

Questo modulo, **Modulo 3.3 – Elaborazione statistica dei dati**, prevede una durata di 14 ore ed è pensato per fornire una solida base nell'analisi statistica dei dati, sia dal punto di vista teorico che pratico.

### **Prerequisiti:**

Anche se non vengono indicati prerequisiti specifici, è utile avere una conoscenza base di matematica e di Excel, oltre a familiarità con i concetti fondamentali della statistica e della probabilità.

### **Contenuti:**

#### **1. Introduzione alla Statistica e alla Probabilità:**

- Nozioni di base per comprendere come analizzare i dati e calcolare le probabilità.

#### **2. Tipologia di variabili:**

- Variabili qualitative vs. quantitative, variabili discrete vs. continue.

#### **3. Analisi descrittiva mediante distribuzione di frequenza (e diagrammi):**

- Come organizzare i dati in distribuzioni di frequenza e rappresentarli graficamente (istogrammi, grafici a barre, ecc.).

#### **4. Tabelle di frequenza (tramite tabelle pivot su Excel):**

- Creazione di tabelle pivot su Excel per analizzare i dati attraverso la distribuzione delle frequenze.

#### **5. Sintesi e descrizione delle variabili mediante valori medi:**

- Calcolo della media, mediana, moda come sintesi centrale dei dati.

#### **6. Indici di posizione e variabilità:**

- Indici come quartili, percentili, range interquartile.

#### **7. Sintesi e descrizione delle variabili mediante misure di variabilità:**

- Varianza, deviazione standard, range e altre misure di dispersione dei dati.

#### **8. Variabili casuali o aleatorie:**

- Comprensione delle variabili aleatorie e loro utilizzo nelle analisi statistiche.

#### **9. Distribuzione di probabilità:**

- Concetti chiave delle distribuzioni di probabilità, che descrivono come una variabile casuale può assumere diversi valori.
- 10. **Distribuzione Binomiale:**
  - Distribuzione per eventi con due possibili esiti (successo o fallimento).
- 11. **Distribuzione Normale o Gaussiana (caratteristiche e proprietà):**
  - La più importante distribuzione statistica, utilizzata per modellare molti fenomeni naturali.
- 12. **Applicazione di Excel allo studio di una distribuzione:**
  - Utilizzo di Excel per visualizzare e analizzare distribuzioni, calcolare probabilità e creare grafici.

### **Metodologia e strumenti didattici:**

1. **Lezioni frontali:** Forniscono una spiegazione teorica dei concetti chiave.
2. **Esercizi:** Applicazione pratica dei concetti appresi tramite esercizi mirati, sia manualmente che utilizzando Excel.
3. **Laboratorio:** Lavoro pratico in laboratorio, probabilmente con l'uso di software come Excel per l'analisi e l'elaborazione dei dati statistici.

# Lezione 1: Introduzione alla Statistica e Probabilità

**Durata teoria: 1 ora**

---

## **Obiettivo della lezione teorica:**

- Fornire una panoramica generale della statistica.
  - Introdurre i concetti fondamentali della probabilità.
  - Familiarizzare con i tipi di variabili utilizzate in statistica.
  - Presentare l'importanza dell'analisi descrittiva.
- 

## **Argomenti da trattare:**

### **1. Introduzione alla Statistica (15 minuti)**

- **Definizione di Statistica:** Spiega che la statistica è la scienza che si occupa della raccolta, organizzazione, analisi e interpretazione dei dati.
  - **Esempio pratico:** Descrivi la statistica come una lente attraverso la quale possiamo osservare e comprendere fenomeni complessi come i voti degli studenti, l'andamento delle vendite, i sondaggi di opinione o lo studio dei fenomeni naturali.

### **Definizione di Statistica**

La **statistica** è la scienza che si occupa della raccolta, organizzazione, analisi e interpretazione dei dati. Attraverso metodi quantitativi e qualitativi, la statistica consente di trasformare dati grezzi in informazioni significative. Questo processo è fondamentale per comprendere fenomeni complessi, identificare tendenze, fare previsioni e prendere decisioni informate in vari campi come l'economia, la medicina, le scienze sociali e naturali.

