STATISTICA

La statistica è una disciplina di carattere quantitativo che studia i fenomeni di interesse attraverso la raccolta, l'organizzazione, l'analisi e l'interpretazione di dati. Essa svolge un ruolo fondamentale in una vasta gamma di ambiti, come le scienze sociali, le scienze naturali e l'economia, fornendo gli strumenti necessari per prendere decisioni informate e comprendere meglio il mondo che ci circonda. In particolare, la statistica permette di estrapolare informazioni significative dai dati raccolti, identificando relazioni e pattern nascosti. Attraverso l'utilizzo di metodi statistici, è possibile fare previsioni, testare ipotesi e valutare l'affidabilità dei risultati ottenuti. Grazie a questi concetti e tecniche, la statistica si è evoluta nel corso degli anni diventando una scienza indispensabile per il progresso e lo sviluppo delle società moderne.

Concetti fondamentali della statistica

Per comprendere appieno la statistica, è necessario partire dalla definizione di alcuni concetti chiave.

L'**unità statistica** è l'elemento base di un'indagine statistica, ossia l'oggetto o individuo su cui si raccolgono i dati.

L'insieme di tutte le unità statistiche che interessano un'indagine è la **popolazione**. Ad esempio, se si vuole studiare l'altezza degli studenti di una scuola, la popolazione è composta da tutti gli studenti della scuola.

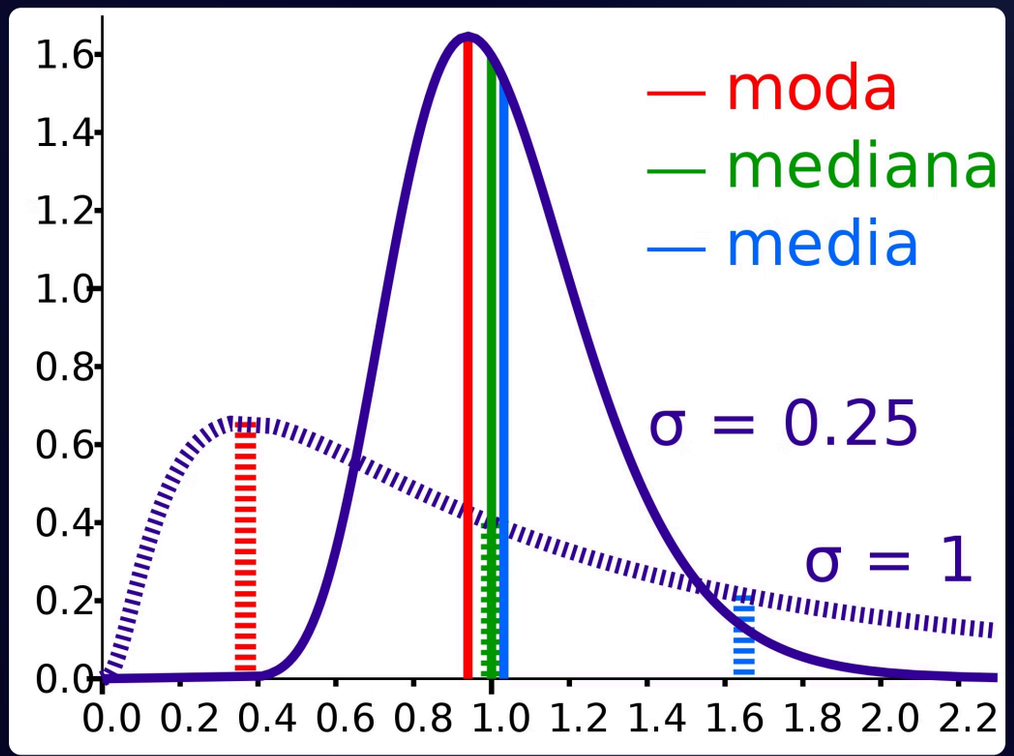
Le **variabili** sono le caratteristiche delle unità statistiche che possono assumere valori differenti. Le variabili possono essere **qualitative**, se esprimono qualità o attributi (es. colore degli occhi), o **quantitative**, se esprimono quantità numeriche (es. altezza).

