CORSO DI PREPARAZIONE PER RESPONSABILE DEL SISTEMA DI GESTIONE DELLA SICUREZZA

PRINCIPI DI STATISTICA



INTRODUZIONE

Nelle prossime quattro ore discuteremo insieme dei concetti fondamentali della statistica descrittiva e della probabilità, strumenti essenziali per analizzare, interpretare e presentare dati raccolti in vari contesti.

L'obiettivo principale di questa sessione è di trasferirvi gli elementi principali della statistica di base, consentendovi di comprendere e applicare in modo efficace quanto appreso.

CORSO DI PREPARAZIONE PER RESPONSABILE DEL SISTEMA DI GESTIONE DELLA SICUREZZA

La statistica è una disciplina scientifica fondamentale che **si occupa della raccolta, organizzazione, analisi, interpretazione e presentazione dei dati**.

È suddivisa in due principali rami: la statistica descrittiva e la statistica inferenziale.

La **STATISTICA DESCRITTIVA** consente di riassumere i dati raccolti attraverso metodi come grafici e indici numerici, fornendo un quadro chiaro e sintetico.

CORSO DI PREPARAZIONE PER RESPONSABILE DEL SISTEMA DI GESTIONE DELLA SICUREZZA

La **STATISTICA INFERENZIALE**, invece, permette di effettuare previsioni e generalizzazioni a partire dai dati di un campione, applicando specifiche tecniche probabilistiche.

Un esempio concreto potrebbe essere la raccolta e analisi di dati provenienti da un questionario aziendale, mirato a valutare aspetti importanti come per esempio il livello di sicurezza sul lavoro, identificando eventuali criticità o punti di forza.

CORSO DI PREPARAZIONE PER RESPONSABILE DEL SISTEMA DI GESTIONE DELLA SICUREZZA

Chiariamo qualche definizione di base:

La **POPOLAZIONE STATISTICA** rappresenta l'intero insieme di elementi sui quali intendiamo condurre uno studio o una ricerca.

Ognuno degli elementi di questo insieme è denominato **UNITÀ STATISTICA**.

Eseguire un'analisi su tutti gli elementi della popolazione può risultare difficile o addirittura impossibile, soprattutto quando questa è molto grande per cui utilizziamo un **CAMPIONE**

CORSO DI PREPARAZIONE PER RESPONSABILE DEL SISTEMA DI GESTIONE DELLA SICUREZZA

Un **CAMPIONE** è un sottoinsieme della popolazione statistica scelto in modo rappresentativo.

Per permettere generalizzazioni attendibili, il campione deve rispecchiare le caratteristiche principali della popolazione.

Per esempio, se una popolazione aziendale è composta da 500 dipendenti, un campione rappresentativo potrebbe includere 50 dipendenti **selezionati in modo casuale o stratificato** per ottenere risultati accurati e generalizzabili.

CORSO DI PREPARAZIONE PER RESPONSABILE DEL SISTEMA DI GESTIONE DELLA SICUREZZA

Scegliere un campione in modo stratificato significa suddividere la popolazione statistica in sottogruppi omogenei (detti strati o strati statistici) prima di estrarre il campione, e poi selezionare un certo numero di elementi da ciascuno di questi strati.

In pratica: Si divide la popolazione in «strati» in base a una o più caratteristiche rilevanti e si estrae un campione da ciascuno strato in modo proporzionale.

MA PERCHÉ LO SI FA?

CORSO DI PREPARAZIONE PER RESPONSABILE DEL SISTEMA DI GESTIONE DELLA SICUREZZA

Per rappresentare meglio la popolazione (soprattutto se è molto eterogenea) e per aumentare l'accuratezza delle stime garantendo che nel campione siano rappresentati tutti i gruppi, anche quelli più piccoli.

Per esempio: abbiamo una popolazione di 1.000 persone di cui 600 donne e 400 uomini di cui vogliamo studiare delle caratteristiche significative (es. incidenza dei fumatori) e decidiamo di eseguire uno studio su un campione di 100 persone.

CORSO DI PREPARAZIONE PER RESPONSABILE DEL SISTEMA DI GESTIONE DELLA SICUREZZA

Se eseguiamo il campionamento in maniera del tutto casuale, senza stratificare i dati, potremmo trovarci una popolazione con, ad esempio, 70 donne e 30 uomini, oppure un campione ancora più sbilanciato dal punto di vista della rappresentatività.

Il modo più corretto è di **stratificare il campionamento in modo proporzionale**, scegliendo casualmente 60 donne fra le 600 che compongono la popolazione e 40 uomini sui 400 di partenza.

In questo modo il campione sarà perfettamente rappresentativo.

CARATTERI E MODALITÀ

Quando analizziamo una serie di dati, dobbiamo considerare alcune caratteristiche chiamate «caratteri», che possono essere distinti in **QUALITATIVI** e **QUANTITATIVI**.

I **CARATTERI QUALITATIVI** rappresentano attributi non misurabili numericamente, come ad esempio il ruolo aziendale (operaio, impiegato, dirigente) o il livello di istruzione.

CARATTERI E MODALITÀ

I **CARATTERI QUANTITATIVI** sono invece grandezze misurabili numericamente e possono essere discreti o continui, come ad esempio l'età, il numero di figli o il reddito.

Le diverse varianti o possibili manifestazioni di ciascun carattere prendono il nome di "modalità".

Ad esempio, considerando il carattere "ruolo aziendale", le modalità saranno "operaio", "impiegato", "dirigente«, ecc.

CORSO DI PREPARAZIONE PER RESPONSABILE DEL SISTEMA DI GESTIONE DELLA SICUREZZA

Per quanto abbiamo detto finora, **per analizzare statisticamente un fenomeno in maniera adeguata è necessario capire a priori la natura del fenomeno che vogliamo studiare e stabilirne le caratteristiche salienti**.

È importante comprendere chiaramente queste distinzioni perché influenzano direttamente il modo in cui raccogliamo e analizziamo i dati statistici ed in ultima analisi influenzano i risultati a cui arriviamo.

FREQUENZE ASSOLUTE

La **FREQUENZA ASSOLUTA** rappresenta il numero di volte in cui una specifica modalità compare in un insieme di osservazioni.

È uno strumento essenziale per quantificare i dati raccolti e per iniziare un'analisi descrittiva efficace.

Ad esempio, se consideriamo una piccola azienda composta da 50 dipendenti, possiamo rilevare quanti di essi occupano ciascun ruolo lavorativo.

FREQUENZE ASSOLUTE

Se 25 dipendenti sono operai, 15 impiegati e 10 dirigenti, queste cifre costituiscono rispettivamente le frequenze assolute per ciascuna categoria. Attraverso la frequenza assoluta, possiamo comprendere rapidamente la distribuzione dei dati e identificare immediatamente quale modalità è più o meno frequente, fornendo una base solida per ulteriori analisi descrittive e interpretative.

FREQUENZE RELATIVE

La **FREQUENZA RELATIVA è ottenuta dividendo la frequenza assoluta per il numero totale delle osservazioni**.

Per consentire una rapida comprensione della distribuzione dei dati, la frequenza relativa viene spesso espressa in percentuale.

Prendendo in considerazione l'esempio precedente, con 50 dipendenti, di cui 25 operai, possiamo calcolare che la frequenza relativa degli operai è il 50% (25 diviso 50).

FREQUENZE RELATIVE

Questo valore (50%) ci consente di interpretare immediatamente che metà della forza lavoro aziendale appartiene alla categoria degli operai.

Le frequenze relative, quindi, offrono un'immagine chiara e immediata delle proporzioni e possono essere particolarmente utili quando confrontiamo gruppi diversi o quando presentiamo i risultati della nostra analisi statistica a un pubblico non tecnico.

RAPPRESENTAZIONI GRAFICHE: DIAGRAMMA A BARRE

La **rappresentazione grafica dei dati qualitativi** tramite diagrammi a barre è uno dei metodi più semplici ed efficaci per visualizzare chiaramente la distribuzione delle modalità.

Nel nostro esempio aziendale, potremmo creare un grafico a barre con tre barre separate per operai, impiegati e dirigenti, ciascuna alta proporzionalmente alla frequenza assoluta o relativa della categoria rappresentata.

RAPPRESENTAZIONI GRAFICHE: DIAGRAMMA A BARRE

Questo tipo di grafico facilita l'identificazione visiva delle differenze numeriche tra categorie, permettendo anche ai meno esperti di comprendere rapidamente e chiaramente i dati raccolti. Utilizzando un diagramma a barre possiamo evidenziare facilmente quali gruppi siano maggioritari o minoritari, semplificando la comunicazione e la comprensione delle informazioni presentate.

RAPPRESENTAZIONI GRAFICHE: DIAGRAMMI A TORTA

Un altro metodo molto efficace per visualizzare i dati qualitativi è rappresentato dai **DIAGRAMMI A TORTA**.

In questo grafico ogni modalità viene rappresentata da un settore della «torta», la cui grandezza (angolo al centro) è proporzionale alla frequenza relativa di quella modalità. Se, ad esempio, il 50% dei dipendenti sono operai, metà della «torta» rappresenterà gli operai.

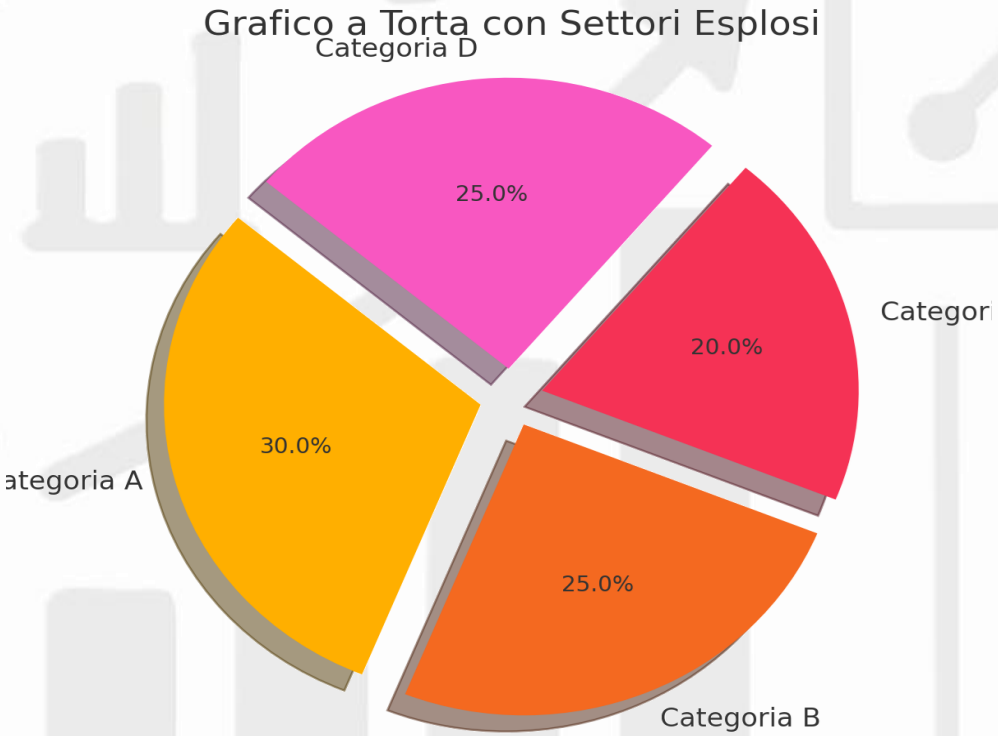
RAPPRESENTAZIONI GRAFICHE: DIAGRAMMI A TORTA

I diagrammi a torta sono particolarmente utili per mostrare chiaramente la composizione percentuale dei dati e sono ideali per evidenziare immediatamente l'importanza relativa di ciascuna categoria.

Sono particolarmente efficaci quando il numero delle modalità è limitato, permettendo una visione immediata e intuitiva della struttura dei dati raccolti.

CORSO DI PREPARAZIONE PER RESPONSABILE DEL SISTEMA DI GESTIONE DELLA SICUREZZA

RAPPRESENTAZIONI GRAFICHE: DIAGRAMMI A TORTA



RAPPRESENTAZIONI GRAFICHE: DIAGRAMMA DI PARETO

Il **DIAGRAMMA DI PARETO** è uno strumento grafico particolarmente efficace per analizzare dati qualitativi, ordinando le modalità in ordine decrescente di frequenza.

Esso combina un diagramma a barre con una curva cumulativa sovrapposta, che mostra la proporzione cumulativa delle frequenze.

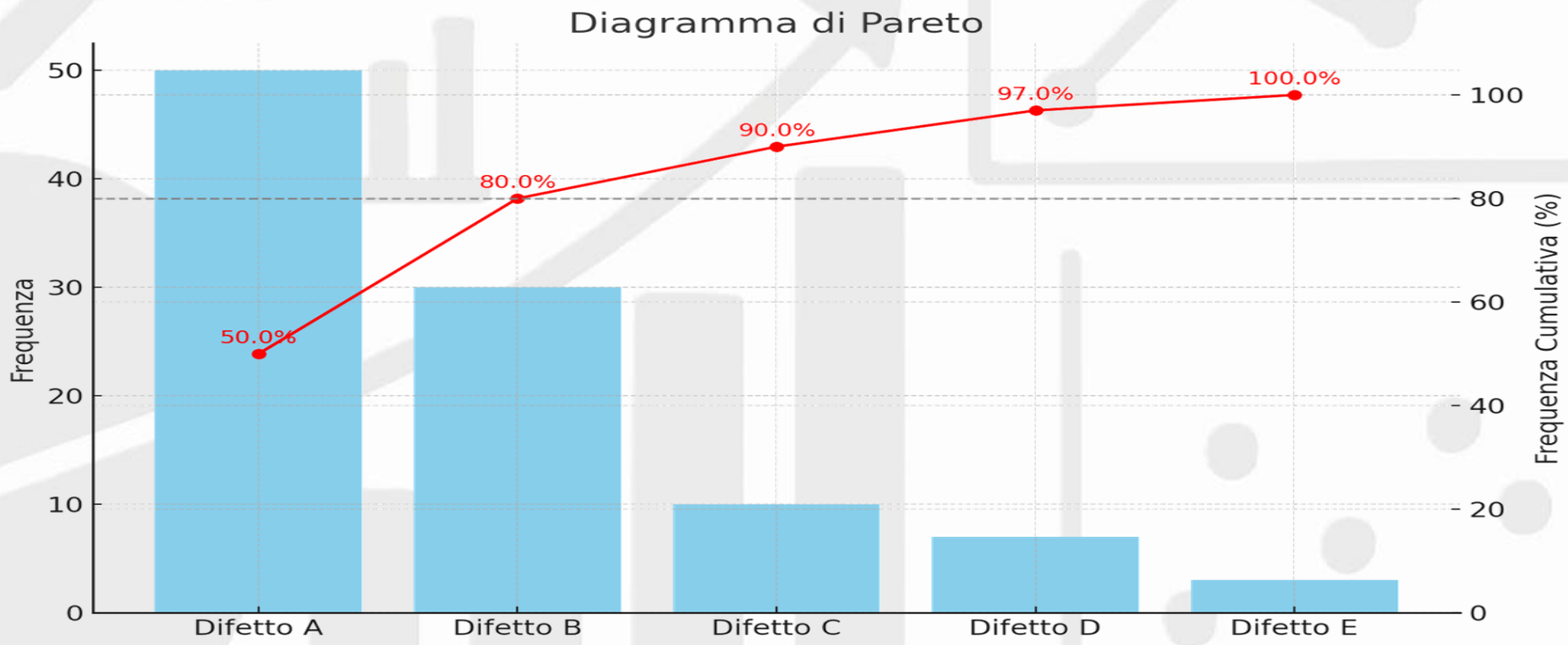
CORSO DI PREPARAZIONE PER RESPONSABILE DEL SISTEMA DI GESTIONE DELLA SICUREZZA

Il DIAGRAMMA DI PARETO è molto utile, ad esempio, nell'analisi delle cause degli infortuni sul lavoro, poiché evidenzia chiaramente quali siano i fattori più rilevanti che contribuiscono al fenomeno studiato.

