MADDALENA CAVICCHIOLI • MARIA FRANCO VILLORIA • PATRIZIO FREDERIC • ISABELLA MORLINI

CHE COS'È LA STATISTICA?

UNA PRIMA INTRODUZIONE ALLA SCIENZA DEI DATI

UNIMORE
UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

Dipartimento di Economia
Marco Biagi

Sommario

Prefazione

Michele Lalla

L'insegnamento della probabilità e della statistica nelle scuole ha incontrato diverse difficoltà di natura sia concettuale, perché tratta nozioni di non facile comprensione da parte dei discenti, sia professionale, perché ha destato spesso un certo malessere tra i docenti per l'aumento del programma da svolgere senza il corrispondente aumento del numero di lezioni e per la carenza di preparazione in probabilità e statistica nella loro formazione concernente l'insegnamento di matematica (Ottaviani, 2011).
Le discipline della probabilità e della statistica svolgono un ruolo importante nella formazione dei discenti perché promuovono lo sviluppo del pensiero critico, necessario per comprendere la realtà nella sua dimensione quantitativa, e aiutano a prendere decisioni razionali anche nel quotidiano (tra tanti, Ottaviani, 2008; Rampichini, 2008; Da Valle et al., 2015).
Probabilità e statistica sono discipline relativamente recenti; entrambe hanno iniziato il loro sviluppo nel XVII secolo, ma la seconda ha avuto una sistemazione/ formalizzazione un po' più lenta della prima. L'introduzione del loro insegnamento, per semplicità non si distingue tra le due, perfino all'università, all'inizio è stato episodico e altalenante (Meusnier, 2006). Il dibattito sull'inserimento dell'insegnamento dei concetti di base di queste due discipline in tutte le scuole secondarie superiori è cominciato agli inizi degli anni cinquanta del XX secolo (Dutka e Kafka, 1950), se si escludono i percorsi formativi che includevano la trattazione delle assicurazioni, le quali necessariamente dovevano includere nozioni di probabilità, come accadeva in alcuni istituti tecnici commerciali.

Il presente testo è sicuramente un buon ausilio per gli studenti della scuola secondaria di primo grado e del primo biennio della Scuola secondaria superiore e sembra adeguato alla loro maturità e preparazione, ma può essere anche un utile strumento per i docenti che devono insegnare i concetti di probabilità e statistica. Il testo potrebbe essere idoneo anche agli studenti della Scuola secondaria inferiore, ma è presumibile che non desti opportunamente il loro interesse, mentre è certamente utile agli insegnanti e perfino ai genitori che intendono seguire i loro figli nello studio dei concetti di base qui esaminati. Per la Scuola primaria e per la scuola dell'infanzia può suscitare l'interesse del corpo insegnante, perché offre molti spunti per trattare i vari concetti in classe.
Una delle indicazioni normative più recente e esaustiva è il Decreto del Ministero dell'Istruzione, dell'Università e della Ricerca (DM 254/2012), che riporta le indicazioni dei curricoli della scuola dell'infanzia e del primo ciclo d'istruzione, pubblicato nella Gazzetta Ufficiale n. 30 del 05-02-2013 e entrato in vigore dal 20-02-2013.

1. Scuola dell'Infanzia

La scuola dell'infanzia "si pone la finalità di promuovere nei bambini lo sviluppo dell'identità, dell'autonomia, della competenza e li avvia alla cittadinanza" (DM 254/2012, p.18). Dal complesso discorso affrontato dal documento (DM 254/2012), si estrae qualche elemento chiave dai riquadri "Traguardi per lo sviluppo della competenza" nei paragrafi di interesse per l'argomento in oggetto:

- dal paragrafo "Il sé e l'altro" si considera che il bambino "sa argomentare, confrontarsi, sostenere le proprie ragioni, … sviluppa il senso dell'identità personale, percepisce i propri sentimenti, sa esprimerli in modo sempre più adeguato"; insomma, ascolta gli altri e dà spiegazioni del proprio comportamento e del proprio punto di vista;
- dal paragrafo "Il corpo in movimento" si considera il sintagma "valuta il rischio";
- dal paragrafo "I discorsi e le parole" si legge: "comprende i discorsi e fa ipotesi sui significati … usa il linguaggio per progettare attività e per definire regole" sicché è in grado di formulare un piano di azione, individualmente e in gruppo, e scegliere con attenzione materiali e strumenti in relazione al progetto da realizzare;
- da "La conoscenza del mondo" si ha "Il bambino raggruppa e ordina oggetti e materiali secondo criteri diversi, ne identifica alcune proprietà, confronta e valuta quantità; utilizza simboli per registrarle; esegue misurazioni usando strumenti alla sua portata. …. Osserva con attenzione [e sistematicità] … gli organismi viventi e i loro ambienti, i fenomeni naturali, accorgendosi dei loro cambiamenti [sulla base di criteri o ipotesi]; in altre parole, è curioso, esplora l'ambiente, pone domande, discute, confronta ipotesi, valuta spiegazioni, esamina soluzioni e azioni.

Alla fine del triennio della scuola dell'infanzia, "è ragionevole attendersi che ogni bambino abbia sviluppato alcune competenze di base" e abbia migliorato l'uso del linguaggio appropriato per descrivere le osservazioni o le esperienze.

2. Scuola del primo ciclo

La scuola del primo ciclo di istruzione comprende la scuola primaria e la scuola secondaria di primo grado (DM 254/2012, p. 26) e ha l'obiettivo di conseguire "l'acquisizione delle conoscenze e delle abilità fondamentali per sviluppare le competenze culturali di base nella prospettiva del pieno sviluppo della persona"; insomma, si è cercato di attivare un sistema denominato anche integrazione verticale.

2.1. Scuola primaria

La scuola primaria deve mirare "all'acquisizione degli apprendimenti di base… [allo sviluppo delle] dimensioni cognitive, emotive, affettive, sociali, corporee, etiche e religiose, e di acquisire saperi irrinunciabili… ponendo le premesse per lo sviluppo del pensiero riflessivo e critico" (DM 254/2012, p. 27). Dal complesso discorso affrontato dal documento (DM 254/2012), si estrae qualche elemento chiave dal riquadro "Traguardi per lo sviluppo della competenza" nel paragrafo di interesse per l'argomento in oggetto che concerne l'insegnamento di matematica (DM 254/2012, p. 51) il quale deve fornire "strumenti per la descrizione scientifica del mondo e per affrontare problemi utili nella vita quotidiana":

- ricerca dati per ricavare informazioni e costruisce rappresentazioni (tabelle grafici). Ricava informazioni anche da dati rappresentati in tabelle e grafici;
- riconosce e quantifica, in casi semplici, situazioni di incertezza.

2.2. Scuola Secondaria di Primo Grado

La scuola secondaria di primo grado dovrebbe realizzare un approccio alle varie discipline come "punti di vista della realtà e come modalità di conoscenza, interpretazione e rappresentazione del mondo" (DM 254/2012, p. 27), evitando la suddivisione in compartimenti stagni dei saperi e favorendo le loro interrelazioni che si hanno proprio nelle zone di contatto e sovrapposizione. Bisogna favorire una padronanza più approfondita "delle discipline e una articolata organizzazione delle conoscenze". Si estrae qualche elemento chiave dal riquadro "Traguardi per lo sviluppo della competenza" nel paragrafo dell'insegnamento di matematica (DM 254/2012, p. 53):
- analizza e interpreta rappresentazioni di dati per ricavarne misure di variabilità e prendere decisioni;
- riconosce e risolve problemi in contesti diversi valutando le informazioni e la loro coerenza;
- spiega il procedimento seguito, anche in forma scritta, mantenendo il controllo sia sul processo risolutivo, sia sui risultati;
- confronta procedimenti diversi e produce formalizzazioni che gli consentono di passare da un problema specifico a una classe di problemi;
- nelle situazioni di incertezza (vita quotidiana, giochi, ...) si orienta con valutazioni di probabilità;
- ha rafforzato un atteggiamento positivo rispetto alla matematica attraverso esperienze significative e ha capito come gli strumenti matematici appresi siano utili in molte situazioni per operare nella realtà.

Sono stati definiti, in particolare, gli obiettivi specifici di apprendimento al termine della classe terza della scuola secondaria di primo grado: si riportano quelli che interessano la probabilità e la statistica si trovano nel titolo "Dati e Previsioni" e sono i seguenti (DM 254/2012, p. 54).
- Rappresentare insiemi di dati, anche facendo uso di un foglio elettronico. In situazioni significative, confrontare dati al fine di prendere decisioni, utilizzando le distribuzioni delle frequenze e delle frequenze relative. Scegliere ed utilizzare valori medi (moda, mediana, media aritmetica) adeguati alla tipologia ed alle caratteristiche dei dati a disposizione. Saper valutare la variabilità di un insieme di dati determinandone, ad esempio, il campo di variazione.
- In semplici situazioni aleatorie, individuare gli eventi elementari, assegnare a essi una probabilità, calcolare la probabilità di qualche evento, scomponendolo in eventi elementari disgiunti.
- Riconoscere coppie di eventi complementari, incompatibili, indipendenti.

3. Scuola Secondaria di Secondo Grado

Per la scuola seconda di secondo grado ci limita al primo biennio, perché è quello che potrebbe essere interessato a questo prodotto. Si distinguono tra licei, tecnici, e professionali ricordando che nella tabella "Dati e Previsioni" si considerano i nuclei di riferimento riguardanti la Statistica e la Probabilità.

3.1. Scuola Secondaria di Secondo Grado: LICEI primo biennio

Si riportano le indicazioni relative al nucleo "Dati e Previsioni" riguardanti i Licei Scientifici, e si mettono tra parentesi quadre le voci assenti nelle indicazioni delle altre tipologie di liceo.

- Lo studente deve essere in grado di rappresentare e analizzare in diversi modi (anche utilizzando strumenti informatici) un insieme di dati, scegliendo le rappresentazioni più idonee. Deve distinguere tra caratteri qualitativi, quantitativi discreti e quantitativi continui, operare con distribuzioni di frequenze e rappresentarle. Deve conoscere le definizioni e le proprietà dei valori medi e delle misure di variabilità, nonché l'uso di strumenti di calcolo (calcolatrice, foglio di calcolo) per analizzare raccolte di dati e serie statistiche. Lo studio deve essere svolto il più possibile in collegamento con le altre discipline anche in ambiti entro cui i dati siano raccolti direttamente dagli studenti.
- Lo studente è in grado di ricavare semplici inferenze dai diagrammi statistici.
- Egli apprenderà la nozione di probabilità, con esempi tratti da contesti classici e con l'introduzione di nozioni di statistica.
- Si approfondirà in modo rigoroso il concetto di modello matematico, distinguendone la specificità concettuale e metodica rispetto all'approccio della fisica classica.

L'ultimo punto è meno pertinente ai temi della probabilità e della statistica.

3.2. Scuola Secondaria di Secondo Grado: TECNICI primo biennio

Si riportano sempre le indicazioni relative al nucleo "Dati e Previsioni riportiamo" relative al Biennio, che sono comunque comuni agli indirizzi Economico e Tecnologico.

Conoscenze
- Dati, loro organizzazione e rappresentazione. Distribuzioni delle frequenze a seconda del tipo di carattere e principali rappresentazioni grafiche. Valori medi e misure di variabilità.
- Significato della probabilità e sue valutazioni. Semplici spazi (discreti) di probabilità: eventi disgiunti, probabilità composta, eventi indipendenti. Probabilità e frequenza.

Abilità
- Raccogliere, organizzare e rappresentare un insieme di dati. Calcolare i valori medi e alcune misure di variabilità di una distribuzione.
- Calcolare la probabilità di eventi elementari.

3.3. Scuola Secondaria di Secondo Grado: PROFESSIONALI primo biennio

Si riportano sempre le indicazioni relative al nucleo "Dati e Previsioni riportiamo" relative al Biennio, che sono comunque comuni agli indirizzi Servizi e Industria e Artigianato.

Conoscenze

- Dati, loro organizzazione e rappresentazione. Distribuzioni delle frequenze a seconda del tipo di carattere e principali rappresentazioni grafiche. Valori medi e misure di variabilità.
- Significato della probabilità e sue valutazioni. Semplici spazi (discreti) di probabilità: eventi disgiunti, probabilità composta, eventi indipendenti. Probabilità e frequenza.

Abilità

- Raccogliere, organizzare e rappresentare un insieme di dati. Calcolare i valori medi e alcune misure di variabilità di una distribuzione.
- Calcolare la probabilità di eventi elementari.

Si riporta di seguito il sonetto di Trilussa, pseudonimo anagrammatico del cognome di Carlo Alberto Camillo Mariano Salustri, che ha la fronte (le prime due strofe) con rima ABBA ABBA e la sirma con rima CDC EDE, ossia le quartine sono su due rime e le terzine su tre rime, perché è tra i più noti in Italia e uno dei più efficaci per i ragazzi. Non si dà la traduzione, perché si presume che il dialetto romanesco qui usato sia abbastanza accessibile.

