

**Università degli Studi di Ferrara
SSIS- VIII ciclo
CL. A049 MATEMATICA E FISICA**

11 Febbraio 2008

**“Calcolo Combinatorio-Triangolo di Tartaglia e
Binomio di Newton.**

Elementi di Calcolo delle Probabilità.

**Elementi di Statistica Descrittiva, correlazione e
regressione lineare”**

**Supervisor Tirocinio
Proff. Minni – Neri -Tomasì**

**Specializzanda
Anna Chiara De Candia**

INDICE

<i>Cosa dicono i programmi</i>	3
I programmi ministeriali del Liceo Scientifico.....	3
PNI programmi per il biennio	3
PNI programmi per il triennio	4
Da Matematica 2003 (UMI)	5
<i>Il percorso didattico</i>	6
Destinatari.....	6
Obiettivi generali	6
Obiettivi trasversali	6
Obiettivi specifici	6
Metodologie didattiche.....	7
Materiali e strumenti utilizzati	7
Controllo dell'apprendimento	7
Griglia per la valutazione	8
Attività di recupero	8
Tempi dell'intervento didattico	8
<i>U.D. 1 CALCOLO COMBINATORIO</i>	9
Obbiettivi specifici U.D.1.....	9
Contenuti U.D.1	9
Tempi U.D.1.....	9
Sviluppo dei contenuti U.D.1	9
Dagli insiemi ai raggruppamenti	9
Le disposizioni semplici.....	12
Le disposizioni con ripetizione	13
Le permutazioni semplici.....	15
Le permutazioni con ripetizione	16
Le combinazioni semplici	18
Coefficiente binomiale e triangolo di Tartaglia	19
Triangolo di Tartaglia e binomio di Newton.....	20
Le combinazioni con ripetizione	22
Il calcolo combinatorio con l'uso di software	22
Verifica sommativa di fine modulo (2h di tempo)	24
<i>U.D. 2 ELEMENTI DI STATISTICA DESCRITTIVA</i>	27
Obbiettivi specifici U.D.2.....	27
Prerequisiti U.D. 2.....	27
Contenuti U.D.2	27
Tempi U.D.2.....	27
Sviluppo dei contenuti U.D.2	28
Cos'è la Statistica Descrittiva	28
Indagine statistica.....	29
INDICATORI STATISTICI	36
LA DISTRIBUZIONE NORMALE	45
DISTRIBUZIONE STATISTICA CONGIUNTA.....	48
Retta di regressione e correlazione lineare.....	53
Esempio di esercitazione in laboratorio	57
Verifica sommativa di fine modulo (2h di tempo)	59
<i>U.D. 3 ELEMENTI DI CALCOLO DELLE PROBABILITA'</i>	62
Obbiettivi specifici U.D.3.....	62
Prerequisiti U.D. 3.....	62
Contenuti U.D.3	62
Tempi U.D.3.....	62
Sviluppo dei contenuti U.D.3	63
Introduzione storica al calcolo delle probabilità.....	63
Gli eventi.....	64
La concezione classica della probabilità	64
La probabilità e il calcolo combinatorio.	66
La concezione frequentista della probabilità	68
La concezione soggettiva della probabilità	70
L'impostazione assiomatica della probabilità.....	71
La speranza matematica e i giochi equi.....	81

Esempio di esercitazione di laboratorio	82
Verifica sommativa di fine modulo (2h di tempo)	84
BIBLIOGRAFIA	87

Cosa dicono i programmi

I programmi ministeriali del Liceo Scientifico

V Classe

Massimi e minimi con il metodo delle derivate, applicazioni. Nozione di integrale con qualche applicazione. Disposizioni, permutazioni e combinazioni semplici. Binomio di Newton. Nelle ultime quattro classi: applicazioni dell'algebra alla geometria di 1° e 2° grado con relativa discussione.

PNI programmi per il biennio

“TEMA 3. RELAZIONI E FUNZIONI

- a) Insiemi ed operazioni su di essi. Insiemi finiti: prime nozioni di calcolo combinatorio.
- b) Leggi di composizione ed individuazione di particolari strutture. Prodotto cartesiano. Relazioni binarie: relazioni d'ordine e di equivalenza. Applicazioni (funzioni).
- c) Funzioni $x \rightarrow ax + b$, $x \rightarrow ax$ (alla seconda) $+ bx + c$, $x \rightarrow a/x$. Grafici e zeri di tali funzioni.”

“TEMA 4. ELEMENTI DI PROBABILITA' E DI STATISTICA

- a) Semplici spazi di probabilità: eventi aleatori, eventi disgiunti e "regola della somma".
- b) Probabilità condizionata, probabilità composta. Eventi indipendenti e "regola del prodotto".
- c) Elementi di statistica descrittiva: rilevazione di dati, valori di sintesi, indici di variabilità, regressione e correlazione.”

Dal commento ai temi si legge:

“TEMA 3..... Lo studio del calcolo combinatorio sarà limitato alle disposizioni, permutazioni, combinazioni e loro proprietà principali; il docente ne approfitterà per abituare, tra l'altro, l'allievo a dimostrazioni di tipo algebrico.....”

” TEMA 4. Al concetto di probabilità si perverrà da vari punti di vista, avvalendosi di opportune esemplificazioni tratte da situazioni reali.

L'analisi dei problemi sarà facilitata da appropriate rappresentazioni: diagrammi di Eulero-Venn e, soprattutto, grafici di vario tipo.

Il programma di statistica è limitato ad elementi di statistica descrittiva, ma occorre tener presente che anche nella componente descrittiva vi sono numerosi aspetti di tipo induttivo, che l'insegnante metterà opportunamente in risalto. Gli esempi e i problemi saranno scelti in modo da sottolineare l'importanza della statistica nei vari ambiti scientifici e nella realtà in genere.”

PNI programmi per il triennio

“Tema n. 2 - Insiemi numerici e strutture

2.a **Calcolo combinatorio: disposizioni, permutazioni, combinazioni**

....

Suddivisione per anno

Classe terza: 2.a ...”

“Tema n. 4 - Probabilità e statistica

4.a **Statistica descrittiva bivariata: matrice dei dati, tabelle a doppia entrata, distribuzioni statistiche** (congiunte, condizionate, marginali). **Coefficienti di correlazione**

4.b **Valutazioni e definizioni di probabilità in vari contesti**

4.c **Correlazione, indipendenza, formula di Bayes. Variabili aleatorie in una e *in due dimensioni*** (casi finiti)

4.d ***Variabili aleatorie discrete: distribuzioni binomiale, geometrica, di Poisson***

Suddivisione per anno

Classe terza: 4.a

Classe quarta: 4.b - 4.c

Classe quinta: 4.d”

Nel commento

“Tema n. 2 - Insiemi numerici e strutture

Lo studio del calcolo combinatorio si può limitare alle disposizioni, permutazioni e combinazioni semplici e alle loro proprietà principali; esso contribuirà, tra l'altro, ad abituare lo studente a dimostrazioni di tipo algebrico.....

Tema n. 4 - Probabilità statistica

Gli elementi di calcolo delle probabilità e statistica rispondono all'esigenza di abituare l'alunno ad effettuare modellizzazioni di situazioni in condizioni di incertezza.

A questo fine è preferibile che la **statistica descrittiva (studio dei fenomeni collettivi) preceda il calcolo delle probabilità**, in quanto atta a fornire semplici modelli capaci di aprire la problematica concettuale delle probabilità. Inoltre la statistica descrittiva bivariata è così largamente utilizzata nella pubblicistica quotidiana che appare molto opportuno e naturale il suo inserimento precoce nella scuola.

Per quanto riguarda il calcolo delle **probabilità l'allusione ai vari contesti in cui si valutano queste probabilità conduce alle diverse definizioni di probabilità** che sono state storicamente proposte; definizioni che, opportunamente riprese, **non verranno viste come antitetiche l'una dell'altra**, potendosi usare in ogni contesto applicativo quella che appare più opportuna nello stato di informazione in cui si sta operando.

Una possibile sintesi tra le varie definizioni, che potrà essere effettuata **all'ultimo anno, sta nella formalizzazione assiomatica della teoria**, che va presentata e motivata sia da un punto di vista storico, sia secondo una

giustificazione di comodità per lo sviluppo dell'intera teoria, sia per fornire un ulteriore esempio di teoria matematica espressa in forma ipotetico-deduttiva.

Questo esempio potrà utilmente essere accostato a quelli di geometria e di altri contesti per consentire quella sintesi finale che è il ripensamento del metodo matematico.

Le semplici distribuzioni di probabilità, che saranno trattate se il docente lo ritiene opportuno, sono sufficienti a dare indicazioni non banali sulla problematica di questa parte del calcolo delle probabilità, anche perché sono particolarmente ricche di applicazioni in vari contesti: fisico, biologico, economico, applicazioni che saranno utilizzate per meglio mettere in luce gli aspetti peculiari dei diversi modelli (binomiale, poissoniano ecc.).

Particolare cura sarà posta nel ricordare le basi storiche e filosofiche (Pascal, empirismo inglese, ecc.).“

Da Matematica 2003 (UMI)

Dati e previsioni: primo biennio

Abilità	Conoscenze
• Comprendere la differenza fra caratteri qualitativi, quantitativi discreti e continui. • Predisporre la struttura della matrice dei dati grezzi rispetto a una rilevazione pianificata e inserire i dati rilevati anche in un foglio elettronico (cfr. <i>Laboratorio di Matematica</i>). • Passare dai dati grezzi alle distribuzioni statistiche di frequenze ed alle corrispondenti rappresentazioni grafiche. • Calcolare i principali indici di posizione e di dispersione per caratteri quantitativi. (1) • Costruire lo spazio degli eventi in casi semplici e determinarne la cardinalità. • Valutare la probabilità in diversi contesti problematici. • Distinguere tra eventi indipendenti e non.	• Distribuzioni delle frequenze a seconda del tipo di carattere. Frequenze assolute, relative, percentuali e cumulate. • Principali rappresentazioni grafiche per le distribuzioni di frequenze (2). Serie storiche e loro rappresentazione. • Proprietà dei principali indici statistici di posizione e di dispersione (2). • Eventi e operazioni con gli eventi. Eventi incompatibili; eventi esaustivi. Significato della probabilità e sue valutazioni (3). • Probabilità condizionata, probabilità composta; probabilità totale.

(1) Occorre distinguere con chiarezza, attraverso opportuni esempi, il caso in cui il calcolo degli indici si basa sui singoli valori effettivamente rilevati o su una distribuzione statistica con valori raggruppati in classi.

(2) Riprendere le rappresentazioni/indici già introdotti nella scuola media in maniera ampliata e approfondita.

(3) Si suggerisce di dare spazio anche alla valutazione statistica della probabilità, non presentando cioè solo esempi riconducibili alla valutazione classica.