L'utilizzo del diagramma di Pareto consente di identificare rapidamente quali problematiche meritano maggiore attenzione e aiuta a dare priorità ad interventi mirati per risolvere le cause principali di un determinato problema o fenomeno osservato.

CORSO DI PREPARAZIONE PER RESPONSABILE DEL SISTEMA DI GESTIONE DELLA SICUREZZA

RAPPRESENTAZIONI GRAFICHE: DIAGRAMMA DI PARETO



RAGGRUPPAMENTO DEI DATI IN CLASSI

Quando si trattano **dati quantitativi continui**, è frequente raggruppare i dati in intervalli definiti chiamati "**CLASSI**".

Questa tecnica facilita la gestione e l'analisi dei dati, soprattutto quando il numero di osservazioni è elevato o quando il range di valori è ampio.

Ad esempio, possiamo raggruppare le età dei lavoratori in classi tipo: 20-30 anni, 31-40 anni e 41-50 anni.

RAGGRUPPAMENTO DEI DATI IN CLASSI

Dopo aver stabilito le classi, calcoliamo le frequenze assolute e relative **per ciascuna classe**, ottenendo così una distribuzione chiara e ordinata dei dati. Questo metodo semplifica ulteriori analisi statistiche e rende possibile l'uso di rappresentazioni grafiche come l'istogramma, che forniscono un quadro immediatamente comprensibile della distribuzione e della variabilità dei dati raccolti.

DENSITÀ DI FREQUENZA:

La DENSITÀ DI FREQUENZA è una misura statistica fondamentale, particolarmente utile quando si analizzano dati quantitativi continui suddivisi in classi di diversa ampiezza.

Si utilizza principalmente quando le classi non hanno tutte la stessa ampiezza; in questo caso un confronto diretto delle classi basato soltanto sulla frequenza assoluta è inadeguato.

Infatti, per rappresentare correttamente i dati, è necessario considerare non solo quante osservazioni si trovano in ciascuna classe, ma anche quanto «ampia» è la classe stessa.

DENSITÀ DI FREQUENZA:

La densità di frequenza si calcola mediante una formula molto semplice: si divide la frequenza assoluta (cioè il numero di osservazioni) di ogni classe per la propria ampiezza (cioè l'intervallo di valori compresi nella stessa classe).

Questo rapporto consente di ottenere un valore che rappresenta la **concentrazione delle osservazioni per unità di ampiezza**, rendendo così possibile un confronto equo tra classi di dimensioni diverse.

DENSITÀ DI FREQUENZA:

Ad esempio, se abbiamo una classe di età compresa tra 20 e 25 anni che contiene 10 lavoratori, l'ampiezza della classe è pari a 5 anni (25-20).

La **densità di frequenza sarà quindi pari a 2** (lavoratori per anno), ottenuta dividendo 10 per 5.
Questo valore ci dice che, mediamente, ci sono due lavoratori per ogni anno di età in quella classe.

DENSITÀ DI FREQUENZA:

Utilizzare **la DENSITÀ DI FREQUENZA è particolarmente importante per costruire grafici come L'ISTOGRAMMA**, uno strumento visivo essenziale nell'analisi esplorativa dei dati.

In un istogramma basato sulla densità di frequenza, **l'altezza di ciascun rettangolo rappresenta appunto la densità e non la frequenza assoluta**: questo permette di rispettare le proporzioni tra le varie classi e di ottenere una rappresentazione visivamente coerente della distribuzione.

DENSITÀ DI FREQUENZA:

La densità di frequenza è uno strumento indispensabile per una corretta analisi dei dati raggruppati, poiché consente di:

1. Rappresentare i dati in modo equo anche con classi non uniformi.
2. Evitare distorsioni visive nei grafici.
3. Facilitare il confronto tra classi diverse per ampiezza.
4. Evidenziare con maggiore chiarezza le zone di maggiore o minore concentrazione dei dati.

ISTOGRAMMA:

L'ISTOGRAMMA è una rappresentazione grafica dei dati quantitativi continui organizzati in classi.

Le barre adiacenti rappresentano le classi, e l'altezza di ciascuna barra corrisponde alla densità di frequenza, o alla frequenza assoluta o relativa.

Questa rappresentazione permette di evidenziare rapidamente la distribuzione e la concentrazione dei dati, facilitando la comprensione di dove si trovano i valori più frequenti.

ISTOGRAMMA:

Per esempio, se analizziamo l'età dei lavoratori in un'azienda, possiamo visualizzare rapidamente in quali intervalli d'età si concentra la maggioranza dei dipendenti.

L'istogramma è dunque uno strumento indispensabile per identificare eventuali anomalie, concentrazioni o dispersioni che possono orientare ulteriori analisi e decisioni aziendali.

GRAFICO A BARRE O ISTOGRAMMA?

La differenza tra un **GRAFICO A BARRE** e un **ISTOGRAMMA** è fondamentale in statistica, anche se a prima vista possono sembrare due rappresentazioni molto simili.

La differenza fondamentale fra le due rappresentazioni sta nel tipo di dati con cui vengono utilizzate:

Il grafico a barre viene usato per rappresentare **dati qualitativi o dati discreti** (cioè dati che si contano, non si misurano);

L'istogramma è utilizzato per rappresentare **dati quantitativi continui suddivisi in classi di valori** (es. età, peso, altezza, ecc.).

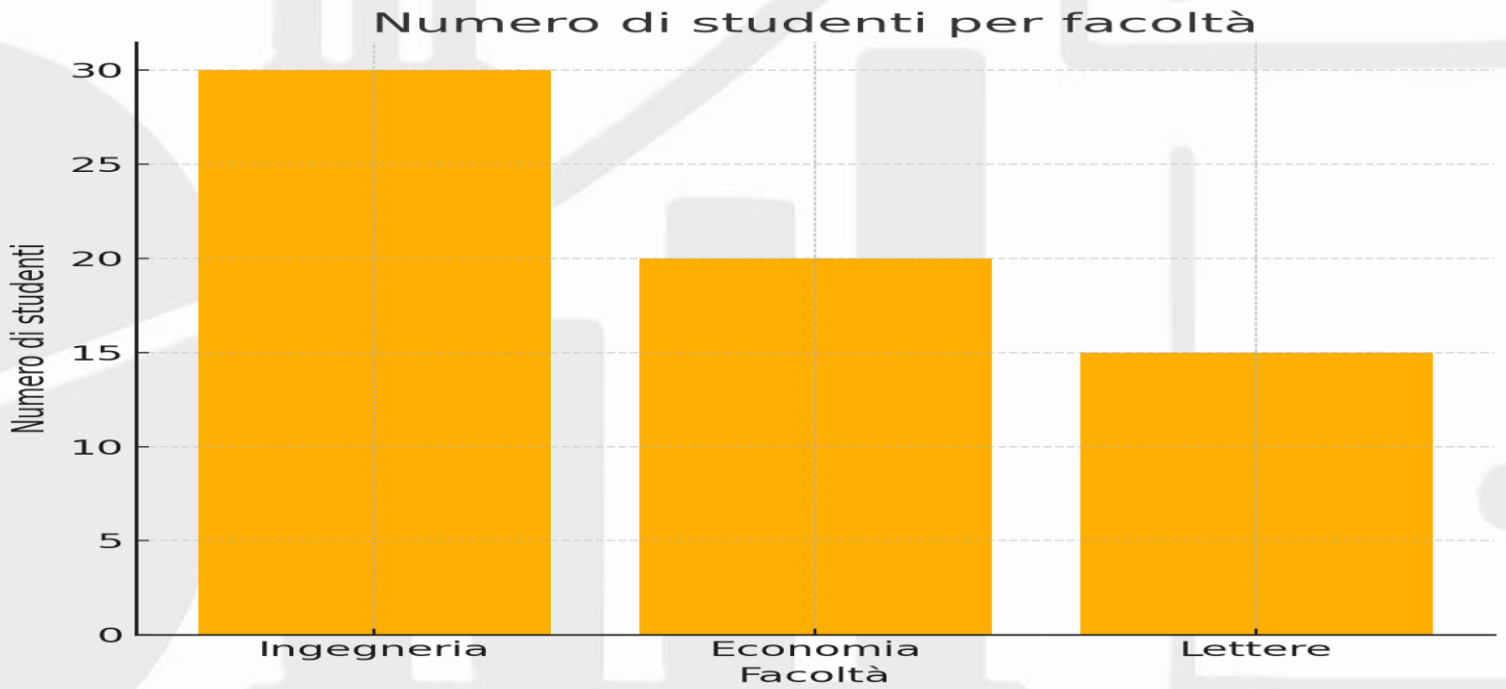
GRAFICO A BARRE

- **Tipi di dati**: viene usato per rappresentare dati qualitativi o dati discreti (cioè dati che si contano, non si misurano).
- **Spazi tra le barre**: le barre sono separate da spazi, proprio per indicare che le categorie sono distinte e non hanno continuità tra loro.
- **Etichette sull'asse x**: sull'asse orizzontale troviamo categorie (es. colori, professioni, regioni, ecc.).
- **Altezza delle barre**: rappresenta la frequenza assoluta o relativa di ogni categoria.

CORSO DI PREPARAZIONE PER RESPONSABILE DEL SISTEMA DI GESTIONE DELLA SICUREZZA

GRAFICO A BARRE

Esempio: Un grafico a barre per il numero di studenti per facoltà: Ingegneria (30), Economia (20), Lettere (15), ecc.

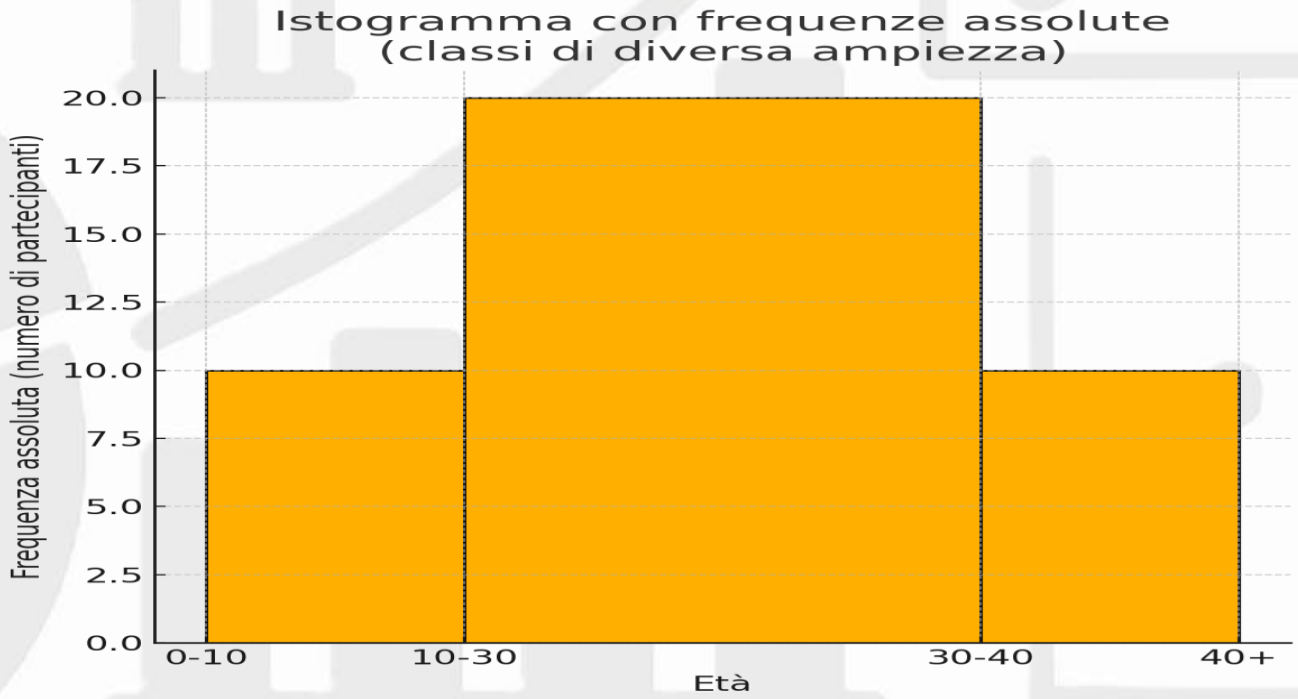


ISTOGRAMMA

- **Tipi di dati**: è utilizzato per rappresentare dati quantitativi continui suddivisi in classi di valori (es. età, peso, altezza, ecc.).
- **Spazi tra le barre**: le barre sono attaccate, perché rappresentano intervalli continui e non categorie separate.
- **Etichette sull'asse x**: sull'asse orizzontale troviamo categorie (es. età, altezza, reddito, ecc.).
- **Altezza delle barre**: rappresenta la frequenza assoluta o relativa di ogni categoria.