Sai ched'è la statistica? È 'na cosa
che serve pe fa' un conto in generale
de la gente che nasce, che sta male,
che more, che va in carcere e che sposa.

Ma pe' me la statistica curiosa
è dove c'entra la percentuale,
pe' via che, lì, la media è sempre eguale
puro co' la persona bisognosa.

Me spiego: da li conti che se fanno
seconno le statistiche d'adesso
risurta che te tocca un pollo all'anno:

e, se nun entra nelle spese tue,
t'entra ne la statistica lo stesso
perché c'è un antro che ne magna due.

Bibliografia

- Da Valle S., Rampichini C., Tinelli C. (2015). "Insegnare Statistica: idee e strumenti" una proposta di aggiornamento per gli insegnanti della Toscana, Statistica & Società, Anno IV, N. 3, Strumenti.
- Dutka S. e Kafka F. (1950). Statistical Training below the College Level, The American Statistician, 4, 1, pp. 6-7.
- DM 254/2012 (2013). Regolamento recante indicazioni nazionali per il curricolo della scuola dell'infanzia e del primo ciclo d'istruzione, a norma dell'articolo 1, comma 4, del decreto del Presidente della Repubblica 20 marzo 2009, n. 89, Gazzetta Ufficiale, Serie Generale, n. 30, 05-02-2013, 1-77.
- Meusnier N. (2006). Sur l'histoire de l'enseignement des probabilités et des statistiques, Journ@l Electronique d'Histoire des Probabilités et des Statistique, 2, 2, 1-20.
- Ottaviani M. G. (2011). Insegnare ed apprendere statistica e probabilità a scuola: il problema dell'aggiornamento degli insegnanti. Periodico di matematiche, 3, 2011, 33-44.
- Ottaviani M. G. (2008). Statistica e Matematica a scuola: due discipline e un solo insegnamento. Confronto culturale e opportunità interdisciplinare. Induzioni, 36, 1, 17-38.
- Rampichini C. (2008), La statistica incontra gli studenti e le scuole; esperienza con gli studenti delle scuole superiori toscane. Sis-Magazine.
- Eastaway Rob, Wyndham Jeremy (2003). Probabilità, numeri e code. La matematica nascosta nella vita quotidiana, Edizioni Dedalo, Bari.
- Gigerenzer Gerd (2003). Quando i numeri ingannano. Imparare a vivere con l'incertezza, Raffaello Cortina, Milano.
- Kac Mark (1995). Gli enigmi del caso. Vicissitudini di un matematico, Bollati Boringhieri, Torino.
- Krämer Walter (2009). Le bugie della statistica, Mimesis, Sesto San Giovanni (MI).
- Paulos John A. (2009). Un matematico legge i giornali. Difendersi con la logica dai trucchi dell'informazione, Rizzoli, Milano.

Sitografia

https://www.indire.it/
https://www.disia.unifi.it/vp-104-l-insegnamento-della-statistica-nelle-scuole.html
https://local.disia.unifi.it/gmm/scuola/

Introduzione

"Il portiere della nazionale di calcio ha parato nell'ultimo anno 6 rigori, portando a 0.75 il suo numero medio di rigori parati a partita", "la maggioranza dei ragazzi dai 6 ai 13 anni non fa un'adeguata colazione la mattina", "il programma trasmesso ieri in prima serata sul canale 66 del digitale terrestre ha avuto il 15% di share", "oggi la probabilità di precipitazioni è del 60%". Sono frasi che inevitabilmente ogni giorno ci vengono proposte nella nostra vita quotidiana. Ovunque, sempre, ci vengono presentati numeri e termini tecnici che possono sembrare strani e di difficile comprensione. Se non impariamo a trasformare le cifre che ci circondano in informazione utile, interpretata e a nostra volta comunicata in modo appropriato, non riusciremo mai a rispondere a domande importanti per le nostre decisioni e a soddisfare le nostre esigenze conoscitive.

E se vi dicessimo che la statistica è proprio la materia che permette di decodificare i dati numerici e non numerici e che vi aiuta a valutare in modo intelligente le informazioni che i dati contengono?

E se vi dicessimo che la statistica è fondamentale per poter sfruttare appieno le fonti di informazioni facilmente accessibili in questa era digitale? E se vi dicessimo che la statistica è ormai diventata una parte indispensabile della vita e non conoscerla e utilizzarla significa consegnare inevitabilmente il proprio futuro a qualche altra persona? Forse non avreste una preclusione a priori verso questa materia ... E magari sareste anche incuriositi di sapere di cosa si occupa la statistica e quali sono le principali nozioni per iniziare a trasformare i dati in informazione utile! Questa pubblicazione, che raccoglie i contributi di docenti universitari ma che utilizza un linguaggio informale e discorsivo, cercherà di soddisfare queste vostre nuove curiosità.

Il libro nasce da un'iniziativa di public engagement promossa dall'Dipartimento di Economia Marco Biagi dell'Università di Modena e Reggio Emilia (Unimore) per condividere anche con gli studenti più giovani la scoperta dei benefici e del fascino della statistica, in tutte le sue declinazioni.

Che cos'è la Statistica?

La **Statistica** può essere definita come un insieme di tecniche che hanno come scopo la **conoscenza quantitativa** dei **fenomeni collettivi.**

Conoscenza quantitativa significa che la statistica si basa su numeri e si serve della matematica

Per fenomeno intendiamo qualsiasi evento o accadimento o caratteristica di persone o cose

Se il fenomeno è collettivo allora una singola persona, un singolo accadimento o evento NON è materia di studio della statistica!

Operazioni tipiche delle analisi statistiche sono:

- il **conteggio**
- la **classificazione**
- la **misurazione**
- la **sintesi** e la "spiegazione" dei fenomeni reali tramite modelli semplificativi o indici sintetici

Una possibile definizione della parola Statistica

Anche come raccogliere i dati è un problema statistico

dati reali

La statistica è un METODO per la **raccolta**, la classificazione e l'elaborazione dei **dati di fatto**, utilizzati nelle **scienze empiriche** e per la generalizzazione dei risultati, in termini **probabilistici**, ai casi non osservati.

Quando non si conoscono tutti i dati, le conclusioni che possiamo trarre hanno sempre un margine di incertezza misurato con la probabilità di errore

Quindi la statistica è una scienza utilizzata in tutte le scienze empiriche. Sono scienze empiriche (come la medicina, l'economia, le scienze sociali, ...) le scienze che si avvalgono dei dati per MOSTRARE conclusioni (conclusioni che possono essere smentite o confermate sulla base di nuovi dati). Qual è il contrario di scienza empirica??? Il contrario è scienza esatta. Qual è un esempio di **scienza esatta**???? La matematica! Nella matematica, dati gli assiomi, tutto si DIMOSTRA sulla base di teoremi e corollari. Quello che si dimostra è esatto e rimane invariato nel tempo.

Quali sono i termini che utilizzano frequentemente gli statistici?

- **unità statistiche:** sono le persone, gli oggetti, le aziende, … tutto quello che interessa al fine di un'indagine statistica.
- **variabili:** sono gli aspetti che vengono rilevati in ciascuna unità statistica al fine dell'indagine.

\>>> Esempio: se vogliamo sapere quante ore al mese gli studenti della classe 3B dedicano allo studio, quante ore dedicano a guardare la TV e a fare sport, ogni studente (o studentessa) della classe 3B sarà un'unità statistica e le variabili saranno: numero di ore dedicate a studiare, numero di ore dedicate allo sport, numero di ore dedicate a guardare la tv, e, eventualmente, per fare un'indagine più completa, il genere (se studente o studentessa), l'età, il numero di fratelli o sorelle. Con questa indagine potremmo mostrare se le studentesse sono più studiose degli studenti, se il fatto di avere fratelli o sorelle aiuta a guardare meno la tv, quanto in media gli studenti della 3ªB dedicano allo sport.

- **indagini complete o campionarie:** le indagini sono compete (o CENSUARIE) quando si rilevano tutte le unità appartenenti alla popolazione di riferimento, sono campionarie quando si rileva solo un sotto-insieme della popolazione, chiamato campione.

\>>> Esempio: se rileviamo tutti gli studenti della 3ªB (la popolazione di riferimento) effettuo un'indagine censuaria. Se invece qualcuno è assente il giorno in cui mi reco in classe per chiedere (con un questionario) come gli studenti distribuiscono le ore al pomeriggio tra studio, tv e sport, rilevo solo una parte della popolazione e quindi l'indagine diventa campionaria.

In sintesi, l'indagine

È completa quando si esaminano tutte le unità statistiche che compongono la popolazione oggetto di studio. **Pregi:** **Ricchezza delle informazioni** **Esaustività** **Difetti:** **Costo elevato** **Tempi di elaborazione dei dati molto lunghi** **Qualità dei dati non elevata (è più difficile controllare gli errori di trascrizione del dato o di rilevazione)**	**È parziale quando ci si limita a studiare un sottoinsieme, detto campione dell'insieme di riferimento.** **Pregi:** **Economicità** **Indagini più mirate e approfondite** **Dato maggiormente controllato (se i dati sono pochi, si riesce a controllarli e a verificarli)** **Difetti:** **I risultati su tutta la popolazione sono affetti da margine di errore perché si è rilevato solo parte dei dati.**

Al lavoro!

Per ciascuno dei seguenti casi, indicare se abbiamo a che fare con un campione (indagine campionaria) o con la popolazione (indagine censuaria):

a. Rilevazione di tutta la popolazione in Italia in età lavorativa per definire il tasso di occupazione in Italia
b. Rilevazione di tutti gli iscritti alla scuola secondaria di primo grado Amedeo d'Aosta per definire la percentuale di studenti e studentesse all'interno della scuola.
c. Rilevazione di tutti gli iscritti alla scuola secondaria di primo grado Leonardo Da Vinci per definire la percentuale di studenti e studentesse presenti nelle scuole secondarie di primo grado italiane
d. Rilevazione del numero di gol segnati da una squadra di calcio nel campionato per definire il numero di gol medi a partita della squadra.
e. Rilevazione del numero di gol segnati da due squadre di calcio nel campionato per definire il numero di gol medi a partita delle squadre nel campionato

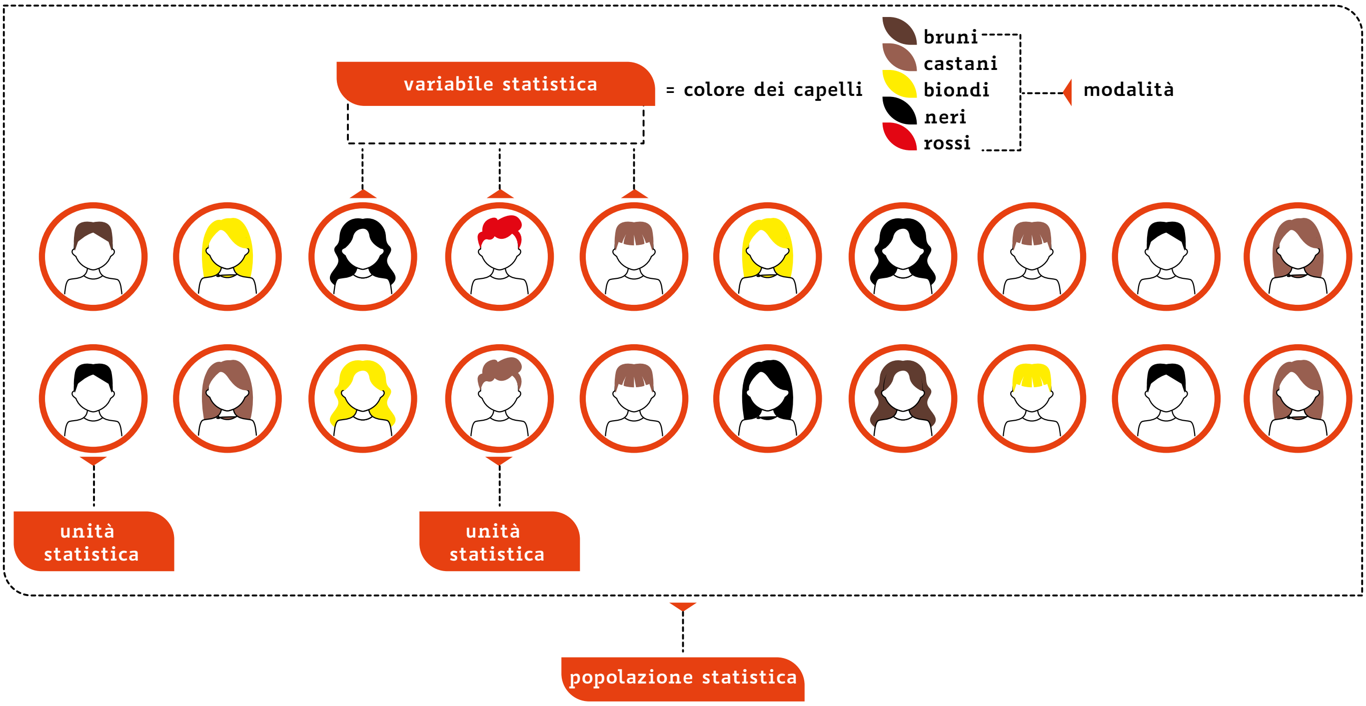
Fasi di un'indagine statistica

- **definizione degli obiettivi**
- **definizione delle unità, delle variabili da rilevare e scelta del periodo di riferimento**
- **individuazione della popolazione**
- **definizione del piano di campionamento**
- **raccolta dei dati**
  - scelta della tecnica di rilevazione (con un rilevatore, attraverso email o indagine telefonica o postale)
  - formulazione del questionario
  - rilevazione sul campo (ad esempio, se si vuole analizzare l'inquinamento nella città di Modena, i dati saranno rilevati dalle centraline per il monitoraggio della qualità dell'aria)
- **registrazione dei dati**
  - registrazione su supporto magnetico, controllo e correzione
- **elaborazione e analisi dei dati**
  - la parte di calcolo e analisi è solo l'ultimo step! Prima ci sono passaggi importanti e spesso trascurati …