Dati e previsioni: secondo biennio

Abilità	Conoscenze
• Identificare situazioni che richiedono di rilevare lo stesso carattere su una unità statistica formata da 2 elementi, o 2 caratteri diversi sulla stessa unità statistica. • Impostare una tabella a doppia entrata; classificare i dati secondo due caratteri e riconoscere in essa i diversi elementi individuabili. • Selezionare, produrre ed usare appropriate rappresentazioni grafiche delle distribuzioni doppie. • Utilizzare la formula di Bayes. • Valutare criticamente le informazioni fornite dai	• Distribuzione doppia di frequenze e tabella a doppia entrata. Distribuzioni condizionate e marginali. • Principali rappresentazioni grafiche per le distribuzioni doppie rispetto a caratteri di qualsiasi natura. • Concetto e significato di modello: correlazione e regressione. • Formula di Bayes e suo significato. • Semplici distribuzioni di probabilità • Il concetto di gioco equo.

media, con riferimento particolare ai giochi di sorte e ai sondaggi.	Il ragionamento induttivo e le basi concettuali dell'inferenza (1).
--	---

- (1) Trattare il ragionamento induttivo e le basi concettuali dell'inferenza a livello informativo, partendo da esempi significativi come, per esempio, exit-poll per un ballottaggio, risultati di sondaggi.

Il percorso didattico

Il presente percorso didattico si articola in tre unità didattiche svolte nel triennio di un Liceo Scientifico PNI:

U.D.1 CALCOLO COMBINATORIO

U.D. 2 ELEMENTI DI STATISTICA DESCRITTIVA

U.D. 3 ELEMENTI DI CALCOLO DELLE PROBABILITA'

Destinatari

U.D.	Classe	Ore di matematica
U.D.1	Classe III	5 (2 di laboratorio)
U.D. 2	Classe IV	5 (2 di laboratorio)
U.D. 3	Classe V	5 (2 di laboratorio)

Obiettivi generali

- Acquisire le conoscenze, le competenze e le capacità previste dalle singole unità.
- Acquisire consapevolezza dell'utilità logica delle proprietà degli argomenti trattati.
- Condurre ad un appropriato lessico matematico.
- Imparare ad operare con la simbologia opportuna.
- Sviluppare la capacità di utilizzare metodi, strumenti e modelli matematici in situazioni diverse.
- Promuovere la padronanza dei processi di analisi, sintesi e valutazione.
- Sollecitare la padronanza a fronteggiare una impostazione ciclica dello studio in cui le stesse nozioni e gli stessi concetti possono essere affrontati a più riprese, con approfondimenti via via maggiori.
- Sviluppare l'interesse per gli aspetti storico-epistemologici della matematica.
- L'uso di software, servirà ad abituare l'allievo ad operare consapevolmente all'interno di diversi sistemi, dotati di loro regole formali e limiti operativi.
- Riconoscere l'importanza della conoscenza degli argomenti trattati nella vita di tutti i giorni.
- Confrontare schematizzazioni matematiche diverse di uno stesso fenomeno

Obiettivi trasversali

- Sviluppare attitudine alla comunicazione ed ai rapporti interpersonali, favorendo lo scambio di opinione tra il docente e allievo e tra gli allievi stessi.
- Proseguire ed ampliare il processo di preparazione scientifica e culturale degli studenti.
- Contribuire a sviluppare lo spirito critico e l'attitudine a riesaminare criticamente ed a sistemare logicamente le conoscenze acquisite.
- Contribuire a sviluppare capacità logiche e argomentative.
- Imparare a rispettare i tempi di consegna dei lavori da svolgere.

Obiettivi specifici

Si rimanda agli obiettivi specifici di ogni U.D.

Metodologie didattiche

Per affrontare gli argomenti di questo percorso didattico si **alterneranno lezioni frontali-dialogiche**, quali **l'insegnamento per problemi** che stimolano gli alunni a formulare ipotesi di soluzione ricorrendo non solo alle conoscenze già possedute ma anche alla intuizione e alla fantasia, a **metodologie didattiche attive** quali le **discussioni guidate e il lavoro di gruppo**. Il lavoro di gruppo rappresenta una risorsa ad alto potenziale, sia come strumento che consente l'emergere di aspetti emozionali, sia come veicolo che, facilitando lo scambio di idee ed esperienze induce un **maggiore coinvolgimento e assunzione di impegno da parte degli alunni**. Si ritiene più proficuo affrontare gli argomenti, oggetto delle lezioni, **utilizzando il metodo induttivo** che consiste nel **partire da esempi concreti** della vita quotidiana (es. partite di calcio, le estrazioni del gioco del lotto, giochi sociali, ecc.) per **passare poi, gradatamente, alla formalizzazione dei concetti introdotti**.

A sostegno si farà uso del **laboratorio di informatica**. Questo momento dovrà essere visto prevalentemente come attività diretta degli allievi, interazione allievo-docente, e dovrà essere armonicamente inserita nella trattazione dei temi affrontati di volta in volta. **Molti degli esempi che serviranno da introduzione ai nuovi concetti potranno essere svolti direttamente in laboratorio con l'ausilio di un foglio elettronico (Microsoft Excel o altri)**.

Si dovranno utilizzare in modo corretto gli strumenti di calcolo e di elaborazione. L'attività di laboratorio integra gli elementi di contenuto dei vari temi e costituisce essa stessa, un momento di riflessione teorica. Essa consisterà in:

- analisi dei problemi e loro risoluzione informatica attraverso l'utilizzo di programmi disponibili e di software di utilità;

- esplorazione e verifiche di proprietà matematiche, rappresentazioni grafiche e calcoli, simulazioni, come momenti costitutivi del processo di apprendimento della matematica e delle sue successive sistematizzazioni.

Gli argomenti delle unità verranno presentati mediante opportuni esempi svolti, che ne agevoleranno la comprensione; a conclusione di ciascuno di essi si cercherà di proporre un buon numero di esercizi, risolti assieme in classe in modo che siano momento immediato di sostegno e anche di recupero della teoria.

Agli studenti verranno assegnati esercizi per casa, in modo che questi possano acquisire una maggiore familiarità con l'argomento. In classe la discussione permetterà di chiarire eventuali dubbi.

Infine la **storia della matematica** come strumento metodologico per inquadrare da un punto di vista storico le nozioni ed i concetti introdotti, con brevi accenni, affinché la matematica non sembri una scienza data ma frutto di una evoluzione.

Materiali e strumenti utilizzati

- Lavagna tradizionale.
- Libri di testo.
- Materiale concreto quali monete, dadi, carte, fiches.
- Uso del computer con foglio elettronico tipo Microsoft Excel
- Schede di lavoro per il laboratorio di informatica.

Controllo dell'apprendimento

Durante lo svolgimento della lezione il docente concretizzerà la **valutazione formativa mediante brevi colloqui, esercitazione scritta in classe costituita da quesiti a risposta multipla e non, correzione degli esercizi per casa**, verificando l'acquisizione progressiva delle conoscenze, competenze e capacità previste come obiettivi specifici. Questo aiuterà a comprendere meglio le difficoltà che hanno ostacolato l'apprendimento e quindi permetterà di **progettare l'attività di recupero in itinere**.

Un altro mezzo di controllo saranno le **verifiche orali**, nelle quali gli alunni saranno chiamati ad eseguire alla lavagna alcuni esercizi motivando, con riferimenti alla teoria, le strategie utilizzate per la loro risoluzione. Questo momento permetterà di valutare non solo la preparazione, ma anche la capacità di ragionamento e la loro proprietà di linguaggio.

Al termine di ogni unità didattica, verrà assegnata una verifica sommativa in cui si proporrà agli studenti la risoluzione di esercizi simili a quelli visti in classe, ma non solo, con lo scopo di appurare se gli argomenti trattati sono stati capiti e assimilati e se gli studenti hanno raggiunto un sufficiente grado di autonomia nel risolvere gli esercizi. La correzione della verifica sommativa sarà occasione ulteriore per appurare l'efficacia del metodo didattico utilizzato.

Come momento di sintesi la **valutazione finale** svolgerà il compito di certificare il raggiungimento o meno degli obiettivi generali di apprendimento prefissati, integrando i risultati delle verifiche formative-sommativa con le

osservazioni sistematiche dell'insegnante, che avranno tenuto conto sia delle conoscenze, competenze e capacità raggiunte dai singoli (anche in relazione al loro livello di partenza) sia dell'impegno profuso nella partecipazione al dialogo educativo.

Griglia per la valutazione

Per determinare gli esiti delle verifiche sommative assegniamo ad ogni esercizio un punteggio. La diversità di punteggio tra i vari esercizi rispecchia livelli diversi di difficoltà in termini di conoscenze, competenze e capacità richieste per svolgerli. Nell'attribuire il punteggio completo, nullo o nella frazione intermedia, teniamo conto dei seguenti indicatori suggeriti dal Ministero dell'Istruzione per la correzione della prova scritta di Matematica:

- i) Conoscenze specifiche.
- ii) Competenze nell'applicare le procedure ed i concetti acquisiti.
- iii) Capacità logico e argomentative.

Si evidenzia che verranno prese in considerazione anche la completezza nella risoluzione e la correttezza della risoluzione e dell'esposizione, ma nella griglia di valutazione, entreranno tacitamente nelle altre voci.

E' chiaro che, nel caso di errore nello svolgimento degli esercizi si presenterà la necessità di attribuire solo parte del punteggio completo previsto per essi. Per fare questo, stabiliremo, di volta in volta, a seconda della gravità dell'errore commesso, quanto farlo pesare e di quanto abbassare il punteggio. Una volta soppesati così gli errori commessi dalla classe, applicheremo la stessa diminuzione di punteggio a ciascuno studente che sarà incorso nello stesso errore. La griglia di valutazione relativa alla prova sommativa è inserita al termine delle unità didattiche.

Attività di recupero

Affinché l'attività didattica risulti efficace e completa, si prevede di svolgere eventuali attività di recupero così articolate:

- ✓ Recupero da effettuare in classe durante le ore curricolari, attraverso la ripresa dei concetti non ben compresi e lo svolgimento di esercizi riguardanti tali argomenti.
- ✓ Attività pomeridiane con gli studenti interessati.
- ✓ Assegnazione al singolo studente di esercizi mirati, in modo da risolvere i suoi problemi e superare le sue difficoltà.

Per individuare gli argomenti che necessitano di recupero, sia a livello collettivo sia a livello individuale, ci si avvarrà della verifica formativa, delle prove orali e dell'attività di collaborazione insegnante-allievo

Tempi dell'intervento didattico

Si rimanda alle singole unità didattiche.

U.D. 1 CALCOLO COMBINATORIO

Obbiettivi specifici U.D.1

Conoscenze

- Conoscere le nozioni di permutazione, disposizione, combinazione.
- Definire e calcolare la funzione fattoriale. Definire e calcolare i coefficienti binomiali, conoscere le proprietà dei coefficienti binomiali.
- Conoscere il Triangolo di Tartaglia e sua relazione con i coefficienti binomiali.
- Conoscere e usare lo sviluppo della potenza di un binomio nella forma data dal binomio di Newton.