## Esempio pratico

Immagina la statistica come una lente attraverso la quale osserviamo e comprendiamo il mondo che ci circonda. Ad esempio:

- **Voti degli studenti:** In una scuola, la semplice raccolta dei voti non basta per capire il rendimento complessivo degli studenti. Utilizzando strumenti statistici, è possibile calcolare la media dei voti, individuare le materie in cui gli studenti eccellono o hanno difficoltà, e confrontare i risultati tra diverse classi o anni scolastici. Questo aiuta insegnanti e amministratori a migliorare i programmi educativi.
- **Andamento delle vendite:** Un'azienda può raccogliere dati sulle vendite dei suoi prodotti. Attraverso l'analisi statistica, può identificare quali prodotti vendono di più, in quali periodi dell'anno le vendite aumentano o diminuiscono, e prevedere le tendenze future. Queste informazioni sono cruciali per strategie di marketing e gestione dell'inventario.
- **Sondaggi di opinione:** Quando si vuole conoscere l'opinione pubblica su un determinato argomento, si conducono sondaggi. La statistica permette di analizzare le risposte raccolte, determinare la percentuale di persone che supportano una certa idea e valutare il margine di errore. Questo è essenziale per politici, aziende e ricercatori che vogliono comprendere le esigenze e le preferenze della popolazione.
- **Studio dei fenomeni naturali:** Nelle scienze naturali, la statistica aiuta a interpretare dati complessi come quelli relativi al clima, ai terremoti o alla diffusione di malattie. Ad esempio, gli epidemiologi utilizzano modelli statistici per prevedere l'andamento di un'epidemia e valutare l'efficacia di interventi sanitari.

In tutti questi casi, la statistica non solo organizza e analizza i dati, ma offre anche una comprensione più profonda dei fenomeni, permettendo di agire in modo informato e strategico. Senza la statistica, saremmo sommersi da informazioni non strutturate, incapaci di cogliere il quadro generale e le connessioni nascoste tra i dati.

## Definizione Rigorosa di Statistica

La **statistica** è una disciplina matematica che si occupa dello studio dei fenomeni caratterizzati da incertezza e variabilità, attraverso la raccolta, l'organizzazione, l'analisi, l'interpretazione e la presentazione dei dati. Si basa su metodi scientifici e matematici per estrarre informazioni significative da dati numerici o categoriali, con l'obiettivo di descrivere fenomeni, testare ipotesi, fare previsioni e supportare decisioni informate.

Formalmente, la statistica può essere suddivisa in due branche principali:

1. **Statistica Descrittiva:** Si occupa della sintesi e descrizione dei dati raccolti. Utilizza misure numeriche (come media, mediana, varianza, deviazione standard) e rappresentazioni grafiche (come istogrammi, box-plot, diagrammi a dispersione) per riassumere le caratteristiche fondamentali di un insieme di dati.
2. **Statistica Inferenziale:** Basandosi su un campione di dati, mira a trarre conclusioni o inferenze sulla popolazione più ampia da cui il campione è stato estratto. Include metodi come la stima puntuale e per intervalli, il test di ipotesi, l'analisi della regressione e l'analisi della varianza (ANOVA).

La statistica si appoggia alla **teoria della probabilità** per modellare e quantificare l'incertezza associata ai fenomeni aleatori. Attraverso modelli probabilistici, è possibile:

- **Formulare ipotesi:** Proporre congetture sui parametri o sulle distribuzioni delle popolazioni.
- **Stimare parametri:** Utilizzare dati campionari per ottenere stime dei parametri sconosciuti della popolazione.
- **Verificare ipotesi:** Applicare test statistici per accettare o rifiutare le ipotesi formulate, controllando il rischio di errore.
- **Predire valori futuri:** Costruire modelli predittivi per stimare valori non osservati o futuri basandosi sui dati esistenti.

In un contesto rigoroso, la statistica implica anche:

- **Definizione di Popolazione e Campione:** La popolazione è l'insieme completo degli elementi di interesse, mentre il campione è un sottoinsieme rappresentativo utilizzato per l'analisi.
- **Validità Statistica:** Assicurare che i metodi e le conclusioni siano statisticamente validi, considerando aspetti come la dimensione del campione, la casualità nella selezione e l'assenza di bias.
- **Inferenza Statistica:** Applicare metodi deduttivi per generalizzare i risultati dal campione alla popolazione, quantificando l'incertezza tramite intervalli di confidenza e livelli di significatività.

La statistica è fondamentale in numerosi campi scientifici e applicativi, poiché fornisce gli strumenti necessari per affrontare problemi complessi dove la variabilità e l'incertezza giocano un ruolo cruciale. Attraverso un approccio rigoroso e metodologico, la statistica permette di trasformare dati grezzi in conoscenza utile e affidabile.