>>> Esempio:

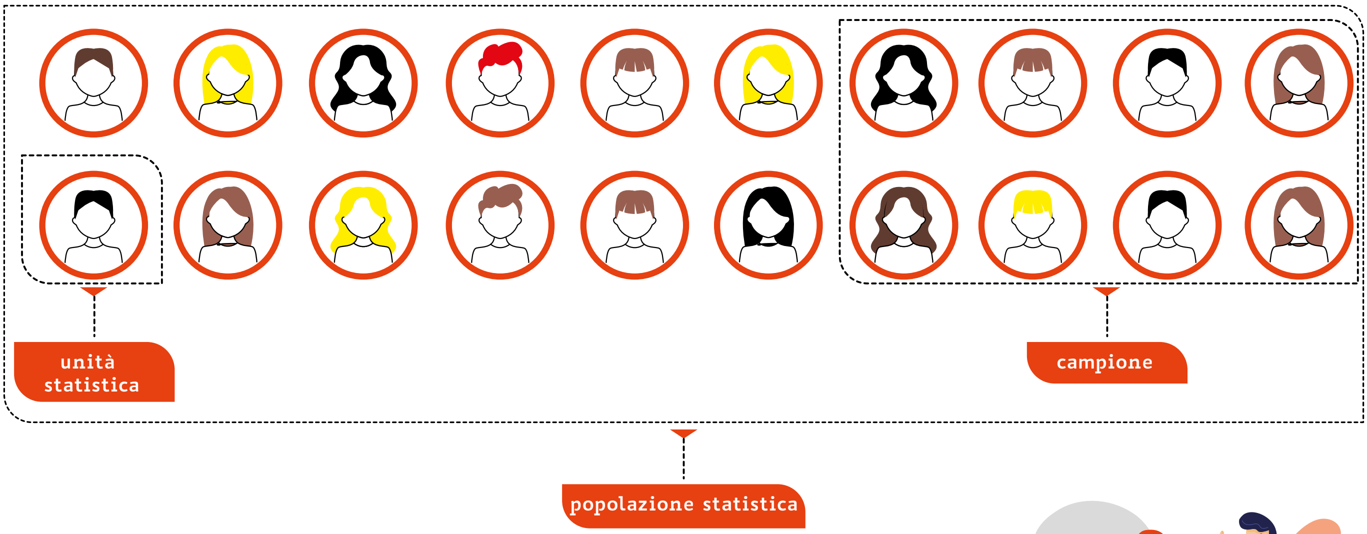
Vorremmo sapere quante persone nella provincia di Modena, nate nel 2010, si chiamano Sofia e quanti Luca, dove abitano e che colore di capelli hanno: ho definito **gli obiettivi**.

Le unità sono tutti i residenti a Modena nati nel 2010, il periodo di riferimento è il corrente anno solare, le variabili sono il nome, la data di nascita, il comune di residenza, il colore dei capelli: ho definito **le unità, le variabili da rilevare e la scelta del periodo di riferimento**.



La popolazione di riferimento è quella presente nelle anagrafi dei comuni di Modena (dove vengono registrate tutte le persone residenti). La lista delle unità appartenenti alla popolazione è presente nelle anagrafi: ho **individuato la popolazione** e lista delle unità statistiche.

Non abbiamo la possibilità di consultare tutte le anagrafi e quindi facciamo un'indagine campionaria, scegliendo a caso quattro comuni (uno montano, due di pianura e uno collinare) e consultando le anagrafi presso questi comuni. Alternativamente, potremmo selezionare a caso delle scuole primarie di secondo grado dove andare a rilevare personalmente i dati: abbiamo **definito due possibili piani di campionamento.**



Se chiediamo i dati alle anagrafi selezionai, ci avvaliamo delle cosiddette fonti statistiche. Se andiamo presso le scuole selezionate e intervisto gli alunni nati nel 2010 effettuiamo la cosiddetta rilevazione tramite questionario. In entrambi i casi abbiamo **raccolto i dati!**

Inseriamo i dati in un foglio elettronico per la loro **registrazione** e finalmente possiamo calcolare quante ragazze si chiamano Sofia sul totale delle ragazze rilevate e quanti ragazzi si chiamano Luca, sul totale dei ragazzi rilevati. Iniziamo ad **elaborare i dati**. Possiamo anche calcolare la percentuale di ogni nome separatamente per i comuni montani, di collina e di pianura.

Analizziamo i risultati: se il 10% delle ragazze si chiama Sofia ed il 5% dei ragazzi si chiama Luca, allora Sofia è più utilizzato di Luca. Se la maggioranza dei ragazzi che si chiamano Luca abitano nel comune montano, quel nome è utilizzato soprattutto in montagna ed è meno comune in pianura e collina. Se la maggioranza delle bimbe che si chiama Sofia ha i capelli biondi, allora il nome Sofia è maggiormente comune per le bionde!

Scala di misura delle variabili

Le variabili oggetto di studio possono essere molto diverse tra di loro. Ad esempio, se rilevo il colore dei capelli degli studenti, ottengo come risposta una parola come marrone o castano, se rilevo l'età ottengo invece un valore numerico. Se rilevo la temperatura ottengo un valore numerico che può essere anche negativo (in inverno in montagna si può arrivare anche a -10° centigradi!!!!!).

Se la riposta ad una variabile è un valore numerico, la variabile si dice quantitativa, se invece è una parola, la risposta si chiama modalità e la variabile si dice qualitativa.

In base al tipo di risposta, esistono quattro scale di misura: due qualitative (nominale e ordinale) e due quantitative (su scala di intervalli e su scala di rapporti). Elenchiamo le caratteristiche di ogni scala (in ordine crescente di complessità).

1) Variabile qualitativa nominale: tra le modalità di risposta valgono solo le relazioni di uguale o diverso.

>>> Esempi:

- il genere della persona: fra le modalità *maschio* e *femmina* valgono solo le relazioni di uguale o diverso;
- il colore dei capelli: se entrambi i miei fratelli sono castani, hanno uguale colore di capelli, se uno è castano e l'altro biondo, hanno un colore diverso.

2) Variabile qualitativa ordinale: tra le modalità di risposta valgono le relazioni di uguale o diverso ed anche di maggiore e minore.

>>> Esempi:

- il giudizio ottenuto in una prova: *sufficiente* è minore di *buono* e *ottimo* è maggiore di buono;
- il titolo di studio: se mio fratello ha ottenuto la laurea ed io ho ottenuto la licenza superiore, mio fratello ha ottenuto un titolo di studio maggiore del mio; se un nostro amico ha solo la licenza della scuola primaria di secondo grado, ha un titolo di studio inferiore a quello mio e di mio fratello.

3) Variabile quantitativa su scala di intervalli: tra i valori numerici che assume la variabile, è lecito non solo utilizzare le relazioni di uguale e diverso, maggiore e minore ma anche calcolare la distanza tra due valori (distanza intesa come differenza in modulo). In queste variabili lo zero è arbitrario (cambia da variabile a variabile).

>>> Esempi:

- Nelle scale Celsius e Farenheit per la misurazione della temperatura, uguali differenze su ciascuna di queste scale rappresentano uguali differenze in temperatura, anche se non si può affermare che una temperatura di 30 gradi sia il doppio di 15 gradi. Lo zero nelle due scale identifica due temperature diverse.
- Sia in Celsius sia in Farenheit, se oggi la massima è stata di 21° gradi e la minima di 11°, la differenza tra la massima e la minima è stata di 10°.
- Le date del calendario: il periodo intercorso tra il primo Novembre ed il 6 Novembre è di 5 giorni, quello tra il primo Novembre ed il 10 Novembre è di 9 giorni.

4) Variabile quantitativa su scala di rapporti: tra i valori numerici che assume la variabile, è lecito non solo utilizzare le relazioni di uguale e diverso, maggiore e minore, calcolare la distanza tra due valori, ma anche calcolare i rapporti. IL calcolo dei rapporti si può fare perché il valore zero significa quantità nulla! Si parla di zero assoluto.

>>> Esempi:

- il voto ottenuto in una prova: se io ho ottenuto 5 (sigh....) ed il mio compagno 10, posso dire non solo che abbiamo ottenuto due voti diversi, che il mio voto è minore del suo, ma anche che la differenza tra i nostri due voti è di 5 punti e che il mio compagno ha ottenuto un voto doppio del mio (doppio sigh...);
- l'età: se Sofia ha 15 anni, Mario ha 5 anni e Aurora ha 10 anni, la differenza di età tra Sofia e Mario è di 10 anni, quella tra Sofia e Aurora è di 5 anni, Sofia ha tre volte gli anni di Mario e Mario ha la metà degli anni di Aurora.

Al lavoro!

Per ciascuna delle seguenti variabili, indica se la scala di misura corrispondente è nominale, ordinale, intervallo o rapporto:

a. Tempo necessario per svolgere la prova di matematica
b. Password del cellulare
c. Taglia della felpa (XS, S, M, L, XL)
d. Anni scolastici
e. Categorie sportive (pulcini, ragazzi, cadetti, ...)
f. Anni di istruzione
g. Quoziente di intelligenza

Quale ente in Italia si occupa di rilevare dati di interesse nazionale? Istat

l'Istituto nazionale di statistica (Istat o ISTAT) è un ente pubblico di ricerca italiano che si occupa dei

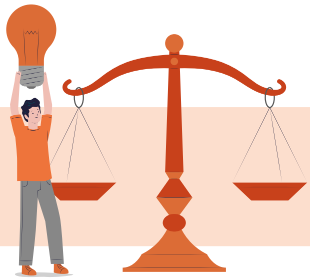
- **censimenti** generali della popolazione, dei servizi e dell'industria, e dell'agricoltura;
- **di indagini campionarie** sulle famiglie, sulla società e le istituzioni, su istruzione e lavoro, su ambiente e territorio ed economia a **livello nazionale**.

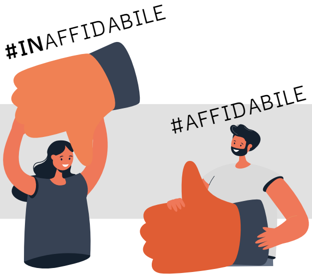
L'operato dell'istituto è controllato dalla Commissione per la garanzia dell'informazione statistica della Presidenza del Consiglio dei ministri. La commissione ha il compito di garantire l'imparzialità e la completezza dei dati raccolti e pubblicati.

L'Istat è il produttore di statistica ufficiale in Italia. Realizza indagini, studi e analisi per fornire informazioni ufficiali mirate a soddisfare il bisogno informativo della popolazione italiana. Le rilevazioni di pubblico interesse sono stabilite dal Programma statistico nazionale, il documento che regola l'attività di produzione statistica.

L'Istituto fa parte del Sistema statistico europeo e produce e diffonde informazioni ispirate ai 6 principi fondamentali della statistica ufficiale:

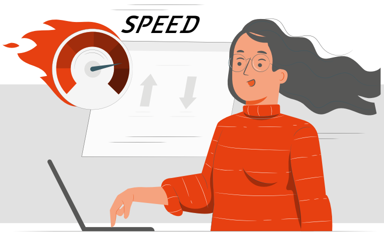
1 IMPARZIALITÀ
Le informazioni NON devono essere mirate a mostrare qualche idea pre-costituita, ma devono essere oneste e neutrali





2 AFFIDABILITÀ
Le informazioni (i dati raccolti ed i risultati delle elaborazioni) devono essere vere e non ambigue

3 PERTINENZA
Le informazioni devono soddisfare i bisogni della collettività e riguardare gli obiettivi scritti nel programma statistico nazionale. Solo se pertinenti ai bisogni le informazioni sono utili



4 EFFICIENZA
Le informazioni devono essere rese disponibili velocemente. I risultati devono essere diffusi tempestivamente per essere realmente utili

5 RISERVATEZZA
Il dato relativo ad ogni singola unità statistica NON può essere divulgato

6 TRASPARENZA
Deve essere chiaro come sono stati raccolti i dati e come sono stati elaborati

L'Istat, a conclusione del processo di produzione dell'informazione statistica, mette a disposizione dei cittadini, delle imprese e delle istituzioni i risultati delle rilevazioni. Tutte le informazioni pubblicate sono accompagnate dai metadati, ovvero descrizioni dei dati che permettono la loro ricerca, selezione e localizzazione.