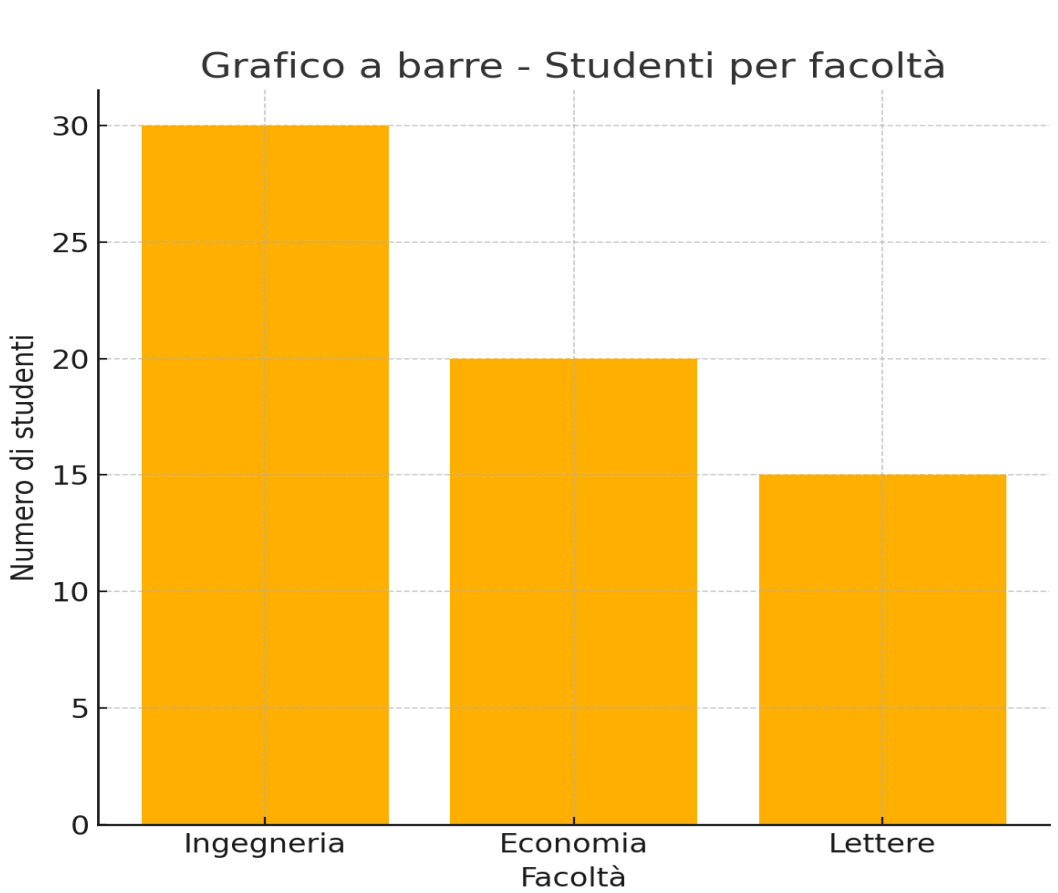
ISTOGRAMMA

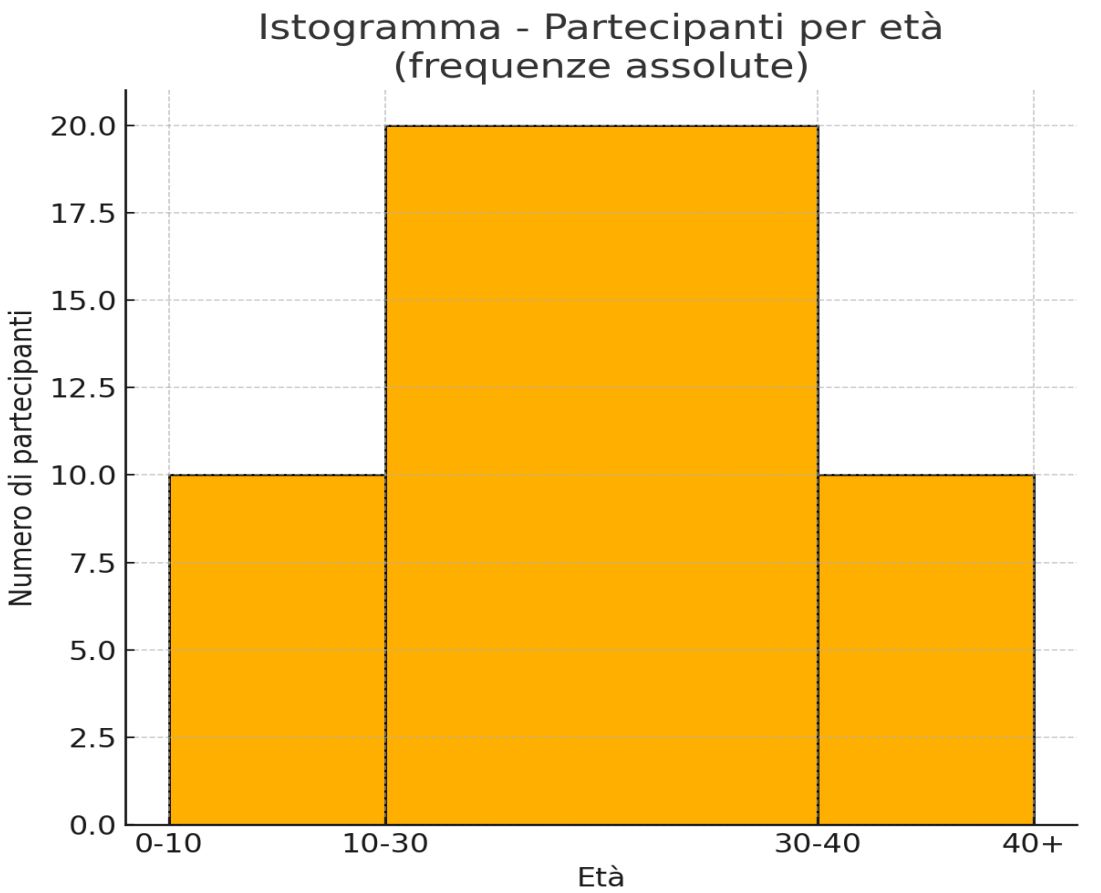
Esempio: Un istogramma per età dei partecipanti a un corso:
Classi:0–10 anni (10 persone), 10–30 anni (20), 30–40 anni (10).



CORSO DI PREPARAZIONE PER RESPONSABILE DEL SISTEMA DI GESTIONE DELLA SICUREZZA

GRAFICO A BARRE VS ISTOGRAMMA





LA MODA

LA MODA rappresenta il valore che compare più frequentemente in un insieme di dati.

Questo indicatore statistico è particolarmente utile con dati qualitativi e quantitativi discreti. Per esempio, se tra i lavoratori di un'azienda l'età più frequente è 41 anni, questa è la moda dell'età.

Identificare la moda aiuta a individuare rapidamente quale sia la caratteristica predominante nella popolazione analizzata.

LA MODA

La determinazione della moda può essere utile anche per adottare decisioni pratiche, come scegliere il prodotto più venduto o individuare il problema più ricorrente.
Sebbene sia un valore di semplice determinazione, la moda, che non deve essere confuso con altri indicatori sintetici (tipo la media aritmetica o la mediana), offre informazioni immediate e pratiche per una prima valutazione dei dati e per impostare analisi più approfondite.

LA MEDIA ARITMETICA

La **MEDIA ARITMETICA rappresenta il valore medio di un insieme di dati**, calcolata sommando tutti i valori e dividendo per il loro numero complessivo.

Ad esempio, se vogliamo conoscere l'età media dei dipendenti di un'azienda, sommiamo tutte le età e dividiamo per il numero di dipendenti. Se le età fossero 30, 35 e 40 anni, la media sarebbe (30 + 35 + 40) / 3 = 35 anni.

LA MEDIA ARITMETICA

La media aritmetica è molto utilizzata per descrivere una caratteristica centrale di una popolazione.

È importante notare, però, che può essere influenzata dalla presenza, nella popolazione o nel campione rappresentativo considerato, di valori estremi – che possono essere in numero esiguo ma molto differenti dal resto dei dati – e pertanto va interpretata con attenzione facendo uso di altri **indicatori di dispersione** che aiutano la corretta interpretazione della distribuzione statistica

LA MEDIANA

La MEDIANA è una misura di tendenza centrale che rappresenta il valore centrale di una distribuzione di dati quando questi sono ordinati in modo crescente o decrescente.

A differenza della media aritmetica, che può essere influenzata da valori estremi (outlier), la mediana è più robusta e fornisce una rappresentazione più fedele della "posizione centrale" di un insieme di dati asimmetrici o distribuiti in modo irregolare.

LA MEDIANA

Per calcolare la mediana è necessario ordinare preventivamente i dati.

- Se il numero di osservazioni è dispari, la mediana corrisponde al valore che si trova esattamente nel mezzo.
- Se invece il numero di osservazioni è pari, la mediana è la media aritmetica dei due valori centrali.

Ad esempio, nella serie ordinata 3, 5, 7, 9, 11 la mediana è 7; nella serie 2, 4, 6, 8, la mediana è (4+6)/2 = 5.

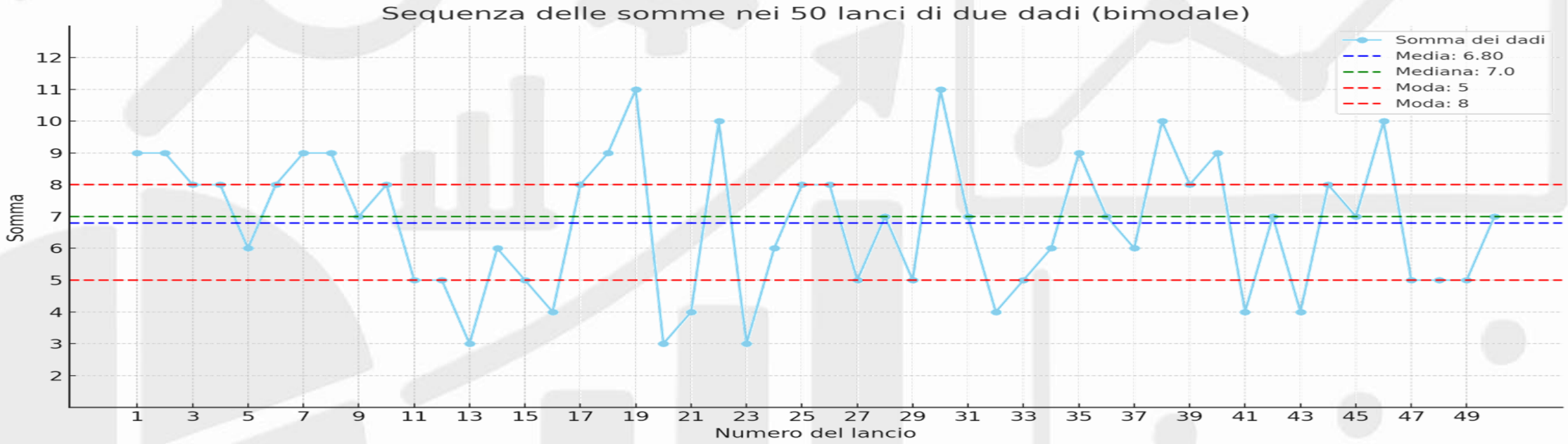
LA MEDIANA

La mediana è particolarmente utile in presenza di dati con valori anomali o fortemente asimmetrici, come nel caso dei redditi o dei prezzi immobiliari, perché fornisce un'indicazione del valore "tipico" senza essere distorta.

In statistica descrittiva viene utilizzata insieme ad altre misure (come media e moda) per avere una visione più completa della distribuzione dei dati.

CORSO DI PREPARAZIONE PER RESPONSABILE DEL SISTEMA DI GESTIONE DELLA SICUREZZA

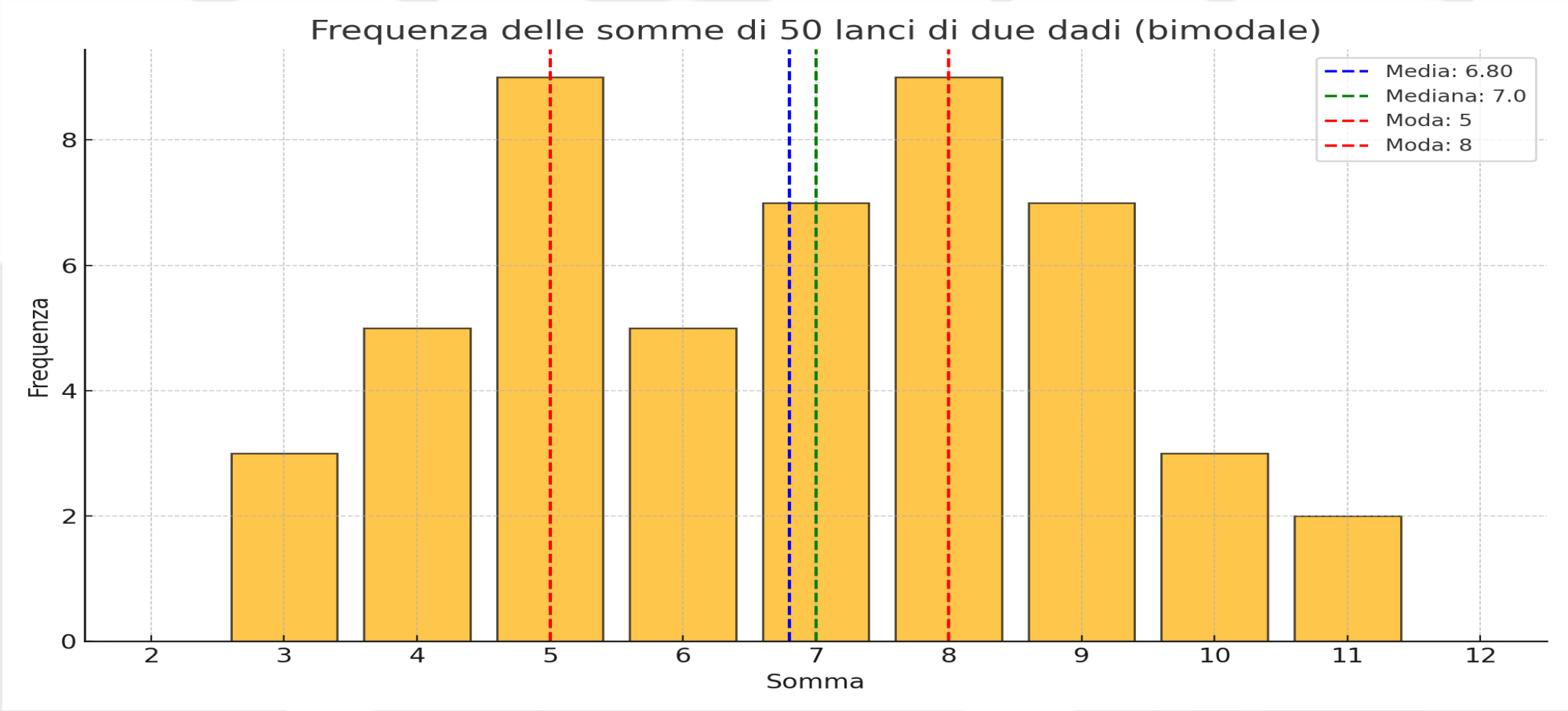
LA MEDIANA



	3	4	5	6	7	8	9	10	11
Frequenza Assoluta	3.0	5.0	9.0	5.0	7.0	9.0	7.0	3.0	2.0
Frequenza Relativa	0.06	0.1	0.18	0.1	0.14	0.18	0.14	0.06	0.04

CORSO DI PREPARAZIONE PER RESPONSABILE DEL SISTEMA DI GESTIONE DELLA SICUREZZA

LA MEDIANA



LA MEDIA PESATA (O PONDERALE)

La MEDIA PONDERALE tiene conto dell'importanza relativa di ciascun valore attraverso l'attribuzione di pesi diversi alle osservazioni.