## Due principali rami della Statistica:

- **Statistica descrittiva:** Riguarda la sintesi e la rappresentazione dei dati. Introduci concetti come grafici, tabelle, media, varianza, ecc. (Esempio: riassumere i voti di una classe con la media e il grafico a barre).
- **Statistica inferenziale:** Usa un campione di dati per fare previsioni o inferenze su una popolazione più ampia. Introduci in modo semplice il concetto di inferenza (Esempio: stimare l'altezza media di una popolazione sulla base di un campione).

## Statistica Descrittiva

La **statistica descrittiva** si occupa della raccolta, organizzazione, sintesi e rappresentazione dei dati. Il suo scopo principale è descrivere le caratteristiche fondamentali di un insieme di dati in modo chiaro e comprensibile. Questo ramo della statistica utilizza strumenti matematici e grafici per riassumere le informazioni, rendendo più facile l'interpretazione dei dati grezzi.

### Concetti chiave della statistica descrittiva:

- **Tabelle:** Organizzano i dati in modo strutturato, facilitando la lettura e il confronto tra diverse categorie o gruppi.
- **Grafici:** Rappresentano visivamente i dati per evidenziare tendenze, distribuzioni e anomalie. Esempi comuni includono:
  - **Grafici a barre:** Utilizzati per confrontare quantità tra diverse categorie.
  - **Istogrammi:** Mostrano la distribuzione di una variabile continua suddivisa in classi.
  - **Grafici a torta:** Rappresentano le proporzioni di un insieme rispetto al totale.
  - **Diagrammi a dispersione:** Illustrano la relazione tra due variabili quantitative.
- **Misure di tendenza centrale:**

- **Media:** Somma di tutti i valori divisa per il numero totale di osservazioni; rappresenta il valore medio del dataset.
- **Mediana:** Il valore centrale in un dataset ordinato; divide il dataset in due parti uguali.
- **Moda:** Il valore che si verifica con maggiore frequenza nel dataset.
- **Misure di dispersione:**
  - **Varianza:** Indica quanto i dati sono sparsi intorno alla media; è la media dei quadrati delle differenze tra ogni dato e la media.
  - **Deviazione standard:** La radice quadrata della varianza; misura la dispersione dei dati rispetto alla media in unità originali.
  - **Intervallo (Range):** Differenza tra il valore massimo e il valore minimo del dataset.



### Esempio pratico:

Supponiamo di voler riassumere i voti di matematica di una classe composta da 20 studenti. I voti ottenuti sono:

6, 7, 8, 9, 7, 8, 6, 10, 7, 9, 8, 7, 6, 9, 8, 7, 6, 10, 7, 8

### Passi per l'analisi descrittiva:

#### 1. Calcolo della media:

$$\text{Media} = \frac{\sum \text{voti}}{\text{numero di studenti}} = \frac{6 + 7 + 8 + \dots + 8}{20} = \frac{157}{20} = 7.85$$

La media dei voti è 7.85.

#### 2. Determinazione della mediana:

- Ordiniamo i voti in ordine crescente: 6, 6, 6, 6, 7, 7, 7, 7, 7, 7, 8, 8, 8, 8, 8, 8, 9, 9, 9, 10, 10
- Poiché il numero di osservazioni è pari (20), la mediana è la media dei due valori centrali (10° e 11° valore):  
$$\text{Mediana} = \frac{7+8}{2} = 7.5$$

#### 3. Identificazione della moda:

- Il voto che compare più frequentemente è 7.

#### 4. Calcolo della deviazione standard:

- Si utilizza la formula della deviazione standard per quantificare la dispersione dei voti attorno alla media.

#### 5. Creazione di un grafico a barre:

- Sull'asse orizzontale (asse delle ascisse) si riportano i voti possibili (6, 7, 8, 9, 10).
- Sull'asse verticale (asse delle ordinate) si indica il numero di studenti che hanno ottenuto ciascun voto.
- Ogni barra rappresenta la frequenza di un voto specifico.

Questo grafico permette di visualizzare rapidamente la distribuzione dei voti nella classe, evidenziando quale voto è più comune e come sono distribuiti gli altri voti.

### Interpretazione:

La statistica descrittiva ci fornisce una panoramica immediata dei risultati della classe. La media di 7.85 indica un rendimento generale buono. La moda di 7 suggerisce che il voto più frequente è 7, mentre la deviazione standard ci informa sulla variabilità dei voti rispetto alla media.

