



Cofinanziato dal
programma Erasmus+
dell'Unione europea



Il supporto della Commissione Europea per la produzione di questa pubblicazione non costituisce un avallo del contenuto che riflette solo il punto di vista degli autori, e la Commissione non può essere ritenuta responsabile per qualsiasi uso che potrebbe essere fatto delle informazioni ivi contenute.

2. Statistica

L'Oxford Learner's Dictionary definisce la **statistica** come un insieme di informazioni rappresentate da numeri. La statistica è un campo matematico che si occupa dei metodi di raccolta dei dati e della loro analisi. Un esempio molto semplice di tale trattamento dei dati è la rubrica sportiva in un giornale. Diciamo che si gioca a Wimbledon e sul giornale troviamo le seguenti informazioni:

Final - Center Court		Final
 11 S. Williams		2 2
 7 S. Halep		6 6 ◀

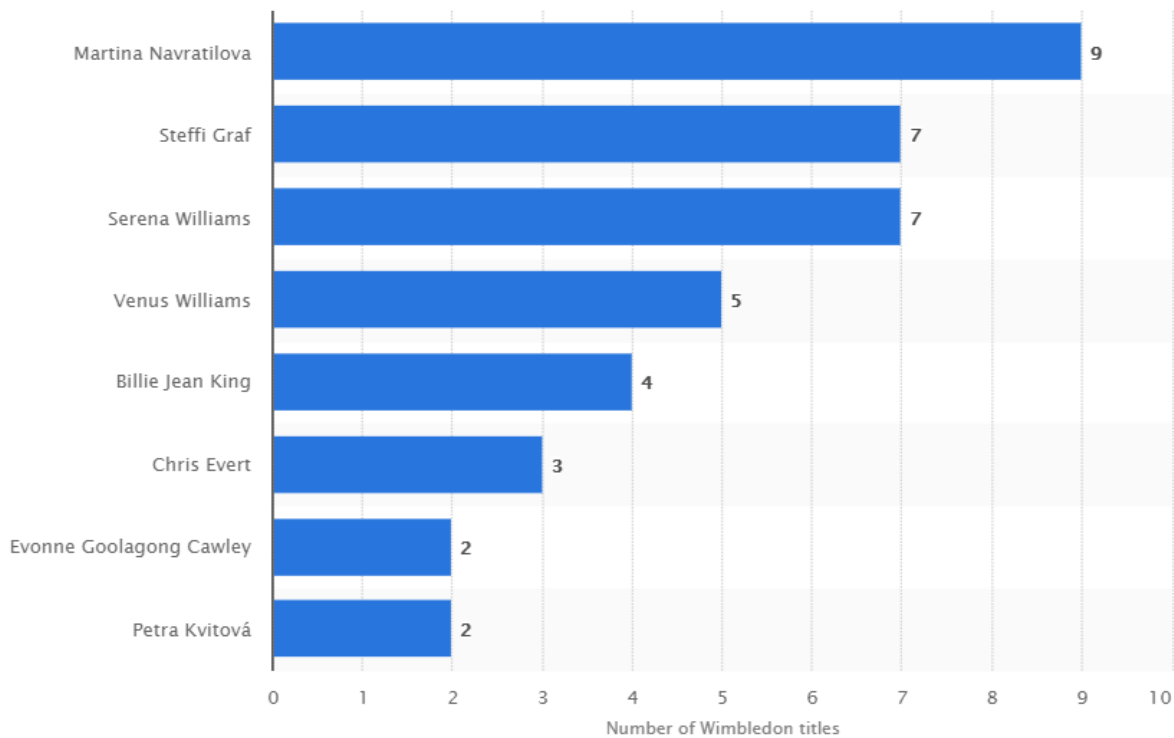
 Serena Williams		Simona Halep 
2	Aces	1
1	Double faults	0
68%	First serve %	76%
59%	Win % on 1st serve	83%
50%	Win % on 2nd serve	45%
0/1	Break points	4/5
0	Tiebreaks won	0
12	Receiving points won	21
38	Points won	55
4	Games won	12
1	Max games won in a row	5
5	Max points won in a row	7

