

Elementi di Statistica



Cosa studia la Statistica?

- * La statistica si interessa del rilevamento, dell'elaborazione e dello studio dei dati.
- * Il termine statistica ricorda proprio l'origine di questa scienza nata come strumento per ben governare lo stato.
- * La statistica studia ciò che accade o come è fatto un gruppo numeroso di oggetti.
- * Cerca, attraverso l'uso della matematica e della geometria, di rappresentare e di dare informazioni semplici su un gruppo numeroso ed eterogeneo di oggetti.

La popolazione

Nelle indagini statistiche si prendono in esame fenomeni di vario tipo (per esempio, demografici, economici, sociali ecc.) che riguardano delle popolazioni, ossia insiemi di elementi che presentano tutti delle caratteristiche comuni.

ad esempio, l'insieme delle persone, delle abitazioni o delle aziende oggetto di un censimento

Gli elementi di una data popolazione si chiamano individui o unità statistiche.

ad esempio, le possibili cinque estraibili nel gioco del superenalotto

Si distingue anche tra:

- * popolazione reale o empirica: un insieme le cui unità possono essere tutte concretamente osservate ;
- * popolazione virtuale o teorica: un insieme definibile con accuratezza ma non concretamente osservabile .

Esempi

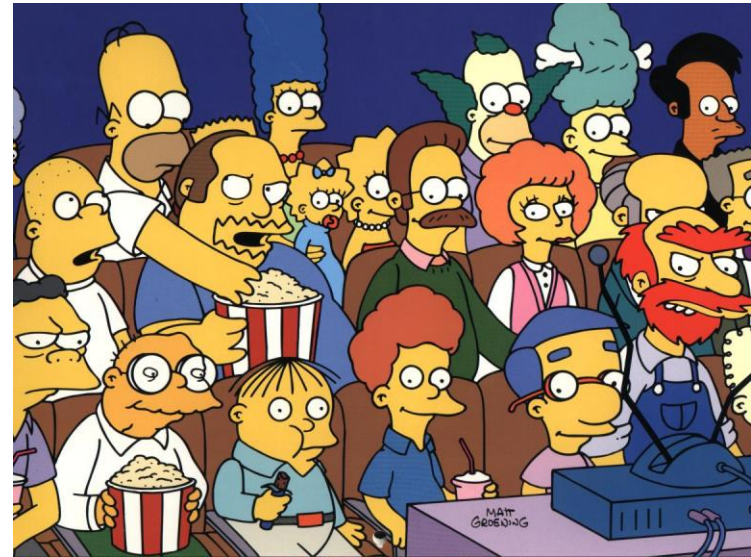
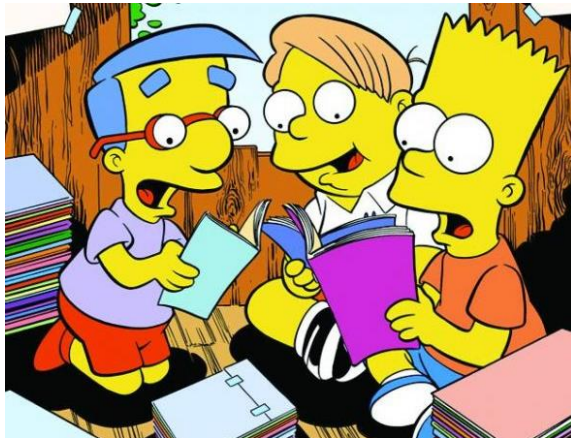


- L'insieme delle persone che in questo istante vivono in Italia (popolazione in senso demografico);
- I cittadini che hanno diritto al voto nelle elezioni per il Parlamento;
- Le aziende agrarie della Lombardia;
- Le pile elettriche di una certa ditta;
- Le precipitazioni atmosferiche giornaliere in una certa località e misurate in una stazione meteorologica;
- Le autovetture in circolazione attualmente in Italia;
- I lanci di una moneta durante un certo intervallo di tempo.