Competenze e Abilità

- Sapere calcolare il numero di permutazioni, disposizioni, combinazioni
- Discriminare il ruolo di permutazioni e combinazioni nel descrivere raggruppamenti ordinati o non ordinati.
- Formulare semplici problemi in termini delle nozioni del calcolo combinatorio(permutazioni, disposizioni, combinazioni) e risolverli discutendo le soluzioni trovate.

Prerequisiti U.D. 1

- Insiemistica.
- Prodotti notevoli.
- Equazioni di grado superiore al primo.
- Principio di Induzione.

Contenuti U.D.1

- Permutazioni
- Disposizioni
- Combinazioni
- Coefficienti binomiali
- Potenza di un binomio
- Triangolo di Tartaglia

Tempi U.D.1

Lezioni frontali e svolgimento degli esercizi	5h
Verifica formativa + discussione	2h
Attività di laboratorio	2h
Verifica sommativa + discussione	3h
TOTALE	12h

Per un totale di 12 ore che, tenuto conto delle 5 ore settimanali di matematica, equivalgono a circa tre settimane di lavoro. La previsione è da intendersi elastica, perché occorre tener conto dell'andamento e dei processi di apprendimento della classe.

Sviluppo dei contenuti U.D.1

Dagli insiemi ai raggruppamenti

Un ragazzo ha a disposizione due paia di pantaloni e quattro magliette. Ci domandiamo in quanti modi diversi può vestirsi. Fissato un paio di pantaloni, a questo può accostare, una alla volta, ognuna delle quattro magliette e quindi sono quattro possibilità. Ma a questo numero di possibilità dobbiamo aggiungere ancora le possibilità che

si ottengono con il secondo paio di pantaloni e di nuovo ognuna delle quattro magliette. Quindi le possibilità sono in totale otto.

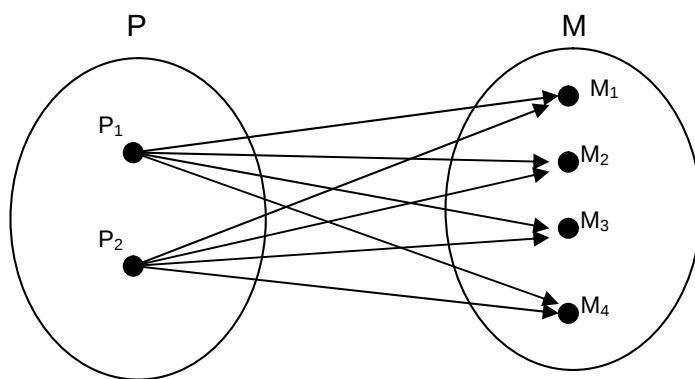
Indichiamo le due paia di pantaloni con P_1 e P_2 , le quattro magliette M_1, M_2, M_3, M_4 e consideriamo gli insiemi $P = \{P_1, P_2\}$ e $M = \{M_1, M_2, M_3, M_4\}$

Elenchiamo tutte le possibili coppie. Esse non sono altro che il prodotto cartesiano fra l'insieme dei pantaloni P e l'insieme delle magliette M .

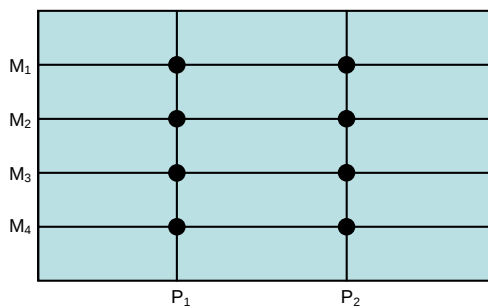
$$P \times M = \{ (P_1; M_1), (P_1; M_2), (P_1; M_3), (P_1; M_4), (P_2; M_1), (P_2; M_2), (P_2; M_3), (P_2; M_4) \}$$

Per visualizzare la situazione possiamo utilizzare diverse rappresentazioni grafiche

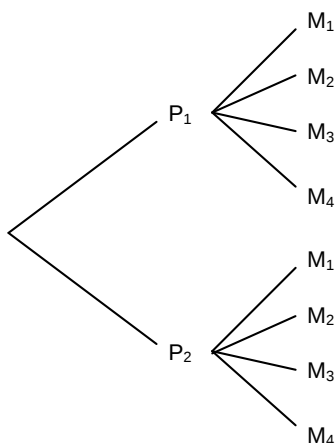
- a. Nel **diagramma sagittale** colleghiamo mediante frecce ogni elemento del primo insieme a ogni elemento del secondo insieme.



- b. Nel **diagramma cartesiano** gli elementi di P sono rappresentati su una semiretta orizzontale, quelli di M su una semiretta verticale, mentre le possibili coppie sono i punti di intersezione della griglia



- c. Nel **diagramma ad albero** al termine dei primi due rami mettiamo un paio pantaloni. Da ognuno di questi punti (nodi) partono poi quattro rami, al termine dei quali mettiamo una maglietta.



Il diagramma ad albero suggerisce un metodo per determinare il numero di tutti i gruppi che è possibile formare

Le 2 possibilità corrispondenti ai rami dei pantaloni indicano quante volte vengono ripetute le 4 possibilità corrispondenti ai rami delle magliette.

Quindi in totale abbiamo $2 \cdot 4 = 8$ gruppi.

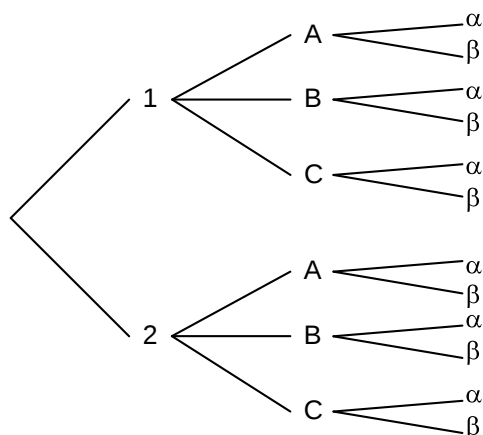
Quando si prendono in esame gruppi di più di due oggetti, il diagramma ad albero è lo strumento di rappresentazione più efficace. Inoltre esso fornisce un metodo per essere sicuri di elencare tutti i raggruppamenti.

Esempio 1

Elenchiamo tutte le sigle di tre elementi che possiamo scrivere utilizzando le cifre 1 e 2 per il primo posto, le lettere A, B, C per il secondo e le lettere greche α e β per l'ultimo posto. Calcoliamo poi quante sono.

Disegniamo il diagramma ad albero.

Osservazione: è possibile e pratico disegnare il diagramma solo se le possibilità non sono molte, altrimenti si può pensare di percorrerne i rami solo mentalmente.



Calcoliamo il numero delle sigle che possiamo scrivere:

2 sono le possibilità per la prima posizione

3 sono le possibilità per la seconda posizione

2 sono le possibilità per la terza posizione

Complessivamente quindi abbiamo: $2 \times 3 \times 2 = 12$ gruppi.

In generale per determinare quanti gruppi si possono formare assegnando il primo posto a un elemento di un insieme A con n elementi, il secondo a uno di un insieme B con m elementi, il terzo a uno di un insieme C con k elementi,... occorre calcolare il prodotto $n \cdot m \cdot k \dots$

Possiamo chiamare "metodo delle possibilità" il metodo usato in questo esempio per il calcolo dei raggruppamenti.

Le disposizioni semplici

Una persona possiede 5 quadri, ma può appenderne solo 3 lungo una parete. Ha importanza anche l'ordine con cui vengono appesi per questioni di colore e di luce. Calcoliamo in quanti modi può appenderli.

Indichiamo i 5 quadri con le lettere P, Q, R, S, T.

Per scrivere tutti i raggruppamenti possibili iniziamo considerando ognuno dei quadri:

P Q R S T 5 possibilità

Ad ognuno affianchiamo poi, uno alla volta gli altri elementi procedendo da sinistra verso destra.

PQ	QP	RP	SP	TP
PR	QR	RQ	SQ	TQ
PS	QS	RS	SR	TR
PT	QT	RT	ST	TS

5x4=20 possibilità

Per ottenere i raggruppamenti finali affianchiamo a ogni coppia, uno alla volta, gli altri elementi di P,Q,R,S,T. Sempre procedendo da sinistra a destra.

PQR	PQS	PQT	QPR	QPS	QPT	RPQ	RPS	RPT	SPQ	SPR	SPT
PRQ	PRS	PRT	QRP	QRS	QRT	RQP	RQS	RQT	SQP	SQR	SQT
PSQ	PSR	PST	QSP	QSR	QST	RSP	RSQ	RST	SRP	SRQ	SRT
PTQ	PTR	PTS	QTP	QTR	QTS	RTP	RTQ	RTS	STP	STQ	STR

TPQ	TPR	TPS
TQP	TQR	TQS
TRP	TRQ	TRS
TSP	TSQ	TSR

5X4X3 = 60 possibilità

Notiamo che ogni terna di quadri differisce dalle altre per:

- almeno un elemento
- ordine degli elementi

oppure per entrambi i motivi.

Per arrivare rapidamente al calcolo consideriamo che per il primo posto le possibilità sono 5, per il secondo si sono ridotte a 4 e per il terzo a 3. Complessivamente i gruppi sono $5 \times 4 \times 3 = 60$.

DEFINIZIONE

Le **disposizioni semplici di n elementi distinti di classe k** (con $k \leq n$) sono tutti i gruppi di k elementi scelti fra gli n, che differiscono per almeno un elemento o per l'ordine con cui gli elementi sono collocati.

Indichiamo il numero delle **disposizioni semplici di n elementi distinti di classe k** con $D_{n,k}$

Nell'esempio precedente quindi $D_{5,3}=60$

Generalizziamo il procedimento seguito per i quadri considerando n oggetti distinti e disposizioni di classe k:

$k = 1$	$D_{n,1}=n,$
$k = 2$	$D_{n,2} = n \cdot (n-1),$
$k = 3$	$D_{n,3} = n \cdot (n-1) \cdot (n-2)$

In generale $D_{n,k} = n \cdot (n-1) \cdot (n-2) \cdot \dots \cdot (n-k+1)$

Osservazione: facciamo notare che nei testi stranieri e sulle calcolatrici le disposizioni semplici sono indicate con il simbolo $n P k$.

Esempio 1

Vogliamo determinare quante sono le sigle di 4 elementi diversi che si possono formare utilizzando le prime 7 lettere dell'alfabeto.

L'insieme degli elementi è $A = \{a, b, c, d, e, f\}$

I gruppi che si formano sono distinti per almeno un elemento oppure per l'ordine in cui compaiono gli elementi, sono quindi disposizioni semplici di 7 elementi di classe 4.

Dunque si avrà:

$$D_{7,4} = 7 \cdot 6 \cdot 5 \cdot 4 = 840$$

Esempio 2

Nell'assegnazione dei posti su un aereo sono rimasti solo 4 posti vicino al finestrino. In quanti modi possono essere assegnati questi posti se le persone che li hanno richiesti sono 10?