## Statistica Inferenziale

La **statistica inferenziale** si occupa di fare previsioni o trarre conclusioni su una popolazione più ampia basandosi su un campione di dati. Poiché spesso è impraticabile o impossibile analizzare ogni membro di una popolazione, si utilizza un campione rappresentativo per inferire le caratteristiche della popolazione intera.

### Concetti chiave della statistica inferenziale:

- **Popolazione:** L'insieme completo di tutti gli individui o elementi di interesse in uno studio statistico.
- **Campione:** Un sottoinsieme della popolazione selezionato per l'analisi; deve essere rappresentativo per garantire l'affidabilità delle inferenze.
- **Inferenza statistica:** Il processo di utilizzare i dati campionari per fare stime, testare ipotesi o fare previsioni sulla popolazione.
- **Stima puntuale:** Una singola valore numerico che rappresenta la migliore stima di un parametro della popolazione (es. la media campionaria come stima della media della popolazione).
- **Intervallo di confidenza:** Un intervallo di valori entro il quale ci si aspetta che il parametro della popolazione cada con una certa probabilità (es. 95%).
- **Test di ipotesi:** Un procedimento che permette di valutare se le evidenze presenti nel campione supportano una certa ipotesi sulla popolazione.

### Esempio pratico:

Supponiamo di voler stimare l'altezza media degli adulti in una città con una popolazione di 100.000 persone. Misurare l'altezza di ogni individuo sarebbe impraticabile. Invece, selezioniamo un **campione casuale** di 1.000 adulti.

### Passi per l'inferenza statistica:

#### 1. Raccolta dei dati campionari:

- Misuriamo l'altezza di ciascuno dei 1.000 adulti selezionati.
- 2. Calcolo della media campionaria:**
  - Supponiamo che la media delle altezze nel campione sia 170 cm.
- 3. Stima della media della popolazione:**
  - Utilizziamo la media campionaria come stima puntuale della media della popolazione.
- 4. Calcolo dell'intervallo di confidenza:**
  - Determiniamo l'intervallo entro il quale la vera media della popolazione probabilmente cade.
  - Supponiamo di calcolare un intervallo di confidenza al 95% che va da 169 cm a 171 cm.
  - Questo significa che siamo fiduciosi al 95% che la vera altezza media della popolazione sia compresa tra questi valori.
- 5. Inferenza sulla popolazione:**
  - Basandoci sul campione, concludiamo che l'altezza media degli adulti nella città è probabilmente intorno ai 170 cm.

### **Spiegazione del concetto di inferenza:**

L'inferenza statistica permette di estendere i risultati ottenuti dal campione alla popolazione intera, tenendo conto dell'incertezza intrinseca dovuta al fatto che non tutti gli individui sono stati misurati.

- **Perché utilizzare un campione?**
  - È spesso impossibile o troppo costoso misurare ogni individuo in una popolazione.
  - Un campione ben progettato può fornire informazioni affidabili sulla popolazione.
- **Come assicurare l'affidabilità delle inferenze?**
  - **Campionamento casuale:** Ogni individuo ha la stessa probabilità di essere selezionato, riducendo bias.
  - **Dimensione del campione:** Campioni più grandi tendono a produrre stime più precise.
  - **Controllo degli errori:** Si considerano errori casuali e sistematici per valutare l'accuratezza delle stime.

### **Applicazioni pratiche:**

- **Sondaggi di opinione:**
    - Per conoscere le preferenze politiche di una popolazione, si intervistano un campione di elettori.
    - I risultati vengono poi estesi alla popolazione generale con un certo margine di errore.
  - **Ricerca medica:**
    - Per testare l'efficacia di un nuovo farmaco, si conduce uno studio su un campione di pazienti.
    - I risultati aiutano a inferire l'efficacia del farmaco sull'intera popolazione dei pazienti affetti da una certa malattia.
  - **Controllo di qualità:**
    - In un'industria, si esamina un campione di prodotti per valutare la qualità dell'intera produzione.
    - Questo permette di identificare problemi senza dover testare ogni singolo prodotto.
- 

## Conclusione

La **statistica descrittiva** e la **statistica inferenziale** sono complementari nell'analisi dei dati:

- **Statistica descrittiva:** Fornisce strumenti per organizzare e sintetizzare i dati raccolti, rendendo più semplice la comprensione delle caratteristiche principali del dataset.
- **Statistica inferenziale:** Permette di estendere le conclusioni oltre il dataset, facendo previsioni e prendendo decisioni informate sulla popolazione basandosi su un campione.