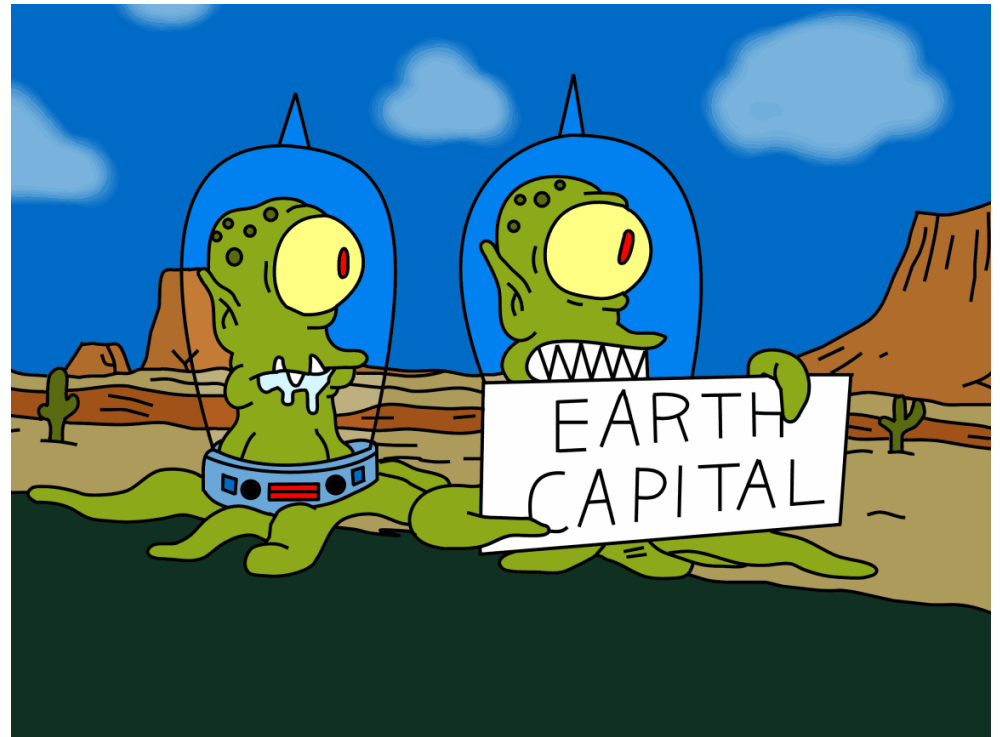
La statistica studia le caratteristiche di una popolazione o di alcuni suoi gruppi



La statistica studia ad esempio cosa piace fare alle persone



La statistica studia quanto sono diverse le persone



Caratteri di una popolazione

I caratteri di una popolazione statistica che sono oggetto di un'indagine statistica possono essere:

Qualitativi

il colore, la nazionalità, la religione, lo stato civile, l'affidabilità, l'attitudine agli studi ecc.

In tal caso, le modalità di un carattere qualitativo saranno espresse da aggettivi.

Per esempio:

- il carattere colore degli occhi ha le modalità celesti, grigi, neri, . . . ;
- il carattere stato civile ha le modalità celibe, nubile, coniugato;
- il carattere titolo di studio ha le modalità laurea, diploma, . . .

quantitativi

la statura, il peso, il numero di stanze di un appartamento, ecc.

Le modalità di un carattere quantitativo saranno, allora, espresse da numeri, che si chiamano anche valori del carattere

Le frequenze

DEFINIZIONE

Frequenza (o frequenza assoluta)

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

DEFINIZIONE

Frequenza relativa

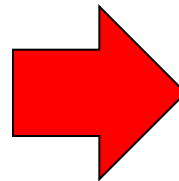
La frequenza relativa di una modalità è il rapporto fra la frequenza delle modalità e il numero totale delle unità statistiche.

frequenza
relativa

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

frequenza

totale delle
unità statistiche



$$F = f \cdot T$$

Le frequenze

ESEMPIO

Abbiamo chiesto, in un questionario, a 26 studenti di una classe di indicare con una delle seguenti lettere l'attività a cui la maggior parte del tempo libero:

S sport; A amici; C cinema; H hobby; N altre attività.

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%

Dire che la frequenza associata a sport è 6, significa che 6 studenti hanno risposto di praticare lo sport. Dire che la frequenza relativa è $3/13$ vuol dire che 6 su 26 alunni praticano lo sport. Moltiplicando per 100 si ottiene la frequenza relativa percentuale