Si tratta di distribuire 10 oggetti diversi (le persone) a gruppi di 4 (i posti disponibili).

Si avrà quindi

$$D_{10,4} = 10 \cdot 9 \cdot 8 \cdot 7 = 5040$$

Esempio 3

Quanti sono i numeri interi di 4 cifre diverse che si possono formare con le 10 cifre decimali?

Se calcoliamo $D_{10,4} = 10 \cdot 9 \cdot 8 \cdot 7 = 5040$

Abbiamo incluso così anche i numeri che iniziano con la cifra 0 che però non sono numeri interi di 4 cifre diverse.

Occorre quindi calcolare quante delle 5040 disposizioni iniziano con 0 e sottrarle per trovare il numero richiesto.

Consideriamo allora le 9 cifre diverse da 0 e calcoliamo le disposizioni di classe 3:

$$D_{9,3} = 9 \cdot 8 \cdot 7 = 504$$

Se a queste poniamo lo 0 davanti abbiamo le disposizioni da eliminare.

Il numero richiesto sarà pertanto $5040 - 504 = 4536$.

Avremmo anche potuto ragionare con il metodo delle possibilità:

per la prima posizione ci sono 9 possibilità (le cifre da 1 a 9), per la seconda ancora 9 perché non usiamo la cifra del primo posto ma aggiungiamo lo 0, per la terza 8 e per la quarta 7.

Dunque: $9 \cdot 9 \cdot 8 \cdot 7 = 4536$.

Esempio 4

Nell'ippica è denominata "corsa tris" una corsa in cui gli scommettitori devono indovinare i cavalli che arriveranno al primo, secondo e terzo posto. Supponendo che partano 10 cavalli, quanti sono i possibili ordini d'arrivo nelle prime tre posizioni?

Il problema proposto si riduce al calcolo del numero di modi diversi in cui si possono disporre in ordine 3 cavalli scelti in un insieme di 10. Tale numero è perciò

$$D_{10,3} = 10 \cdot 9 \cdot 8 = 720$$

Esempio 5

Ogni incontro di calcio viene indicato mediante i nomi, posti in un certo ordine, delle due squadre che vi partecipano. Perciò ogni incontro del campionato di calcio corrisponde a una disposizione, di classe 2, di elementi dell'insieme delle squadre che partecipano al campionato. Se il campionato si svolge con il "girone all'italiana", ogni squadra deve incontrare due volte, una sul proprio campo e una in trasferta, ciascuna delle avversarie e si svolgono così tutti gli incontri possibili. Qual è il numero di partite da giocare in un campionato a 18 squadre?

Esso è ovviamente

$$D_{18,2} = 18 \cdot 17 = 306$$

Le disposizioni con ripetizione

Lanciamo una moneta 4 volte e cerchiamo di prevedere tutti i modi con cui si succedono le due facce.

Indichiamo con T testa e C croce.

Se lanciamo una moneta una volta possiamo ottenere:

T C

Se la lanciamo 2 volte possiamo ottenere:

TT TC CT CC

Se la lanciamo 3 volte:

TTT TCT CTT CCT

TTC TCC CTC CCC

Per avere tutti i possibili risultati per 4 lanci:

TTTT TCTT CTTT CCTT

TTTC TCTC CTTC CCTC

TTCT TCCT CTCT CCCT

TTCC TCCC CTCC CCCC

I gruppi così ottenuti differiscono per l'ordine degli elementi contenuti, ma un elemento può comparire più di una volta.

I gruppi trovati si chiamano **disposizioni con ripetizione**

Per determinare il loro numero possiamo ricorrere al "metodo delle possibilità". Per il primo posto abbiamo 2 possibilità, che restano 2 anche per il secondo, il terzo e il quarto in quanto un elemento già utilizzato può essere utilizzato nuovamente

$$2 \cdot 2 \cdot 2 \cdot 2 = 2^4 = 16.$$

Utilizziamo la notazione $D'_{2,4} = 16$

Generalizziamo il procedimento considerando n oggetti distinti e determiniamo la formula per raggruppamenti di classe k:

$$k=1 \quad D'_{n,1}=n,$$

$$k=2 \quad D'_{n,2}=n \cdot n = n^2$$

$$k=3 \quad D'_{n,3}=n \cdot n \cdot n = n^3$$

$$k=4 \quad D'_{n,4}=n \cdot n \cdot n \cdot n = n^4$$

Osserviamo che il numero delle disposizioni con ripetizione di n oggetti distinti di classe k si ottiene moltiplicando fra loro tanti fattori n quanti sono gli oggetti che formano una classe

Formalizziamo la definizione

DEFINIZIONE

Le **disposizioni con ripetizione** di n oggetti distinti di classe k (con k maggiore, uguale o minore di n) sono tutti i gruppi di k elementi, anche ripetuti, scelti fra gli n, che differiscano per almeno un elemento o per il loro ordine

Dal ragionamento precedente si ha: $D'_{n,k} = n^k$

Esempio 1

Calcoliamo quante sigle di quattro elementi, anche ripetuti, si possono formare utilizzando le prime sette lettere dell'alfabeto.

Abbiamo $n=7$ e $k=4$.

Quindi il numero richiesto è $D'_{7,4}=7^4=2401$

Osservazione didattica: osserviamo confrontando con analogo esempio per le disposizioni semplici che il numero in questo caso è maggiore, come è logico.

Esempio 2

Quante sigle di cinque elementi, anche non distinti, si possono formare, tali che i primi due posti siano indicati da due cifre e gli ultimi tre da lettere?

Per i primi due posti abbiamo: $D'_{10,2}=10^2=100$

Per gli ultimi tre posti abbiamo: $D'_{21,3}=21^3=9261$

A ogni disposizione con ripetizione di due cifre accompagniamo una disposizione con ripetizione di tre lettere:

Applicando il metodo delle possibilità il numero richiesto sarà pertanto: $D'_{10,2} \cdot D'_{21,3}=926100$

Le permutazioni semplici

Vogliamo calcolare in quanti modi 3 persone, che indicheremo con A, B e C possono sedersi sulle 3 poltrone di prima fila di un palco a teatro.

Per risolvere il problema è sufficiente determinare il numero delle disposizioni semplici di 3 elementi di classe 3: $D_{3,3}=3 \cdot 2 \cdot 1=6$

In casi di questo tipo ogni gruppo contiene tutti gli elementi dell'insieme, cioè la classe di ogni gruppo coincide con il numero degli elementi e ogni gruppo differisce dagli altri solo per l'ordine.

In questi casi parliamo di permutazioni semplici, o più comunemente di permutazioni. Indichiamo con P_n il loro numero nel caso di n oggetti.

La formula generale per calcolare il numero di permutazioni si ottiene ponendo nella formula delle disposizioni semplici $n=k$:

$$P_n = D_{n,n} = n \cdot (n-1) \cdot (n-2) \cdot \dots \cdot (n-n+1) = n \cdot (n-1) \cdot (n-2) \cdot \dots \cdot 1$$

Osservazione: in tale prodotto i fattori sono tutti e soli i numeri naturali a partire da 1 fino ad n .

DEFINIZIONE

Se k è un numero intero maggiore di 1, chiamasi **fattoriale** del numero k , e si indica con $k!$, il numero che risulta prodotto dei primi k numeri naturali a partire da 1, cioè:

$$k! = 1 \cdot 2 \cdot 3 \cdot \dots \cdot (k-1) \cdot k$$

Se poi è $k=1$, oppure $k=0$ si pone sempre per definizione:

$$1! = 1, \quad 0! = 1.$$

Possiamo quindi formalizzare la definizione di permutazione.

DEFINIZIONE

Le **permutazioni semplici di n elementi distinti**, sono tutti i gruppi formati dagli n elementi che differiscono per

il loro ordine.

Si ha

$$P_n = n! = n \cdot (n-1) \cdot (n-2) \cdot \dots \cdot 3 \cdot 2 \cdot 1$$

Esempio 1

Quanti numeri di 6 cifre distinte possiamo scrivere utilizzando gli elementi dell'insieme $A = \{2, 3, 4, 7, 8, 9\}$?

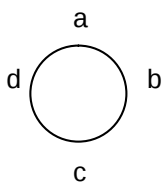
$$P_6 = 6! = 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 720$$

Esempio 2

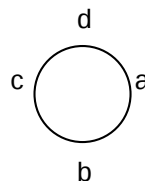
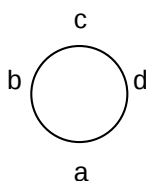
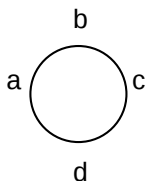
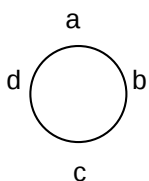
4 ragazzi hanno a disposizione 4 sedie. In quanti modi possono sedersi se le sedie sono disposte attorno ad un tavolo rotondo?

Indichiamo i ragazzi con a, b, c, d

E' evidente che la situazione



coincide, in questo contesto, con ciascuna delle seguenti:



per cui il numero richiesto è uguale al numero delle permutazioni di 4 oggetti, diviso per 4:

$$\frac{P_4}{4} = \frac{4!}{4} = \frac{4 \cdot 3 \cdot 2 \cdot 1}{4} = 3 \cdot 2 \cdot 1 = 3! = P_3$$

Questo è un esempio di **permutazione circolare**, in quanto gli elementi non sono in fila ma disposti intorno a una circonferenza.

Le permutazioni con ripetizione

Calcoliamo quanti anagrammi (anche privi di significato) si possono formare con le lettere della parola MENTO.

Dobbiamo considerare le permutazioni di 5 elementi:

$$P_5 = 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 120.$$

Calcoliamo ora gli anagrammi della parola TETTO.

Possiamo partire anche questa volta dalle permutazioni P_5 , ma dobbiamo considerare che al posto di tre lettere distinte (la M, la N e la T) abbiamo tre lettere uguali (le tre T). Se con MENTO le parole che si ottengono permutando le tre consonanti fra di loro sono diverse, con TETTO sono invece uguali. Per esempio, lasciando ferme le vocali nella prima e seconda posizione, nell'anagramma di MENTO si ottengono le parole: EOMNT, EOMTN, EONMT, EONIM, EOTMN, EOTNM.

Abbiamo 6 casi distinti corrispondenti alle permutazioni di 3 lettere:

Tutti i casi precedenti sono invece indistinguibili se a M e ad N sostituiamo due T. In tutti i sei gruppi otteniamo infatti EOTTT. Se consideriamo le 120 permutazioni di 5 lettere, in questo caso troviamo ogni raggruppamento ripetuto 6 volte. Quindi per ottenere il numero degli anagrammi di TETTO dobbiamo dividere 120 per 6:

$$\frac{120}{6} = 20$$

Per indicare che dei cinque elementi tre corrispondono a uno stesso elemento ripetuto usiamo il simbolo $P_5^{(3)}$ che si legge: "permutazioni di 5 elementi di cui 3 ripetuti". Abbiamo che:

$$p_5^{(3)} = \frac{p_5}{p_3} = \frac{5!}{3!}$$

Chiamiamo i raggruppamenti di questo tipo permutazioni con ripetizione
In generale:

$$P_n^{(k)} = \frac{n!}{k!}$$

La formula si generalizza ulteriormente quando gli elementi ripetuti sono $k, h, \dots, r$ dove $k + h + \dots + r \leq n$. Si ottiene:

$$P_n^{(h,k,\dots)} = \frac{n!}{h! \cdot k! \cdot \dots}$$

Formalizzando:

Definizione

Le **permutazioni con ripetizione** di n elementi, di cui $h, k, \dots$ ripetuti, sono tutti i gruppi formati dagli n elementi che differiscono per l'ordine in cui si presentano gli elementi distinti e la posizione che occupano gli elementi ripetuti.