**Utilità della statistica:** Sottolinea l'importanza della statistica nel mondo reale, come strumento per prendere decisioni informate in vari settori (es. aziendale, medico, sociale).

**Esempi pratici:** Analisi dei dati aziendali per ottimizzare la produzione, valutazione dei risultati elettorali, studio degli effetti di nuovi farmaci.

## 2. Tipi di Dati e Variabili (20 minuti)

### Definizione di Variabili

Una **variabile** è una caratteristica o attributo che può assumere diversi valori. Le variabili sono ciò che misuriamo o osserviamo in uno studio statistico e possono rappresentare informazioni su individui, oggetti o eventi.

### Esempi di variabili:

- Età
- Altezza
- Colore degli occhi
- Voto di un esame
- Sesso
- Professione

### Classificazione delle Variabili

Le variabili possono essere classificate in due categorie principali: **qualitative** e **quantitative**.

#### a) Variabili Qualitative (o Categoricali)

Le **variabili qualitative** descrivono qualità o categorie e non sono misurate numericamente. Servono a classificare gli individui o gli oggetti in gruppi distinti.

#### Sottotipi di variabili qualitative:

- **Nominali:**
  - **Definizione:** Variabili categoriali senza un ordine intrinseco.
  - **Esempi:**
    - **Colore degli occhi:** blu, verde, marrone.
    - **Città di nascita:** Roma, Milano, Napoli.
    - **Genere:** maschio, femmina.
- **Ordinali:**
  - **Definizione:** Variabili categoriali con un ordine logico, ma senza una distanza misurabile tra le categorie.
  - **Esempi:**

- **Livello di istruzione:** scuola elementare, scuola media, diploma, laurea.
- **Classi di reddito:** basso, medio, alto.
- **Opinioni:** totalmente d'accordo, d'accordo, neutro, in disaccordo, totalmente in disaccordo.

### Esempi pratici:

- **Genere:** Classificato come maschio o femmina (variabile nominale).
- **Opinione su un prodotto:** Valutata come eccellente, buona, sufficiente, insufficiente (variabile ordinale).

### b) Variabili Quantitative (o Numeriche)

Le **variabili quantitative** sono misurate numericamente e rappresentano quantità o misurazioni.

#### Sottotipi di variabili quantitative:

- **Discrete:**
  - **Definizione:** Variabili che possono assumere solo valori interi o conteggiabili.
  - **Esempi:**
    - **Numero di figli per famiglia:** 0, 1, 2, 3, ...
    - **Numero di auto vendute in un mese:** 15, 20, 25, ...
    - **Voti interi in un esame:** 18, 19, 20, 21, ...
- **Continue:**
  - **Definizione:** Variabili che possono assumere qualsiasi valore all'interno di un intervallo continuo.
  - **Esempi:**
    - **Altezza:** 170.5 cm, 172.3 cm, 168.8 cm.
    - **Peso:** 65.2 kg, 70.5 kg, 80.3 kg.
    - **Temperatura:** 36.6°C, 37.2°C, 38.0°C.

### Esempi pratici:

- **Età:** Può essere considerata come variabile continua (espressa in anni, mesi, giorni).
- **Temperatura corporea:** Misurata con precisione decimale.

### Rappresentazione Grafica dei Dati

La rappresentazione grafica è fondamentale per visualizzare e comprendere i dati. A seconda del tipo di variabile, si utilizzano grafici diversi.

## Grafici per Variabili Qualitative

- **Grafico a barre:**
  - **Descrizione:** Le categorie sono rappresentate sull'asse orizzontale, mentre la frequenza o percentuale sull'asse verticale. Ogni barra rappresenta una categoria e la sua altezza corrisponde alla frequenza.
  - **Uso:** Ideale per confrontare la frequenza tra diverse categorie.
- **Grafico a torta (o diagramma circolare):**
  - **Descrizione:** Il cerchio rappresenta il totale, suddiviso in "fette" proporzionali alla frequenza di ciascuna categoria.
  - **Uso:** Utile per rappresentare la proporzione di ogni categoria rispetto al totale.

### Esempio:

Supponiamo di aver condotto un sondaggio sulle preferenze di gusto di gelato tra vaniglia, cioccolato e fragola. Possiamo rappresentare le preferenze con un grafico a torta per visualizzare la percentuale di persone che preferiscono ciascun gusto.

