.

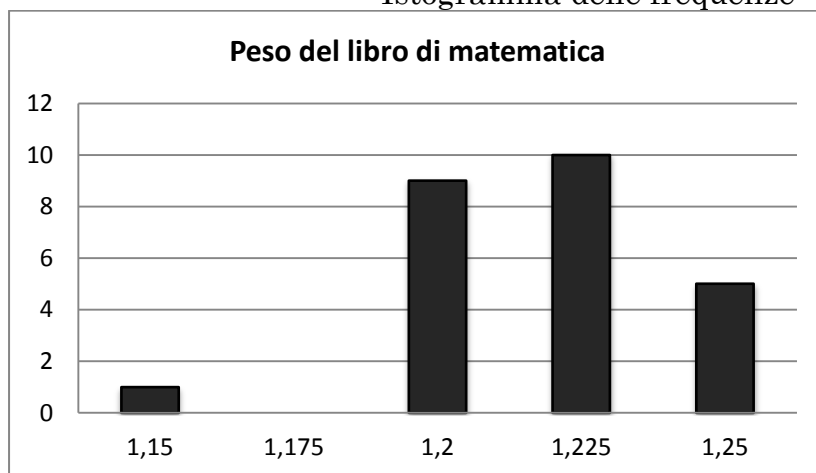
- **Istogramma** è un grafico costituito da rettangoli di uguale base ma di altezza variabile proporzionale alla frequenza assoluta.
- **Areogramma** è un grafico costituito da un cerchio suddiviso in settori circolari la cui ampiezza è proporzionale alla frequenza con cui il valore corrispondente compare;
- **Ideogramma** è un grafico in cui si utilizzano immagini stilizzate che richiamano la natura dei dati di cui ci si occupa (es. spighe di grano per la produzione di frumento, damigiane per la produzione di vino) la grandezza del disegno o il numero di volte che viene rappresentato è proporzionale alla frequenza con cui è stato rilevato il carattere oggetto di indagine.
- **Cartogramma** è una rappresentazione grafica che fa uso di carte geografiche sulle quali vengono riportati i dati rilevati dall'indagine.
- **Diagramma cartesiano** è un grafico in cui i punti hanno come coordinata orizzontale (ascissa) il valore rilevato e come coordinata verticale (ordinata) la sua frequenza assoluta; questi punti si possono unire a formare la **poligonale delle frequenze** (una linea spezzata), nel caso in cui le rilevazioni sono molto fitte e numerose la poligonale assume una forma sempre più arrotondata e nel caso in cui l'andamento della variabile fosse normale, si avrebbe una **forma a campana** detta anche **gaussiana**.
- **ISTAT** è l'Istituto nazionale di statistica un ente di ricerca pubblico italiano che effettua censimenti e indagini campionarie sulle famiglie e sulla società italiana.

Esempio

25 ragazzi di una classe hanno provato a pesare il libro di matematica ottenendo i seguenti risultati in Kg: 1,225; 1,200; 1,250; 1,200; 1,225; 1,225; 1,200; 1,150; 1,200; 1,225; 1,250; 1,250; 1,225; 1,200; 1,250; 1,250; 1,225; 1,225; 1,200; 1,200; 1,225; 1,225; 1,200; 1,225; 1,200

Peso	Frequenza assoluta	Frequenza relativa	Frequenza Percentuale	Frequenze cumulate	
1,150	1	1/25	4%	4%	Moda 1,225 Media aritmetica 1,218 Mediana 1,225 Campo di variazione 0,100
1,175	0	0/25	0%	4%	
1,200	9	9/25	36%	40%	
1,225	10	10/25	40%	80%	
1,250	5	5/25	20%	100%	

Istogramma delle frequenze



LA STATISTICA

1. LE FREQUENZE

DEFINIZIONE

Frequenza (o frequenza assoluta)

La frequenza è il numero delle volte in cui un dato si presenta.