Nella foto vediamo le statistiche dell'incontro finale tra Simona Halepova del Ru,unia e Serena Williamsova degli USA. Ogni numero in queste statistiche ha la sua "storia" ed esprime qualcosa. In alto a destra dell'immagine, c'è una tabella la cui prima e seconda colonna contengono gli stessi valori, ovvero 2 e 6. Ogni colonna rappresenta un insieme dell'incontro e ogni riga indica quante partite (gemme) ogni giocatore ha vinto in esso. Nella seconda parte dell'immagine ci sono ulteriori informazioni sull'incontro. Possiamo interpretare questi numeri come risposte a domande specifiche sulla partita:

- Quanti punti ha segnato Serena Williams?
 - Risposta: 2
- Quanti falli ha fatto Simona Halep?
 - Risposta: 0

- Quanto successo ha avuto il primo servizio di Simona Halep?
 - Risposta : 76%
- Quante partite ha vinto Serena Williams?
 - Risposta: 4 (2 nel primo set e 2 nel secondo set, che è indicato nell'angolo in alto a destra)

L'esempio appena riportato è una dimostrazione dell'elaborazione dei dati ottenuti durante una partita di tennis. Tuttavia, possiamo anche porre altre domande, come quale tennista ha vinto più titoli in singolare a Wimbledon dal 1968 al 2019. Per rispondere a questa domanda, abbiamo bisogno di altri dati:



© Statista .

Questa volta dobbiamo ottenere i dati di chi ha vinto la finale di Wimbledon tra il 1968 e il 2019.

I passaggi comuni che uno statistico esegue nel suo lavoro sono:

1. Raccolta dati:
 - Set di vincitrici delle finali femminili dal 1968 al 2018.
2. Organizzazione dei dati in tabelle e grafici:
 - Classifica dei giocatori in base al numero di titoli vinti in singolare.
3. Trarre conclusioni dall'analisi dei dati:
 - Il grafico sopra mostra che Martina Navrátilová ha vinto il maggior numero di titoli.

Se questi tre passaggi fanno parte dell'elaborazione statistica, allora si parla di **statistica descrittiva**. I dati grezzi sono per lo più sotto forma di elenco, file o database.

Come raccogliere dati grezzi?

Come risulta dagli esempi, i dati con cui lavoriamo possono essere di diverso tipo. Possiamo guardare i dati come risultato:

1. **Questionario**

- Un insieme di domande (elementi) in cui possiamo scrivere la risposta o scegliere una delle opzioni offerte.
- Esempio: censimento della popolazione, elezioni

2. **Record**

- O un diario in cui registriamo regolarmente informazioni.
- Esempio: record ospedalieri (farmaci somministrati ogni 8 anni), temperatura esterna (temperatura dell'aria alle 7:00 tutti i giorni), vittoria di Wimbledon (vincitore della finale femminile, una volta all'anno).

Un censimento della popolazione è un buon esempio di come raccogliamo informazioni da una popolazione sotto osservazione. Sulla base dei dati ottenuti dal censimento della popolazione, la dirigenza statale può prendere decisioni su, ad esempio, come redistribuire denaro ad aree più o meno popolate e simili.

Tuttavia, nella maggior parte dei casi non possiamo ottenere informazioni dall'intera popolazione (set di base). Un buon esempio è un'agenzia pubblicitaria che è interessata a quale programma TV è più visto. Con queste informazioni, l'azienda può guadagnare tempo pubblicitario in questo programma e massimizzare l'impatto della campagna pubblicitaria sulla popolazione. Sarebbe certamente costoso in termini di tempo e denaro procedere come nel caso del censimento della popolazione e quindi cercare di raggiungere l'intero gruppo target. Un modo più pratico per ottenere queste informazioni richiede il cosiddetto **file di selezione**. Il campione è un sottoinsieme (parte) della popolazione da noi scelta.