## Grafici per Variabili Quantitative

- **Istogramma:**
  - **Descrizione:** Simile al grafico a barre, ma le barre sono adiacenti senza spazi, rappresentando la distribuzione di una variabile continua.
  - **Uso:** Mostra la frequenza delle osservazioni all'interno di intervalli (classi) di valori.
- **Diagramma a punti (o grafico scatter):**
  - **Descrizione:** Ogni punto rappresenta un'osservazione su un piano cartesiano, spesso utilizzato per visualizzare la relazione tra due variabili quantitative.
  - **Uso:** Utile per identificare correlazioni o trend tra variabili.

### Esempio:

Se raccogliamo i dati sull'altezza di un gruppo di persone, possiamo creare un istogramma per vedere come le altezze sono distribuite nella popolazione.

## **Esercizio di Esempio**

### **Attività interattiva:**

Chiediamo agli studenti di fornire esempi di variabili qualitative e quantitative.

### **Esempi da annotare:**

- **Variabili qualitative:**
  - **Tipo di abitazione:** appartamento, villa, casale.
  - **Marca di auto posseduta:** Fiat, Ford, Toyota.
- **Variabili quantitative:**
  - **Numero di libri letti in un anno.**
  - **Tempo impiegato per completare una maratona (in ore e minuti).**

### **Discussione:**

Analizziamo insieme se le variabili proposte sono qualitative o quantitative, e se quantitative, se sono discrete o continue. Inoltre, discutiamo su come potremmo rappresentare graficamente ciascuna di queste variabili.



### 3. Introduzione ai Concetti Fondamentali di Probabilità (15 minuti)

#### Definizione di Probabilità

La **probabilità** è una branca della matematica che studia l'incertezza e il caso. Assegna un valore numerico tra 0 e 1 a eventi o risultati, indicando quanto è probabile che un evento si verifichi.

- **Probabilità di 0:** L'evento è impossibile.
- **Probabilità di 1:** L'evento è certo.

#### Esempio pratico:

- **Lancio di una moneta:**
  - Probabilità di ottenere "testa": 0.5
  - Probabilità di ottenere "croce": 0.5

#### Concetti chiave della Probabilità

- **Esperimento casuale:** Un processo che produce un risultato incerto (es. lancio di un dado).
- **Spazio campionario (S):** L'insieme di tutti i possibili risultati di un esperimento casuale.
- **Evento:** Un sottoinsieme dello spazio campionario; uno o più risultati specifici.
- **Probabilità di un evento (P):** Misura la possibilità che l'evento si verifichi.

#### Esempio:

- **Lancio di un dado a sei facce:**
  - **Spazio campionario S:** {1, 2, 3, 4, 5, 6}
  - **Evento A (ottenere un numero pari):** {2, 4, 6}
  - **Probabilità di A:**  $P(A) = \text{numero di risultati favorevoli} / \text{numero totale di risultati} = 3/6 = 0.5$

#### Relazione tra Statistica e Probabilità

La probabilità fornisce la base teorica per la statistica inferenziale. Mentre la statistica si occupa di analizzare dati reali, la probabilità si concentra sui modelli teorici di eventi casuali.

### **Esempio pratico:**

- **Test di ipotesi:** Utilizziamo la probabilità per determinare se un risultato ottenuto da un campione è compatibile con un'ipotesi sulla popolazione.

## 4. Importanza dell'Analisi Descrittiva (10 minuti)

### Perché l'Analisi Descrittiva è Fondamentale

- **Comprensione iniziale dei dati:** Prima di procedere con analisi più complesse, è essenziale capire le caratteristiche fondamentali del dataset.
- **Identificazione di errori o anomalie:** L'analisi descrittiva può evidenziare dati mancanti, errori di inserimento o outlier.
- **Comunicazione efficace:** Grafici e tabelle aiutano a presentare i dati in modo chiaro a un pubblico non specializzato.

### Processo dell'Analisi Descrittiva

1. **Raccolta dei dati:** Assicurarsi che i dati siano accurati e completi.
2. **Organizzazione dei dati:** Creare tabelle e ordinare i dati per categoria o valore.
3. **Calcolo delle misure statistiche:** Calcolare media, mediana, moda, varianza, ecc.
4. **Visualizzazione dei dati:** Creare grafici appropriati per rappresentare le informazioni in modo visivo.
5. **Interpretazione dei risultati:** Analizzare ciò che le misure e i grafici indicano sulle caratteristiche dei dati.