Esempio

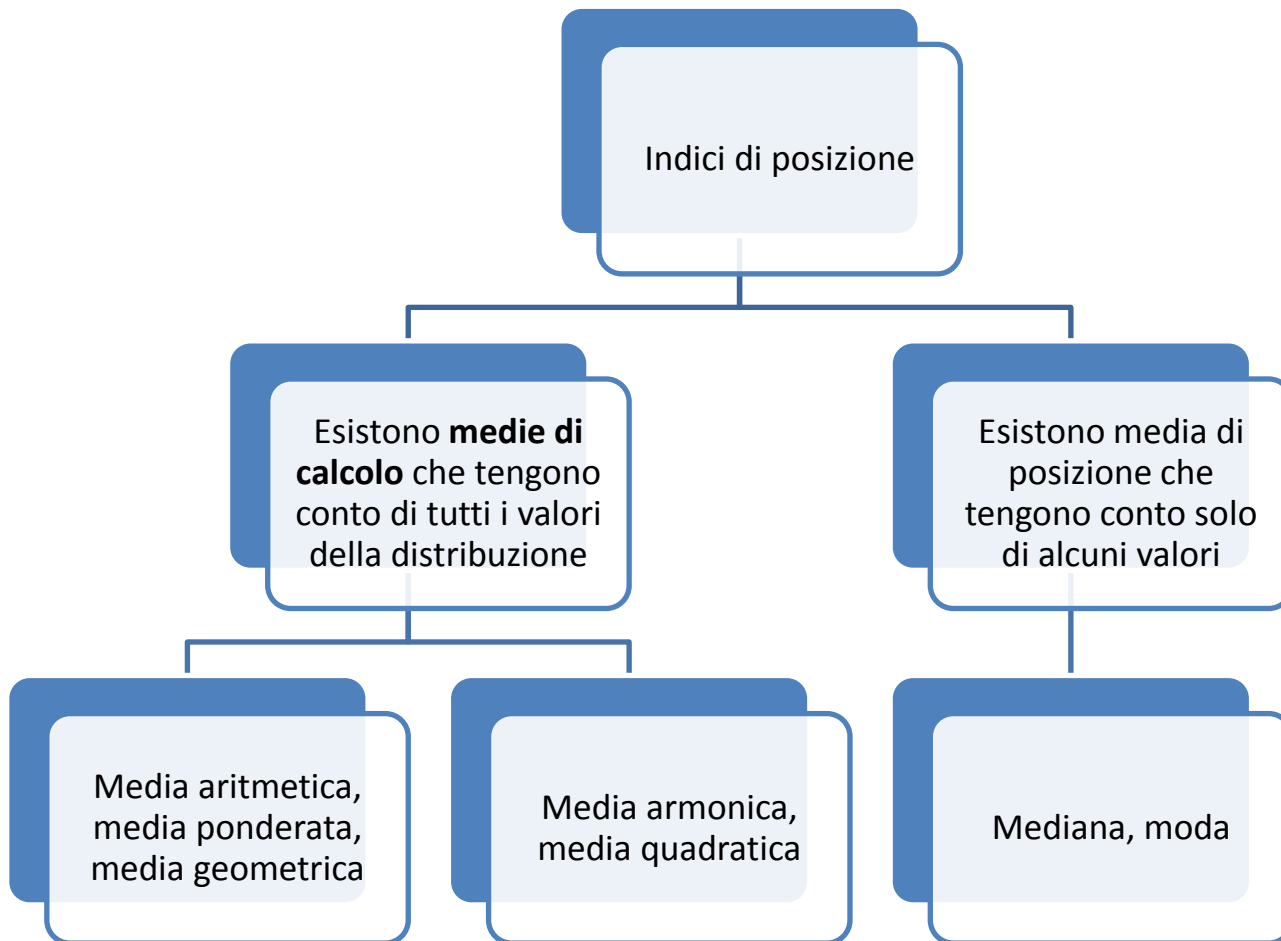
In un sondaggio fatto all'interno di una facoltà composta da 250 studenti (la popolazione statistica), si intende rilevare il carattere "gradimento dei professori", secondo le cinque modalità "molto deluso", "insoddisfatto", "parzialmente soddisfatto", "soddisfatto", "entusiasta". 10 studenti si dicono entusiasti dell'operato dei professori, 51 si dicono soddisfatti, 63 mediamente soddisfatti, 90 insoddisfatti, 36 molto delusi.