La scelta del file campione deve soddisfare determinati criteri:

Tuttavia, nella maggior parte dei casi non siamo in grado di raggiungere l'intera popolazione di interesse. Considera, per esempio un'agenzia pubblicitaria interessata a quale programma televisivo è il più visto. Con queste informazioni, l'agenzia può guadagnare tempo pubblicitario nei programmi dati e massimizzare l'impatto sulla popolazione. Sarebbe certamente molto costoso e dispendioso in termini di tempo imitare un censimento e cercare di raggiungere tutti i presenti. Questo studio richiede quello che chiamiamo **campione**. Un campione è una parte della "popolazione" che stiamo considerando.

1. Il campione deve **rappresentare l'intera popolazione in studio**,
2. Il file di selezione deve contenere un **numero sufficientemente elevato** di
3. Dovrebbero essere utilizzati **metodi di campionamento appropriati e selezione casuale**.

TROVA QUALCHE BUON ESEMPIO DI FILE DI SELEZIONE E DI VOTO (IL SITO YUGOV PUÒ ESSERE UTILIZZATO) E SEMPLICEMENTE SPIEGATELO!!!!

L'agenzia pubblicitaria ha incaricato l'agenzia di ricerca YouGov di aiutarli a capire quale prima serata sarebbe appropriato per pubblicizzare un determinato prodotto nel Regno Unito. YouGov creerà un questionario online, che poi invierà ai suoi utenti registrati, che rappresentano tutte le fasce d'età, i gruppi socio-economici e altri gruppi demografici. Il gruppo di utenti è attualmente composto da circa 1 milione di utenti nel Regno Unito.

Che aspetto hanno i dati grezzi?

Il metodo di organizzazione dei dati è stato presentato all'inizio del capitolo. Abbiamo utilizzato due esempi di presentazione dei dati basati sui risultati delle finali femminili di Wimbledon.

Ora diamo un'occhiata a un esempio più tradizionale. Una classe di 32 studenti ha scritto il test con i seguenti risultati:

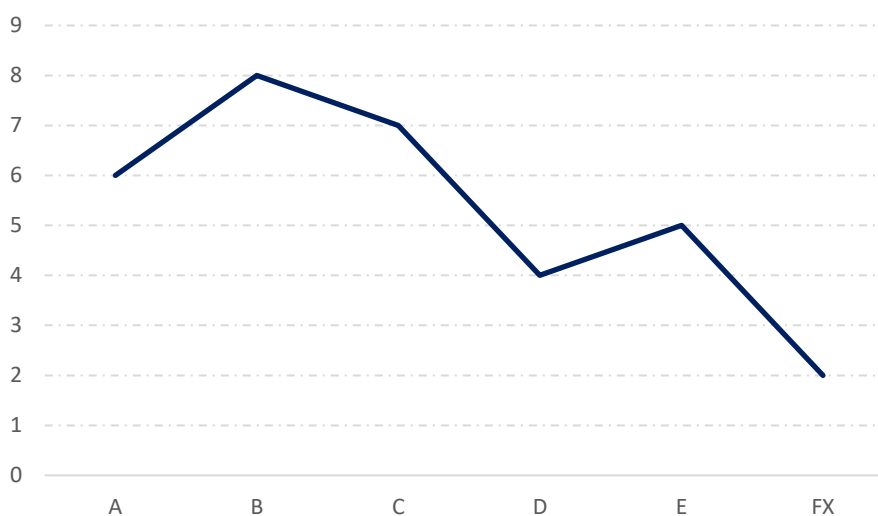
82, 90, 61, 80, 88, 96, 58, 74, 69, 82, 53, 78, 72, 58, 54, 95, 99, 86, 80, 95, 99, 88, 83, 51, 65, 77, 28, 82, 70, 73, 100, 19

La soglia per il superamento del test è di 50 punti. Ogni studente che ottiene più di 50 punti supera la materia e ottiene la valutazione secondo la chiave data nella tabella. Di seguito gli strumenti per l'organizzazione e la rappresentazione dei dati:

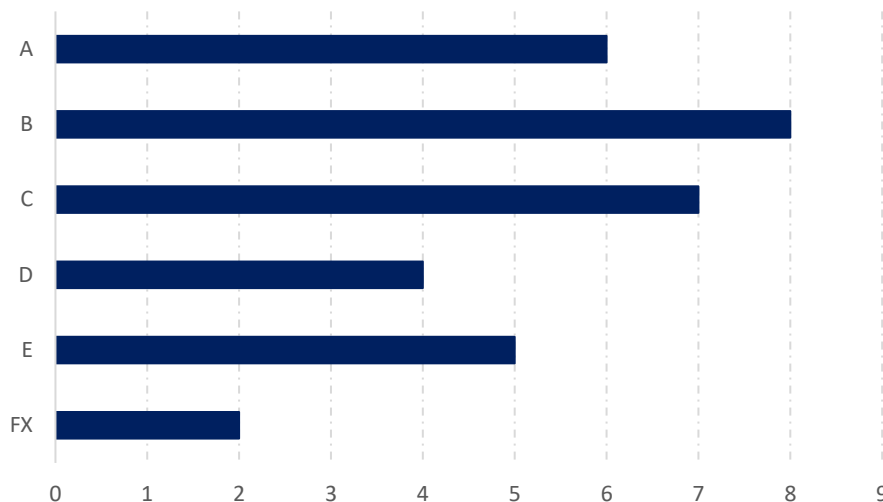
1. Tavolo:

Valutazione	Il punteggio	# degli studenti	% degli studenti
UN	100 - 91	6	19%
B	90-81	8	25%
C	80-71	7	22%
D	70-61	4	13%
e	60-51	5	16%
FX	50 - 0	2	6%

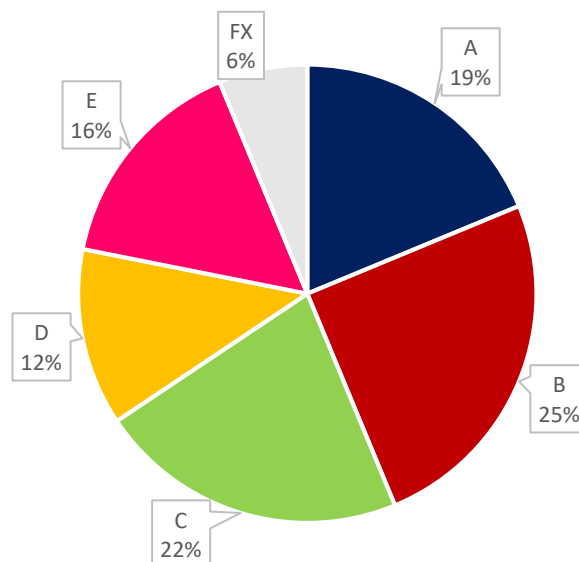
2. Grafico a linee:



3. Grafico a barre:



4. Grafico a torta:



Cosa significano, mediana e moda rappresentano?

Abbiamo già discusso di come raccogliere, organizzare e presentare i dati. Semplicemente guardando i grafici, possiamo trovare informazioni sulla valutazione più frequente. Per comprendere altre informazioni, dobbiamo prima spiegare concetti come media, moda e mediana.

1. **Media:**

- In statistica, la media aritmetica è chiamata **media**.
- Nel nostro caso, potremmo essere interessati a qual è stato il numero medio di punti ottenuti. nel test o quale fosse il punteggio medio di ciascun livello scolastico.
- **Calcoliamo la media di un insieme di N numeri sommando tutti i numeri e dividendoli per il loro numero.**
- La prima media richiederà quindi di sommare i punteggi di tutti gli studenti che hanno scritto il test = 2385) e di dividere tale somma per il loro numero (= 32); il numero medio di punti è **74,53**.
- Il numero medio di punti per ogni livello richiederà un passaggio in più.
- In questa fase è necessario identificare il numero di punti ottenuti dagli studenti con una determinata valutazione (per la valutazione A: 96, 95, 99, 95, 99 e 100).

- Fatto ciò, procediamo come nel caso precedente; sommiamo i punteggi e li dividiamo per il numero di studenti con il punteggio dato (cioè $(96+95+99+95+99+100)/6 = 97,33$).