Le variabili quantitative possono essere **discrete**, se possono assumere solo valori interi (es. numero di figli), o **continue**, se possono assumere qualsiasi valore all'interno di un intervallo (es. altezza).

Statistica descrittiva e inferenziale

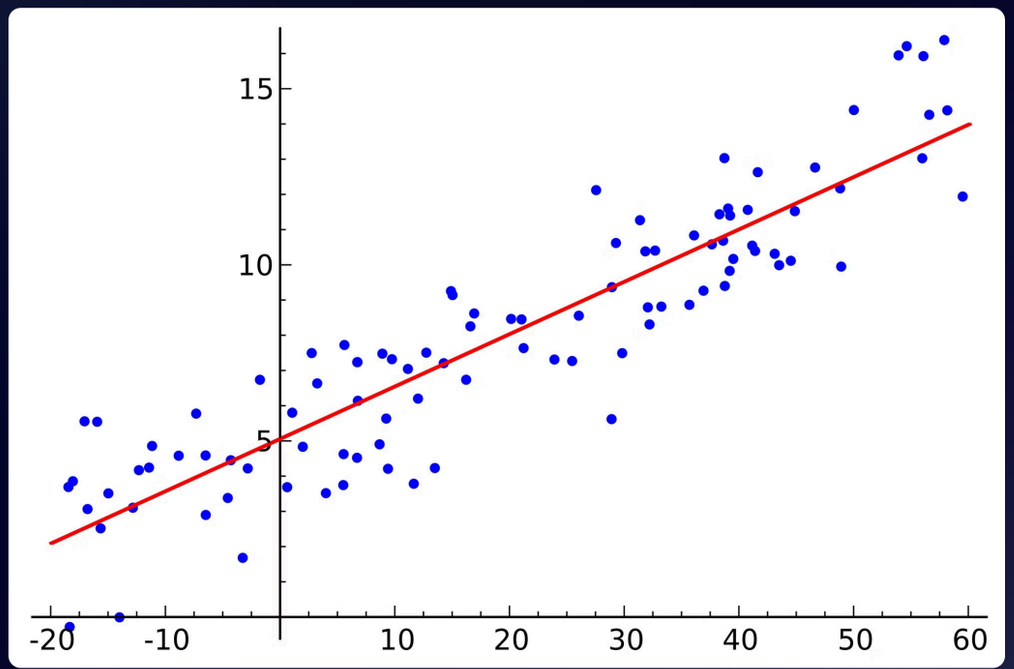
Statistica descrittiva

La statistica descrittiva si occupa di raccogliere, organizzare, sintetizzare e presentare i dati in modo chiaro e comprensibile. Questa fase è cruciale per comprendere le caratteristiche di un fenomeno, come la sua distribuzione, la tendenza centrale e la dispersione.



Statistica inferenziale

La statistica inferenziale, invece, si concentra sull'utilizzo dei dati per trarre conclusioni e fare previsioni su una popolazione a partire da un campione. Questa parte della statistica permette di formulare ipotesi, testare teorie e generalizzare i risultati ottenuti.



Rappresentazione dei dati: tabella delle frequenze

Tabella delle frequenze

La tabella delle frequenze è uno strumento fondamentale per organizzare e presentare i dati in modo chiaro e conciso. Essa mostra le diverse categorie o classi presenti nei dati, insieme alle rispettive frequenze assolute e relative.