Esempio1

Calcoliamo il numero degli anagrammi delle seguenti parole:

$$\text{MASSA} \quad P_5^{(2,2)} = \frac{5!}{2! \cdot 2!} = \frac{120}{2 \cdot 2} = 30$$

$$\text{ATTRAENTE} \quad P_9^{(2,3,2)} = \frac{9!}{2! \cdot 3! \cdot 2!} = \frac{362880}{2 \cdot 6 \cdot 2} = 15120$$

Esempio2

Una partita di calcio tra la squadra A e la squadra B è terminata con il risultato di 4 a 3 a favore della squadra A. In quanti modi diversi possono essersi succedute le reti?

Se indichiamo con il simbolo a le reti segnate dalla squadra A e con b le reti fatte dalla squadra B, possiamo rappresentare ogni possibile successione di reti nell'incontro considerato con una permutazione di 7 simboli di cui 4 uguali ad a e 3 uguali a b .

Ad esempio $abababab$ indica che la squadra A ha sempre segnato per prima mentre la B ha poi pareggiato successivamente tranne che per il definitivo vantaggio di A alla fine.

Invece $aaaabbb$ indica che la squadra A ha segnato portandosi fino al 4 a 0 per poi subire la parziale rimonta di B e vincendo con il minimo scarto.

La risposta al quesito si ottiene calcolando il numero di permutazione di 7 oggetti di cui 4 uguali ad a e 3 uguali a b :

$$P_7^{(4,3)} = \frac{7!}{4!3!} = 35$$

Osservazione importante:

a questo punto, dopo aver introdotto la funzione fattoriale è opportuno ritornare al concetto di disposizione per esprimerla in termini fattoriali. Allo stesso tempo osserviamo alcune proprietà della funzione $n!$

Osserviamo che dalla definizione di $n!$ discende la relazione:

$$n! = n \cdot (n-1)!$$

Questa relazione applicata per $2!$ Giustifica l'aver posto per definizione $1! = 1$.

Infatti: $2! = 2 \cdot (2-1)! = 2 \cdot 1! = 2$, si ritrova così il valore corretto secondo la definizione di $2!$.

Sempre dalla definizione discende anche che:

$$(n+1)! - n! = n \cdot n!$$

Infatti: $(n+1)! - n! = (n+1) \cdot n! - n! = n \cdot n! + n! - n! = n \cdot n!$

Ritorniamo quindi a considerare le disposizioni

$$D_{5,3} = 5 \cdot 4 \cdot 3 = \frac{5 \cdot 4 \cdot 3 \cdot (2 \cdot 1)}{(2 \cdot 1)} = \frac{5!}{2!}$$

In generale possiamo ipotizzare $D_{n,k} = \frac{n!}{(n-k)!}$

Verifichiamo che questo è vero, sfruttando le proprietà del fattoriale viste sopra:

$$D_{n,k} = n \cdot (n-1) \cdot (n-2) \cdot \dots \cdot (n-k+1) = \frac{n \cdot (n-1) \cdot (n-2) \cdot \dots \cdot (n-k+1) \cdot (n-k)!}{(n-k)!} = \frac{n!}{(n-k)!}$$

Osserviamo che ritroviamo anche la formula delle permutazioni:

$$D_{n,n} = \frac{n!}{(n-n)!} = \frac{n!}{0!} = n!$$

Le combinazioni semplici

I problemi di raggruppamento di oggetti che abbiamo analizzato fino ad ora avevano un assunto importante: raggruppamenti che avevano gli stessi oggetti con ordine diverso erano considerati diversi.

Ad esempio nel gioco del lotto si estraggono cinque numeri ciascuno imbussolato in una piccola sfera (da 1 a 90). In tale gioco importa sapere quali sono i cinque numeri estratti, ma non conta l'ordine in cui essi si presentano. Estrarre questi cinque numeri equivale a scegliere un sottoinsieme di 5 elementi dell'insieme $\{1; 2; 3; \dots; 89; 90\}$: si parla, in tal caso, di combinazioni dei 90 elementi presi a 5 a 5 (o di classe 5).

Quindi, dati n elementi distinti e un numero $k \leq n$, si dicono combinazioni di n elementi di classe k (o a k a k) tutti i raggruppamenti che si possono formare con k degli n elementi in modo che i diversi raggruppamenti differiscano tra loro almeno per un elemento. Il numero di combinazioni di classe k di n elementi dipende solo da k e da n , e si indica con $C_{n,k}$.

Definizione

Si chiama **combinazione**, o anche *combinazione semplice*, **di classe k di n elementi** (con $k \leq n$), o anche *combinazione di n elementi presi a k a k* , un qualunque sottoinsieme di k elementi, di un dato insieme di n elementi, tutti distinti tra loro.

Osservazione didattica: è importante ricordare la differenza tra combinazione e disposizione, anche in un modo che ne favorisca la memorizzazione.

“Disposizioni: c'entra l'ordine

Combinazioni: NON c'entra l'ordine”

Per calcolare $C_{n,k}$ consideriamo un esempio. Vogliamo formare tutte le combinazioni di classe 2 dell'insieme $A = \{a; b; c; d\}$. Formiamo prima di tutto le disposizioni (semplici) di classe 2 degli elementi di tale insieme; esse sono:

$$D_{4,2} = 4 \cdot 3 = 12 \text{ e precisamente}$$

$ab; ac; ad; ba; bc; bd; ca; cb; cd; da; db; dc$.

Si può osservare che le disposizioni ab e ba sono formate con gli elementi di uno stesso sottoinsieme di A , e precisamente $\{a; b\}$. Lo stesso si può dire delle disposizioni ac e ca , ad e da e così via. Perciò, se si prescinde dall'ordine, ossia se si identificano quelle disposizioni che differiscono solo per l'ordine dei loro elementi, dalle disposizioni precedenti si possono ottenere le combinazioni seguenti: $ab; ac; ad; bc; bd; cd$.

Viceversa, da tali combinazioni si possono ottenere le disposizioni: basta permutare in tutti i modi possibili i due elementi che compongono ciascuna combinazione: in questo caso i modi possibili sono 2 ($P_2 = 2! = 2$). E così da ogni combinazione si ottengono due disposizioni e si avrà quindi che il numero delle disposizioni delle 4 lettere a 2 a 2 è uguale al numero delle combinazioni, $C_{4;2}$, moltiplicato per il numero delle permutazioni, P_2 .

In generale, le combinazioni di classe k di un insieme di n elementi si possono ottenere dalle disposizioni della stessa classe, identificando tra loro quelle che differiscono solo per l'ordine dei loro elementi.

Viceversa, le disposizioni si possono ottenere dalle combinazioni disponendo in tutti i modi possibili, cioè permutando, gli elementi che formano ciascuna combinazione. Poiché il numero di modi in cui si possono permutare i k elementi che compongono ciascuna combinazione è P_k , da ogni combinazione si possono ottenere P_k disposizioni; perciò il numero delle disposizioni $D_{n;k}$ si ottiene moltiplicando $C_{n;k}$ per P_k , ossia si ha

$$D_{n;k} = C_{n;k} \cdot P_k$$

Perciò in generale abbiamo che il numero di combinazioni di classe k di n elementi è

$$C_{n;k} = \frac{D_{n;k}}{P_k}$$

In altri termini possiamo esprimere in forma fattorizzata $C_{n;k}$

$$C_{n;k} = \frac{D_{n;k}}{P_k} = \frac{\overbrace{n(n-1)(n-2)\cdots(n-k+1)}^{k \text{ fattori}}}{k(k-1)\cdots 2 \cdot 1} = \frac{n(n-1)(n-2)\cdots(n-k+1)}{k!}$$

Ovvero: il numero delle combinazioni di classe k che si possono formare con n elementi distinti è uguale al prodotto di k numeri interi decrescenti a partire da n , diviso per il prodotto dei primi k numeri interi, cioè per il fattoriale di k .

Coefficiente binomiale e triangolo di Tartaglia

I numeri $C_{n;k}$ sono detti coefficienti binomiali (il motivo risulterà evidente nel seguito) e vengono indicati con il

simbolo $\binom{n}{k}$ che si legge "n su k".

$$\text{Possiamo allora definire } \binom{n}{k} = \frac{n(n-1)(n-2)\cdots(n-k+1)}{k!}$$

Si decide per comodità di definire anche $\binom{0}{0} = 1$

Utilizzando la formula per le disposizioni semplici espressa come rapporto di fattoriali si ottiene la formula seguente, detta Legge dei tre fattoriali.

$$C_{n;k} = \frac{D_{n;k}}{P_k} = \frac{n!}{(n-k)! \cdot k!} = \frac{n!}{(n-k)! k!}$$

Legge dei tre fattoriali

Osservazione: si tratta della stessa formula usata per le permutazioni con ripetizione di n elementi dei quali k e $n-k$ sono ripetuti:

$$P_n^{(k, n-k)} = \frac{n!}{k! (n-k)!}$$

Generalizzando il risultato

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

Overo si ha la seguente proprietà detta **Legge delle classi complementari**:

$$C_{n,k} = C_{n,n-k} = \binom{n}{k} = \binom{n}{n-k}$$

Osservazione didattica: la verità di questa relazione è evidente perché $\binom{n}{k}$ è il numero di modi in cui si possono scegliere k elementi tra n dati, ma questo equivale a sceglierne n-k tra gli stessi n, il che si può fare in $\binom{n}{n-k}$ modi.