Questa misura statistica è utile quando non tutte le osservazioni hanno lo stesso valore. Ad esempio, se si vogliono calcolare i voti medi di uno studente che ha ottenuto 18 con peso 3 e 24 con peso 6, la media pesata sarà calcolata moltiplicando ciascun voto per il rispettivo peso, sommando i risultati ottenuti, e infine dividendo per la somma dei pesi.

LA MEDIA PESATA (O PONDERALE)

La formula per calcolare la media ponderale è la seguente:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i \cdot w_i}{\sum_{i=1}^{n} w_i}$$

dove:

- x_i = valori osservati
- w_i = pesi associati ai valori

LA MEDIA PESATA (O PONDERALE)

Facciamo un esempio semplice per fissare il concetto:

Uno studente universitario ha tre voti:

- 28 in un esame da 6 CFU
- 30 in un esame da 12 CFU
- 25 in un esame da 3 CFU

La media (ponderata) dei suoi voti sarà:

$$\bar{x} = \frac{28 \cdot 6 + 30 \cdot 12 + 25 \cdot 3}{6 + 12 + 3} = \frac{168 + 360 + 75}{21} = \frac{603}{21} \approx 28.7$$

LA VARIANZA

La **VARIANZA è una misura statistica che rappresenta la dispersione dei dati intorno alla media aritmetica**.
In altre parole, essa quantifica quanto i dati raccolti siano distanti, in media, dal valore medio.

Una varianza alta indica una grande dispersione, cioè che i dati sono molto variabili, mentre una varianza bassa indica che i dati sono più concentrati attorno alla media.

LA VARIANZA

Per calcolare la varianza, s^2, si sommano i quadrati delle differenze tra ciascun valore e la media, e poi si divide per il numero delle osservazioni, meno uno, come nella formula seguente:

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

dove:

- $\bar{x}$ è la media aritmetica del campione
- x_i sono i singoli valori
- n è il numero totale di osservazioni

LA VARIANZA - Esempio

Supponiamo di avere questi 5 valori: $x = [4, 7, 6, 2, 5]$

1. Calcola la media:

$$\bar{x} = \frac{4+7+6+2+5}{5} = \frac{24}{5} = 4.8$$

2. Calcola gli scarti dalla media (cioè $x_i - \bar{x}$):

- $4 - 4.8 = -0.8$
- $7 - 4.8 = 2.2$
- $6 - 4.8 = 1.2$
- $2 - 4.8 = -2.8$
- $5 - 4.8 = 0.2$

3. Eleva al quadrato ogni scarto:

- $(-0.8)^2 = 0.64$
- $2.2^2 = 4.84$
- $1.2^2 = 1.44$
- $(-2.8)^2 = 7.84$
- $0.2^2 = 0.04$

4. Somma i quadrati degli scarti:

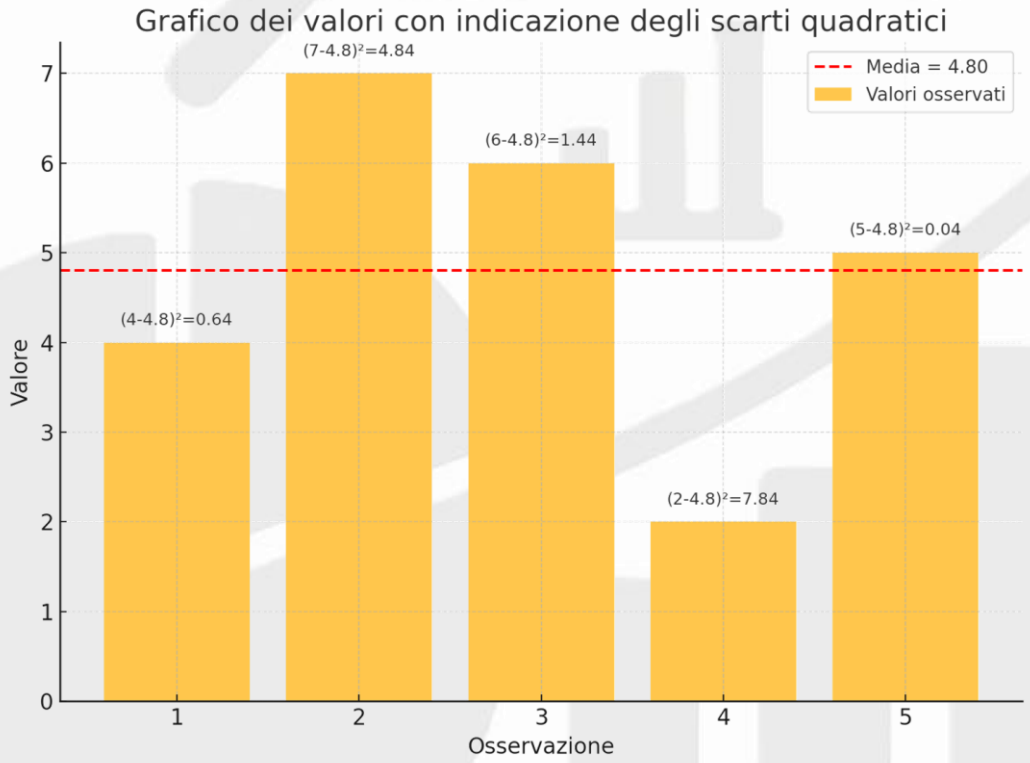
$$0.64 + 4.84 + 1.44 + 7.84 + 0.04 = 14.8$$

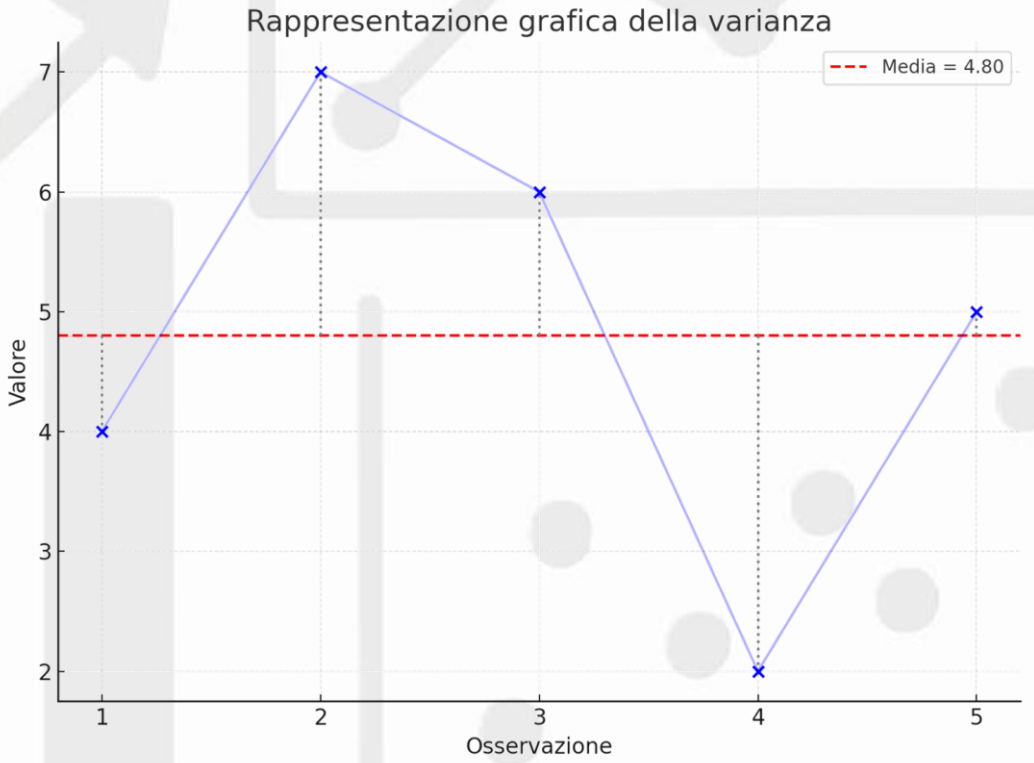
5. Dividi per $n - 1 = 4$:

$$s^2 = \frac{14.8}{4} = 3.7$$

CORSO DI PREPARAZIONE PER RESPONSABILE DEL SISTEMA DI GESTIONE DELLA SICUREZZA

LA VARIANZA – Grafici esplicativi





CORSO DI PREPARAZIONE PER RESPONSABILE DEL SISTEMA DI GESTIONE DELLA SICUREZZA

15 Ottobre 2025 – BARI
PRINCIPI DI STATISTICA



LA DEVIAZIONE STANDARD

La **DEVIAZIONE STANDARD** è la radice quadrata della varianza e rappresenta una misura molto utilizzata per descrivere la dispersione dei dati rispetto alla media, nello stesso ordine di grandezza delle osservazioni.

La deviazione standard permette di comprendere immediatamente la variabilità reale dei dati.

Riprendendo i dati precedenti con varianza pari a 3,7 la deviazione standard ne sarà la radice quadrata, cioè circa 1,92

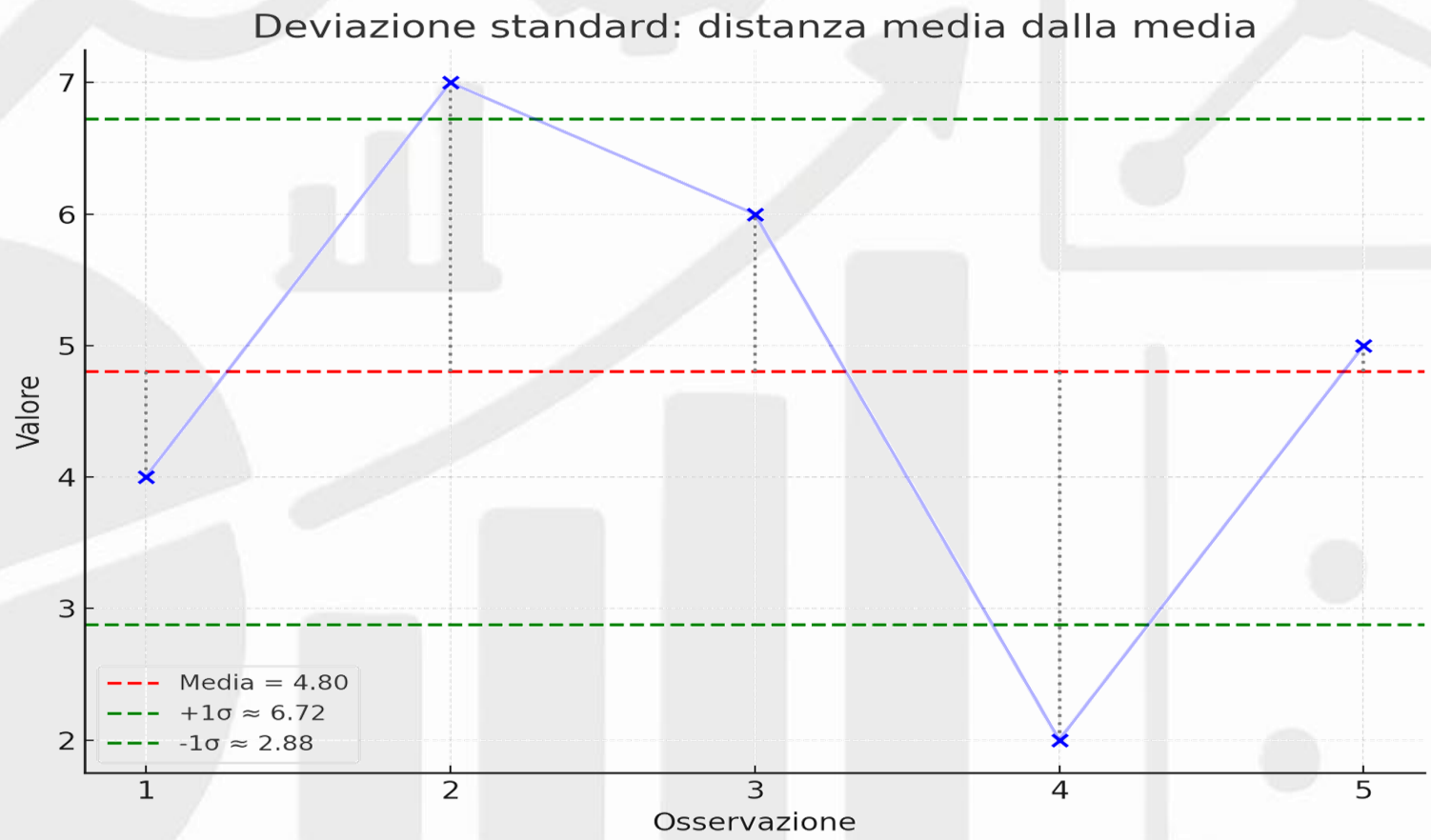
LA DEVIAZIONE STANDARD

La deviazione standard ci dice che, in media, i dati si discostano di circa 1,92 unità dalla media (che è 4,8).

A differenza della varianza, la deviazione standard è direttamente confrontabile con i valori originali, perché non è elevata al quadrato.

Misura	Valore
Media $\bar{x}$	4,8
Varianza s^2	3,7
Dev. standard s	$\approx$ 1,92

LA DEVIAZIONE STANDARD – Grafico esplicativo



IL COEFFICIENTE DI VARIAZIONE

Il coefficiente di variazione è un indicatore statistico fondamentale che misura la dispersione relativa dei dati rispetto alla loro media.

A differenza della deviazione standard, che esprime la variabilità in valori assoluti, il coefficiente di variazione fornisce una misura standardizzata, espressa in percentuale, rendendo più semplice il confronto tra insiemi di dati anche molto diversi tra loro.

IL COEFFICIENTE DI VARIAZIONE

Il **COEFFICIENTE DI VARIAZIONE** è uno strumento particolarmente utile in ambito statistico quando si desidera confrontare la variabilità di più distribuzioni che hanno medie di entità molto diverse, oppure quando i dati sono espressi in unità di misura differenti (ad esempio euro, chilogrammi, metri o anni).

Viene spesso utilizzato in settori come l'economia, la finanza, le scienze sociali e la biologia, dove è fondamentale valutare la coerenza interna di un insieme di dati rispetto alla sua media.

IL COEFFICIENTE DI VARIAZIONE

Il coefficiente di variazione si calcola mediante una formula semplice ed intuitiva:

$$CV = \left(\frac{\text{Deviazione Standard}}{\text{Media}} \right) \times 100$$

In questo modo, il valore ottenuto rappresenta la variabilità relativa espressa come percentuale della media.

Un valore più alto indica una maggiore dispersione dei dati, mentre un valore più basso suggerisce che i dati sono più concentrati intorno alla media.

IL COEFFICIENTE DI VARIAZIONE

Ad esempio, supponiamo di avere una popolazione in cui la media delle età sia pari a 40 anni, e la deviazione standard sia di 8 anni. Applicando la formula, otteniamo:

$$CV = \left(\frac{8}{40}\right) \times 100 = 20\%$$

Questo risultato ci dice che la dispersione delle età è pari al 20% della media, fornendo un'indicazione chiara e interpretabile del grado di variabilità.