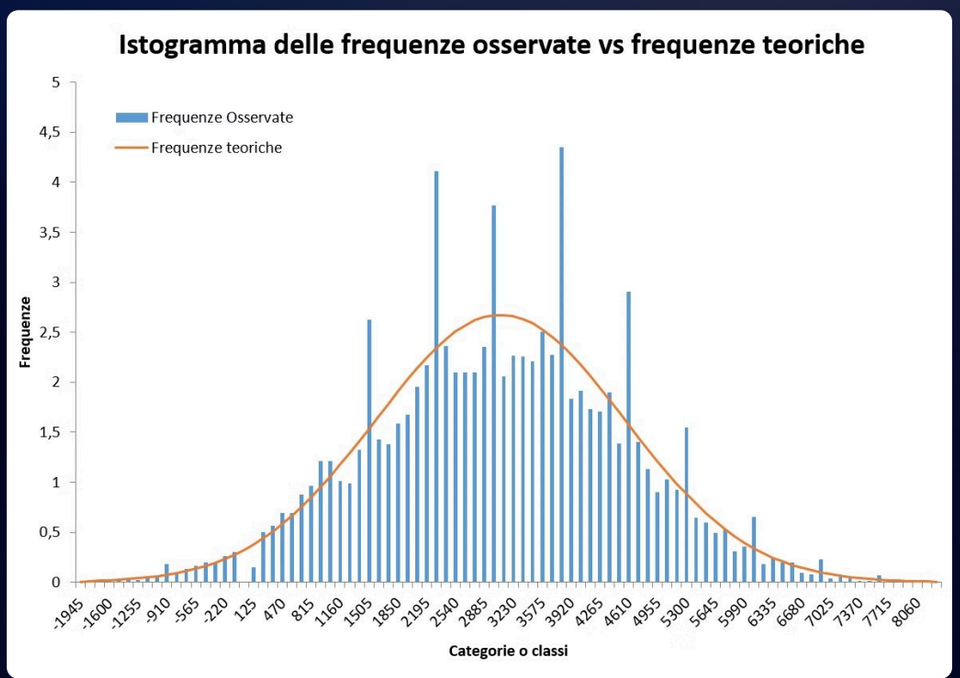
Analisi delle frequenze

L'analisi delle frequenze permette di comprendere la distribuzione dei dati, identificando le categorie più e meno rappresentate. Questo approccio è essenziale per interpretare correttamente i risultati e trarre conclusioni significative.

Visualizzazione grafica

La tabella delle frequenze può essere ulteriormente arricchita attraverso l'utilizzo di rappresentazioni grafiche, come istogrammi o diagrammi a torta. Queste visualizzazioni comunicano in modo chiaro ed efficace le informazioni contenute nei dati.

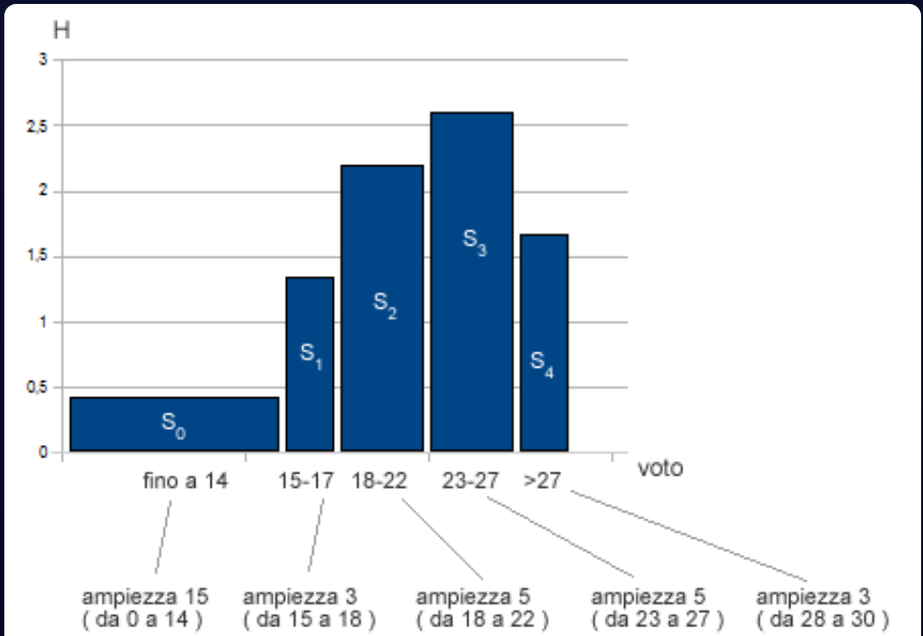
Altezza in metri	Frequenza f_i	Frequenze relativa f_i/n	Percentuale $[fi/n *(100)]\%$
1.40	5	5/28=0.1786	17.86%
1.41	0	0/28=0.0	0%
1.42	0	0/28=0.0	0%
1.43	3	3/28=0.1071	10.71%
1.44	2	2/28=0.0714	7.14%
1.45	0	0/28=0.0	0%
1.46	5	5/28=0.1786	17.86%
1.47	0	0/28=0.0	0%
1.48	2	2/28=0.0714	7.14%
1.49	0	0/28=0.0	0%
1.50	6	6/28=0.2142	21.42%
1.51	1	1/28=0.0357	3.57%
1.52	2	2/28=0.0714	7.14%
1.53	2	2/28=0.0714	7.14%
Σ	28		