Osservazione didattica: La legge delle classi complementari è utile nel calcolo delle combinazioni quando $k \geq n/2$. Ad esempio invece di calcolare $C_{10,8} = \binom{10}{8} = \frac{10 \cdot 9 \cdot 8 \cdot 7 \cdot \dots \cdot 3}{8 \cdot \dots \cdot 2 \cdot 1}$ è più rapido calcolare

$$C_{10,2} = \binom{10}{2} = \frac{10 \cdot 9}{2}$$

Altra importante proprietà dei coefficienti binomiali è la cosiddetta **formula di ricorrenza**:

$$\binom{n}{k+1} = \binom{n}{k} \cdot \frac{n-k}{k+1}$$

Verifichiamone la validità:

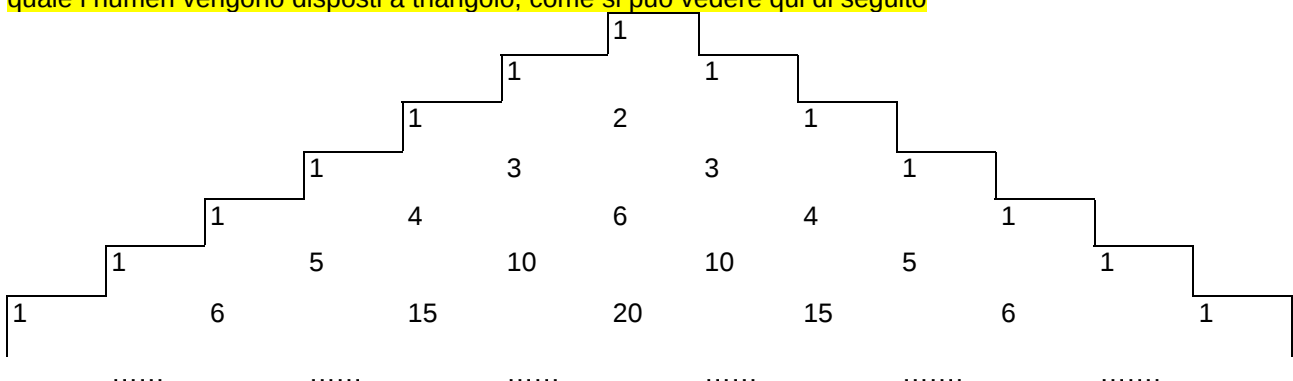
$$\binom{n}{k+1} = \frac{n \cdot (n-1) \cdot \dots \cdot (n-k+1) \cdot (n-k)}{k!(k+1)} = \frac{n \cdot (n-1) \cdot \dots \cdot (n-k+1)}{k!} \cdot \frac{(n-k)}{(k+1)} = \binom{n}{k} \cdot \frac{n-k}{k+1}$$

Triangolo di Tartaglia e binomio di Newton

Consideriamo lo sviluppo della potenza di un binomio.

Questo è un polinomio omogeneo ordinato di grado n con (n+1) termini ognuno moltiplicato per un certo coefficiente.

I coefficienti dello sviluppo si trovano mediante una costruzione "geometrica" detta *triangolo di Tartaglia*, nella quale i numeri vengono disposti a triangolo, come si può vedere qui di seguito



Regola costruttiva

- Il primo e l'ultimo numero di ciascuna riga è 1
- Ogni riga è composta da $n+1$ numeri, se n è il numero di componenti la riga superiore.
- Ogni numero diverso da 1 di una riga qualunque viene ottenuto sommando il numero che gli sta immediatamente sopra a sinistra con quello che gli sta immediatamente sopra a destra

Osservazione didattica: il nome triangolo di Tartaglia per questa costruzione è tipicamente italiano, in onore del matematico Nicolò Fontana (detto Tartaglia perché balbuziente). All'estero però è più conosciuta come triangolo di Pascal, da Blaise Pascal, che un secolo dopo Tartaglia, nel 1654, ne fece grande uso nei suoi studi sulla probabilità. In Germania invece è comunemente attribuita a Stiefel che ne scrisse nel 1544.

I coefficienti che compaiono nel triangolo di Tartaglia sono proprio i coefficienti binomiali. Nella n -esima riga del triangolo al posto k si ha $\binom{n}{k}$.

Tramite il calcolo combinatorio è possibile dimostrare la formula per lo sviluppo del binomio, detta **formula del binomio di Newton**:

$$(a+b)^n = \binom{n}{0}a^n + \binom{n}{1}a^{n-1}b + \binom{n}{2}a^{n-2}b^2 + \dots + \binom{n}{n-1}ab^{n-1} + \binom{n}{n}b^n = \sum_{k=0}^n \binom{n}{k}a^{n-k}b^k$$

Dimostrazione della formula.

$(a+b)^n = (a+b)(a+b) \dots (a+b)$ dove a secondo membro abbiamo n fattori.

Si può pensare di effettuare la moltiplicazione scegliendo, da ciascun fattore $(a+b)$, o il termine a , o il termine b , in tutti i modi possibili, per poi sommare algebricamente i prodotti così ottenuti.

Ora, se si sceglie, ad esempio, k volte il fattore b e $n-k$ volte il fattore a , avrò il monomio $a^{n-k}b^k$

Quante volte comparirà, questo monomio, nella somma finale?

Tante volte quanti sono i modi coi quali, fra gli n fattori, posso selezionare quei k dai quali scegliere b .

E tali modi sono $\binom{n}{k}$

Di qui la formula.

Dalle proprietà del triangolo di tartaglia osserviamo si possono dedurre/ritrovare le proprietà dei coefficienti binomiali:

Dalla simmetria delle righe si ritrova la Legge delle classi complementari:

$$\binom{n}{k} = \binom{n}{n-k}$$

Dalla regola di formazione del triangolo si trova invece la **Formula di Stifel**:

$$\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}$$

E' comunque facile verificarne la validità usando la legge dei tre fattoriali.

$$\binom{n-1}{k-1} + \binom{n-1}{k} = \frac{(n-1)!}{(k-1)!(n-1-k+1)!} + \frac{(n-1)!}{k!(n-1-k)!} = \frac{k(n-1)! + (n-k)(n-1)!}{k(k-1)!(n-k)(n-k-1)!}$$

$$= \frac{n(n-1)!}{k!(n-k)!} = \frac{n!}{k!(n-k)!} = \binom{n}{k}$$

Le combinazioni con ripetizione

Riprendiamo il problema affrontato nello studio delle disposizioni con ripetizione.

Lanciamo successivamente una moneta e segniamo la successione di uscita di testa (T) e croce (C). Questa volta non importa l'ordine di uscita ma solo la composizione dei gruppi.

Se i lanci sono due il numero delle possibilità confronto alle disposizioni si riduce a 3:

TT TC CC k=2

Per 3 lanci a 4:

TTT TTC TCC CCC k=3

Per 4 lanci a 5:

TTTT TTTC TTCC TCCC CCCC k=4

Chiamiamo questi raggruppamenti **combinazioni con ripetizione**.

Osserviamo che in ogni gruppo un elemento può ripetersi fino a k volte e non interessando l'ordine, pur contenendo gli stessi elementi, ogni gruppo li contiene un numero di volte diverso.

Per indicare le combinazioni con ripetizione usiamo la seguente notazione

$C'_{n,k}$

Definizione

Le **combinazioni con ripetizione di n oggetti di classe k** sono i raggruppamenti che si possono formare con gli n elementi di un insieme dato, in modo che ogni gruppo ne contenga k, ma ogni elemento possa trovarsi ripetuto nel gruppo 1. 2 k volte e in modo che ogni gruppo differisca dall'altro per almeno un elemento.

Si dimostra che vale:

$$C'_{n,k} = C'_{n+k-1,k} = \binom{n+k-1}{k} = \frac{(n+k-1) \cdot (n+k-2) \cdot \dots \cdot (n-1) \cdot n}{k!}$$

Il calcolo combinatorio con l'uso di software

Derive

Esistono due funzioni di Derive che consentono di calcolare il numero delle disposizioni e delle combinazioni semplici:

PERM(n,k) calcola il numero delle disposizioni semplici

COMB (n,k) calcola il numero delle combinazioni semplici

Per esempio: PERM(7,3) restituisce il valore 210; infatti $D_{7,3} = 7 \cdot 6 \cdot 5 = 210$

COMB(8,3) restituisce il valore 56; infatti $C_{8,3} = \binom{8}{3} = \frac{8 \cdot 7 \cdot 6}{3!} = 56$

La funzione PERM calcola poi il numero delle permutazioni se si indica lo stesso valore per n e per k; per esempio: PERM(5,5) restituisce 120; infatti $P_5 = 5! = 120$

Derive non possiede altre funzioni predefinite per calcolare il numero delle disposizioni e delle combinazioni con ripetizione; cerchiamo allora di trovare un metodo per calcolarle.
Ricordiamo innanzi tutto le formule:

$$D_{n,k}^r = n^k \quad C_{n,k}^r = \frac{n(n+1)(n+2)\dots(n+k-1)}{k!} = \binom{n+k-1}{k}$$

Allora per calcolare il numero delle disposizioni con ripetizione possiamo creare una funzione DISPR così definita:

DISPR(n,k):=n^k

Una volta definita la funzione, possiamo calcolare il numero delle disposizioni sostituendo al posto di n e di k i valori desiderati; per esempio

DISPR(3,5) restituisce 243; infatti $3^5=243$

DISPR(4,8) restituisce 65536; infatti $4^8=65536$

Per calcolare $C_{n,k}^r$ possiamo servirci ancora della funzione COMB ponendo come parametri n+k-1 e k; per esempio per calcolare $C_{5,4}^r$ dobbiamo inserire la funzione COMB(8,4) e il programma restituisce 70.

Excel

Anche Excel possiede delle funzioni di calcolo per determinare il numero delle disposizioni e delle combinazioni di n elementi di classe k; vediamo la sintassi:

- PERMUTAZIONE(n;k)

Per calcolare il numero delle disposizioni. Se n e k sono uguali viene calcolato il numero delle permutazioni di n elementi

- COMBINAZIONE(n;k)

Per calcolare il numero delle combinazioni

Come Derive anche excel non possiede funzioni predefinite per calcolare il numero delle disposizioni o di combinazioni con ripetizione; occorre quindi costruire di volta in volta una formula appropriata.

Per le disposizioni sarà sufficiente impostare la formula tramite la funzione POTENZA(n;k).