Le informazioni sono pubblicate gratuitamente sul sito web dell'Istat sotto forma di comunicati stampa, opuscoli divulgativi, pubblicazioni cartacee, banche dati, sistemi informativi, file di dati.

Esistono altri enti pubblici e privati che rilevano dati e pubblicano periodicamente report. Questi enti si dicono FONTI STATISTICHE. Sono, ad esempio, le banche centrali, l'organizzazione mondiale della sanità, L'UNESCO, le Camere di Commercio, le associazioni di categoria. E non scordiamoci di Internet, che negli ultimi anni è diventata una fonte inesauribile di dati!

Al lavoro!

- Una società sportiva, dopo aver fatto un'indagine sul livello di preparazione atletica dei 200 ragazzi iscritti al primo anno di corso di atletica leggera, pubblica il voto assegnato dall'istruttore in ogni prova test di fianco al nome e cognome del ragazzo. Quale principio della statistica trasgredisce?
- Dopo sei mesi dall'inizio del corso la Federazione Sportiva regionale chiede i tempi migliori di ogni iscritto nella prova dei 60 metri piani, per mandare una rappresentanza ai campionati nazionali giovanili. I campionati nazionali si svolgono il 21 Marzo. La società sportiva manda i tempi il 22 Marzo. Quale principio della statistica trasgredisce?
- Un genitore chiede all'istruttore come ha assegnato il voto di ogni prova test (se considerando come è stato svolto l'esercizio oppure in quanto tempo oppure il numero di errori effettuati nell'esercizio, oppure una combinazione di tutte queste valutazioni). Se l'istruttore non riesce a spiegare in modo chiaro come ha assegnato i voti, quale principio della statistica trasgredisce?
- L'istruttore vorrebbe che gran parte dei ragazzi iscritti scegliessero la corsa come specialità, invece del salto in lungo. Nelle prove test iniziali, per incentivare i ragazzi a scegliere la corsa, inserisce prove di corsa molto semplici dove i ragazzi ottengono senza problemi buoni punteggi e prove di salto molto complesse, dove è difficile svolgere in maniera corretta e veloce l'esercizio. Quale principio della statistica trasgredisce?
- Durante alcune prove test, il cronometro si rompe e l'istruttore decide quindi di registrare i tempi di esecuzione della prova contando i secondi a mente, senza orologio. Quale principio trasgredisce?
- La Federazione Sportiva Regionale, dopo sei mesi dall'inizio del corso, chiede alla società di atletica di mandare i migliori tempi degli iscritti sul salto in lungo, per scegliere una rappresentativa regionale che gareggi su questa specialità nei campionati nazionali giovanili del 22 Marzo. La società manda i tempi dei 60 metri piani. Quale principio trasgredisce?

Tabelle e Grafici

I **Dati** possono essere tantissimi e per prima cosa bisogna imparare a saperli organizzare

La matrice dei Dati

È un modo efficace per organizzare tantissime tipologie di dati, si tratta di una tabella nella quale le righe indicano l'individuo (l'unità statistica) rilevata, e le colonne le variabili di interesse.

>>> Esempio:

Nel piccolo comune C vivono 200 persone maggiorenni, ad ogni individuo (l'unità statistica) sono state poste alcune domande:

- **Genere:** sei maschio o femmina? La risposta genera una variabile qualitativa nominale.
- **Età:** quanti anni hai? La risposta genera una variabile quantitativa discreta su scala di rapporti espressa in anni compiuti.
- **Titolo di studio:** qual è il tuo massimo titolo di studio, le medie inferiori, le medie superiori o la laurea? La risposta genera una variabile qualitativa ordinale.
- **Statura:** quanti metri sei alto? La risposta genera una variabile quantitativa continua su scala di rapporti.

- **Reddito:** quanto guadagni in un anno? Il reddito si misura in euro e per comodità divideremo per mille. La risposta genera una variabile quantitativa continua su scala di rapporti.
- ... potremmo fare tante altre domande: sei sposato? Hai figli? Ecc.

La tabella A ci mostra le prime cinque righe della matrice dei dati.

Tabella A *Le prime 5 righe della matrice dei dati, ogni riga corrisponde ad un individuo, ordinati secondo l'ordine di intervista. Ogni colonna corrisponde ad una variabile, ognuna con la propria codifica.*

Intervista	Genere	Età	Titolo di Studio	Statura	Reddito (x 1000€)
1	F	26	Laurea	1,76	34,90
2	M	32	Superiori	1,63	189,20
3	M	18	Medie	1,72	24,00
4	F	54	Superiori	1,67	18,53
5	F	75	Medie	1,84	8,30
...	...	...	...	...	...

Operazioni preliminari

Prima di fare le interviste devi decidere alcune cose che ti aiuteranno ad organizzare meglio i dati.

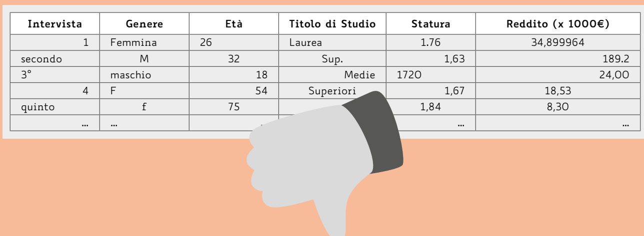
1. **Codifica:** decidi le etichette da attribuire alle modalità, per esempio usa M per indicare maschio e F per indicare Femmina.
2. **Unità di misura:** per le variabili quantitative scegli un'unità di misura appropriata. Ad esempio, per la statura abbiamo scelto di usare i metri, ma avremmo anche potuto usare i centimetri. Per il reddito abbiamo deciso di dividere per 1000 per semplificare a velocizzare la lettura: anziché scrivere 12 400€, scriveremo 12,4 sottintendendo la moltiplicazione per mille.
3. **Numero di decimali:** fissa il numero di decimali con cui raccoglierai le misure numeriche.

Attenzione!!!

>>> Usa sempre la stessa codifica!

>>> Allinea i numeri a destra e usa lo stesso numero di decimali!

Intervista	Genere	Età	Titolo di Studio	Statura	Reddito (x 1000€)
1	F	26	Laurea	1,76	34,90
2	M	32	Superiori	1,63	189,20
3	M	18	Medie	1,72	24,00
4	F	54	Superiori	1,67	18,53
5	F	75	Me	1,84	8,30
...	...	...	...	...	...



Intervista	Genere	Età	Titolo di Studio	Statura	Reddito (x 1000€)
1	Femmina	26	Laurea	1.76	34,899964
secondo	M	32	Sup.	1,63	189.2
3°	maschio	18	Medie	1720	24,00
4	F	54	Superiori	1,67	18,53
quinto	f	75		1,84	8,30
...	...			...	...

Il Foglio Elettronico

Se i dati sono tanti, lavorare a carta e penna diventa molto difficile. Aiutati con un **foglio elettronico**. È un programma (un'app) che ti consente di compilare molto velocemente la matrice dei dati. Ma non solo, un foglio elettronico è una calcolatrice evoluta che ti consente di realizzare calcoli lunghi e complessi, con pochi click.

I **fogli elettronici** disponibili sono tanti, alcuni a pagamento, altri gratuiti. Il foglio elettronico a pagamento più noto è *Excel* di Microsoft, tra i fogli elettronici gratuiti ci sono i *Fogli* di Google e *Calc* di OpenOffice. Parla con il tuo referente di informatica per capire quale è lo strumento più adatto alla tua classe.

La distribuzione di Frequenza

È una tabella che consente di osservare in modo **sintetico** una colonna specifica della matrice dei dati. Consiste nell'identificare tutte le **modalità** che competono alla variabile, **contare** quanti individui possiedono un determinata modalità (**frequenze assolute**) e, infine, associare in una tabella le modalità con le loro frequenze.

>>> Esempio:

Analizziamo il titolo di studio dei maggiorenni del comune C. La variabile titolo di studio prevede tre sole modalità:

1. Medie
2. Superiori
3. Laurea

La **frequenza** della modalità *Medie* si ottiene contando tutti quelli che hanno come massimo titolo di studio il diploma medio inferiore. Nel nostro caso 40.

Eseguiamo il conteggio per tutte le modalità e otteniamo la tabella B.

Tabella B *Distribuzione di frequenza della variabile Titolo di Studio.*

Modalità	Frequenza Assoluta
Medie	40
Superiori	98
Laurea	62
Totale	200

E leggeremo: 40 su 200 hanno come massimo titolo di studio il diploma medio inferiore, 98 su 200 hanno il diploma medio superiore e 62 su 200 hanno conseguito la laurea.

Osserva: *200 informazioni sul titolo di studio sono contenute in sole tre righe!*

Le **frequenze percentuali** si ottengono dividendo le frequenze assolute per il numero totale degli individui e moltiplicando per 100. La tabella C ci mostra la distribuzione delle frequenze percentuali ottenuta a partire dalla tabella B.

Tabella C *Distribuzione delle frequenze percentuali della variabile Titolo di Studio*

Modalità	Frequenza Percentuale
Medie	$\frac{40}{200} \times 100 = 20\%$
Superiori	$\frac{98}{200} \times 100 = 49\%$
Laurea	$\frac{62}{200} \times 100 = 31\%$
Totale	100%

E si legge, il 20 per cento del collettivo ha come massimo titolo di studio le scuole medie, il 49 percento del collettivo ha il diploma superiore e il restante 31 per cento è laureato.

Osserva!!!

>>> La somma delle frequenze assolute fa il numero totale di individui osservati

40 + 98 + 62 = 200

>>> La somma delle frequenze percentuali fa sempre 100

20% + 49% + 31% = 100%

Al lavoro!

1. Prepara un questionario per la tua classe che preveda:
   - Almeno una variabile qualitativa sconnessa
   - Almeno una variabile qualitativa ordinata
   - Almeno una variabile quantitativa discreta
   - Almeno una variabile quantitativa continua
2. Inserisci i dati in un foglio elettronico
   - Controlla che tutte le codifiche siano uniformi
   - Controlla la formattazione dei numeri
3. Estrai una colonna di una variabile qualitativa (sconnessa o ordinata)
   - Costruisci la distribuzione delle frequenze assolute e percentuali

Dati raccolti in classi

Costruire una tabella di frequenza se i dati sono quantitativi può richiedere un'operazione preliminare. Quando il numero di modalità è troppo grande la tabella non riesce a compiere alcuna sintesi. L'obiettivo è di raggruppare i dati numerici in **intervalli contigui**.

>>> Esempio:

Se dobbiamo analizzare la colonna dei Redditi abbiamo 200 numeri tutti diversi:

34,90; 189,20; 24,00; 18,53; 8,30; 149,99; 54,66; 3,84; 12,04; 30,18; 15,44; 4,60; 14,10; 64,61; 55,21; 17,21; 23,20; 87,18; 39,14; 25,69...

Costruiamo intervalli (**classi**), per esempio:

1. Quanti individui hanno un reddito compreso tra 0 e 10 (mila euro annui)?
   >>> 61 individui guadagnano tra 0 e 10 (mila euro annui)
2. Quanti individui hanno un reddito compreso tra 10 e 20 (mila euro annui)?
   >>> 65 individui guadagnano tra 10 e 20 (mila euro annui)
3. Quanti individui hanno un reddito compreso tra 20 e 50 (mila euro annui)?
   >>> 59 individui guadagnano tra 20 e 50 (mila euro annui)
4. Quanti individui hanno un reddito compreso tra 50 e 100 (mila euro annui)?
   >>> 13 individui guadagnano tra 50 e 100 (mila euro annui)
5. Quanti individui hanno un reddito maggiore di 100 (mila euro annui)?
   >>> 2 individui guadagnano più di 100 (mila euro annui)

Osserva!!!

>>> **abbiamo costruito cinque nuove modalità e abbiamo ricavato le frequenze assolute, possiamo anche costruire le frequenze percentuali.**

La tabella D riporta le **classi**, le frequenze assolute e le frequenze percentuali.