### Esempio pratico:

Supponiamo di avere i risultati di un questionario sulla soddisfazione dei clienti di un'azienda.

- **Passo 1:** Organizzare i dati in una tabella, suddividendo le risposte per categoria (es. molto soddisfatto, soddisfatto, insoddisfatto).
  - **Passo 2:** Calcolare la percentuale di clienti in ciascuna categoria.
  - **Passo 3:** Creare un grafico a torta per visualizzare le proporzioni.
  - **Passo 4:** Interpretare i risultati, identificando se la maggior parte dei clienti è soddisfatta o se ci sono aree di miglioramento.
-

## Conclusione della Lezione

In questa lezione, abbiamo introdotto i concetti fondamentali della statistica e della probabilità, concentrandoci sui tipi di variabili e sull'importanza dell'analisi descrittiva. Abbiamo visto come la statistica sia essenziale per comprendere e interpretare i dati nel mondo reale, fornendo strumenti per prendere decisioni informate.

### **Punti chiave da ricordare:**

- La statistica è la scienza della raccolta, analisi e interpretazione dei dati.
- Esistono due principali rami della statistica: descrittiva e inferenziale.
- Le variabili possono essere qualitative (nominali o ordinali) o quantitative (discrete o continue).
- La rappresentazione grafica è fondamentale per visualizzare i dati e comunicare le informazioni.
- La probabilità fornisce la base teorica per l'inferenza statistica.
- L'analisi descrittiva è il primo passo essenziale in qualsiasi studio statistico.

## Laboratorio pratico (2 ore)

Durante queste due ore di laboratorio, gli studenti applicheranno i concetti teorici di statistica usando Microsoft Excel. Attraverso esercizi pratici, creeranno un dataset, classificheranno le variabili e utilizzeranno strumenti di analisi descrittiva.

---

### Attività 1: Creazione di un Dataset su Excel

**Obiettivo:** Creare un piccolo database con dati raccolti dai compagni di classe, includendo variabili sia qualitative che quantitative.

#### Passaggi dettagliati

##### 1. Preparazione dell'ambiente di lavoro:

- **Avvio di Excel:**
  - Aprire Microsoft Excel e creare un nuovo foglio di lavoro.
- **Salvataggio del file:**
  - Salvare immediatamente il file con nome "Dataset\_Statistica\_NomeCognome.xlsx" per evitare perdite di dati.

##### 2. Definizione delle variabili da raccogliere:

- **Discussione delle variabili:**
  - Insieme all'insegnante, gli studenti decideranno quali variabili raccogliere, come:
    - Età (quantitativa, discreta)
    - Genere (qualitativa, nominale)
    - Voto in Matematica (quantitativa, discreta o continua)
    - Materia preferita (qualitativa, nominale)
    - Numero di ore di studio settimanali (quantitativa, continua)
    - Sport praticato (qualitativa, nominale)
    - Numero di fratelli/sorelle (quantitativa, discreta)
    - Colore preferito (qualitativa, nominale)
    - Possesso di animali domestici (qualitativa, nominale)

##### 3. Creazione dell'intestazione delle colonne in Excel:

- Inserire nella prima riga del foglio di lavoro i nomi delle variabili come nell'esempio sottostante:

A	B	C	D	E	F	G	H
ID Studente	Età	Genere	Voto Matematica	Materia Preferita	Ore di Studio Settimanali	Sport Praticato	Numero di Fratelli/Sorelle

##### 4. Inserimento dei dati:

- **Assegnazione di un ID univoco:** Ogni studente assegna a se stesso un numero ID univoco.
- **Compilazione dei dati personali:** Ogni studente inserisce i propri dati nelle colonne. Di seguito un esempio di compilazione:

ID Studente	Età	Genere	Voto Matematica	Materia Preferita	Ore di Studio Settimanali	Sport Praticato	Numero di Fratelli/Sorelle
1	16	F	8	Inglese	10	Nuoto	1
2	17	M	7	Storia	8	Calcio	2
3	16	F	9	Matematica	12	Pallavolo	0

#### 5. Classificazione delle variabili nel dataset:

- Aggiungere una riga o colonna con la classificazione delle variabili. Ecco un esempio:

Variabile	Tipo	Sottotipo
Età	Quantitativa	Discreta
Genere	Qualitativa	Nominale
Voto Matematica	Quantitativa	Discreta
Materia Preferita	Qualitativa	Nominale
Ore di Studio Settimanali	Quantitativa	Continua
Sport Praticato	Qualitativa	Nominale
Numero di Fratelli/Sorelle	Quantitativa	Discreta

#### 6. Verifica e salvataggio dei dati:

- **Controllo:** Verificare che i dati inseriti siano corretti e privi di errori.
- **Salvataggio:** Salvare nuovamente il file per garantire che tutte le modifiche siano memorizzate.