IL COEFFICIENTE DI VARIAZIONE

Il vantaggio principale del coefficiente di variazione è che consente di confrontare efficacemente la variabilità tra gruppi diversi anche quando le scale di misura non sono comparabili. Ad esempio, si può confrontare la variabilità dei redditi tra due paesi con valute diverse, o la variabilità dei pesi di due gruppi di individui con medie molto distanti. In sintesi, **il coefficiente di variazione rappresenta una misura sintetica e comparabile della dispersione**, particolarmente preziosa in analisi statistiche descrittive e comparative.

I QUARTILI E I BOX PLOT

I quartili sono valori statistici che dividono un insieme di dati ordinati in quattro parti uguali, ciascuna contenente il 25% delle osservazioni.
Il primo quartile (Q1) separa il 25% inferiore dei dati, il secondo quartile (Q2) è la mediana e divide a metà i dati, mentre il terzo quartile (Q3) separa il 25% superiore.
Il Box-Plot è una rappresentazione grafica dei quartili, che evidenzia chiaramente la distribuzione dei dati, la presenza di eventuali valori estremi e la simmetria o asimmetria dei dati raccolti.

I QUARTILI E I BOX PLOT

I QUARTILI sono valori statistici fondamentali utilizzati per descrivere la distribuzione di un insieme di dati.

Essi **suddividono i dati ordinati in quattro parti uguali, ciascuna delle quali contiene esattamente il 25% delle osservazioni**.

Sono molto utili per capire la dispersione, la simmetria e l'eventuale presenza di valori anomali in un dataset.

I QUARTILI E I BOX PLOT

Il primo quartile (Q1) rappresenta il valore sotto il quale si trova il 25% dei dati. È detto anche "quartile inferiore".

Il secondo quartile (Q2) coincide con la mediana, ovvero il valore che divide il dataset esattamente a metà: il 50% dei dati si trova al di sotto, e il 50% al di sopra.

Il terzo quartile (Q3) rappresenta il valore sotto il quale si trova il 75% dei dati; in altre parole, solo il 25% dei dati è superiore a Q3.

La **distanza tra Q1 e Q3** è detta **intervallo interquartile (IQR)** e misura la dispersione centrale dei dati: $\mathrm{IQR} = Q3 - Q1$

I BOX PLOT

Per rappresentare graficamente i quartili e fornire un colpo d'occhio immediato sulla distribuzione, si utilizza il Box-Plot (o diagramma a scatola).

Questo tipo di grafico visualizza:

- Il minimo e il massimo (escludendo i valori anomali).
- I quartili Q1, Q2 (mediana) e Q3;
- Eventuali «outlier» (valori anomali), rappresentati come punti isolati.

I BOX PLOT

Nel Box-Plot osserviamo queste caratteristiche:

- Il "box" centrale va da Q1 a Q3;
- Una linea all'interno del box rappresenta la mediana (Q2);
- Le "whiskers" (i baffi) si estendono fino ai valori minimo e massimo entro un intervallo definito (spesso 1.5 volte l'IQR);
- I valori anomali oltre questi limiti vengono rappresentati come puntini o cerchietti.

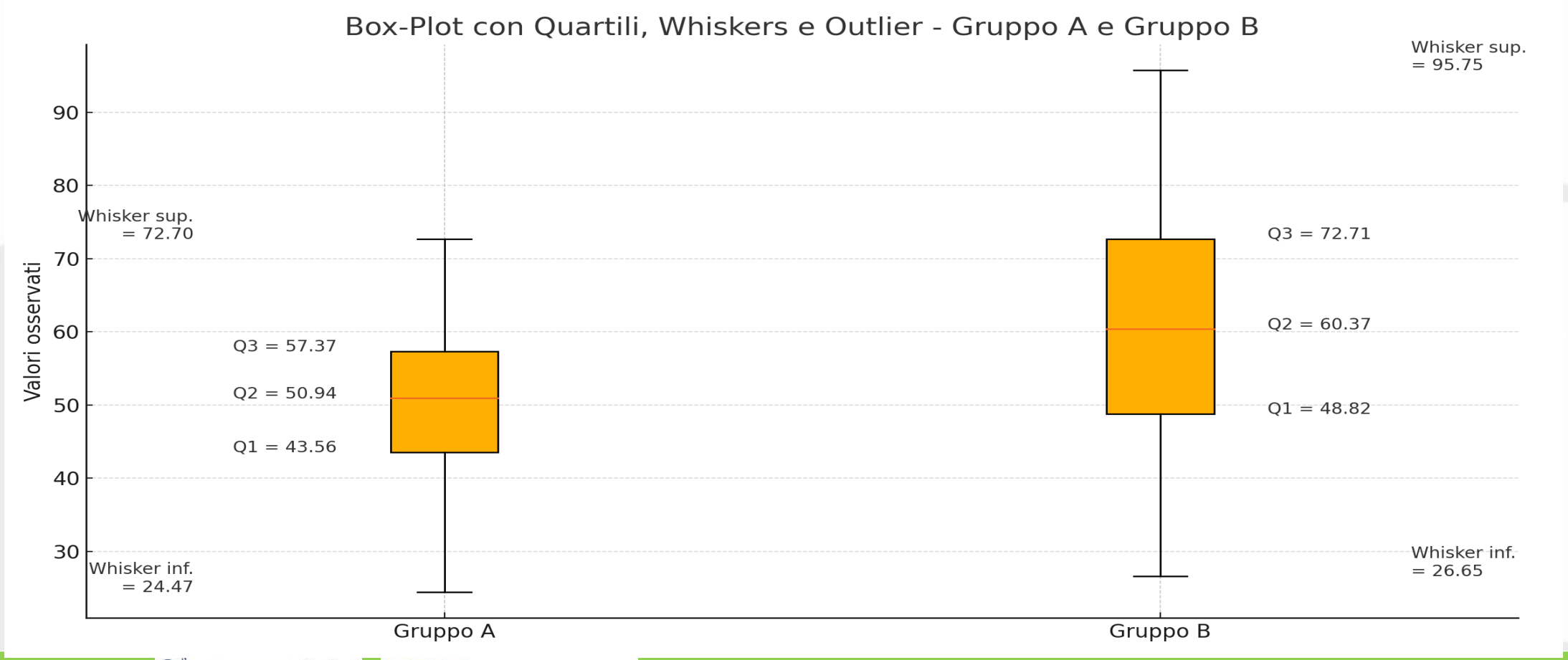
I BOX PLOT

Il Box-Plot è uno strumento efficace e intuitivo per:

- Rilevare rapidamente asimmetrie nella distribuzione;
- Confrontare la variabilità tra diversi gruppi di dati;
- Individuare valori estremi che potrebbero influenzare la media o altre statistiche descrittive.

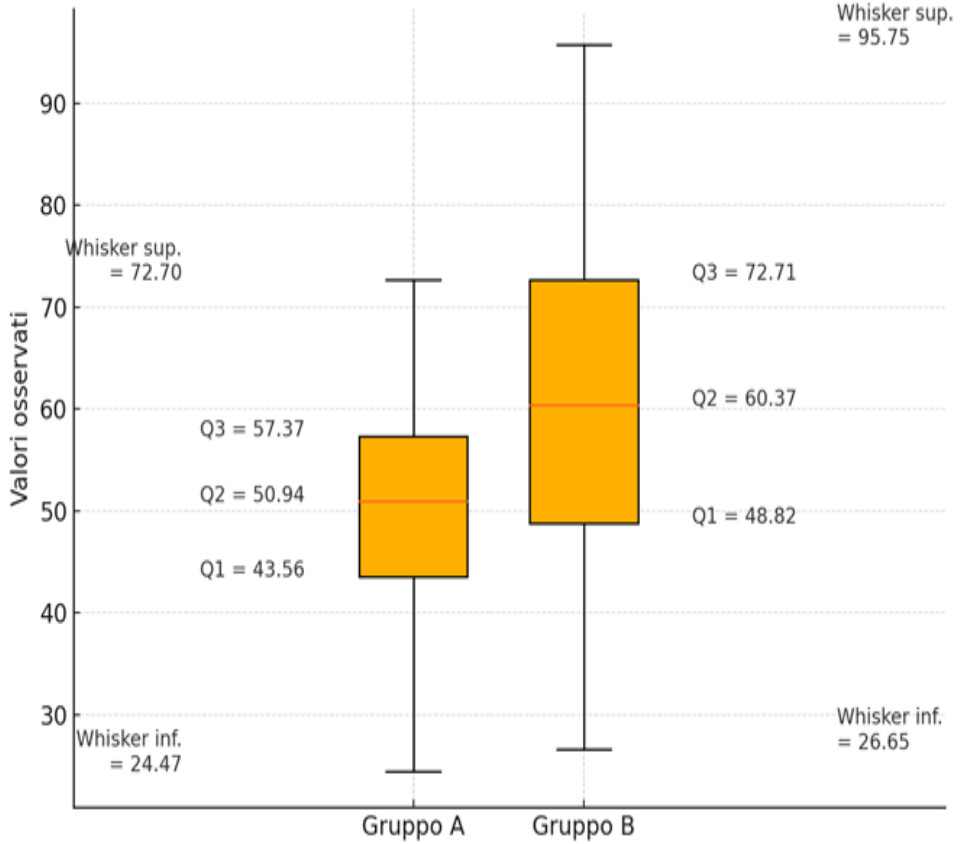
CORSO DI PREPARAZIONE PER RESPONSABILE DEL SISTEMA DI GESTIONE DELLA SICUREZZA

I BOX PLOT



Posizione centrale (mediana - Q2)

- Il Gruppo A ha una mediana (Q2) di circa 50,94.
- Il Gruppo B ha una mediana più alta, circa 60,37.
- Questo indica che i valori del Gruppo B, in media, sono più alti rispetto a quelli del Gruppo A.



CORSO DI PREPARAZIONE PER RESPONSABILE DEL SISTEMA DI GESTIONE DELLA SICUREZZA

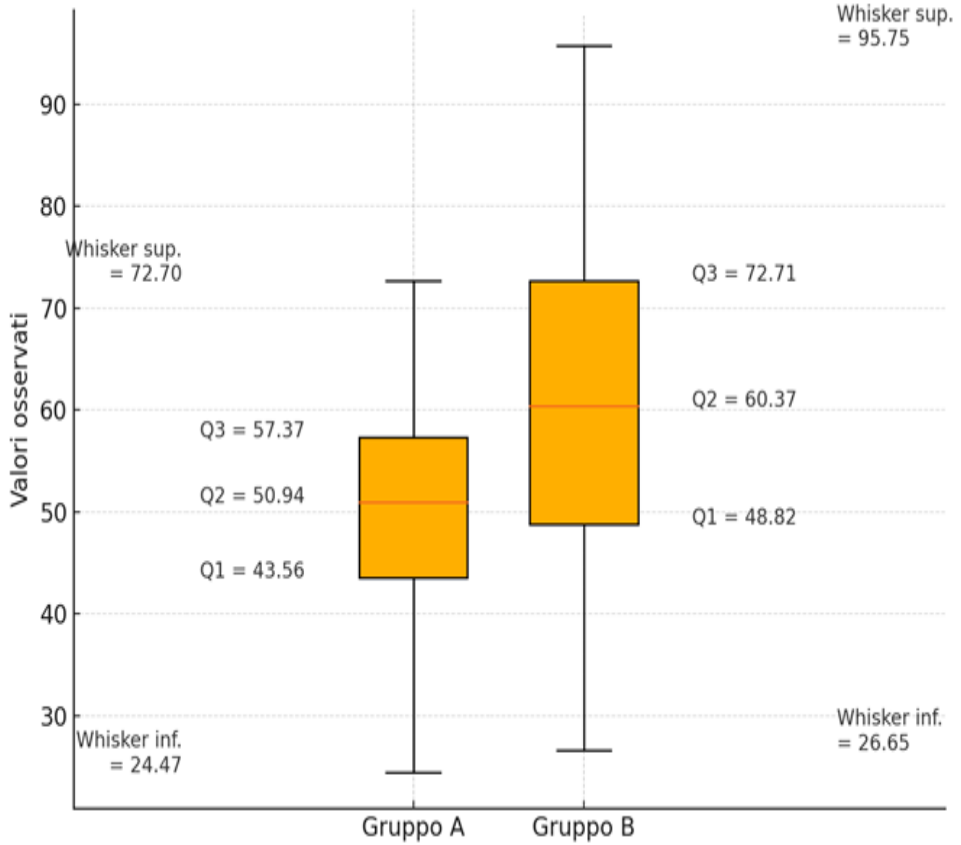
Intervallo interquartile - IQR

Gruppo A:

Q1 ≈ 43,56
Q3 ≈ 57,37
IQR ≈ 13,81

Gruppo B:

Q1 ≈ 48,82
Q3 ≈ 72,71
IQR ≈ 23,89



Il Gruppo B ha una variabilità più ampia nella parte centrale dei dati: l'intervallo interquartile è quasi il doppio di quello del Gruppo A. Ciò suggerisce una maggiore dispersione tra i dati centrali nel Gruppo B.