Tabella D *Distribuzioni in classi della variabile Reddito*

Da	a	Frequenze Assolute	Frequenze Percentuali
0	10	61	30,5%
10	20	65	32,5%
20	50	59	29,5%
50	100	13	6,5%
>100		2	1,0%
		200	100%

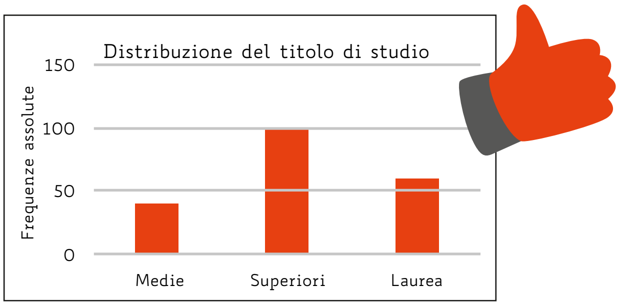
Al lavoro!

4. Estrai una colonna di una variabile quantitativa continua
   - Prova diverse divisioni in classi
   - Per ogni divisione in classi costruisci la distribuzione delle frequenze assolute e di quelle percentuali
5. Confronta le tabelle che hai ottenuto
   - Osserva che un numero troppo piccolo di classi (2 o 3) offre una grande sintesi ma si perde tanta informazione
   - Osserva che un numero troppo elevato di classi preserva informazione sui dati ma non semplifica la lettura

I Grafici

Un grafico statistico è spesso la rappresentazione grafica di una tabella di frequenza. L'obiettivo è di facilitare la lettura dei numeri attraverso forme e colori.

Intervista	Genere	Età	Titolo di Studio	Statura	Reddito (x 1000€)
1	F	26	Laurea	1,76	34,90
2	M	32	Superiori	1,63	189,20
3	M	18	Medie	1,72	24,00
4	F	54	Superiori	1,67	18,53
5	F	75	Medie	1,84	8,30
...	...	...	...	...	...



Esistono molti tipi di rappresentazione facili da realizzare sia a mano che usando un foglio elettronico.

Il diagramma a barre. È un grafico che consiste nell'affiancare alle modalità rettangoli di altezza pari alle frequenze. Si può costruire sia dalle frequenze assolute che da quelle percentuali.

- Quali tipi variabili si possono analizzare?
  - Tutte le variabili qualitative, sia sconnesse che ordinate
- Quali tipi variabili non si possono analizzare?
  - Le variabili quantitative continue raccolte in classi

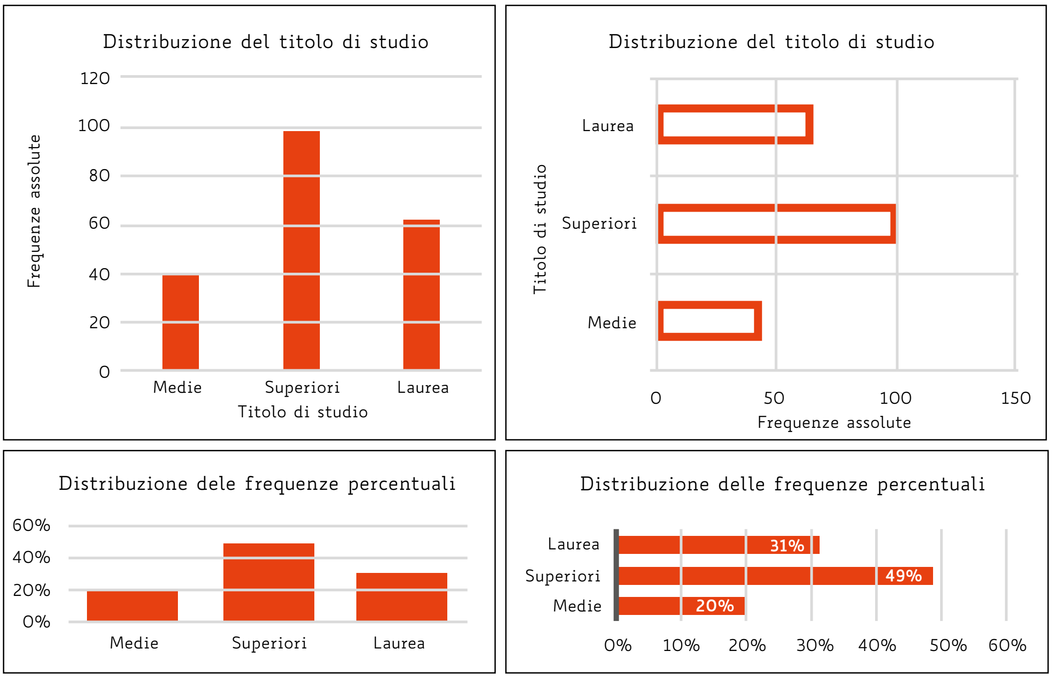


Figura 1 sopra: il diagramma a barre verticali delle frequenze assolute del titolo di studio e diagramma a barre orizzontali, sotto il diagramma a barre verticali e orizzontali delle frequenze percentuali per la stessa variabile. La leggibilità è la stessa, anche se abbiamo usato decorazioni diverse.

La figura 1 mostra il diagramma a barre per la variabile titolo di studio, lo abbiamo costruito sia orizzontale che verticale, sia per le frequenze assolute che per quelle relative.

Ci siamo divertiti a cambiare i colori e gli effetti, ma tutti e quattro i grafici riassumono nello stesso modo la distribuzione del titolo di studio.

Il diagramma a torta. È un grafico che consiste nel rappresentare le frequenze all'interno di una circonferenza. Si può costruire sia dalle frequenze assolute sia da quelle percentuali.

- Quali tipi variabili si possono analizzare?
    - Tutte le variabili qualitative, sia nominali sia ordinali
    - I conteggi
    - Le variabili quantitative continue raggruppate in classi (se le classi sono di uguale ampiezza)

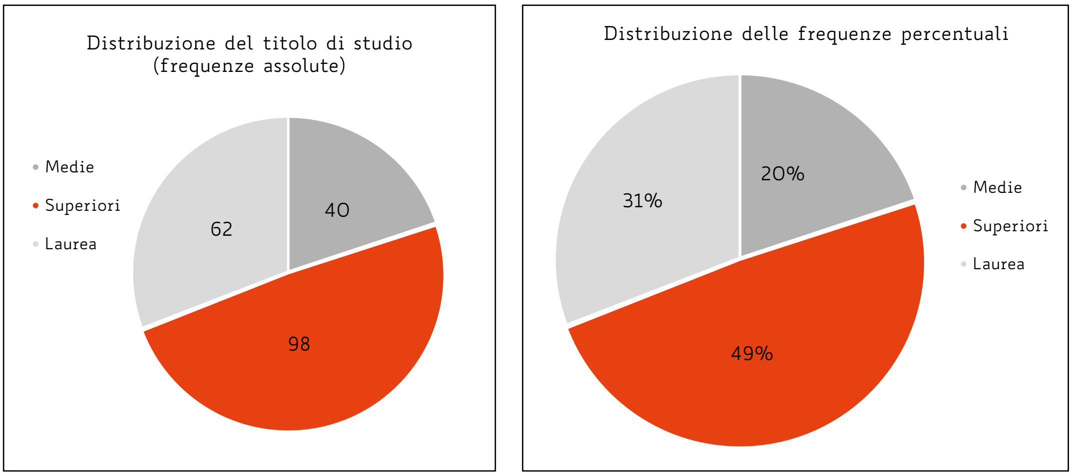


Figura 2 Diagramma a torta, a sinistra le frequenze assolute, a destra quelle percentuali. Le frequenze sono rappresentate come angoli. La leggibilità è la stessa.

La figura 2 mostra il diagramma a torta per la variabile titolo di studio, lo abbiamo costruito sia per le frequenze assolute che per quelle percentuali. Ci siamo divertiti a cambiare i colori e gli effetti, ma i due grafici riassumono nello stesso modo la distribuzione del titolo di studio.

Come costruire un diagramma a torta!!!

>>> Le frequenze percentuali sommano a 100, i gradi degli spicchi della circonferenza sommano a 360. Dobbiamo fare una proporzione:

$$\textit{Frequenza}: 100 = \text{Angolo } (x).360$$

>>> e quindi

$$\text{Angolo } (x) = \frac{\textit{Frequenza}}{100} \times 360$$

>>> Nel nostro esempio

$$\text{Medie} \quad \text{Angolo } (x) = \frac{20}{100} \times 360 = 72{,}0°$$

$$\text{Superiori} \quad \text{Angolo } (x) = \frac{49}{100} \times 360 = 176{,}4°$$

$$\text{Laurea} \quad \text{Angolo } (x) = \frac{62}{100} \times 360 = 111{,}6°$$

>>> Adesso usa il compasso per disegnare una circonferenza e utilizza il goniometro per realizzare gli spicchi.

L'istogramma di densità di frequenza. È un grafico che consiste nell'affiancare alle classi delle variabili quantitative rettangoli di area pari alle frequenze. Si può costruire sia dalle frequenze assolute che da quelle percentuali.

- Quali tipi variabili si possono analizzare?
  - Le variabili quantitative continue raccolte in classi

Attenzione!!!

>>> Sui fogli elettronici non è sempre disponibile come grafico preimpostato.

Tabella E *Calcolo delle altezze dei rettangoli che compongono l'istogramma*

Da	a	Frequenze Assolute	Frequenze Percentuali	Ampiezza	Densità Percentuali
0	10	61	30,5%	10-0=10	$\frac{30,5}{10}=3,05$
10	20	65	32,5%	20-10=10	$\frac{32,5}{10}=3,25$
20	50	59	29,5%	20-50=30	$\frac{29,5}{30}=0,98$
50	100	13	6,5%	100-50=50	$\frac{6,5}{5}=0,13$
>100		2	1,0%	La classe aperta non si disegna	
		200	100%		

>>> Esempio: se dobbiamo analizzare la colonna dei Redditi abbiamo imparato a mettere i dati in classi come, per esempio, in tabella D. Vogliamo costruire un grafico composto da rettangoli di base e altezza diversi. La base è data dall'ampiezza della classe, l'area dalla frequenza. L'altezza del rettangolo è chiamata **densità** e si ottiene dividendo la frequenza per l'ampiezza della classe

$$densità = \frac{frequenza}{ampiezza}$$

La tabella E riporta il calcolo delle densità di frequenza partendo dall'esempio del reddito. Per comodità abbiamo diviso il calcolo in due fasi, abbiamo prima trovato l'ampiezza delle basi e poi trovato la densità.

In figura 3 possiamo vedere l'istogramma di densità di frequenza percentuale. Ogni rettangolo ha **area** pari alla frequenza percentuale della classe. Osserviamo che la densità maggiore si ha nell'intervallo 10-20, ovvero il 32,5% del collettivo ha un reddito molto simile in un intervallo molto piccolo, mentre il 29,5% del collettivo ha un reddito nella classe 20-50, che è molto più ampia e ne consegue una densità molto inferiore.

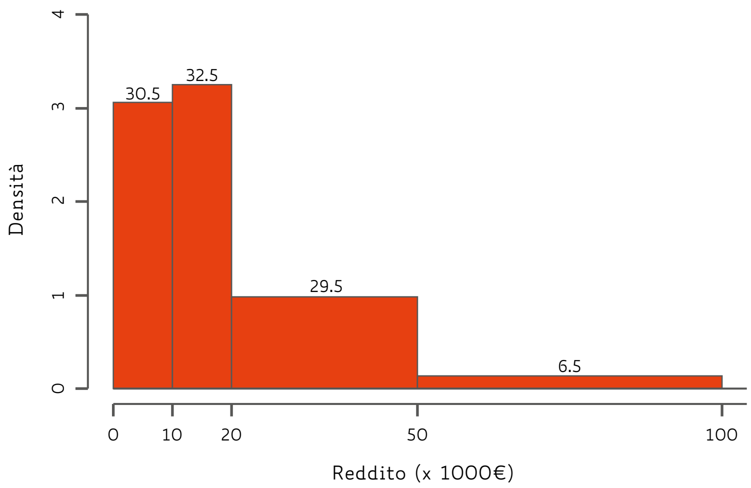


Figura 3 Istogramma di densità di frequenza percentuale della variabile Reddito. Ogni rettangolo ha un'area pari alla frequenza percentuale, che è indicata sopra ad ognuno di essi.

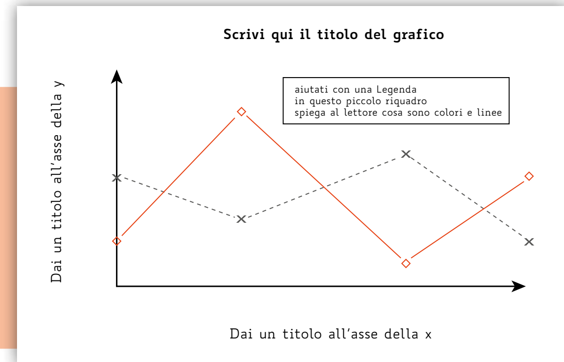
Regole per un buon grafico!!!

>>> Dai sempre un titolo al grafico.