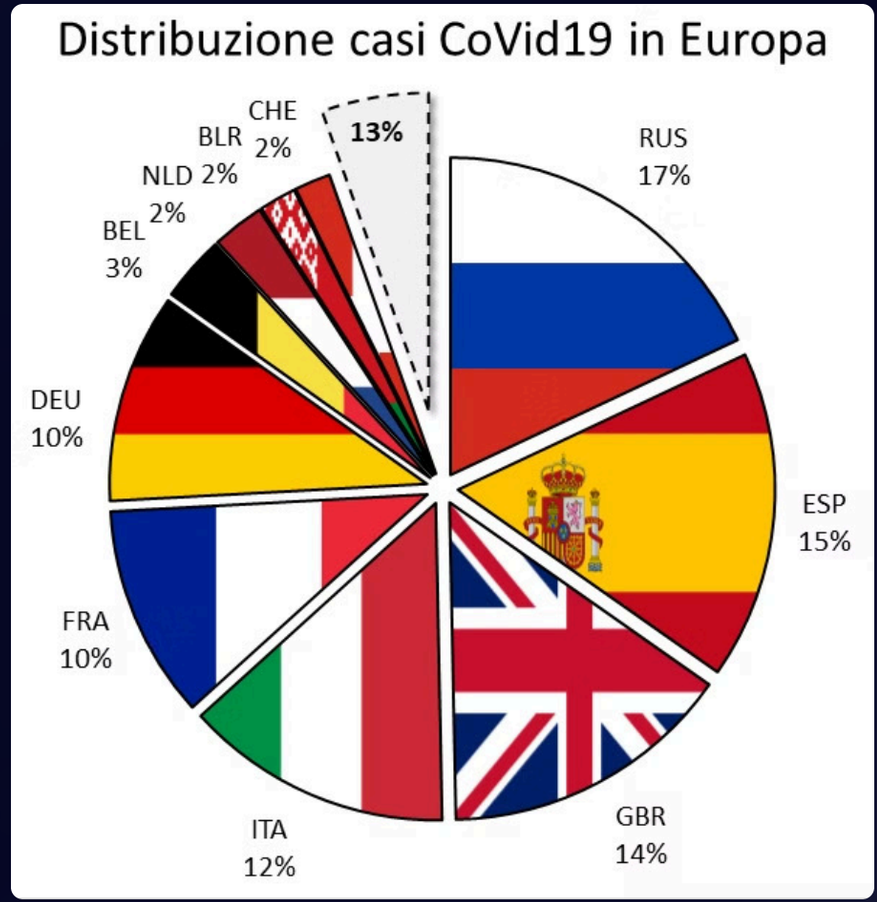
Istogramma e Diagramma a torta

L'istogramma è un grafico a barre che rappresenta la distribuzione di frequenze di un dato quantitativo. Le barre sono disposte in modo contiguo e l'area di ciascuna barra rappresenta la frequenza assoluta della classe corrispondente.

Il diagramma a torta, invece, è un grafico circolare che rappresenta la composizione percentuale di un dato qualitativo. Ogni fetta rappresenta una categoria e la dimensione della fetta è proporzionale alla frequenza relativa della categoria corrispondente.



	frequenza	ampiezza	H =F/A	S
fino a 14	6	15	0,4	6
15-17	4	3	1,33	4
18-22	11	5	2,2	11
23-27	13	5	2,6	13
Oltre 27	5	3	1,67	5
	39			



Indicatori di centralità (1)

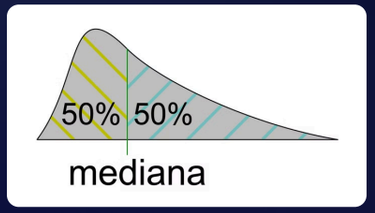
Media aritmetica

La media aritmetica è l'indicatore di tendenza centrale più comunemente utilizzato. Essa rappresenta il valore medio di un insieme di dati, calcolato come somma di tutti i valori divisa per il numero totale di osservazioni.

$$\bar{x} = \frac{\sum x_i}{n}$$

Mediana

La mediana è il valore centrale di un insieme di dati ordinati in modo crescente o decrescente.