CORSO DI PREPARAZIONE PER RESPONSABILE DEL SISTEMA DI GESTIONE DELLA SICUREZZA

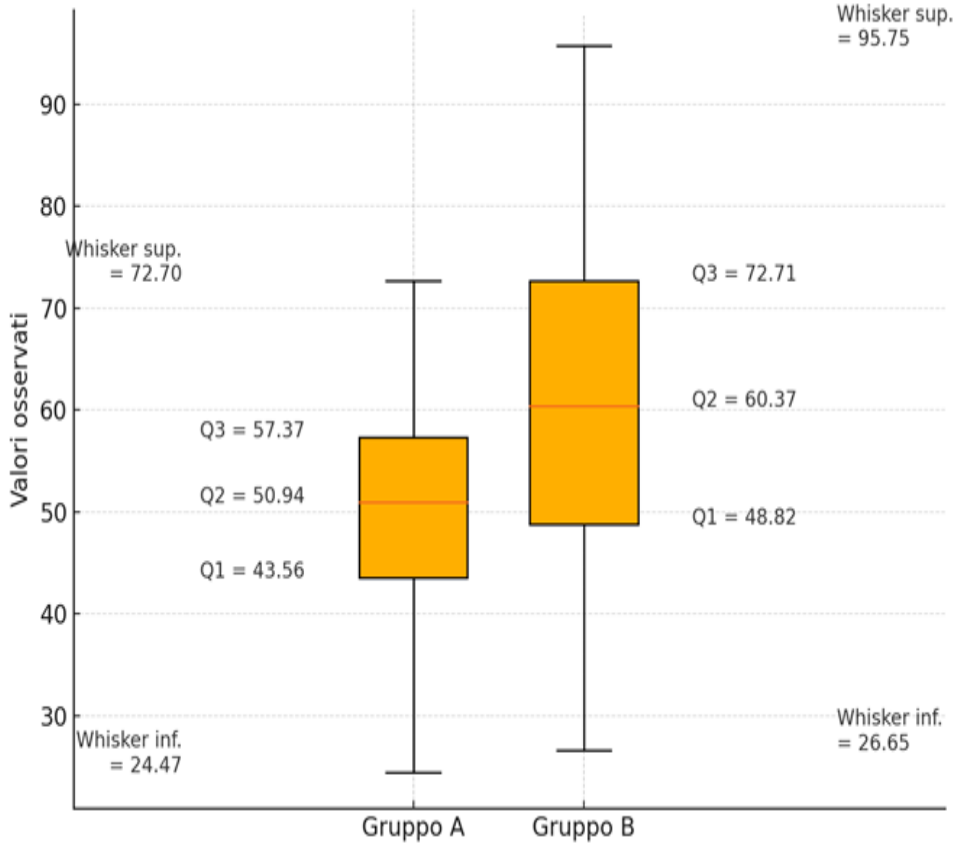
WHISKER (max e min «normali»)

Gruppo A:

W. inf. ≈ 24,47
W. sup. ≈ 72,70

Gruppo B:

W inf. ≈ 26.65
W sup. ≈ 95,75

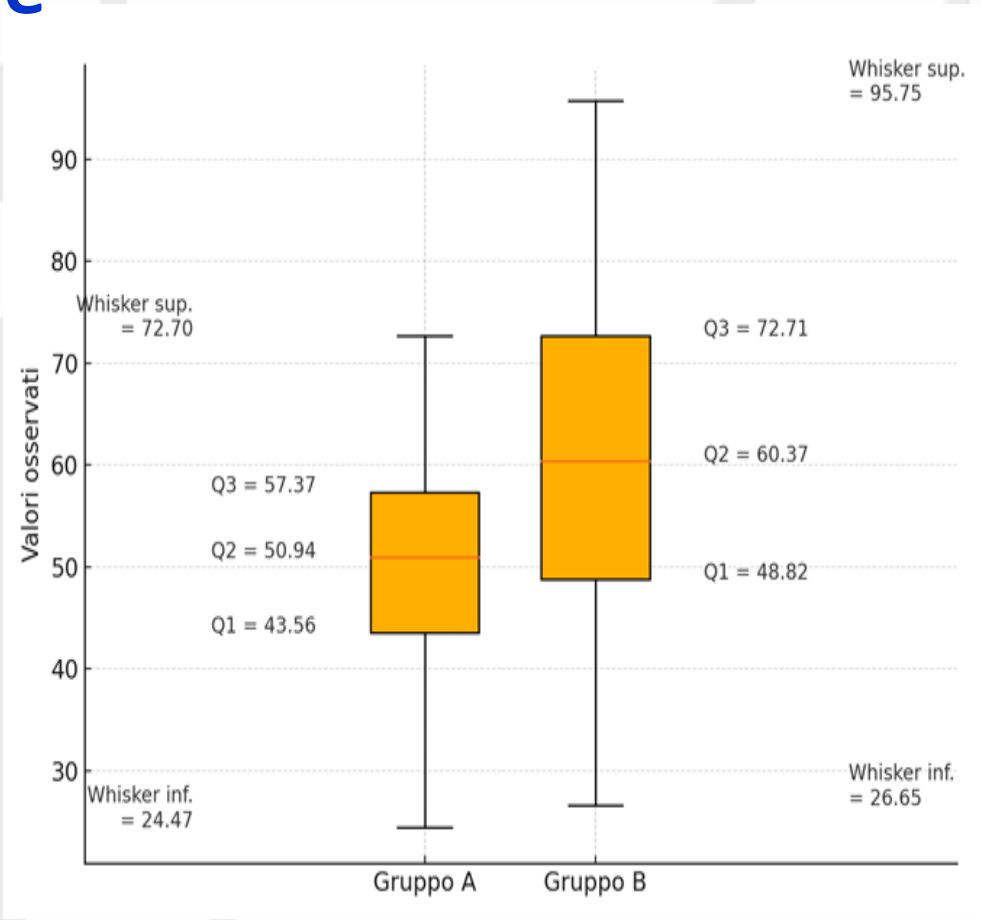


I whisker del Gruppo B coprono un intervallo molto più ampio, suggerendo che i dati si estendono su un range più vasto, anche escludendo gli outlier.

Simmetria e forma della distribuzione

Il Gruppo A mostra una distribuzione relativamente simmetrica, con la mediana abbastanza centrata nel box.

Il Gruppo B presenta una leggera asimmetria verso destra (coda lunga), dovuta alla presenza di valori molto alti e outlier superiori.

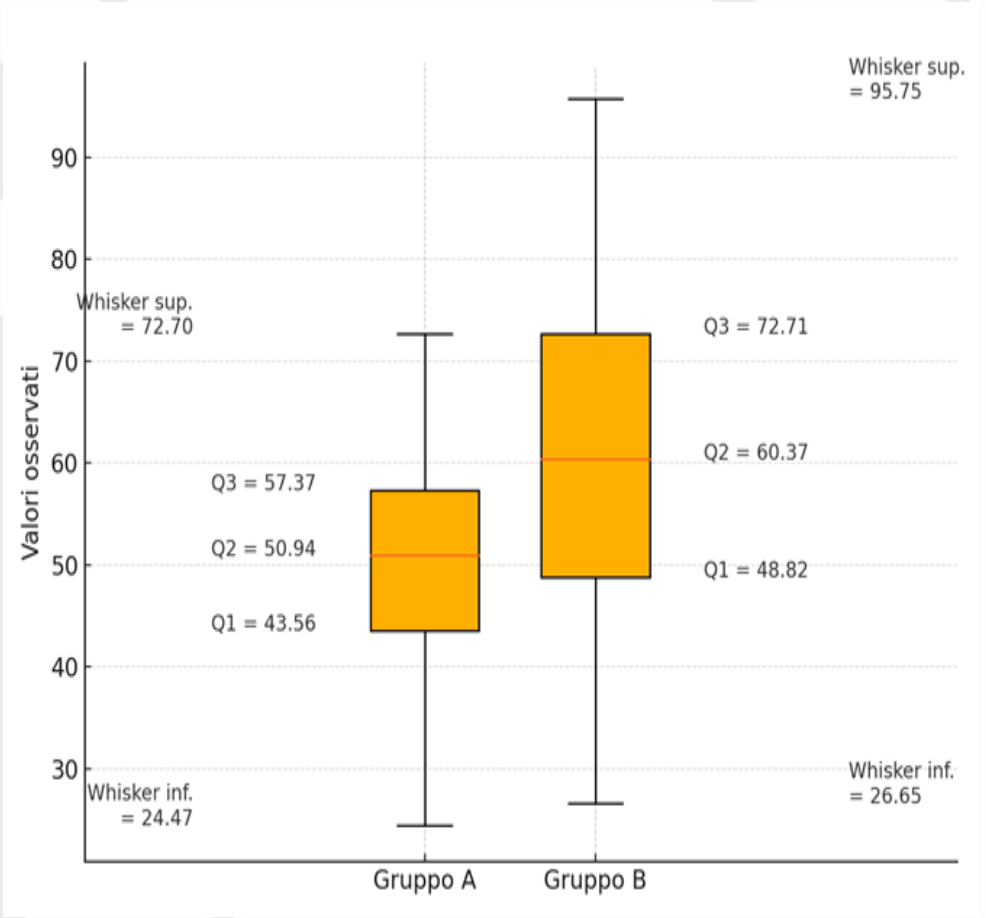


BOX PLOT: CONCLUSIONI

Il Gruppo A è più compatto e uniforme, con una minore variabilità e meno valori anomali.
Il Gruppo B ha valori più alti in media, ma anche molta più dispersione.

Se si cerca stabilità e bassa variabilità, il Gruppo A è più omogeneo.

Se si accettano valori estremi e si è interessati alla gamma più ampia, il Gruppo B offre più varietà ma anche maggiore incertezza.



CORSO DI PREPARAZIONE PER RESPONSABILE DEL SISTEMA DI GESTIONE DELLA SICUREZZA

15 Ottobre 2025 – BARI
PRINCIPI DI STATISTICA



DISTRIBUZIONE DI PROBABILITÀ

Una distribuzione di probabilità è una funzione matematica che descrive la probabilità associata a ciascun valore possibile assunto da una variabile casuale.

Le distribuzioni di probabilità possono essere discrete, quando la variabile casuale assume un insieme numerabile di valori, o continue, quando assume valori in un intervallo continuo.

DISTRIBUZIONE DI PROBABILITÀ

Un esempio comune di distribuzione discreta è la distribuzione binomiale, che rappresenta la probabilità di ottenere un certo numero di successi in una serie di prove indipendenti, ciascuna con la stessa probabilità di successo.

Comprendere le distribuzioni di probabilità è fondamentale per effettuare previsioni e prendere decisioni basate su dati statistici.

DISTRIBUZIONE DI PROBABILITÀ

A seconda del tipo di variabile casuale, la distribuzione può essere: **DISCRETA**, quando la variabile può assumere un numero finito o numerabile di valori distinti (es. numero di figli, lanci di dado) oppure **CONTINUA**, quando la variabile può assumere infiniti valori all'interno di un intervallo (es. altezza, tempo, peso).
In pratica, una distribuzione di probabilità assegna una probabilità a ciascun possibile risultato.
La somma di tutte le probabilità deve essere uguale al 100%.

DISTRIBUZIONE DI PROBABILITÀ

La distribuzione di probabilità è importante perché:

- Permette di modellare fenomeni reali casuali (es. durata di vita, tempo di attesa, vincita al gioco).
- È la base per calcolare valori attesi, varianza, probabilità di eventi, intervalli di confidenza, e per effettuare test statistici.
- È fondamentale per costruire simulazioni, modelli predittivi e machine learning.

DISTRIBUZIONE DI PROBABILITÀ DISCRETA

Immaginiamo di lanciare un dado equo a 6 facce.

La variabile casuale X rappresenta il risultato del lancio.

I valori possibili sono: 1,2,3,4,5,6

La probabilità di ciascun valore è: $P(X = k) = \frac{1}{6}, \quad \text{per } k = 1, 2, 3, 4, 5, 6$

Valore X	1	2	3	4	5	6
Probabilità $P(X)$	1/6	1/6	1/6	1/6	1/6	1/6

Questa è una distribuzione di probabilità uniforme discreta, perché ogni valore ha la stessa probabilità.

DISTRIBUZIONE DI PROBABILITÀ CONTINUA

Consideriamo la variabile casuale Y che rappresenta l'altezza (in cm) di adulti in una popolazione.

Questa variabile è continua, perché può assumere infiniti valori, ad esempio 168,5 cm, 168,53 cm, ecc.

Una possibile distribuzione che modella questa variabile è la distribuzione normale (gaussiana), ad esempio:

- Media $\mu = 170$ cm
- Deviazione standard $\sigma = 10$ cm

DISTRIBUZIONE DI PROBABILITÀ CONTINUA

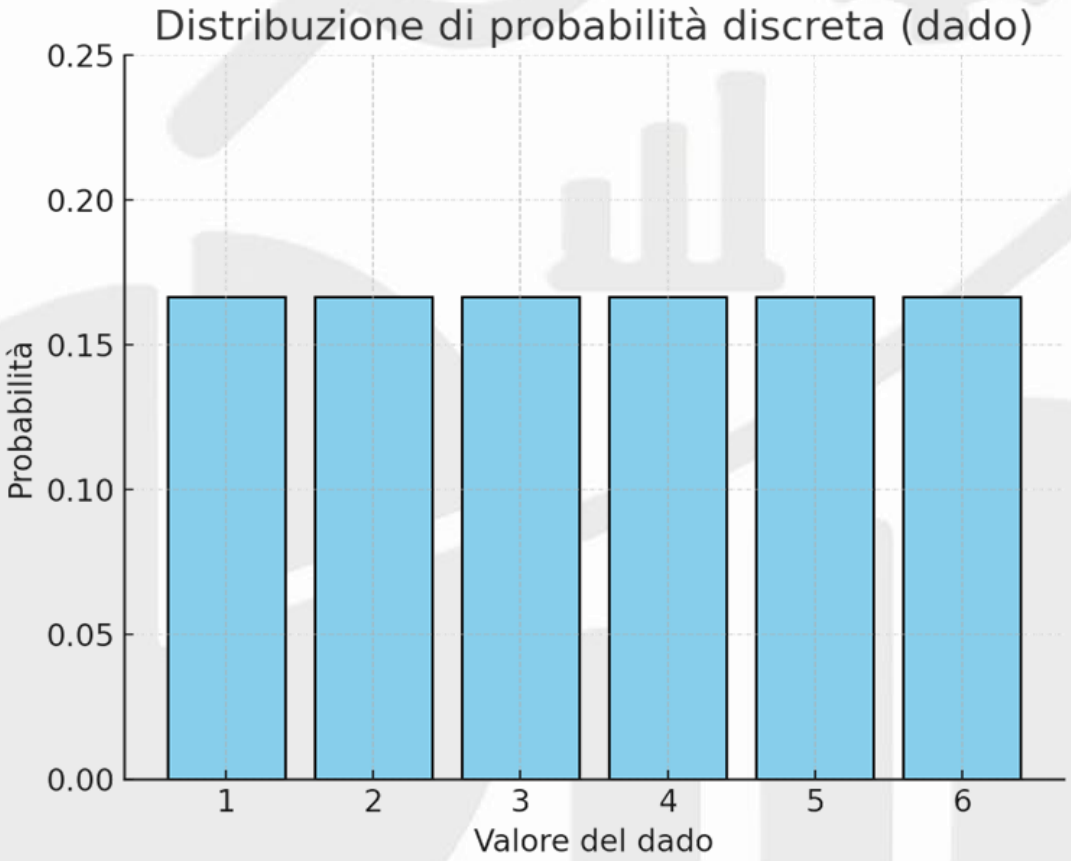
Questa distribuzione ha la classica forma a campana, ed è simmetrica attorno alla media.