>>> Se presenti, dai un titolo agli assi della x e della y.

>>> Aiutati con la Legenda.

>>> Usa la didascalia.



Al lavoro!

6. Seleziona una variabile qualitativa
   - Disegna un istogramma a barre
   - Disegna un istogramma a torta
   - Esplora i grafici che il tuo foglio elettronico ti mette a disposizione per le tue tabelle di frequenza
7. Seleziona una variabile quantitativa suddivisa in classi
   - Disegna l'istogramma di densità percentuale

Attento alle decorazioni!

Un grafico abbellito da decorazioni, colori e forme accattivanti attira l'attenzione ma ricorda, l'obiettivo è semplificare la lettura di una tabella, non complicarla!

Dare tridimensionalità al grafico a torte o a barre rende il grafico gradevole ma l'utilizzo della terza dimensione non aggiunge nessuna interpretazione aggiuntiva, al contrario può complicarne la lettura.

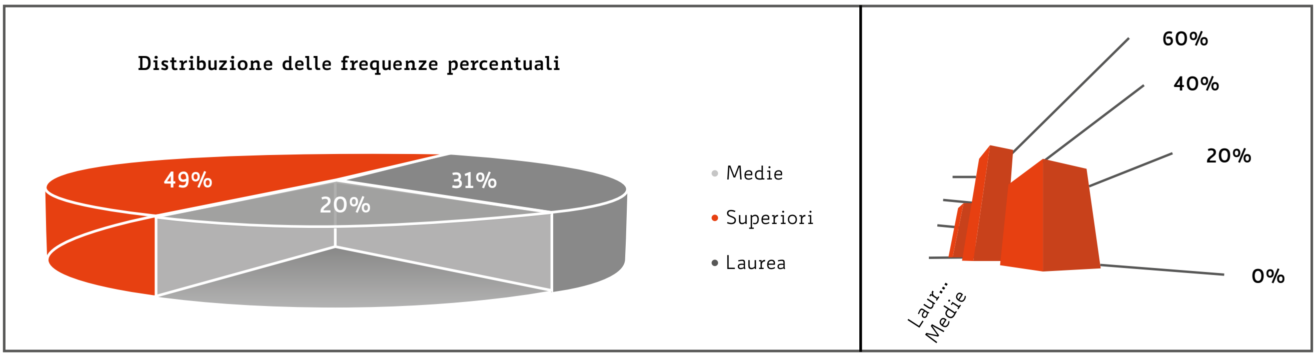


Figura 5 La torta è talmente inclinata che il 20% ci sembra grande quanto il 31% per colpa della prospettiva

I Principali Indici Statistici

Cosa sono gli indici statistici?

Gli indici statistici sono valori numerici utili per fornire informazioni sintetiche sui dati – infatti, con un solo numero permettono di descrivere caratteristiche di un insieme di dati, anche molto grande!

Per cominciare: Alcuni vostri coetanei hanno lasciato recensioni su Internet Movie Database (IMDb) per il film Hunger Games. I loro rating sono i seguenti: 6 8 5 7 7 8 7 5 6

Quale rating si ripete più volte degli altri? [7]

Quale voto si trova nel centro se li ordini in maniera crescente? [7]

Internet Movie Database (IMDb) mette sul suo sito la media di tali rating: sai qual è?

Che differenza c'è tra il rating più alto e quello più basso? [3]

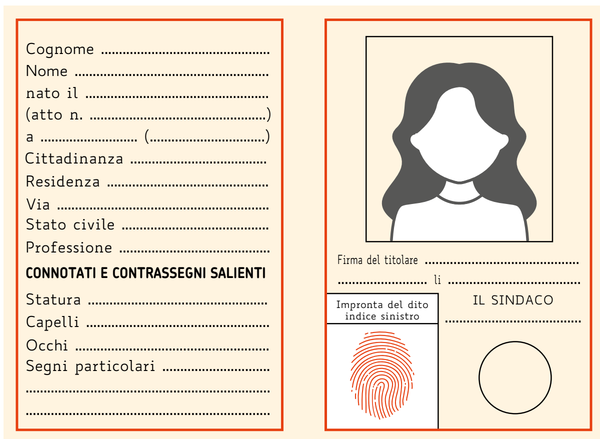
Per descrivere un insieme di dati (come l'elenco dei rating nell'esempio iniziale), si può costruire una carta d'identità di questo insieme che contiene informazioni sui suoi tratti principali, chiamati:

INDICI DI POSIZIONE

INDICI DI DISPERSIONE

INDICI DI FORMA

Questi indici formano la carta d'identità dell'insieme di dati, esattamente come nella tua carta



d'identità trovi i tuoi tratti distintivi: nome, cognome, età, indirizzo, altezza, peso, colore degli occhi, etc.!!

Indici di posizione

>>> *Gli indici di posizione consentono di valutare l'ordine di grandezza dei dati e localizzano la distribuzione dei dati, cioè individuano i valori centrali attorno ai quali si concentra l'insieme dei dati.*

>>> *I principali indici di posizione sono tre: moda, mediana e media.*

La **moda** di un insieme di dati è il valore che compare più volte, ossia esso compare con maggior frequenza.

La moda è l'unico indice che si può determinare anche se la variabile considerata non è quantitativa (ossia non è numerica!).

E se due o più valori compaiono più volte a pari merito? Ci saranno più mode!!! Può anche capitare che un insieme di dati non abbia alcuna moda!

La **mediana** di una serie di dati disposti in ordine crescente è il valore centrale che divide la distribuzione in due gruppi di uguale numerosità.

Se la numerosità del gruppo di dati (n) è dispari, la mediana coincide con il posto centrale: $\frac{n+1}{2}$

Se la numerosità del gruppo di dati (n) è pari, la mediana è la media dei due valori centrali situati nei posti:

$$\frac{n}{2} \quad e \quad \frac{n}{2}+1$$

La mediana è anche detta secondo quantile. I quantili dividono una distribuzione in due parti lasciando da una parte una percentuale di casi e dall'altra la percentuale rimanente. Il secondo quantile, la mediana appunto, lascia sulla sinistra il 50% dei casi e sulla destra il rimanente 50%.

Un'importante proprietà della mediana è che minimizza la somma dei valori assoluti degli scarti da un generico valore.

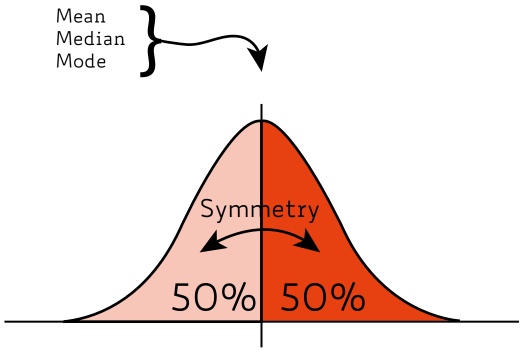
La **media (aritmetica)** di un insieme di dati numerici è la somma dei loro valori divisa per il numero dei dati (n):

$$media = \bar{x} = \frac{\Sigma^{n}_{i=1} x_i}{n} = \frac{x_1 + x_2 + x_3 + \ldots + x_n}{n}$$

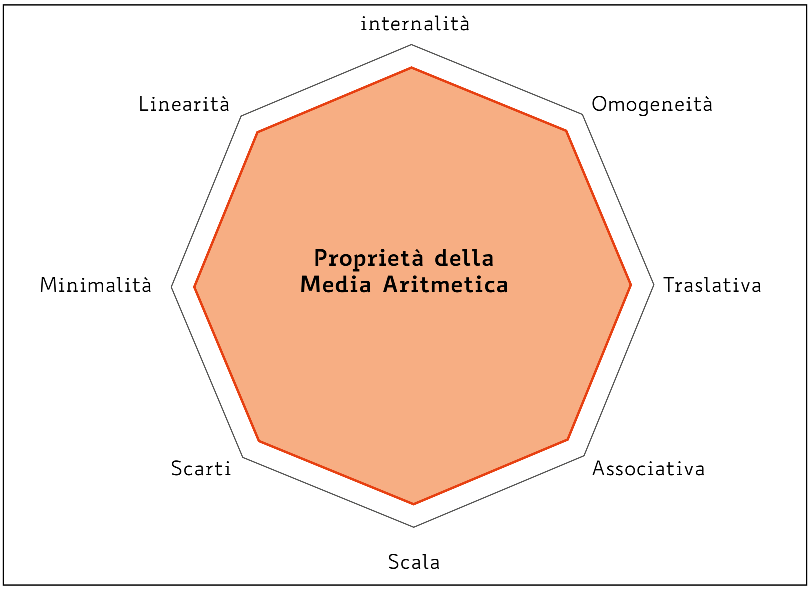
Questo simpatico simbolo Σ è la lettera greca "Sigma" (maiuscola!) e rappresenta la sommatoria! Si legge e si calcola come somma di tutti gli elementi i-esimi con indice i che va da 1 a n.

Qual è l'unità di misura di tali indici di posizione? La stessa del fenomeno che si sta misurando!

Nella curva Normale (curva simmetrica) moda, mediana e media coincidono.



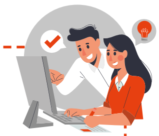
Le proprietà della media aritmetica sono le seguenti:



- **Internalità**: la media è compresa tra i valori minimo e massimo osservati.
- **Omogeneità**: se una serie di dati è ottenuta da un'altra moltiplicata per una costante, anche la media risulta moltiplicata per la medesima costante.
- **Traslativa**: se una serie di dati è ottenuta da un'altra sommando una costante, allora la stessa costante deve essere aggiunta alla media.
- **Associativa**: suddividendo in più gruppi le osservazioni di una serie di dati, la media è uguale alla media aritmetica delle medie parziali dei diversi gruppi ponderata per la numerosità di ciascun gruppo.
- **Scala**: la media conserva l'unità di misura dei valori sui quali viene calcolata.
- **Scarti**: la somma algebrica di tutti gli scarti dalla media è uguale a zero.
- **Minimalità**: la media minimizza la somma dei quadrati degli scarti da un generico valore.
- **Linearità**: se una serie di dati è ottenuta da un'altra mediante una trasformazione lineare, allora la media è soggetta alla stessa trasformazione lineare.

CONFRONTIAMOLI:

	VANTAGGI	SVANTAGGI	DA USARE PER
Moda	- Di facile identificazione - Esiste anche per valori qualitativi - Non è influenzata da valori estremi	- Potrebbe non esistere o essere multipla - Tiene conto solo dei valori più frequenti e non di tutti	- Dati qualitativi - Valore più probabile
Mediana	- Di facile identificazione - Non è influenzata da valori estremi	- Tiene conto solo dei valori centrali e non di tutti	- Dati con valori estremi (anche detti outliers)
Media	- Utilizza tutti i valori dell'insieme dei dati	- Devi calcolarla! - Gli outliers (o valori estremi/eccezionali) incidono e distorcono il risultato	- Dati abbastanza bilanciati nella distribuzione



Al lavoro!

1) Alcuni vostri coetanei hanno sostenuto il test di inglese PET. Gli scores nella parte di writing sono i seguenti: 42 38 56 37 61 53 72 56 45	Trova moda, mediana e media degli scores. Inoltre, qual è il campo di variazione dei voti?
2) Nella prima metà di ottobre 2021 le temperature medie giornaliere rilevate a Modena sono le seguenti:	Calcola la temperatura media della prima metà del mese di ottobre. Calcola l'escursione termica della prima metà del mese di ottobre

	Mese 10 Giorno 14	Mese 10 Giorno 13	Mese 10 Giorno 12	Mese 10 Giorno 11	Mese 10 Giorno 10	Mese 10 Giorno 09	Mese 10 Giorno 08	Mese 10 Giorno 07	Mese 10 Giorno 06	Mese 10 Giorno 05	Mese 10 Giorno 04	Mese 10 Giorno 03	Mese 10 Giorno 02
TEMP.	17,1	18,8	19,1	19,7	17,3	18,4	18,8	16,9	20,1	24,3	23,4	26,6	23,4

Indici di Dispersione

>>> *Gli indici di dispersione consentono di valutare la variabilità dei dati attorno ai valori centrali, indicando l'ampiezza dell'intervallo dei dati.*

>>> *I principali indici di dispersione sono il campo di variazione, la varianza e la deviazione standard.*

Il **campo di variazione (o range)** di un insieme di dati è la differenza fra il valore massimo e il valore minimo dei dati.

$$x_{max} - x_{min}$$

È fortemente influenzato da valori estremi eccezionali e quindi poco utilizzato!

La **varianza** è la media dei quadrati degli scarti dei singoli valori dalla loro media aritmetica.

La varianza si indica con la lettera greca "sigma" (minuscola!) al quadrato.