DOPPIA MEDIANA

1	2	3	4	5	6
N	P	S	M	M	M

A differenza della media, la mediana non risente dell'influenza di valori estremi, rendendola un indicatore più robusto. In caso di doppia mediana, si prende il valor medio dei due.

Moda

La moda è il valore che appare più frequentemente in un insieme di dati. Può essere calcolata individuando il valore con la maggior frequenza. La moda è utile per identificare il valore più rappresentativo o tipico all'interno dei dati.

Peso (kg)	Frequenze
40-50	4
50-60	7
60-70	3
70-80	2
80-90	1

Indicatori di centralità (2)

Media pesata

La media pesata è un'altra metrica di centralità usata in applicazioni statistiche che risulta utile quando alcuni dati hanno maggiore importanza rispetto ad altri. In questo tipo di analisi, ai diversi dati vengono assegnati dei pesi che riflettono la loro importanza nell'insieme dei dati. La media viene quindi calcolata moltiplicando ciascun valore per il relativo peso e dividendo la somma dei prodotti per la somma dei pesi totali.

$$M_p = \frac{\sum_{i=1}^{k} n_i \cdot p_i}{\sum_{i=1}^{k} p_i}$$

Media quadratica

La media quadratica è un'ulteriore metrica di centralità utilizzata per valutare la dispersione dei dati. Questa misura calcola la radice quadrata della somma dei quadrati dei valori. È particolarmente utile quando ci si concentra sulla varianza dei dati rispetto alla media. La media quadratica è spesso utilizzata in ambito matematico e scientifico per descrivere la deviazione di un insieme di dati dalla media.

$$\mu_q = \sqrt{\frac{x_1{}^2 + x_2{}^2 + x_3{}^2 + \ldots + x_{i-1}{}^2 + x_i{}^2}{\mathrm{N}}}$$

Indicatori di centralità (3)

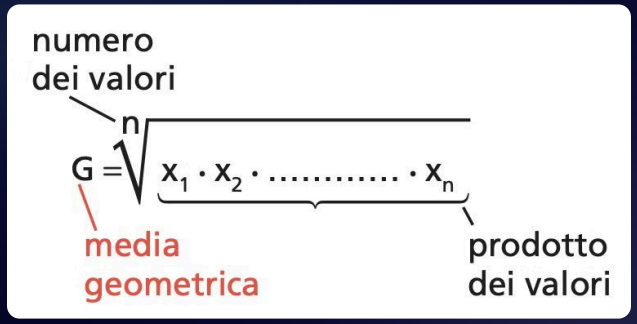
Media armonica

La media armonica è un'altra misura di centralità utilizzata per valutare la distribuzione dei dati. Questa misura calcola il reciproco della media dei reciproci dei valori. La media armonica è spesso utilizzata quando si lavora con dati che hanno una natura reciproca o proporzionale, come ad esempio velocità o tariffe.

$$H = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \ldots + \frac{1}{x_n}}$$

Media geometrica

La media geometrica è un altro indicatore di centralità utilizzato, particolarmente utile quando si hanno dati positivi e distribuiti in modo asimmetrico. La media geometrica è utilizzata quando le variabili non sono rappresentate da valori ottenuti come prodotto o rapporto tra valori lineari. Serve per il confronto di superfici o volumi, di tassi di variazione, cioè valori che sono espressi da rapporti. Essa è calcolata come radice n-esima del prodotto dei valori.



Indicatori di dispersione

Varianza σ²

La varianza è un indicatore di dispersione che misura quanto i valori di un insieme di dati si allontanano dalla loro media. Essa è calcolata come media dei quadrati degli scarti dalla media.