The screenshot shows a Microsoft Excel spreadsheet titled "Microsoft Excel - calc_comb.xls". The spreadsheet is used for combinatorial calculations. It has columns A through O and rows 1 through 35. The data is as follows:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	CALCOLO COMBINATORIO														
2															
3	DISPOSIZIONI		N=	10	K=	3									
4															
5	Disposizione semplice(n,k)		720												
6	Permutazione n		3628800												
7	Disposizione con ripetizione		1000												
8															
9	COMBINAZIONI														
10															
11	Combinazioni semplici		120												
12	Combinazioni con ripetizione		220												
13															
14															
15															
16															
17															
18															
19															
20															
21															
22															
23															
24															
25															
26															
27															
28															
29															
30															
31															
32															
33															
34															
35															

Per le combinazioni con ripetizione ricordando che $C_{n,k}^r = \binom{n+k-1}{k}$ abbiamo utilizzato le proprietà dei

coefficienti binomiali per ricavare la formula $\binom{n+k-1}{k} = \frac{(n+k-1)!}{k!(n-1)!}$

Poiché il fattoriale di n con excel si calcola usando la funzione FATTORIALE(n) la formula che abbiamo inserito in C12 è =FATTORIALE(D3+F3-1)/(FATTORIALE(F3)*FATTORIALE(D3-1))

Verifica sommativa di fine modulo (2h di tempo)

Esercizio 1

Per calcolare quanti differenti valori di 4 cifre si possono ottenere usando le cifre da 0 a 7 devi utilizzare:

- a) La formula per le disposizioni semplici di 8 oggetti di classe 4
- b) La formula per le combinazioni semplici di 8 oggetti di classe 4
- c) La formula per le disposizioni con ripetizione di 8 oggetti di classe 4
- d) La formula per le combinazioni con ripetizione di 8 oggetti di classe 4

Esercizio 2

Devi ripartire un gruppo di 240 partecipanti ad una cena di beneficenza in tavoli da 8 persone ciascuna. Per calcolare in quanti modi puoi distribuire le persone devi utilizzare:

- a) La formula per le disposizioni semplici di 240 oggetti di classe 8
- b) La formula per le combinazioni semplici di 240 oggetti di classe 8
- c) La formula per le disposizioni con ripetizione di 240 oggetti di classe 8
- d) La formula per le combinazioni con ripetizione di 240 oggetti di classe 8

Esercizio 3

In base alle proprietà dei coefficienti binomiali determina quali fra le seguenti uguaglianze sono vere e quali false

- a) $\binom{7}{3} = \binom{7}{4}$ V F
- b) $\binom{12}{5} = \frac{12!}{5!}$ V F
- c) $\binom{8}{2} = \frac{8!}{2!6!}$ V F
- d) $\binom{15}{6} = \binom{15}{4} \cdot \frac{11}{3}$ V F

Esercizio 4

Nello sviluppo del binomio $(2a^2-3b)^5$ quanto vale il coefficiente del termine che ha come parte letterale a^4b^3 ?

Esercizio 5

Quanti anagrammi anche privi di significato si possono fare con la parola libreria? Quanti di essi finiscono per ria?

Esercizio 6

Una password è costituita da: una sequenza di 5 lettere (possibilità di ripetizione di una stessa lettera; sono a disposizione le 26 lettere dell'alfabeto anglosassone), seguita da una sequenza di 3 cifre, non necessariamente distinte.

Quante sono le possibili password?

Esercizio 7

Quante diagonali ha un poligono di 10 lati?

Esercizio 8

Nella mia libreria ho: 10 libri di Storia, 6 libri sugli Animali e 7 libri di Matematica.

Voglio allinearli su di uno scaffale, in modo però che i libri di una medesima materia siano tutti vicini fra loro. In quanti modi diversi posso ordinare i miei 23 libri?

Esercizio 9

In un sacchetto ci sono 9 palline, 3 delle quali recano scritto sulla superficie il numero "1", altre 3 il numero "2", le rimanenti 3 il numero "3".

Si estrae una pallina e si segna la cifra corrispondente. Senza reimpastare la pallina estratta, se ne estrae un'altra, si segna la cifra corrispondente a destra della precedente ... e si prosegue in questo modo fino ad esaurire tutte le palline.

Si costruisce così un numero a 9 cifre.

Quanti numeri distinti è possibile ottenere in questo modo?

Punteggi attribuiti

Esercizio1:	2
Esercizio2:	2
Esercizio3:	4
Esercizio4:	4
Esercizio5:	4
Esercizio6:	3
Esercizio7:	3
Esercizio8:	4
Esercizio9:	5

Griglia di Misurazione

riferita alla verifica sommativa dell'Unità Didattica:

Punteggio Grezzo (Totale 31)	Voto in Decimi (ottenuto con la proporzione)	Voto in decimi (una proposta)
0	0-1	3
1		
2		
3		
4		
5	1-2	
6		
7		
8	2-3	
9		
10		
11	3-4	4
12		
13		
14	4-5	
15		
16		
17	5-6	5
18		
19		
20	6-7	6
21		
22		
23	7-8	
24		
		7
		8

25	8-9	
26		
27		
28	9-10	9
29		
30		10
31	10	

U.D. 2 ELEMENTI DI STATISTICA DESCRITTIVA

Obbiettivi specifici U.D.2

Conoscenze

- Conoscere i concetti di indagine statistica, popolazione, caratteristica, modalità, frequenza....
- Conoscere il concetto di distribuzione
- Conoscere le varie rappresentazioni statistiche
 - o tabelle
 - o diagrammi
 - o istogrammi
- Conoscere i principali indicatori di posizione di una distribuzione
 - o media, media geometrica, moda, mediana
- Conoscere i principali indicatori di dispersione
 - o varianza, scarto quadratico medio
- Conoscere il concetto di distribuzione congiunta
- Conoscere i concetti di dipendenza e correlazione
- Conoscere l'indice del chi quadrato
- Conoscere l'indice di contingenza
- Conoscere il concetto di regressione
- Conoscere il metodo dei minimi quadrati

Competenze e Capacità

- Saper stabilire le caratteristiche principali di un'indagine statistica
- Saper leggere, interpretare e rielaborare i dati statistici forniti in forma tabulare o grafica
- Calcolare indici centrali e di variabilità di una distribuzione statistica
- Stabilire le caratteristiche principali di una indagine statistica
- Valutare l'eventuale dipendenza statistica tra due caratteri attraverso diversi metodi

Prerequisiti U.D. 2

- Familiarità con formule algebriche ed espressioni contenenti radicali
- Coordinate cartesiane nel piano
- Equazioni della retta e della parabola
- Utilizzo del foglio elettronico per l'elaborazione di insiemi di dati più numerosi

Contenuti U.D.2

- Significato statistica descrittiva e introduzione storica
- Cos'è un'indagine statistica e quali sono le sue fasi
- Indici di posizione centrale e di variabilità
- Cenni sulla distribuzione normale
- Distribuzioni congiunte
- Regressione e correlazione lineare

Tempi U.D.2

Lezioni frontali e svolgimento degli esercizi (comprensivi di laboratorio)	8h
Verifica formativa + discussione	2h
Verifica sommativa + discussione	3h
TOTALE	13h

Per un totale di 13 ore che, tenuto conto delle 5 ore settimanali di matematica, equivalgono a circa tre settimane di lavoro. La previsione è da intendersi elastica, perché occorre tener conto dell'andamento e dei processi di apprendimento della classe.

Per questa unità didattica si prevede di svolgere quasi tutte le lezioni con l'ausilio del computer. Si analizzeranno infatti i dati relativi ai vari esempi utilizzati mediante un software foglio elettronico ad esempio excel che permetterà anche la rappresentazione grafica dei dati.

Sviluppo dei contenuti U.D.2

Cos'è la Statistica Descrittiva

E' importante far capire alla classe che la matematica non è solo quella scienza basata su concetti e procedimenti fissati e che opera su dati sicuri, ma che anche realtà non basate su formulazioni esatte sono matematicamente descrivibili.

Ad esempio si può porre un semplice problema:

Ricerca antropologica

Qual è, mediamente, l'altezza delle donne europee?

La risposta a questa domanda non è riferita ad un singolo soggetto ma ad una vastissima moltitudine di casi, questo ci aiuta ad introdurre la statistica, come quella **branca della matematica che ci permette di studiare fenomeni collettivi.**

Note storiche

Anche se la statistica ha avuto una sua autonomia scientifica solo a partire dalla seconda metà del XIX secolo esempi di indagini statistiche risalgono a tempi antichissimi. Già nel 2000 a.C. l'imperatore cinese Yu, avendo diviso il suo impero in province, faceva rilevare attraverso tabelle il grado di sviluppo dell'agricoltura, la qualità e quantità dei raccolti al fine di stabilire, successivamente la "pressione fiscale": ogni regione doveva versare un ammontare di imposte in proporzione al reddito ed allo sviluppo raggiunti.

Gli Egizi usavano il catasto per la rilevazione dei terreni; vi era una dea dei libri e dei conti che si occupava anche della quantità e del valore delle merci importate. Molti autori latini ed anche il Vangelo ci portano testimonianza dei censimenti utilizzati dai Romani per conoscere la consistenza delle popolazioni sulle quali regnavano. Con la caduta dell'Impero Romano cadde anche l'abitudine di compilare statistiche demografiche. Solo nel Medioevo la Chiesa iniziò i censimenti dei beni ecclesiastici, le registrazioni di nascite, battesimi, morti. I dati venivano riportati su appositi libri detti "Registri parrocchiali". Questi furono senza dubbio i più importanti documenti di rilevazione in Italia; divennero obbligatori con il Concilio di Trento a metà del 1500 ed hanno dato un grande contributo alla ricerca pratica e teorica in campo statistico.

E' però Hermann Conring (1606-1681) che ha iniziato a partire dal 1660 un nuovo corso di scienza politica dedicato "descrivere e analizzare le cose più notevoli di uno stato". A tal fine fu istituito da Gottfried Achenwall (1719-1772) il primo corso universitario cui fu dato il nome di **STATISTICA**. Il significato che essa ha mantenuto fino all'inizio del diciannovesimo secolo è stato dunque quello di "descrizione" o "censimento", con successiva analisi dei dati prodotti da tali operazioni. In seguito il significato della statistica si è esteso ben oltre tale interpretazione originaria dando luogo, talvolta, a campi di studio profondamente diversi. Nel corso dei tempi sono state date varie definizioni di statistica (a partire da quella di Achenwall: la descrizione delle cose più notevoli di uno stato) da cui ci si può rendere conto, oltre che della difficoltà di definirla, anche della sua evoluzione, da "aritmetica politica" a "teoria delle decisioni".

Attualmente si parla di due tipi di statistica: una statistica descrittiva o metodologica ed una statistica matematica o inferenziale.

Per statistica descrittiva si intende il complesso delle tecniche (norme) di cui si serve lo sperimentatore per raccogliere, elaborare (organizzare) e rappresentare gruppi di dati osservati, che sono in generale in numero elevato ed in forma disordinata. Lo scopo è quello di ricavare dalle osservazioni opportuni elementi di sintesi.

Si cerca solo di **descrivere il fenomeno oggetto di osservazione** senza cercare di interpretare o generalizzare più di tanto riguardo la totalità dei dati potenzialmente osservabili. La necessità nello studio di un fenomeno è spesso quella di "saperne di più spendendo meno": ciò significa **non dover censire tutto per avere le informazioni che ci servono, ma estrarre le stesse conoscenze dai (pochi) dati osservati** che sono a nostra disposizione. In termini statistici vogliamo **trarre conclusioni su tutta la popolazione partendo dal campione osservato**; ciò implica che molti dei dati saranno a noi ignoti e che dovremo trovare un metodo per averne una qualche conoscenza. Le informazioni di cui sopra rappresentano quindi un passo preliminare, però molto importante, per la successiva **ricerca di un modello interpretativo, che è un aspetto proprio della statistica inferenziale**. Si cerca di studiare gruppi di dati allo scopo di decidere quale tipo di modello è quello che descrive "al meglio" l'intera popolazione. I modelli considerati possono essere diversi fra loro in modo sostanziale o soltanto per alcuni parametri, ovvero caratteristiche della popolazione.

Indagine statistica

La statistica è la scienza che studia i fenomeni collettivi dove un **fenomeno collettivo** è un qualsiasi fatto, avvenimento o situazione costituito da un numero sufficientemente alto di fenomeni singoli simili fra loro.