DEFINIZIONE

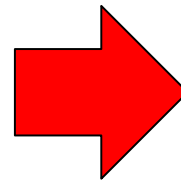
Frequenza relativa

La frequenza relativa di una modalità è il rapporto fra la frequenza delle modalità e il numero totale delle unità statistiche.

frequenza relativa $f = \frac{F}{T}$

frequenza F

totale delle unità statistiche T



$$F = f \cdot T$$

1. LE FREQUENZE

ESEMPIO

Distribuzione delle frequenze relative			
Modalità	Frequenza	Frequenza relativa	Frequenza relativa percentuale
sport	6	$3/13$	23%
amici	8	$4/13$	31%
cinema, TV	5	$5/26$	19%
hobby	4	$2/13$	15%
altre attività	3	$3/26$	12%
Totale	26	1	100%

2. GLI INDICI DI POSIZIONE CENTRALE

DEFINIZIONE

Media aritmetica

La media aritmetica M di n numeri
è il quoziente fra la loro
somma e il numero n .

$$M = \frac{x_1 + x_2 + \dots + x_n}{n}$$

media aritmetica

somma dei valori

numero dei valori

DEFINIZIONE

Media aritmetica ponderata

Dati i numeri e associati a essi i
numeri detti **pesi**, chiamiamo
media aritmetica ponderata P il rapporto
fra la somma dei prodotti dei numeri per
i loro pesi e la somma dei pesi stessi.

$$P = \frac{x_1 p_1 + x_2 p_2 + \dots + x_n p_n}{p_1 + p_2 + \dots + p_n}$$

media aritmetica ponderata

somma dei prodotti dei valori per i loro pesi

somma dei pesi

2. GLI INDICI DI POSIZIONE CENTRALE

ESERCIZIO GUIDA

Un grossista di frutta acquista quattro quantitativi di mele Golden presso aziende agricole diverse che hanno praticato prezzi differenti. La seguente tabella espone i prezzi e le relative quantità e si vuole determinare il prezzo medio al kg.

	Azienda A	Azienda B	Azienda C	Azienda D
Prezzo (€)	0,60	0,55	0,68	0,57
Quantità (kg)	200	300	220	280

Dobbiamo calcolare una media aritmetica ponderata dei prezzi dove i pesi sono le quantità.

$$M = \frac{0,60 \cdot 200 + 0,55 \cdot 300 + 0,68 \cdot 220 + 0,57 \cdot 280}{200 + 300 + 220 + 280} \simeq 0,59.$$

2. GLI INDICI DI POSIZIONE CENTRALE

DEFINIZIONE

Media geometrica

La media geometrica M di n numeri è la radice n -esima del prodotto degli n numeri.

$$G = \sqrt[n]{x_1 \cdot x_2 \cdot \dots \cdot x_n}$$

numero dei valori n
 media geometrica G
 prodotto dei valori $x_1 \cdot x_2 \cdot \dots \cdot x_n$

DEFINIZIONE

Media armonica

La media armonica A di n numeri è il reciproco della media aritmetica dei reciproci dei valori.

$$A = \frac{1}{\frac{1}{x_1} + \frac{1}{x_2} + \dots + \frac{1}{x_n}}$$

media armonica A
 numero dei valori n
 somma dei reciproci dei valori $\frac{1}{x_1} + \frac{1}{x_2} + \dots + \frac{1}{x_n}$

2. GLI INDICI DI POSIZIONE CENTRALE

ESERCIZIO GUIDA

Un capitale è stato investito con i seguenti rendimenti: 7% il primo anno; 5% sia per il secondo sia per il terzo anno; 9% il quarto anno; 8% sia per il quinto sia per il sesto e il settimo anno.

Calcoliamo il rendimento percentuale medio.

Dobbiamo calcolare la media geometrica utilizzando il fattore $(1 + \text{tasso})$ che permette di passare dal valore di un anno a quello dell'anno successivo. Per esempio, poiché il tasso per il primo anno è 7%, ossia 0,07, il primo fattore nella media è $1 + 0,07 = 1,07$:

$$G = \sqrt[7]{(1,07)(1,05)(1,05)(1,09)(1,08)(1,08)(1,08)} = \sqrt[7]{(1,07)(1,05)^2(1,09)(1,08)^3} \simeq 1,0713.$$

Quindi il rendimento medio percentuale è stato del 7,13%.

2. GLI INDICI DI POSIZIONE CENTRALE

ESERCIZIO GUIDA

Un ciclista percorre prima 30 km alla velocità di 25 km/h e successivamente altri 45 km alla velocità di 20 km/h. Calcoliamo la velocità media v_m .