## Attività 2: Analisi Descrittiva dei Dati

**Obiettivo:** Analizzare i dati inseriti utilizzando strumenti descrittivi di base come tabelle di frequenza e grafici semplici.

### Passaggi dettagliati

#### 1. Creazione di tabelle di frequenza:

- **Per variabili qualitative (ad es. Genere):**
  - Creare una tabella di frequenza con due colonne: "Genere" e "Frequenza".
  - Utilizzare la funzione **CONTA.SE** per calcolare il numero di occorrenze di ciascuna categoria.

Genere	Frequenza
M	=CONTA.SE(\$C\$2:\$C\$31,"M")
F	=CONTA.SE(\$C\$2:\$C\$31,"F")

- **Per altre variabili qualitative (ad es. Materia Preferita):**
  - Elencare tutte le materie e contare il numero di studenti che le preferiscono usando la funzione **CONTA.SE**.

Materia Preferita	Frequenza
Inglese	=CONTA.SE(\$E\$2:\$E\$31,"Inglese")
Matematica	=CONTA.SE(\$E\$2:\$E\$31,"Matematica")
Storia	=CONTA.SE(\$E\$2:\$E\$31,"Storia")

## 2. Calcolo di misure di tendenza centrale e dispersione:

- **Per variabili quantitative (ad es. Età):**
  - Media: =MEDIA (B2 : B31)
  - Mediana: =MEDIANA (D2 : D31)
  - Moda: =MODA . UNO (H2 : H31)
  - Deviazione Standard: =DEV . ST (F2 : F31)

Misura	Variabile	Formula
Media	Età	=MEDIA (B2 : B31)
Mediana	Voto Matematica	=MEDIANA (D2 : D31)
Moda	Numero di Fratelli/Sorelle	=MODA . UNO (H2 : H31)
Dev. Standard	Ore di Studio Settimanali	=DEV . ST (F2 : F31)

## 3. Creazione di grafici:

- **Grafico a barre (istogramma) per variabili quantitative discrete (ad es. Numero di Fratelli/Sorelle):**
  - Selezionare la tabella di frequenza e andare su "Inserisci" > "Grafico a colonne" > "Colonne raggruppate".
- **Grafico a torta per variabili qualitative (ad es. Genere):**
  - Selezionare la tabella di frequenza per il genere e andare su "Inserisci" > "Grafico a torta".
- **Istogramma per variabili quantitative continue (ad es. Ore di Studio Settimanali):**
  - Selezionare i dati e utilizzare l'analisi dei dati di Excel ("Dati" > "Analisi dati" > "Istogramma").
- **Grafico a dispersione per analizzare la relazione tra due variabili (ad es. Ore di Studio Settimanali e Voto Matematica):**

- Selezionare le colonne delle due variabili e andare su "Inserisci" > "Grafico a dispersione".
  - 4. **Analisi e interpretazione dei risultati:**
    - **Osservazione dei grafici:**
      - Identificare categorie o valori più frequenti.
    - **Discussione dei risultati:**
      - Qual è la materia preferita più comune? Esiste una relazione tra ore di studio e voti?
- 

### **Attività aggiuntive (se il tempo lo consente):**

- **Esplorazione di correlazioni:** Calcolare il coefficiente di correlazione tra due variabili quantitative utilizzando `=CORRELAZIONE`.
  - **Creazione di tabelle pivot:** Utilizzare le tabelle pivot di Excel per sintetizzare i dati e ottenere nuove prospettive.
  - **Analisi di dati esterni:** Importare un dataset pubblico e applicare le stesse tecniche di analisi.
- 

### **Domande di Riflessione:**

1. Quali difficoltà avete riscontrato durante la raccolta e l'inserimento dei dati?
2. Come ha contribuito la creazione di grafici alla comprensione dei dati?
3. Cosa avete osservato analizzando la relazione tra due variabili?
4. In che modo pensate che la dimensione del campione possa influenzare i risultati?