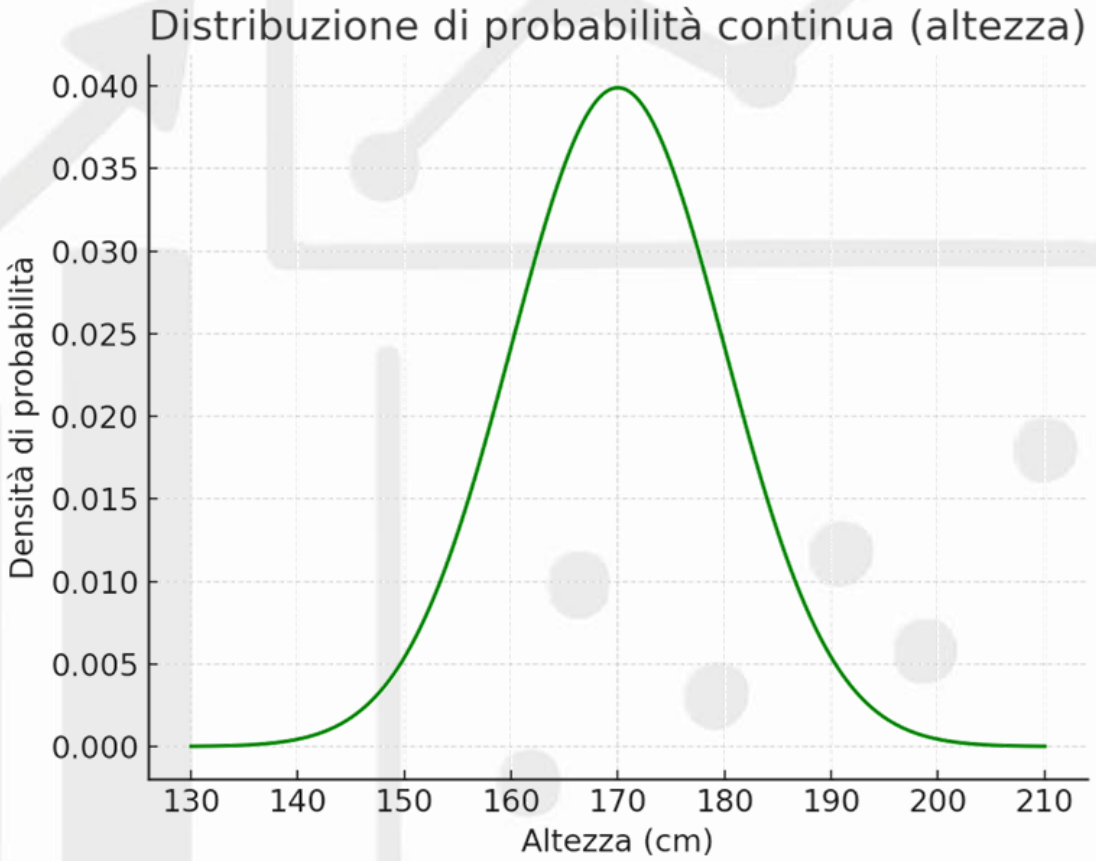
La probabilità che una persona scelta a caso abbia un'altezza tra 160 e 180 cm si calcola integrando la curva della distribuzione normale in quell'intervallo.

La probabilità per un valore esatto (es. P(Y=170)) è zero, perché in una variabile continua le probabilità si calcolano su intervalli e non su valori specifici.

CORSO DI PREPARAZIONE PER RESPONSABILE DEL SISTEMA DI GESTIONE DELLA SICUREZZA

DISTRIBUZIONE DI PROBABILITÀ CONTINUA





DISTRIBUZIONE BINOMIALE

La **DISTRIBUZIONE BINOMIALE è utilizzata quando si effettuano prove ripetute con due possibili esiti: successo o fallimento.** Questa distribuzione descrive la probabilità di ottenere un determinato numero di successi in un numero prefissato di prove indipendenti, tutte con la stessa probabilità di successo.

DISTRIBUZIONE BINOMIALE

Se X è una **variabile casuale binomiale** che conta il numero di successi in n prove indipendenti, ciascuna con probabilità di successo p, allora la **funzione di probabilità** (cioè la probabilità di ottenere esattamente k successi) si calcola come segue:

$$P(X = k) = \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k}$$

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

- n = numero totale di prove (esperimenti indipendenti)
- k = numero di successi desiderati
- p = probabilità di successo in una singola prova
- $1 - p$ = probabilità di insuccesso

DISTRIBUZIONE BINOMIALE – ESEMPIO TESTA O CROCE –

Se Lancio 10 volte una moneta, e voglio sapere la probabilità di ottenere esattamente 6 teste (successi), abbiamo:

- $n = 10$
- $k = 6$
- $p = 0.5$

Applicando la formula:

$$P(X = 6) = \binom{10}{6} \cdot 0.5^6 \cdot 0.5^4 = \binom{10}{6} \cdot 0.5^{10}$$

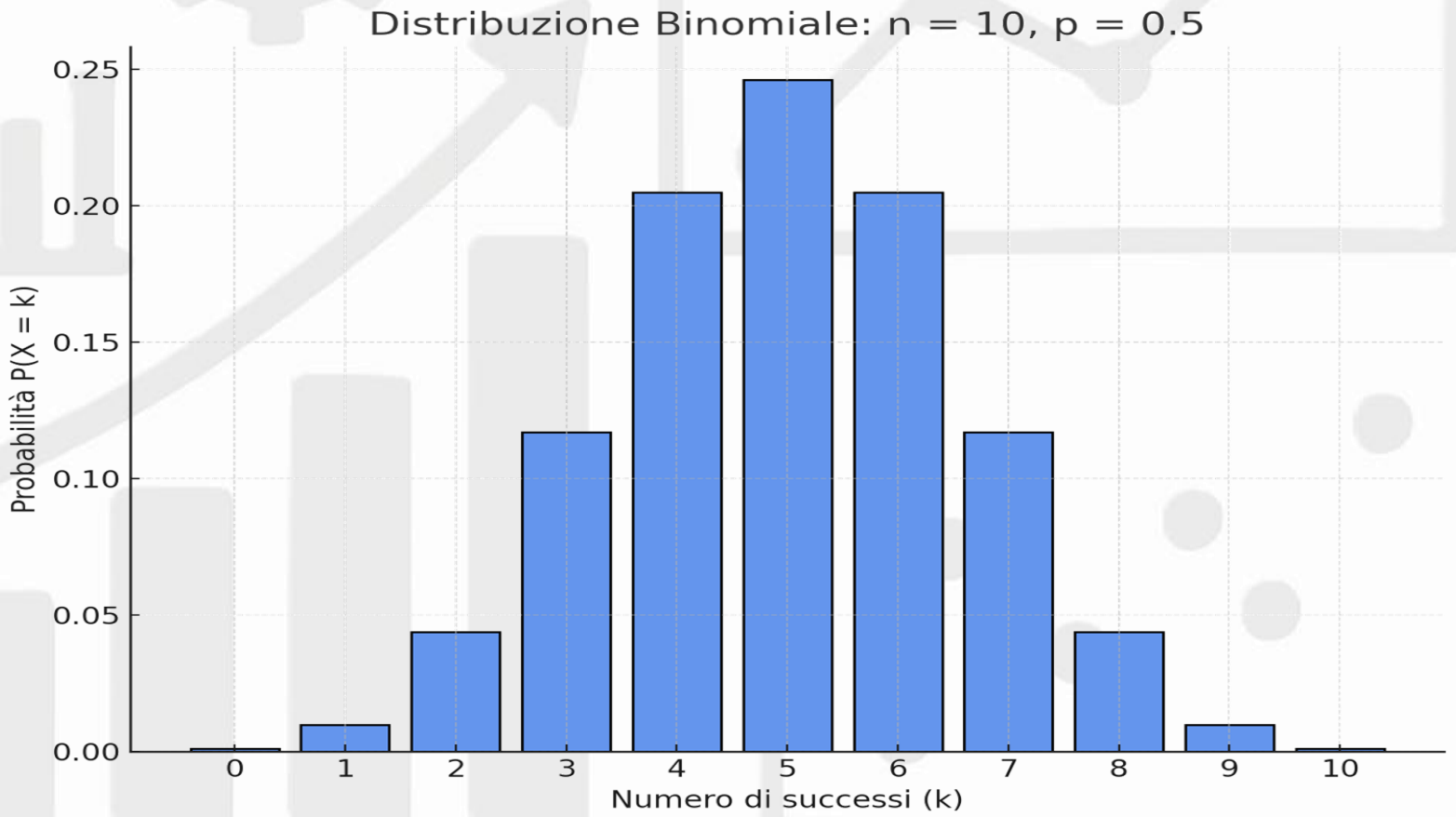
$$P(X = 6) = 210 \cdot 0.0009765625 = 0.205$$

Quindi la probabilità di ottenere **6 teste su 10 lanci** è circa **20,5%**.

CORSO DI PREPARAZIONE PER RESPONSABILE DEL SISTEMA DI GESTIONE DELLA SICUREZZA

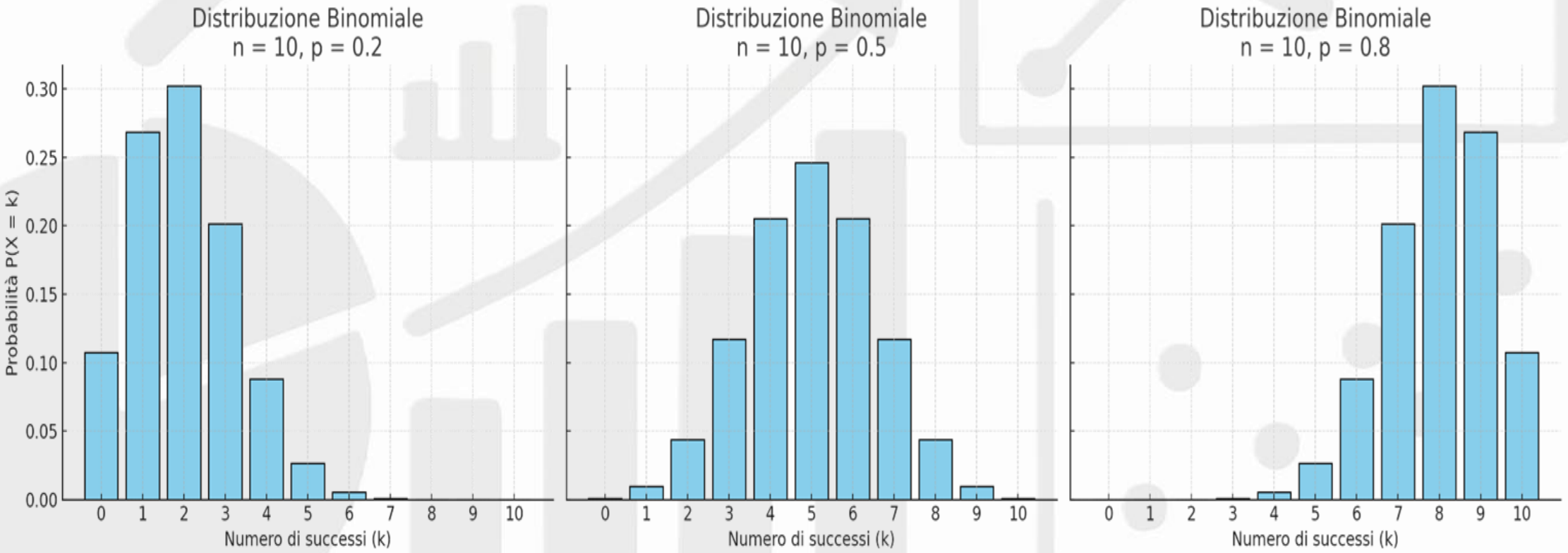
DISTRIBUZIONE BINOMIALE – ESEMPIO TESTA O CROCE –

- $n = 10$
- $k = 6$
- $p = 0.5$



CORSO DI PREPARAZIONE PER RESPONSABILE DEL SISTEMA DI GESTIONE DELLA SICUREZZA

DISTRIBUZIONE BINOMIALE – ESEMPIO GENERICO –



DISTRIBUZIONE DI POISSON

La **DISTRIBUZIONE DI POISSON** descrive la probabilità che un evento accada un certo numero di volte in un intervallo di tempo o spazio definito, assumendo che gli eventi si verifichino in modo indipendente e con una media costante nota.

È particolarmente utile per l'analisi e la predizione di eventi rari.

DISTRIBUZIONE DI POISSON

La distribuzione di Poisson è utilizzata quando analizza il numero di volte in cui un evento avviene in un dato intervallo di tempo, spazio, superficie, volume, ecc. e si verificano le seguenti condizioni:

- Gli eventi sono indipendenti tra loro;
- La probabilità che due eventi avvengano nello stesso istante è trascurabile;
- La media di eventi in un intervallo è costante

DISTRIBUZIONE DI POISSON

La funzione di probabilità della distribuzione di Poisson è:

$$P(X = k) = \frac{\lambda^k \cdot e^{-\lambda}}{k!}$$

- X = variabile casuale che rappresenta il **numero di eventi** osservati
- k = numero specifico di eventi (intero ≥ 0)
- λ = **media** attesa (e anche **varianza**) del numero di eventi nel dato intervallo
- $e \approx 2.71828$ (numero di Eulero)
- $k!$ = fattoriale di k

DISTRIBUZIONE DI POISSON – ESEMPIO CALL CENTER –

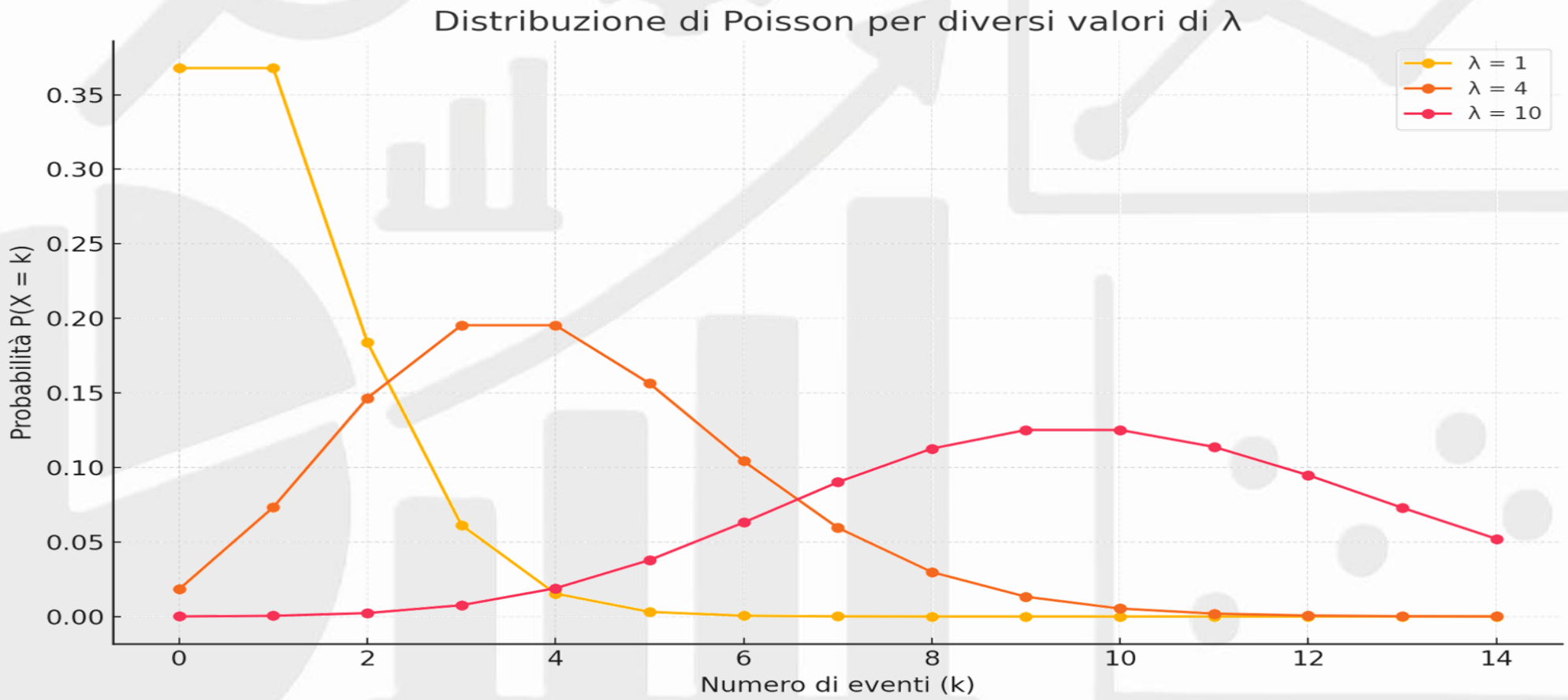
Supponiamo che in un call center arrivino in media λ=4 chiamate ogni ora e vogliamo sapere quale è la probabilità che in un'ora arrivino esattamente 2 chiamate?

$$P(X=2) = \frac{4^2 \cdot e^{-4}}{2!} = \frac{16 \cdot e^{-4}}{2} \approx \frac{16 \cdot 0.0183}{2} \approx 0.146$$

C'è quindi circa il 14,6% di probabilità che arrivino esattamente 2 chiamate in un'ora.