Suddividiamo il percorso di 75 km in 5 tratti uguali ciascuno di 15 km e assegniamo a ogni tratto la sua velocità. Abbiamo i seguenti cinque valori: 25, 25, 20, 20, 20. Calcoliamo la media armonica:

$$A = \frac{5}{\frac{1}{25} + \frac{1}{25} + \frac{1}{20} + \frac{1}{20} + \frac{1}{20}} = \frac{5}{\frac{1}{25} \cdot 2 + \frac{1}{20} \cdot 3} \simeq 21,739 \rightarrow v_m = 21,739 \text{ km/h.}$$

Se moltiplichiamo numeratore e denominatore della formula per 15 otteniamo:

$$A = \frac{75}{\frac{1}{25} \cdot 30 + \frac{1}{20} \cdot 45} \simeq 21,739 \rightarrow v_m = 21,739 \text{ km/h,}$$

che rappresenta la media armonica ponderata dei valori 25 km/h e 20 km/h, ciascuno considerato con il suo «peso».

Osserviamo che il numeratore è la lunghezza totale del percorso e il denominatore è la somma dei tempi impiegati per percorrere i due tratti di strada.

2. GLI INDICI DI POSIZIONE CENTRALE

DEFINIZIONE

Media quadratica

La media quadratica Q di n numeri è la radice quadrata della media aritmetica dei quadrati dei numeri.

$$Q = \sqrt{\frac{x_1^2 + x_2^2 + \dots + x_n^2}{n}}$$

2. GLI INDICI DI POSIZIONE CENTRALE

ESERCIZIO GUIDA

Un orefice ha a disposizione 7 medaglie d'oro, di uguale spessore, da fondere per ricavare altre 7 medaglie, uguali tra loro, dello stesso spessore di quelle fuse. Sappiamo che tre delle medaglie da fondere hanno diametro uguale a 12 mm, due medaglie hanno diametro uguale a 14 mm, una ha diametro di 15 mm e l'ultima di 17 mm. Calcoliamo quale deve essere il diametro delle nuove medaglie.

Possiamo risolvere il problema calcolando prima la superficie totale delle 7 medaglie da fondere. Appliciamo la formula dell'area del cerchio $A = r^2\pi$:

$$S = (6^2\pi) \cdot 3 + (7^2\pi) \cdot 2 + (7,5^2\pi) + (8,5^2\pi) = 334,5\pi \text{ mm}^2.$$

La superficie di ogni medaglia da realizzare è:

$$S = \frac{334,5\pi}{7} \simeq 47,79\pi \text{ mm}^2.$$

Applicando la formula inversa, determiniamo il raggio e quindi il diametro:

$$r = \sqrt{\frac{47,79\pi}{\pi}} \simeq 6,91 \text{ mm}, \quad \text{quindi } 2r \simeq 13,82 \text{ mm}.$$

Possiamo ottenere direttamente il valore del diametro calcolando la media quadratica dei diametri delle medaglie:

$$\begin{aligned} Q &= \sqrt{\frac{12^2 + 12^2 + 12^2 + 14^2 + 14^2 + 15^2 + 17^2}{7}} = \\ &= \sqrt{\frac{12^2 \cdot 3 + 14^2 \cdot 2 + 15^2 + 17^2}{7}} = \quad (\text{media quadratica ponderata}) \\ &= \sqrt{\frac{1338}{7}} \simeq 13,82 \text{ mm}. \end{aligned}$$

2. GLI INDICI DI POSIZIONE CENTRALE

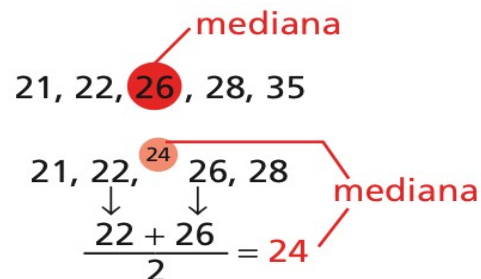
DEFINIZIONE

Mediana

Data la sequenza ordinata di n numeri , la mediana è:

- il valore centrale, se n è dispari;
- la media aritmetica dei due valori centrali, se n è pari.

ESEMPIO



DEFINIZIONE

Moda

Dati i numeri , si chiama moda il valore a cui corrisponde la frequenza massima.

ESEMPIO