Deviazione standard σ

La deviazione standard è la radice quadrata della varianza e rappresenta la misura più comune della dispersione di un insieme di dati. Essa fornisce informazioni sulla distanza media dei valori dalla media.

Covarianza σxy

La covarianza è una misura della relazione lineare tra due variabili x e y. Rappresenta la media del prodotto dei valori scartati dalla media delle due variabili. Una covarianza positiva indica una relazione diretta, mentre una covarianza negativa indica una relazione inversa tra le variabili.

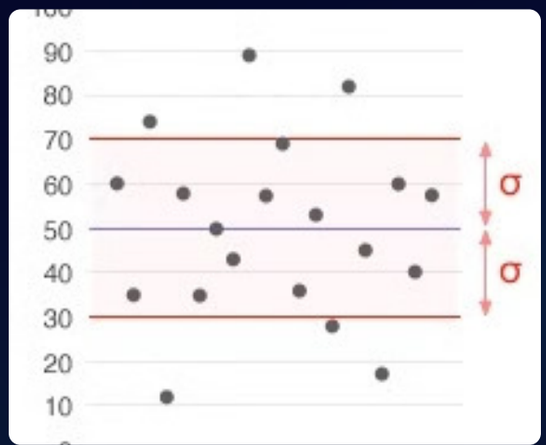
L'analisi congiunta della tendenza centrale e della dispersione dei dati permette di comprendere meglio la distribuzione e le caratteristiche di un fenomeno statistico. Varianza e dev. std campionarie sono stime calcolate sulla base dei dati di un campione e possono essere utilizzata per trarre conclusioni sull'intera popolazione. Tuttavia, tenere presente che queste possono essere influenzate dalle dimensioni del campione utilizzato per stimarle.