Valutazione	Media
UN	97.33
B	85.13
C	76.29
D	66.25
e	54.80
FX	23.50
In totale	74.53

2. Mediano:

- Spesso è vantaggioso scoprire quale valore si trova nel mezzo dei dati.
- **La mediana è il valore medio in un insieme di valori ordinati; nel caso in cui la sequenza abbia un numero pari di membri, la media aritmetica dei due valori medi è considerata valore medio.**
- Perché dovremmo interessarci a questo valore?
- Immaginiamo cinque piatti di biscotti; quattro di loro hanno cinque biscotti ciascuno e il quinto è vuoto.
- Cinque bambini vengono in cucina, ognuno prende un piatto e mangia tutto quello che c'è sopra.
- La media ci dice che un bambino ha ricevuto $(5+5+5+5+0)/5 = 4$ biscotti.
- Sappiamo però che a uno dei bambini non è sfuggito un biscotto e di certo non sarebbe d'accordo con l'affermazione che ha mangiato in media 4 biscotti.
- Il numero mediano di biscotti sui piatti è il valore medio della sequenza : 0, 5, 5, 5, 5 => mediana = 5.
- Immaginiamo cinque piatti di biscotti; ci sono cinque biscotti su quattro piatti e il quinto è vuoto.
- Cinque bambini vengono in cucina, ognuno prende un piatto e mangia tutto quello che c'è sopra.
- La media ci dice che abbiamo avuto $(5+5+5+5+0)/5 = 4$ biscotti per bambino.
- Tuttavia, sappiamo che un bambino non aveva nemmeno un biscotto e di certo non sarebbe d'accordo con l'affermazione che ha mangiato in media 4 biscotti.
- La mediana di queste piastre sarà il valore medio di questa sequenza: 0, 5, 5, 5, 5 => mediana = 5.
- La media ci dice che c'erano meno di cinque biscotti su uno o più piatti e alcuni bambini potrebbero essere ancora affamati.
- Possiamo invece avere quattro piatti vuoti e cinque biscotti su uno; la media dice che ogni bambino ha mangiato un biscotto $(0+0+0+0+5)/5 = 1$, ma la mediana offre una visione più realistica.
- In questo caso, la mediana è pari a zero, quindi dovremmo dare un'occhiata più da vicino a come sono stati distribuiti i biscotti e se per caso alcuni bambini hanno fame.
- Nella tabella seguente, mostriamo il numero mediano di punti del test per ciascun livello di valutazione. Cosa implicano i risultati?
- In questo caso la mediana è uguale a 0 e quindi bisognerebbe dare un'occhiata più da vicino a come sono stati distribuiti i biscotti e quanti bambini hanno ancora fame.
- Di seguito presentiamo la mediana per ogni grado e la mediana complessiva; cosa possiamo leggere dai risultati?

Valutazione	Mediano
UN	97.50
B	84.50
C	77.00
D	67.00
e	54.00
FX	23.50
Complessivamente	79.00

3. Modalità:

- Quando si analizzano i dati, potremmo essere interessati a quale valore si verifica più spesso; tale valore è chiamato **modalità**.
- Disporremo nuovamente i numeri in ordine non decrescente e vedremo quale valore si verifica più spesso in esso.
- Nel nostro caso, è il numero **82** ; tre studenti hanno ottenuto 82 punti, due studenti ciascuno ha segnato 58, 80, 88, 95 e 99 punti.
- Cosa accadrebbe se tre studenti segnavero 82 punti e gli stessi tre studenti segnavero 58 punti?
- In tal caso , dovremmo affermare che l'insieme ha **due modalità** ed è quindi **bimodale**.
- E se nessuno studente ricevesse lo stesso numero di punti? (cioè ciascuno dei valori nella sequenza si trova esattamente una volta).

Fonti esterne di approfondimento:

<https://www.statista.com/statistics/280393/womens-tennis-players-with-the-most-victories-at-wimbledon/>

<https://yougov.co.uk/about/panel-methodology/>

Dressler e Keenan: Matematica Integrata (Seconda Edizione)

https://books.google.sk/books/about/Integrated_Mathematics_Course_1.html?id=h4oiXcjGT3oC&redir_esc=y