Gradimento dei professori	Frequenze assolute	Frequenze relative	Frequenze relative cumulate
molto deluso	36	$36/250 = 0,144$	0,144
insoddisfatto	90	$90/250 = 0,360$	$0,144+0,360 = 0,504$
parzialmente soddisfatto	63	$63/250 = 0,252$	$0,504+0,252 = 0,756$
soddisfatto	51	$51/250 = 0,204$	$0,756+0,204 = 0,960$
entusiasta	10	$10/250 = 0,040$	$0,960+0,040 = 1,000$
Totali	250	$250/250 = 1,000$	

Nel caso ipotizzato, la colonna delle frequenze relative cumulate mostra che è molto deluso il 14,4% degli studenti e che la percentuale degli studenti non pienamente soddisfatti (modalità da "molto deluso" a "parzialmente soddisfatto") arriva al 75,6%

Gli indici di posizione centrale

In statistica si cerca di riassumere una serie di dati con un valore medio che possa esprimere sinteticamente il fenomeno.



Gli indici di posizione centrale

DEFINIZIONE

Media aritmetica

La media aritmetica M di n numeri

x_1, \dots, x_n 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 x_1, \dots, x_n e associati a essi i numeri p_1, \dots, p_n 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

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.

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

Gli indici di posizione centrale

DEFINIZIONE

Media geometrica

La media geometrica M di n numeri x_1, \dots, x_n è la radice n -esima del prodotto degli n numeri.

numero dei valori

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

media geometrica

prodotto dei valori

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

numero dei valori

somma dei reciproci dei valori

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%.

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.

Gli indici di posizione centrale

DEFINIZIONE

Media quadratica

La media quadratica Q di n numeri x_1, \dots, x_n è la radice quadrata della media aritmetica dei quadrati dei numeri.

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

media quadratica

numero dei valori

somma dei quadrati dei valori

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:

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

Gli indici di posizione centrale

DEFINIZIONE

Mediana

Data la sequenza ordinata di n numeri, la mediana è: x_1, \dots, x_n

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

ESEMPIO

21, 22, **26**, 28, 35

mediana

21, 22, **24**, 26, 28

mediana

$$\frac{22 + 26}{2} = 24$$

DEFINIZIONE

Moda

Dati i numeri x_1, \dots, x_n , si chiama moda il valore a cui corrisponde la frequenza massima.

ESEMPIO

50, 100, 200, 200, 200, 300, 300

moda

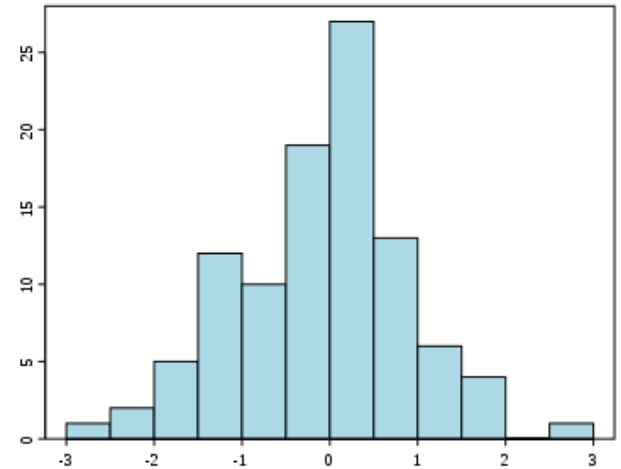
Fasi di una ricerca statistica

- 1. Studio del problema e impostazione della ricerca statistica
- 2. Rilevamento, classificazione e tabulazione dei dati statistici
- 3. Diagrammi e analisi statistica dei dati
- 4. Conclusioni dell'indagine

Rappresentazione grafica dei fenomeni statistici

- Ortogrammi (o grafici a strisce o a colonne)
- Diagrammi circolari (o areogrammi)
- Ideogrammi
- Cartogrammi
- Istogramma
- Poligoni di frequenze

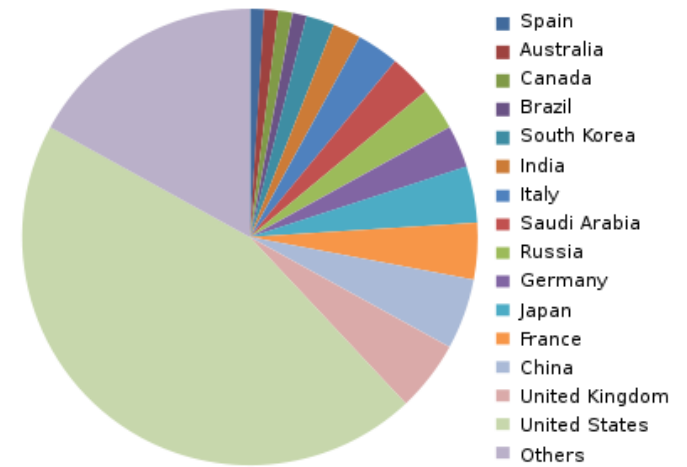
Istogramma



L'**istogramma** è la rappresentazione grafica di una distribuzione in classi di un carattere continuo.