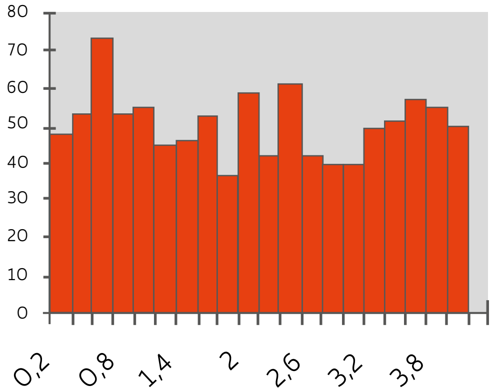
$$\text{varianza}: \sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

La **deviazione standard** è la radice quadrata della varianza: per questo è indicata con σ.

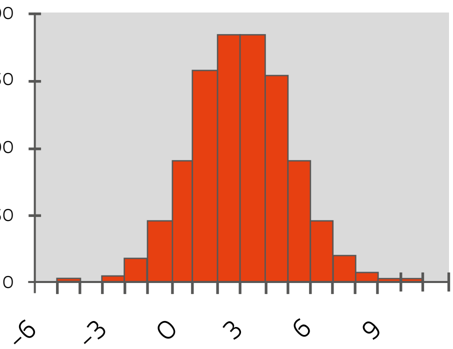
Qual è la loro unità di misura? Dipende dall'indice! Pensiamoci insieme...

Perché è importante far riferimento anche alla varianza e non solo alla media?

È importante perché due insiemi di dati possono avere la stessa media ma essere profondamente diversi per dispersione attorno al valore centrale, come in questo caso:

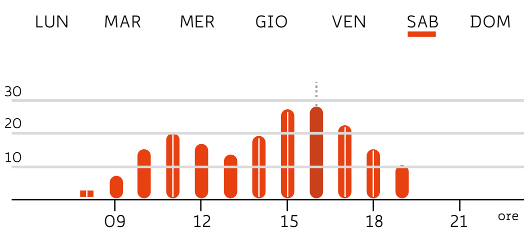


Stessa media = 2
ma varianze uguali a
1.33 (sx) e 4 (dx)



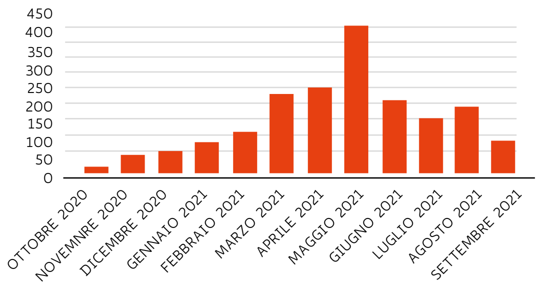
Al lavoro!

1) Nel grafico sono riportati i numeri degli accessi al Museo Estense di Modena nella giornata di sabato tra le ore 8 e le ore 19. In particolare, gli accessi sono:
2 7 14 19 17 13 18 25 26 21 16 10



- Indica la fascia oraria con maggior numero di accessi, il numero totale degli accessi di sabato, la media oraria degli accessi nella giornata di sabato, il campo di variazione degli accessi.
- Calcola la varianza e la deviazione standard degli accessi.
- Pensi che sia una curva Normale?

2) Utilizzando dati da Google Trend, si nota che i numeri di click della ricerca della parola "vaccino" nell'ultimo anno in tutto il mondo sono i seguenti (in miliardi di click):



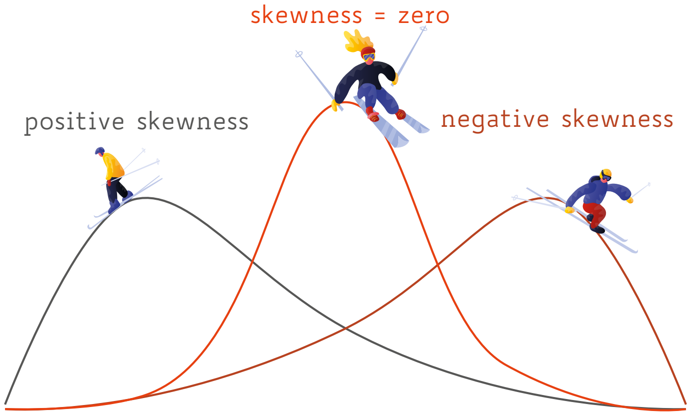
ottobre 2020	20
novembre 2020	55
dicembre 2020	67
gennario 2021	92
febbraio 2021	120
marzo 2021	230
aprile 2021	250
maggio 2021	428
giugno 2021	211
luglio 2021	161
agosto 2021	192
settembre 2021	98

- Indica il mese con maggior numero di click, la media annuale dei click e il campo di variazione.
- Che tipo di distribuzione è? Normale o no? Ti sembra simmetrica o asimmetrica attorno ai valori centrali? Questo ti introduce al prossimo paragrafo!

Indici di Forma

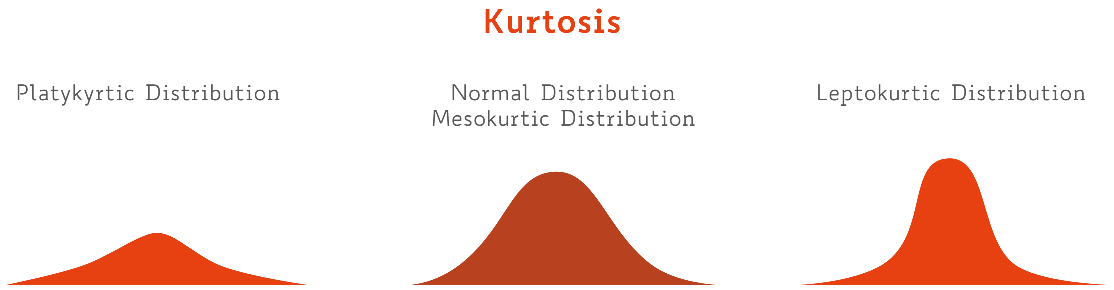
>>> *Gli indici di forma sono indicatori che descrivono ed evidenziano alcune caratteristiche della distribuzione dei dati, ossia il loro grado di simmetria/asimmetria e il loro appiattimento/appuntimento.*

La **skewness** caratterizza il grado di asimmetria di una distribuzione, ossia l'assenza di specularità di una distribuzione attorno al proprio valore centrale.



Per valutare una eventuale asimmetria si confronta la curva della distribuzione con quella Normale (o a campana) che è perfettamente simmetrica!

La **kurtosis** misura il maggiore o minore appuntimento di una distribuzione di dati, rispetto alla distribuzione normale. Di conseguenza essa indica il maggiore o minore peso dei valori posti agli estremi della distribuzione (dette code), rispetto a quelli della parte centrale.



La curva è platicurtica se è più piatta della curva Normale.

La curva è leptocurtica se è maggiormente appuntita della curva Normale.

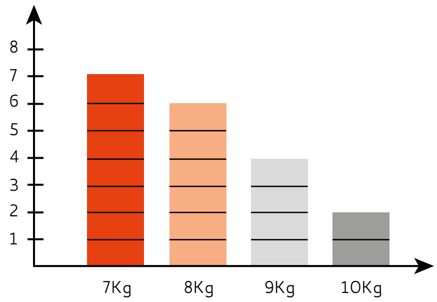
Esiste un indice di kurtosis, chiamato indice di kurtosis di Pearson, che è utile per misurare la kurtosis di una curva di forma campanulare. Esso è esattamente uguale a 3 per la curva Normale, maggiore di 3 per una distribuzione leptocurtica, e inferiore a 3 per una distribuzione platicurtica. Si calcola in questo modo:

$$k = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{x_i - \overline{x}}{\sigma} \right)^4$$

Cosa riconosci? Cosa noti di diverso?

Al lavoro!

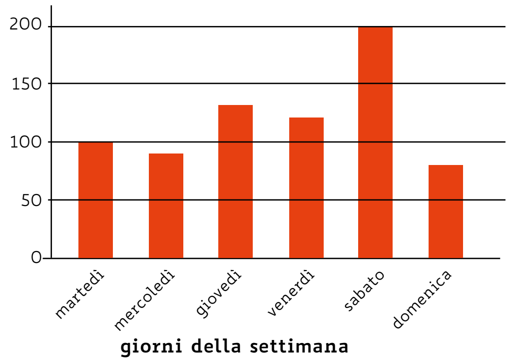
1) Considera la distribuzione dei dati sul peso dei vostri zaini (campione di 19 studenti delle medie):



Che tipo di asimmetria presenta questa distribuzione: positiva o negativa? Perché?

Calcola la media e la moda e posizionali sul grafico. Cosa noti?

2) Considera la distribuzione dei dati sul numero di panini Mc Donald's venduti in uno store in una settimana:



Pensi che la distribuzione dei panini nei giorni della settimana sia platicurtica o leptocurtica? Perché?

L'indice di kurtosis di Pearson sarà maggiore, minore o uguale a 3 per questa distribuzione?

Indici Statistici di Ordine Superiore: i Momenti

In statistica esiste una importante famiglia di indici, chiamati MOMENTI, di cui fa parte anche la media, come descritta sopra.

I momenti si dividono in:

- MOMENTI ORDINARI
- MOMENTI CENTRALI

Questi indici, che possono essere calcolati solo per variabili quantitative, consentono di evidenziare diverse caratteristiche della variabile considerata.

Il **momento ordinario** di ordine r corrisponde alla media delle potenze r-esime dei dati osservati.

Considera n osservazioni (dati): $x_1, x_2, \ldots, x_n$

Considera la stessa sequenza di valori elevata alla potenza r-esima: $x^r_1, x^r_2, \ldots, x^r_n$

Il momento ordinario di ordine r è uguale a

$$m_r = \frac{1}{n} \sum_{i=1}^{n} x^r_i \text{ per } r = 0,1,2, \ldots$$

· Sapresti dirmi cosa risulta il momento ordinario di ordine zero (r=0)?

· E per r=1, cosa riconosci?

Il momento ordinario di ordine r=2 interviene nel calcolo della varianza. Infatti

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2 = \frac{1}{n} \sum_{i=1}^{n} x^2_i - \bar{x}^2 - 2 \frac{1}{n} \sum_{i=1}^{n} x_i \bar{x} = \sigma^2 = \frac{1}{n} \sum_{i=1}^{n} x^2_i - \bar{x}^2 = m_2 - \bar{x}^2$$

Al lavoro!

Data la sequenza di cinque valori di una variabile discreta X: 0 5 11 16 28	Calcola il momento ordinario di ordine 2 e di ordine 3.

Il **momento centrale** di ordine r corrisponde invece alla media delle potenze r-esime degli scarti.

Gli scarti sono le differenze degli n valori dalla loro media!

$$\overline{m}_r = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^r \text{ per } r = 0,1,2, \ldots$$

Nota che il momento centrale di ordine r=1 è la media degli scarti che è uguale a 0.

· Sapresti dirmi cosa risulta il momento centrale di ordine zero (r=0)?

· E per r=2, cosa riconosci?

Al lavoro!

Data la sequenza di cinque valori di una variabile discreta X: 0 5 11 16 28	Calcola il momento centrale di ordine 3. Che caratteristica ha?

Che cos'è la probabilità?

Esistono proposizioni di cui possiamo dire con certezza se siano vere o false, ad esempio "Cristofolo Colombo scoprì l'America nel 1492", oppure "Nel 2020 l'Italia ha vinto gli europei di calcio".

Esistono poi proposizioni incerte di cui ora non sappiamo dire se siano vere o false, ad esempio, "lanciando un dado a sei face ottengo 4".

Fin che non abbiamo lanciato il dado, non sappiamo che numero uscirà! La nostra proposizione diventerà certamente vera o falsa solo dopo avere lanciato il dado. Queste proposizioni affette da **incertezza** si chiamano **eventi**.

Anche se non possiamo sapere il numero che uscirà prima di lanciare il dado, possiamo misurare la certezza/incertezza di ottenere un 4, ovvero possiamo calcolare la **probabilità** che questo evento accada.

Quindi la **probabilità** rappresenta una misura della certezza/incertezza di proposizioni di questo tipo. L'incerto è inevitabilmente presente nella nostra vita, fortuna che la probabilità ci può aiutare a prendere delle decisioni!

A chi sta sistemare il tavolo per cena?

Luca e i suoi 4 fratelli e sorelle danno sempre una mano con i lavori di casa, ma speso fanno fatica a mettersi d'accordo su a chi tocca farlo. Luca ha avuto una idea per decidere chi sistema il tavolo per la cena ogni sera: mettere 5 biglietti con i loro nomi in una scatola, per poi fare pescare uno alla mamma; il nome scritto sul biglietto estratto sarà quello del fratello che deve sistemare la tavola.
A chi toccherà stasera?