σ^2 varianza $\frac{1}{n} \cdot \sum_{i=1}^{n} (x_i - \bar{x})^2$

σ dev. standard $\sqrt{\frac{1}{n} \cdot \sum_{i=1}^{n} (x_i - \bar{x})^2}$

s^2 var. campionaria $\frac{1}{n-1} \cdot \sum_{i=1}^{n} (x_i - \bar{x})^2$

s dev. campionaria $\sqrt{\frac{1}{n-1} \cdot \sum_{i=1}^{n} (x_i - \bar{x})^2}$



$$\sigma_{xy} = \frac{\sum (x_i - \mu_x)(y_i - \mu_y)}{N}$$

$$S_{xy} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{n-1}$$

Regressione lineare

1

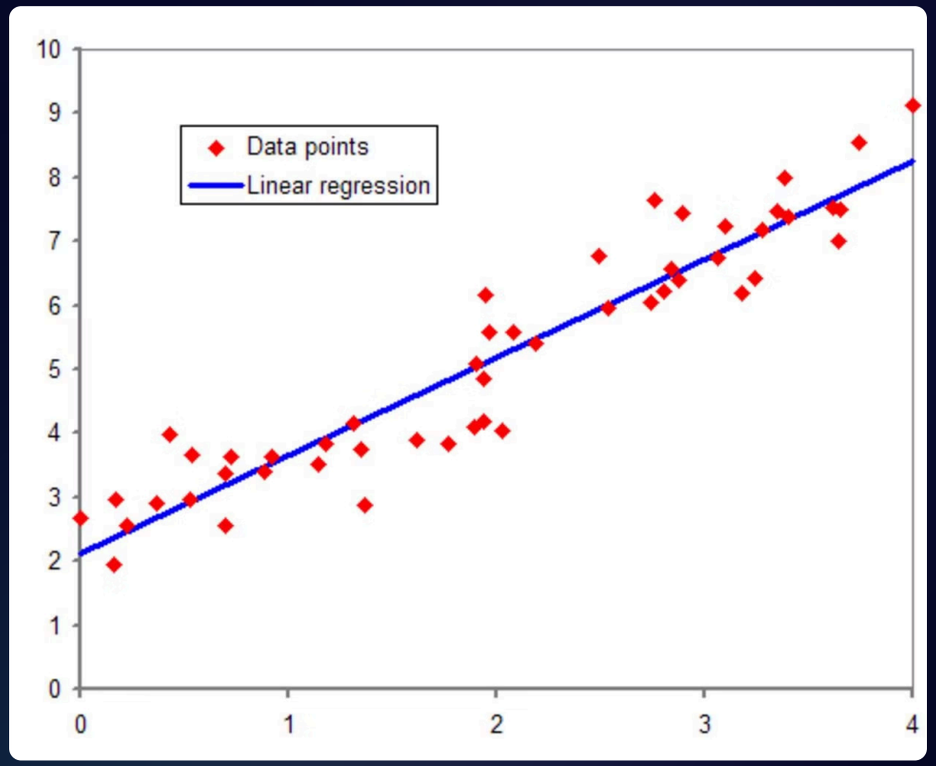
Modello

La regressione lineare è un modello statistico che descrive la relazione tra una variabile dipendente e una o più variabili indipendenti. Il modello esprime la variabile dipendente come combinazione lineare delle variabili indipendenti.

2

Stima dei parametri

I parametri del modello di regressione, ovvero i coefficienti che determinano la retta di regressione, vengono stimati attraverso metodi come il metodo dei minimi quadrati. Questi parametri quantificano l'effetto delle variabili indipendenti sulla variabile dipendente.



Applicazioni

La regressione lineare trova ampia applicazione in diversi ambiti, come l'economia, la sociologia e le scienze naturali, permettendo di studiare e prevedere relazioni tra variabili di interesse.

Metodo dei minimi quadrati

1

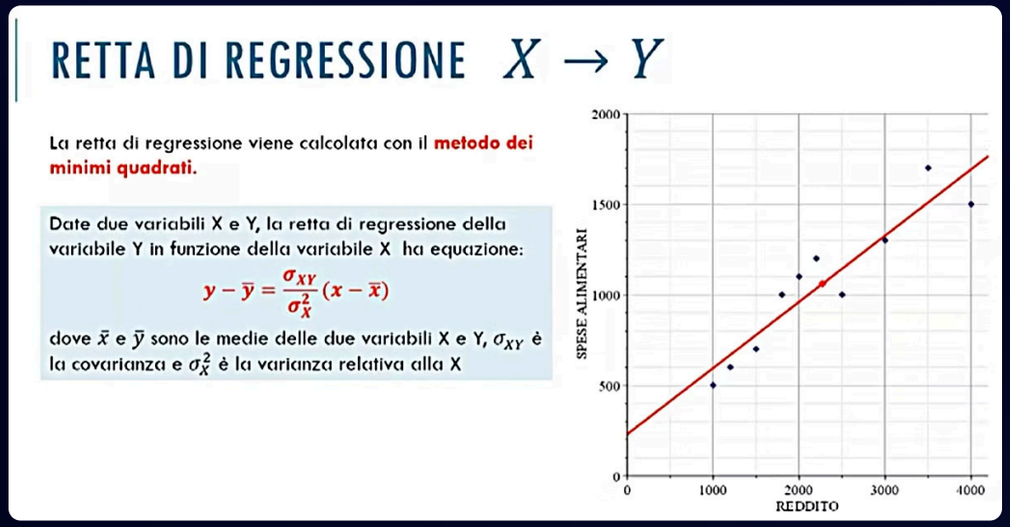
Equazione

La retta di regressione è l'espressione matematica che rappresenta la relazione lineare tra la variabile dipendente e la variabile indipendente. L'equazione della retta di regressione è **y = a + bx**, dove a è l'intercetta e b è il coefficiente angolare.

2

Interpretazione

Il coefficiente angolare b rappresenta la variazione attesa della variabile dipendente y per ogni unità di variazione della variabile indipendente x. L'intercetta sull'asse delle y indica il valore di y quando x è uguale a zero.



Previsione

Una volta stimata la retta di regressione, è possibile utilizzarla per fare previsioni sulla variabile dipendente in base ai valori della variabile indipendente. Questa capacità predittiva è uno degli aspetti più importanti della regressione lineare.

Coefficiente di correlazione lineare

Misurazione

$$\rho = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$$

Il coefficiente di correlazione lineare o Di Bravais-Pearson, indicato con la lettera ρ, è una statistica che misura il grado di associazione lineare tra due variabili. Esso varia tra -1 e 1, con valori vicini a 1 o -1 che indicano una forte relazione.