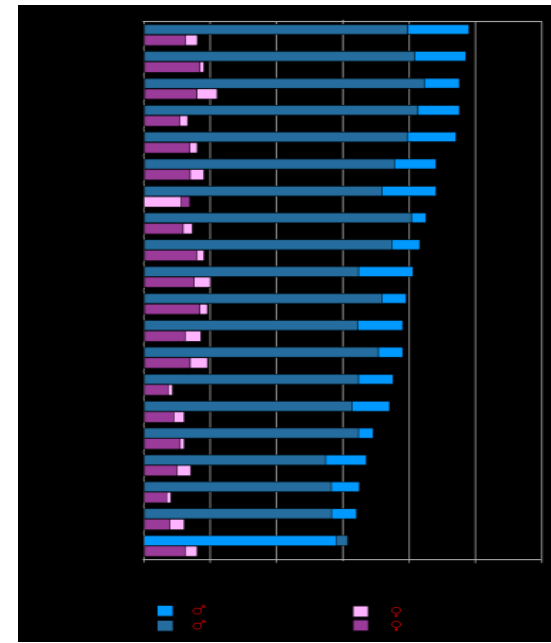
È costituito da rettangoli adiacenti le cui basi sono allineate su un asse orientato e dotato di unità di misura (l'asse ha l'unità di misura del carattere e può tranquillamente essere inteso come l'asse delle ascisse). L'adiacenza dei rettangoli dà conto della continuità del carattere. Ogni rettangolo ha base di lunghezza pari all'ampiezza della corrispondente classe; l'altezza invece è calcolata come densità di frequenza, ovvero essa è pari al rapporto fra la frequenza (assoluta) associata alla classe e l'ampiezza della classe

Aerogramma



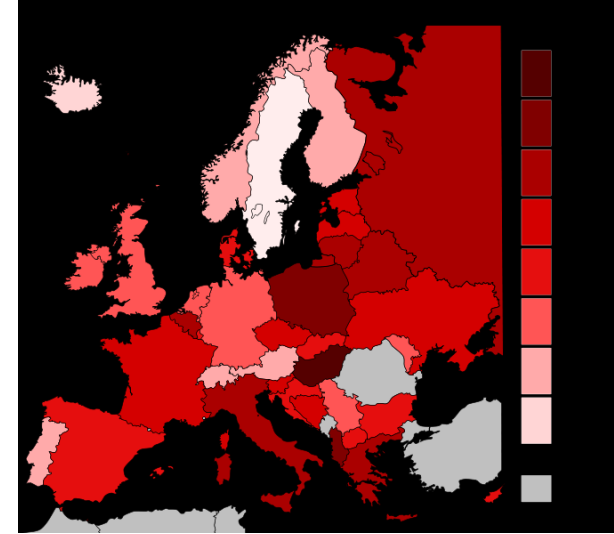
L'**areogramma** è un tipo di rappresentazione grafica in cui le diverse percentuali dei risultati di un'indagine statistica sono visualizzate da aree proporzionali di una figura geometrica piana o tridimensionale. L'unità di misura utilizzata è, spesso, la percentuale. Questo tipo di rappresentazione è particolarmente adatto per operare confronti fra diverse grandezze, senza necessariamente riferirsi ad un totale

Ortogramma



L'**ortogramma** a nastro è un particolare mezzo usato dagli statistici per effettuare un'indagine statistica. Si tratta di una variante dell'istogramma, in cui la frequenza assoluta è posizionata nell'asse delle ascisse, mentre le diverse variabili sono associate all'asse delle ordinate e rappresentate quindi come rettangoli a sviluppo orizzontale

Cartogramma



Un **cartogramma** è una carta geografica sulla quale vengono rappresentati dei dati statistici, con colori e simboli diversi a seconda dei valori del fenomeno osservato, relativi al territorio che la cartina rappresenta.

I cartogrammi vengono utilizzati per dare una visione immediata del rapporto tra territorio e valori del fenomeno. Infatti è più facile osservare un solo grafico già completo che confrontare e sovrapporre una cartina e un istogramma. Come l'ideogramma, questo tipo di grafico è molto approssimativo, sia per quanto riguarda il territorio sia per l'analisi del fenomeno