- *Luca pensa che sarà fortunato e non toccherà a lui visto che l'idea è stata sua!*
- *Andrea ritiene che lui e Tommaso siano i più sfortunati, quindi crede che sarà uno di loro due a dovere sistemare il tavolo.*
- *Sara è convinta che sua mamma non pescherà mai il suo biglietto dato che è la piccolina della casa...*

· *Laura è sicura che non toccherà a lei perché ha già sistemato il tavolo la sera precedente.*

· *Infine, Tommaso è convinto che sarà un maschio a farlo.*

Definizione classica della probabilità

I 5 fratelli e sorelle hanno le idee molto chiare, ma quanta ragione hanno? Cerchiamo di calcolare la probabilità che ogni uno di loro deva sistemare il tavolo. Ci sono 5 biglietti nella scatola, di cui la mamma ne pesca solo uno, per cui ogni biglietto ha una probabilità di 1/5 di essere pescato. Come siamo arrivati a questo numero? **Numero di eventi favorevoli / numero di eventi possibili.**

La estrazione del biglietto si chiama **prova** o **esperimento aleatorio**, possiamo vederlo come un esperimento con possibili risultati, e questi risultati sono incerti. Ogni possibile risultato viene chiamato **evento elementare**. L'insieme di tutti gli eventi elementari forma lo **spazio degli eventi** Ω. Nel nostro esempio:

$$\Omega = \{\text{Luca, Andrea, Sara, Laura, Tommaso}\}$$

In questo caso, parliamo di eventi **equiprobabili**, tutti i fratelli e sorelle hanno esattamente la stessa probabilità di essere scelti per sistemare il tavolo (nonostante quello che loro pensano!). Possiamo formare degli eventi a partire dagli **eventi** elementari. Prendiamo la affermazione di Tommaso, e calcoliamo la probabilità che venga scelto un maschio:

M = *maschio* = {Luca, Andrea, Tommaso}

F = *femmina* = {Sara, Laura}

P(M) = numero di maschi/numero totale = 3/5

mentre P(F)= 2/5

Nota che gli eventi vengono indicati con le lettere maiuscole dell'alfabeto e la loro probabilità con la lettera P

Quindi, Tommaso non può essere convinto della sua scelta, ma comunque è più probabile che sia un maschio che una femmina a dovere sistemare il tavolo!

Consideriamo adesso quello che pensa Andrea, cioè che la mamma pescherà il suo biglietto **oppure** quello di Tommaso. Definiamo gli eventi A = "Andrea" e B = "Tommaso", vogliamo calcolare la probabilità che si verifichi A oppure B, ovvero che almeno uno fra A e B si verifichi. In termini di eventi, questo "oppure" lo traduciamo nell'evento **unione** $A \cup B$, e la probabilità:

$$P(A \cup B) = 2/5$$

Il simbolo ∅ indica l'insieme vuoto, ovvero il fatto che A e B non hanno niente in comune

Nota che i due eventi A e B non possono verificarsi contemporaneamente (la mamma pesca solo un biglietto, quindi non potrà avere entrambi i nomi allo stesso tempo); diciamo che gli eventi A e B sono **incompatibili**, e l'evento **intersezione** $A \cap B = \emptyset$.

Nota che abbiamo "tradotto": A oppure B = evento unione ∪ A e B = evento intersezione ∩

Che proprietà ha la probabilità?

Quanto può valere la probabilità?

- Se pensiamo a come è definita, come un rapporto tra numeri di eventi, è chiaro che la probabilità è sempre un numero positivo. Quando vale 0? Quando il numero di eventi favorevoli è 0, ad esempio, nel caso di Luca e i suoi fratelli, pensate all'evento:

E = "*il padre sistema il tavolo*"

In questo caso, il bigliettino con il nome del papà non è nella scatola, per cui la probabilità di questo evento è 0, è un **evento impossibile.**

- Il valore massimo che può prendere si ottiene quando il numero di eventi favorevoli coincide con il numero di eventi possibile, e quindi al massimo vale 1. Consideriamo l'evento:

A = "*uno dei cinque fratelli/sorelle sistema il tavolo*"

Adesso, il numero di eventi favorevoli è 5 (l'evento A coincide con lo spazio degli eventi Ω), è un **evento certo** e la sua probabilità e pari a 1.

Come calcoliamo la probabilità dell'evento unione?

Prima abbiamo calcolato la probabilità dell'evento unione $A \cup B$ = "*Andrea oppure Tommaso sistema il tavolo*", notando che i due eventi A e B sono incompatibili ($A \cap B = \emptyset$). In questo caso,

$$P(A \cup B) = P(A) + P(B)$$

Abbiamo appena individuo le tre proprietà fondamentali della probabilità, che, in modo formale vengono chiamate gli **assiomi della probabilità**. Questi tre assiomi formalizzano delle idee intuitive ed evidenti e devono sempre verificarsi per qualsiasi definizione della probabilità.

Assiomi della probabilità
1. $P(A) \geq 0$
2. $P(\Omega) = 1$
3. Se $A \cap B = \emptyset$, $P(A \cup B) = P(A) + P(B)$

Al lavoro!

1. Torna all'esempio di Luca e i suoi fratelli e sorelle e calcola la probabilità che sia un maschio oppure una femmina a dovere sistemare il tavolo a cena.
2. L'esame di matematica consiste in un test a risposta multipla dove ogni domanda ha 5 possibili risposte. Non hai avuto modo di studiare tutto e non hai idea di quale sia la risposta corretta all'ultima domanda dell'esame. Qual è la probabilità che indovini la risposta corretta?
3. Nel lancio di un dado, calcola la probabilità che:
   a. esca un numero pari
   b. esca un numero maggiore o uguale a 5
   c. esca un 1 oppure un 3

Serata pizza!

I compagni della terza media si sono trovati per fare una serata pizza insieme. Decidono di ordinare le due specialità del locale, due pizze con vari gusti da condividere. La prima pizza è divisa in 6 spicchi e la seconda in 9. Per rendere la serata più divertenti, decidono di scegliere una pizza e poi prendere un pezzo a caso ad occhi chiusi. Matteo è un po' preoccupato perché a lui non piace molto il piccante... da che pizza gli conviene pescare?

PIZZA 1

2 spicchi margherita

2 spicchi salame piccante

1 spicchio salsiccia

1 spicchio ortolana

PIZZA 2

3 spicchi prosciutto e funghi

2 spicchi melanzane

3 spicchi salame piccante

1 spicchio margherita

Consideriamo l'evento A = "*pescare uno spicchio al salame piccante*" e calcoliamo la probabilità che questo evento si verifichi:

Se Matteo pesca dalla PIZZA 1: P(A) = 2/6 = 1/3

Se Matteo pesca dalla PIZZA 2: P(A) = 3/9 = 1/3

Quindi ha esattamente la stessa probabilità di essere sfortunato e prendere uno spicchio al salame piccante scegliendo sia la pizza 1 che la pizza 2.

Possiamo anche considerare la negazione di un evento A, o evento **complementare**, che denotiamo con $\overline{A}$. La sua probabilità si calcola come $P(\overline{A}) = 1 - P(A)$. Nel caso di Matteo:

$\overline{A}$ = "*pescare uno spicchio diverso da quello al salame piccante*"

$$P(\overline{A}) = 1 - 1/3 = 2/3$$

Indipendenza fra eventi

Due eventi A e B sono **indipendenti** se il fatto che A si sia verificato non modifica la probabilità del verificarsi di B, e viceversa. Più formalmente, due eventi A e B si dicono indipendenti se e solo se:

$$P(A \cap B) = P(A) \times P(B)$$

Andiamo avanti con la serata pizza. Matteo è il primo a prendere due spicchi di pizza, uno di ogn'una, sperando non solo di evitare il salame piccante, ma di beccare i suoi gusti preferiti, salsiccia e melanzane. I due eventi che ci interessano adesso sono:

- A= "*pescare uno spicchio alla salsiccia dalla pizza 1*"
- B= "*pescare uno spicchio alle melanzane dalla pizza 2*"

I due eventi sono indipendenti, Matteo sta pescando di due pizze diverse, quindi quello che trova su una non condiziona quello che troverà sull'altra, ed è il primo a pescare (ovvero le pizze sono ancora intatte). La probabilità che Matteo prenda proprio i suoi due gusti preferiti sarà:

$$P(A \cap B) = P(A) \times P(B) = 1/6 \times 2/9 = 2/54 = 1/27$$

Adesso immaginate che invece di pescare uno spicchio di ogni pizza, Matteo decide di pescare due spicchi dalla pizza 1, sperando di prendere prima uno alla salsiccia ed un secondo spicchio margherita. Che probabilità ha Matteo di soddisfare i suoi desideri? Pensiamo prima agli eventi:

- A = "*pescare il primo spicchio alla salsiccia dalla pizza 1*"
- B = "*pescare il secondo spicchio margherita dalla pizza 1*"

Vogliamo sempre calcolare la probabilità dell'evento intersezione $A \cap B$, ma adesso dobbiamo stare attenti, i due eventi non sono indipendenti. Perché? Una volta che Matteo ha preso il primo spicchio, il numero di spicchi rimasti tra cui scegliere il secondo spicchio cambia e non è più 6 sino 5. Questa idea si formalizza nel concetto di **probabilità condizionata**, ovvero, per l'evento B, dobbiamo calcolare la probabilità che il secondo spicchio di pizza sia margherita condizionatamente al fatto che il primo spicchio preso è alla salsiccia. Formalmente, lo scriviamo in questo modo: $P(B|A)$.

Per calcolare la probabilità dell'evento $B|A$, dobbiamo "dare per certo" che il primo spicchio è alla salsiccia. Questo implica che lo spazio degli eventi Ω ha solo 5 elementi all'ora di prendere il secondo spicchio (perché lo spicchio alla salsiccia non c'è più):

Ω= {margherita, margherita, salame piccante, ortolana}

La probabilità richiesta $P(A \cap B)$, nel caso di eventi A e B dipendenti, si può calcolare come:

$$P(A \cap B) = P(A) \times P(B|A) = 1/6 \times 2/5 = 2/30 = 1/15$$

Al lavoro!

1. Quale dei seguenti eventi è più probabile nel lancio di una moneta 10 volte?
   a. TTTTTTTTTT
   b. TCTCTCTCTC
   c. TTCTCCTCCC
2. Nel lancio di due dadi, calcola la probabilità che esca il numero 6 in entrambi i dadi.
3. Torna all'esempio della serata pizza e calcola la probabilità che, pescando due spicchi dalla pizza 2, Matteo prenda proprio i due spicchi alle melanzane.

Ci sono tanti altri esempi dove la probabilità può essere utile. La definizione classica della probabilità assume che gli eventi elementari siano equiprobabili. Ci sono altre definizioni più utili per il calcolo di probabilità di eventi in problemi concreti. Andiamo alla caccia dei Pokemon per scoprirlo.

Alla caccia dei Pokemon

I compagni della prima media stanno organizzando una gita per andare alla caccia dei Pokemon. Devono decidere il posto dove andare. Così, hanno messo insieme le loro esperienze passate per cercare di trovare quale è il posto migliore. Nella tabella di sotto trovate un riassunto su quante volte sono stati (Numero totale di visite) e quante di queste volte hanno trovato qualche Pokemon (Successi) in diverse città.

	Modena	Reggio-Emilia	Carpi	Vignola	Formigine
Successi	180	108	45	19	42
Numero totale di visite	300	270	125	190	210

Ogni visita ad una città in cerca di Pokemon è una prova, che viene ripetuta (nelle stesse condizioni) un numero di volte, ad esempio 300 volte nel caso della città di Modena. Il risultato di ogni prova (Pokemon trovati si/no) non è sempre lo stesso. In questo caso, possiamo calcolare la probabilità di un evento come:

numero di successi/numero totale di prove

Definizione frequentista della probabilità

Nota che anche questa definizione di probabilità rispetta i tre assiomi!

Nella tabella sotto troviamo la probabilità di cacciare dei Pokemon in ogni città:

	Modena	Reggio-Emilia	Carpi	Vignola	Formigine
Probabilità	0.60	0.40	0.36	0.10	0.20

Attenti, andando a Modena è più probabile cacciare un Pokemon, ma non c'è garanzia! Ricordate che parliamo di eventi incerti...

Al lavoro!

Lancia una moneta 5 volte e segna quante volte esce testa per poi calcolare la probabilità di ottenere testa. Adesso ripete l'esperimento lanciando la moneta 10 volte, 30 volte, e poi 50 volte. Cosa osservi sulla probabilità calcolata su 5, 10, 30 e 50 lanci?

Alcune curiosità sulla probabilità

- Gli inizi della probabilità risalgono al XVII secolo e sono legati ai giochi d'azzardo
- Probabilmente quasi tutti voi avete giocato qualche volta al Monopoly. Sapevate che la casella con la maggiore probabilità è quella del carcere?
- In una classe con 23 studenti, c'è una probabilità del 50% che ce ne siano 2 con lo stesso compleanno! In una classe di 75 studenti, la probabilità è del 99%!

UNIMORE
UNIVERSITÀ DEGLI STUDI DI
MODENA E REGGIO EMILIA

ISBN 978-88-89427-03-3