Si distinguono i seguenti tipi di correlazione:

- Se ρXY > 0, le variabili X e Y si dicono *direttamente correlate*, oppure *correlate positivamente.*
- Se ρXY = 0, le variabili X e Y si dicono *incorrelate.*
- Se ρXY < 0, le variabili X e Y si dicono *inversamente correlate*, oppure *correlate negativamente.*

Inoltre per la correlazione diretta (e analogamente per quella inversa) si distingue:

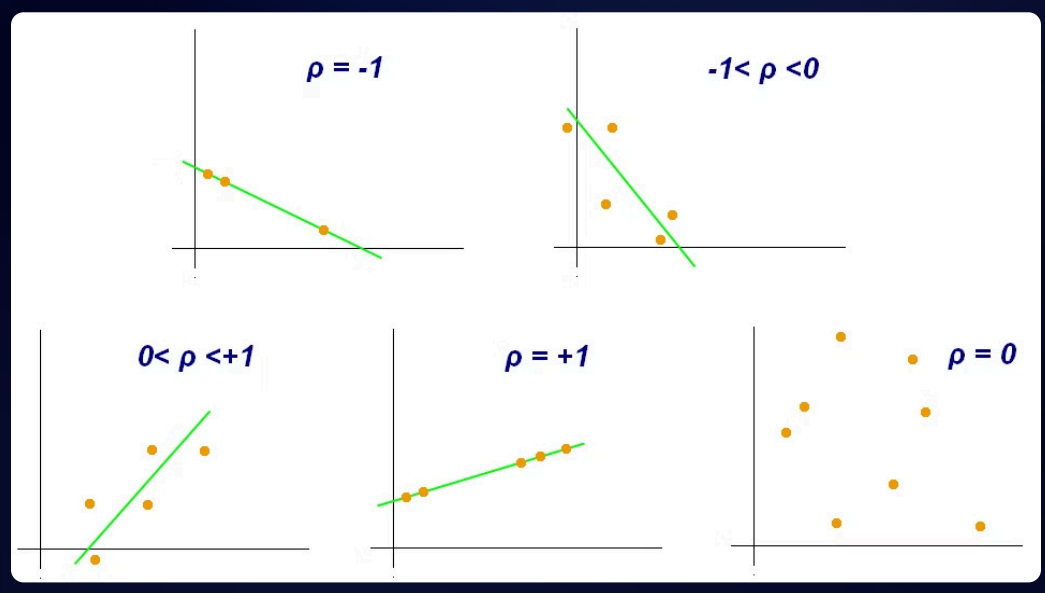
- se 0 < |ρXY| < 0,3 si ha *correlazione debole*;
- se 0,3 < |ρXY| < 0,7 si ha *correlazione moderata*;
- se |ρXY| > 0,7 si ha *correlazione forte*.

Intensità della relazione

Il valore assoluto del coefficiente di correlazione indica l'intensità della relazione lineare tra le variabili: più il valore si avvicina a 1, più forte è la relazione.

Direzione della relazione

Il segno del coefficiente di correlazione (+/-) indica la direzione della relazione: un segno positivo indica una relazione diretta, mentre un segno negativo indica una relazione inversa.



Conclusioni e riepilogo

1

Sintesi

In questo percorso abbiamo esaminato i concetti fondamentali della statistica, dalle misure di tendenza centrale e dispersione alla regressione lineare e al coefficiente di correlazione. Questi strumenti statistici forniscono una solida base per comprendere e interpretare i dati in modo rigoroso e scientificamente valido.

2

Applicazioni pratiche

La statistica è una disciplina estremamente versatile, con applicazioni in una vasta gamma di ambiti, dalla ricerca scientifica all'analisi dei mercati finanziari. L'abilità nell'utilizzo degli strumenti statistici è fondamentale per prendere decisioni informate e approfondire la comprensione dei fenomeni di interesse.

3

Sviluppi futuri

La statistica continua a evolversi, con l'introduzione di nuovi metodi e l'applicazione in contesti sempre più complessi. Mantenere aggiornate le proprie conoscenze e abilità statistiche è essenziale per affrontare le sfide del mondo contemporaneo.