CORSO DI PREPARAZIONE PER RESPONSABILE DEL SISTEMA DI GESTIONE DELLA SICUREZZA

DISTRIBUZIONE DI POISSON – ESEMPIO CALL CENTER –



DISTRIBUZIONE NORMALE

La **DISTRIBUZIONE NORMALE**, o **GAUSSIANA**, è una delle più importanti distribuzioni continue in statistica, caratterizzata dalla famosa curva a campana.
È descritta dalla media (μ) e dalla deviazione standard (σ). La distribuzione normale è simmetrica intorno alla media e indica che la maggior parte dei dati si concentra vicino al valore medio, con una probabilità che diminuisce man mano che ci si allontana da esso.

DISTRIBUZIONE NORMALE

Molti fenomeni naturali e sociali, come l'altezza delle persone o i punteggi di test standardizzati, seguono approssimativamente una distribuzione normale, rendendola estremamente importante per inferenze statistiche e analisi predittive.

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

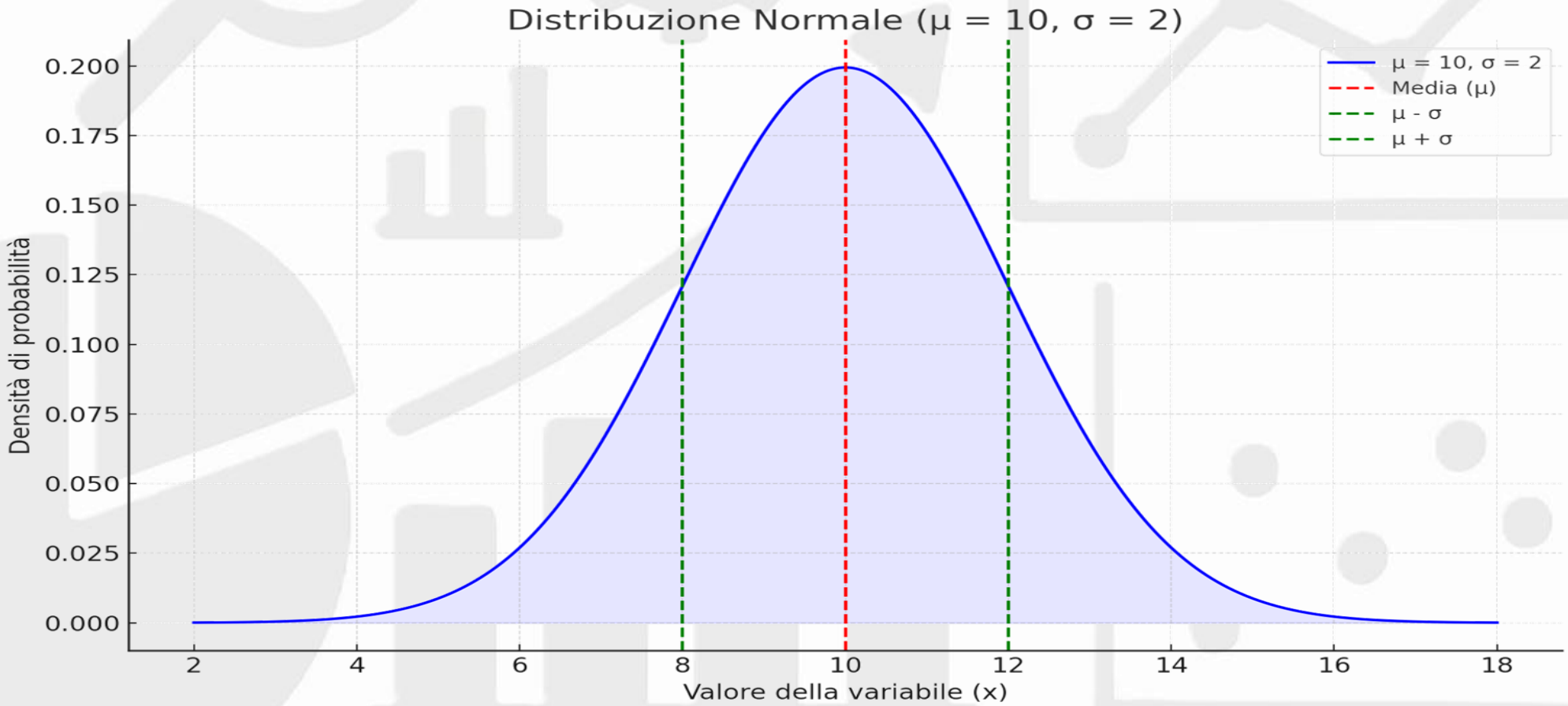
- x = valore della variabile casuale continua
- μ = **media** (valore atteso della distribuzione)
- σ = **deviazione standard** (misura della dispersione)
- σ^2 = **varianza**
- e = numero di Eulero, circa 2.71828
- π = pi greco, circa 3.14159

DISTRIBUZIONE NORMALE – CARATTERISTICHE DELLA CURVA

- La curva è simmetrica rispetto alla media μ
- La media, mediana e moda coincidono
- il 68% dei dati è entro $\mu \pm \sigma$
- il 95% entro $\mu \pm 2\sigma$
- il 99,7% entro $\mu \pm 3\sigma$
- È detta anche “curva a campana” per la sua forma

CORSO DI PREPARAZIONE PER RESPONSABILE DEL SISTEMA DI GESTIONE DELLA SICUREZZA

DISTRIBUZIONE NORMALE – CARATTERISTICHE DELLA CURVA



DISTRIBUZIONE ESPONENZIALE

La **DISTRIBUZIONE ESPONENZIALE** descrive il tempo di attesa tra eventi consecutivi in un processo di Poisson, caratterizzato da una frequenza costante media.

È ampiamente utilizzata per modellare situazioni come il tempo tra arrivi successivi di clienti o guasti in una macchina.

La caratteristica principale è che la la probabilità che un evento accada non dipende da quanto tempo sia già trascorso (mancanza di memoria).

DISTRIBUZIONE ESPONENZIALE

Capire questa distribuzione è cruciale per pianificare strategie efficaci in settori come la manutenzione preventiva e l'organizzazione dei servizi.

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

- x: tempo o distanza (variabile casuale continua)
- λ: **parametro della distribuzione**, detto **tasso di eventi** (rate); rappresenta la **frequenza media degli eventi**
- e: costante di Eulero, circa 2.71828

DISTRIBUZIONE ESPONENZIALE

Le caratteristiche della distribuzione esponenziale sono queste:

- La distribuzione è definita solo per $x \geq 0$
- La **media** è $\frac{1}{\lambda}$
- La **varianza** è $\frac{1}{\lambda^2}$
- È **asimmetrica a destra** (coda lunga)
- Ha la **proprietà di assenza di memoria:** (cioè, il futuro non dipende dal passato)

DISTRIBUZIONE ESPONENZIALE – ESEMPIO –

Supponiamo che in un ospedale, in media, un paziente arrivi ogni 20 minuti.

Allora il **tasso di arrivo** è:

$$\lambda = \frac{1}{20} = 0.05 \text{ arrivi al minuto}$$

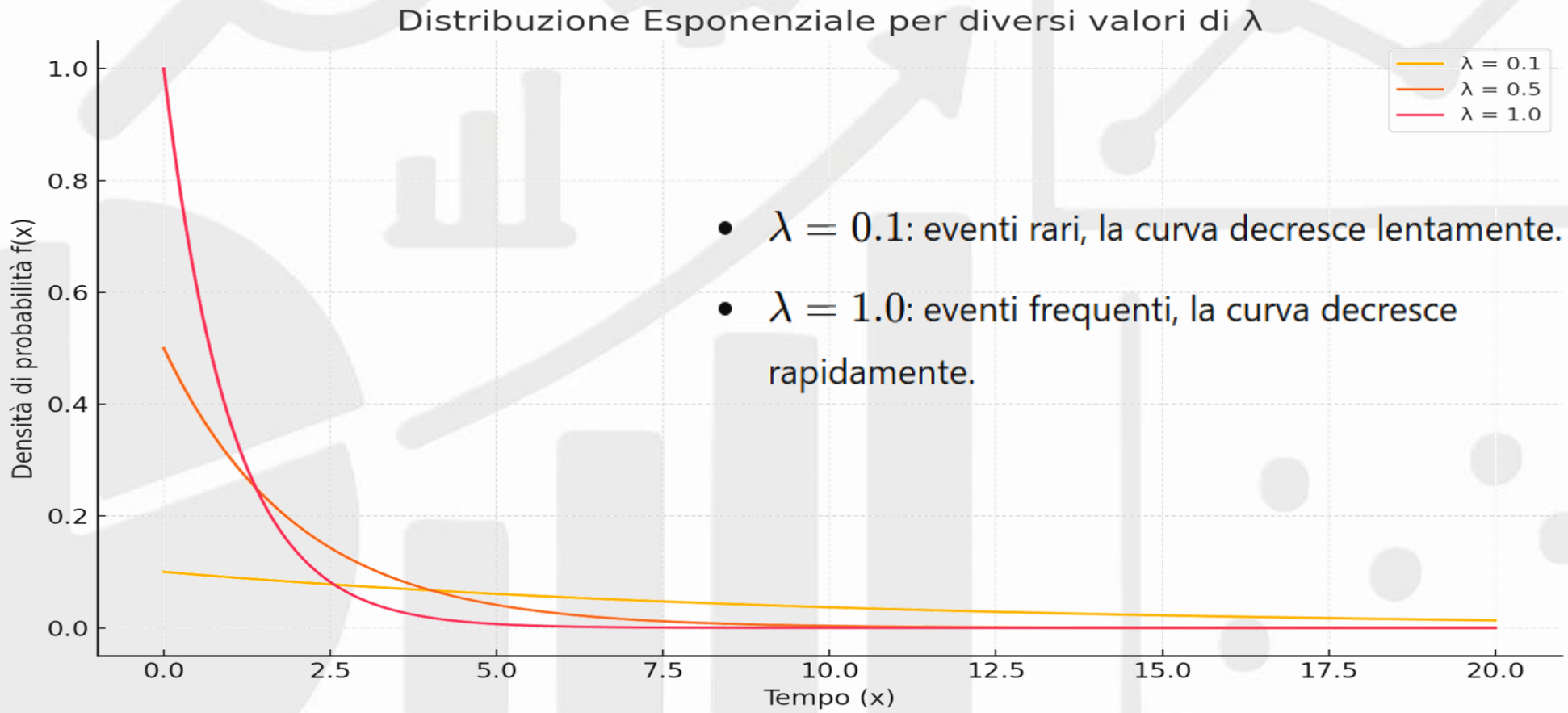
Qual è la probabilità che il prossimo paziente arrivi **entro 10 minuti**?

$$P(X \leq 10) = 1 - e^{-\lambda \cdot 10} = 1 - e^{-0.5} \approx 1 - 0.6065 = 0.3935$$

Quindi c'è circa il **39,35% di probabilità** che un paziente arrivi **entro 10 minuti**.

CORSO DI PREPARAZIONE PER RESPONSABILE DEL SISTEMA DI GESTIONE DELLA SICUREZZA

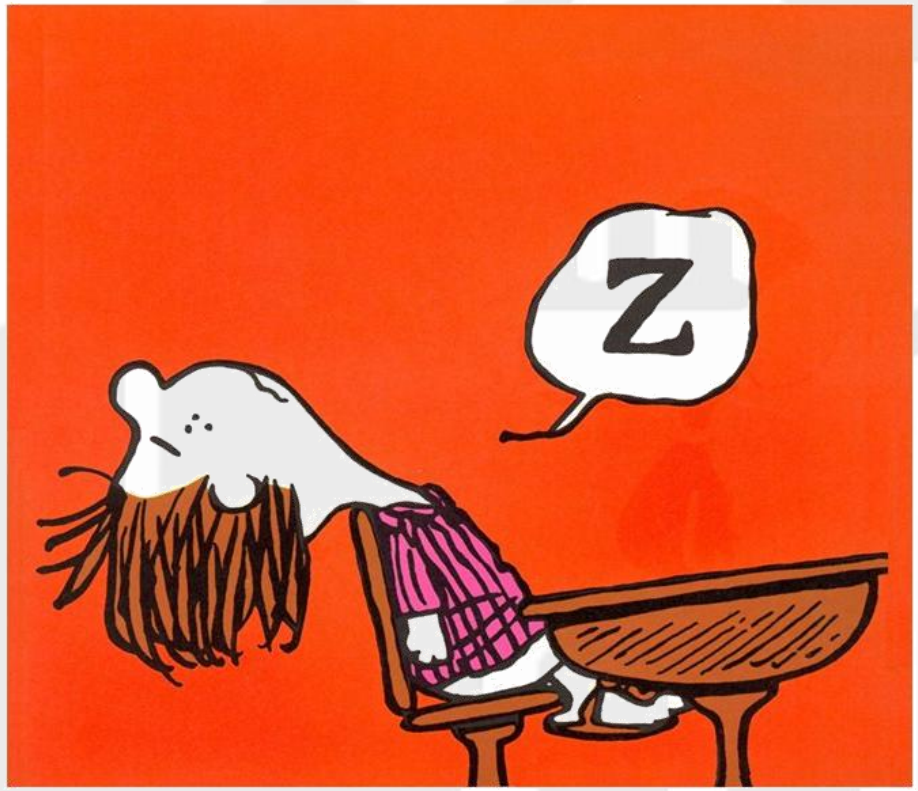
DISTRIBUZIONE ESPONENZIALE – GRAFICO –



- $\lambda = 0.1$: eventi rari, la curva decresce lentamente.
- $\lambda = 1.0$: eventi frequenti, la curva decresce rapidamente.

CORSO DI PREPARAZIONE PER RESPONSABILE DEL SISTEMA DI GESTIONE DELLA SICUREZZA

GRAZIE PER L'ATTENZIONE



Studio di INGEGNERIA MECCANICA Giuseppe Alessandro Corso

http://www.ingcorso.it

Politecnica s.r.l.

Studio di INGEGNERIA MECCANICA Giuseppe Alessandro Corso
http://www.ingcorso.it
Politecnica s.r.l.