Sono fenomeni collettivi, ad esempio:

- il titolo di studio degli impiegati di un'azienda
- gli incidenti stradali verificatisi in una determinata regione
- la statura degli alunni che frequentano una scuola
- le nascite avvenute in una certa città
- lo sport preferito dai ragazzi al di sotto dei 18 anni
- il voto in matematica degli studenti di una scuola

Dall'analisi di un fenomeno collettivo si ottiene una serie di informazioni che permette di comprendere e interpretare il fenomeno e, quando necessario, di fare delle previsioni che riguardano il suo evolversi nel futuro e quindi di programmare in tempo utile gli interventi opportuni.

Quando si vuole indagare su un fenomeno collettivo si effettua su di esso un'indagine statistica.

Il primo passo consiste nel definire in modo preciso la **popolazione (o collettività o universo) statistica**, cioè un insieme di elementi considerati omogenei rispetto ad una o più caratteristiche, su cui si effettua l'indagine. Ogni elemento appartenente a una popolazione è detto **unità statistica**.

In statistica per popolazione non si intende esclusivamente un insieme di persone ma anche un insieme di cose o oggetti.

Esempi di popolazione statistica:

- impiegati di un'azienda
- studenti iscritti ad un certo istituto
- malattie infettive che si sono verificate in una certa regione
- redditi degli abitanti di una determinata città

Se consideriamo ad esempio come popolazione statistica gli studenti di una scuola, al suo interno le unità statistiche (gli studenti) differiscono fra loro per una o più caratteristiche: per il sesso, l'età, il peso, la classe frequentata, la media dei voti in matematica, ecc....

Queste caratteristiche vengono chiamate **variabili statistiche (o caratteri)** ed è rispetto ad una di esse che si effettua l'indagine statistica.

Le variabili statistiche possono essere di due tipi:

- **variabili quantitative** (o numeriche) che possono essere espresse con dei numeri: ad esempio l'età, il peso, la statura...
- **variabili qualitative** che non possono essere espresse mediante numeri, come il sesso, il mezzo di trasporto utilizzato, ecc... (in questo caso sono dette anche **mutabili**)

I **caratteri quantitativi** possono a loro volta essere distinti in **discreti** (descrivibili con numeri interi) e **continui** (o quasi continui) in genere risultai di processi di misura.

Un'indagine statistica è lo studio di un fenomeno collettivo, che consiste nell'analizzare come una popolazione statistica si distribuisce rispetto a una certa variabile statistica.

Supponiamo di volere indagare sulle caratteristiche della popolazione della scuola alla quale si è iscritti (media dei voti in matematica), ciascuno studente è un componente dell'insieme degli alunni della scuola.

Nel nostro caso la collettività statistica o popolazione è l'insieme degli studenti, la caratteristica comune è che frequentano la stessa scuola ed hanno voti in matematica e per unità statistica intendiamo il singolo studente.

Fissato l'obiettivo dell'indagine statistica, cioè la media dei voti in matematica degli studenti, si può effettuare un'indagine statistica totale, o completa, se la scuola non è eccessivamente grande, cioè intervistare ciascuno studente, altrimenti si può effettuare un'indagine per campione cioè considerando solo una parte della popolazione.

Una volta fissato l'obiettivo si passa alla fase di raccolta dati che avviene generalmente tramite un questionario oppure attraverso la consultazione di fonti e documenti di varia natura. Nel nostro esempio l'indagine statistica viene effettuata o prendendo come fonti i dati in possesso della segreteria della scuola, o intervistando direttamente gli studenti oppure con la compilazione di un questionario sempre da parte degli studenti.

Come già abbiamo accennato prima, nella raccolta dei dati si distinguono due diverse modalità di rilevamento:

- in relazione all'ampiezza della popolazione, per cui la rilevazione può essere **completa o parziale**
- rispetto all'intervallo di tempo durante il quale effettuare la rilevazione, per cui può essere **continua o periodica o limitata**

Rilevazioni *complete e continue* sono, per esempio, quelle relative alle nascite, alle morti, alla quantità di merci importate o esportate, alle malattie infettive, ecc.

Rilevazioni *complete e periodiche* sono invece quelle relative ai *censimenti* generali della popolazione e delle abitazioni, dell'industria e del commercio, dell'agricoltura. Tali indagini sono periodicamente effettuate dall'Istituto Centrale di Statistica (ISTAT), il censimento della popolazione viene effettuato con periodicità decennale dal 1961 ad oggi, nel mese di ottobre.

Lo scopo generale dei censimenti è quello di avere una descrizione delle caratteristiche generali di una popolazione.

Tornando al nostro esempio, abbiamo deciso di analizzare la media dei voti in matematica della collettività studenti, ciascuna entità statistica ha un voto in matematica, questa caratteristica presa in considerazione abbiamo già visto che è detta **carattere**, **ciascun carattere si manifesta con diverse modalità**, il carattere "voto in matematica" ha come modalità i numeri compresi tra 1 e 10, se scegliessimo come carattere da studiare "il colore degli occhi degli studenti" questo avrebbe come modalità: neri, castani, azzurri,..., quindi la modalità è l'insieme di valori che può assumere un determinato carattere.

Supponiamo di voler studiare semplicemente il modo in cui si presentano le modalità del carattere "voto in matematica" nella singola unità statistica studente, questa è detta **distribuzione statistica semplice**.

DEFINIZIONE

Si chiama **distribuzione statistica semplice** l'insieme dei dati relativi alle diverse modalità di un singolo carattere

Rappresentazione di una distribuzione statistica semplice

La tabella seguente rappresenta una distribuzione statistica semplice sulla collettività studenti relativa al carattere "voto in matematica".

Matricola studente	1	2	3	4	5	6	7
Voto	7	6	3	6	8	5	6

La distribuzione statistica semplice riproduce in maniera precisa le caratteristiche del fenomeno preso in esame, non dà però una informazione precisa sulle singole modalità. **Ad esempio nella collettività studenti può essere significativo studiare in relazione al carattere "voto in matematica" il numero di volte che la modalità voto =6 compare, dalla tabella dell'esempio questa informazione si può ottenere solo analizzando ogni singola unità statistica. Quando le unità che formano una collettività statistica diventano un numero considerevole, questa operazione diventa quasi impossibile.**

Per ottenere una visione precisa sulle singole modalità di un carattere preso in esame è bene calcolare le frequenze con cui ciascuna modalità di un carattere si è presentata.

DEFINIZIONI

Si chiama **frequenza assoluta** di una modalità il numero di volte che tale modalità compare nel collettivo osservato.

Si chiama **frequenza relativa** di una modalità il rapporto tra la sua frequenza assoluta ed il numero totale di unità statistiche del collettivo osservato. Se il valore viene rapportato a 100, allora si ha la **frequenza relativa percentuale**.

Si chiama **distribuzione di frequenze** l'insieme delle coppie ordinate il cui primo elemento corrisponde alla modalità e il secondo elemento alla sua frequenza, assoluta o relativa.

Quindi la frequenza assoluta ci dice il numero di volte in cui si è presentata una certa situazione nella collettività presa in considerazione.

Nell'esempio precedente, che frequenza assoluta ha la modalità voto = 6 relativa al carattere "voto in matematica"?

Rappresentazione di una distribuzione di frequenze

Data la seguente vogliamo associare alla distribuzione di frequenze assolute la relativa distribuzione di frequenze percentuali e dare una rappresentazione grafica adeguata dei dati.

Scuola XXXX, anno scolastico 2006-2007. Studenti iscritti	
Classi prime	187
Classi seconde	214
Classi terze	225
Classi quarte	176
Classi quinte	182

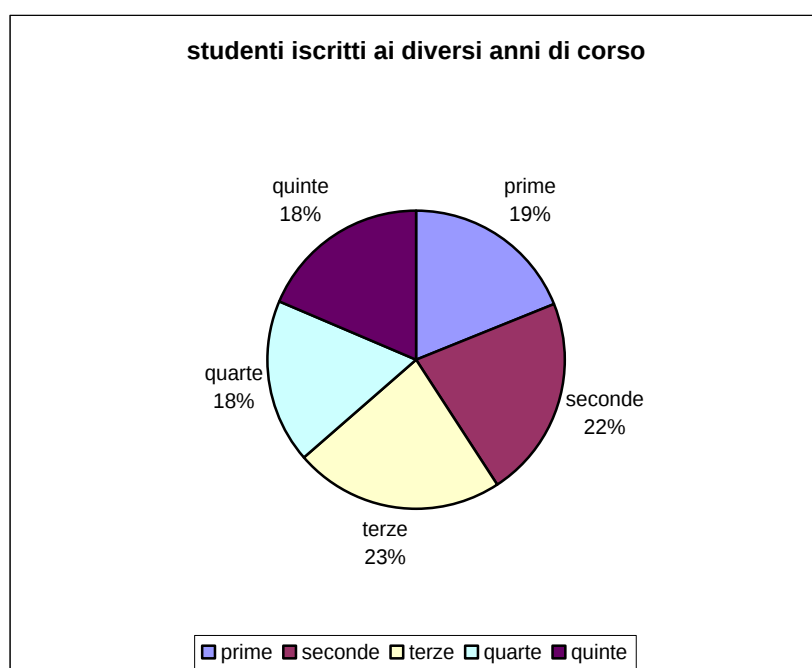
Le due colonne rappresentano rispettivamente le modalità assunte dalla mutabile classe di iscrizione e le frequenze assolute; in questo caso il numero di studenti iscritti a ciascuna classe. Per avere le frequenze percentuali calcoliamo il totale degli iscritti

Scuola XXXX, anno scolastico 2006-2007. Studenti iscritti		
	Frequenze	
	assolute	percentuali
Classi prime	187	19.00%
Classi seconde	214	21.75%
Classi terze	225	22.87%
Classi quarte	176	17.89%
Classi quinte	182	18.50%
Totale	984	100.01%

Il calcolo delle percentuali porta spesso ad arrotondare i risultati, così è possibile che la somma di tutte le percentuali non sia esattamente uguale a 100.

Il diagramma circolare (o a torta) può essere un'utile rappresentazione in questo esempio. Le frequenze percentuali sono rappresentate da settori circolari con ampiezze proporzionali alle frequenze stesse.

Il **Diagramma circolare** è un metodo comunemente utilizzato in statistica descrittiva per rappresentazioni grafiche di variabili quantitative misurate su classi di categorie, al fine di evitare di stabilire, anche involontariamente, un ordine che non esiste nelle categorie, cosa che accadrebbe infatti con l'Istogramma.



Dalle definizioni date derivano le seguenti proprietà per una distribuzione statistica:

- la somma delle frequenze assolute è uguale al numero n di unità statistiche considerate
- la somma delle frequenze relative, espresse in percentuale, è uguale a 100%.

Esempio

Data la seguente tabella dei voti riportati dagli studenti delle classi terze di una scuola, determiniamo le distribuzioni di frequenze percentuali e rappresentiamole graficamente

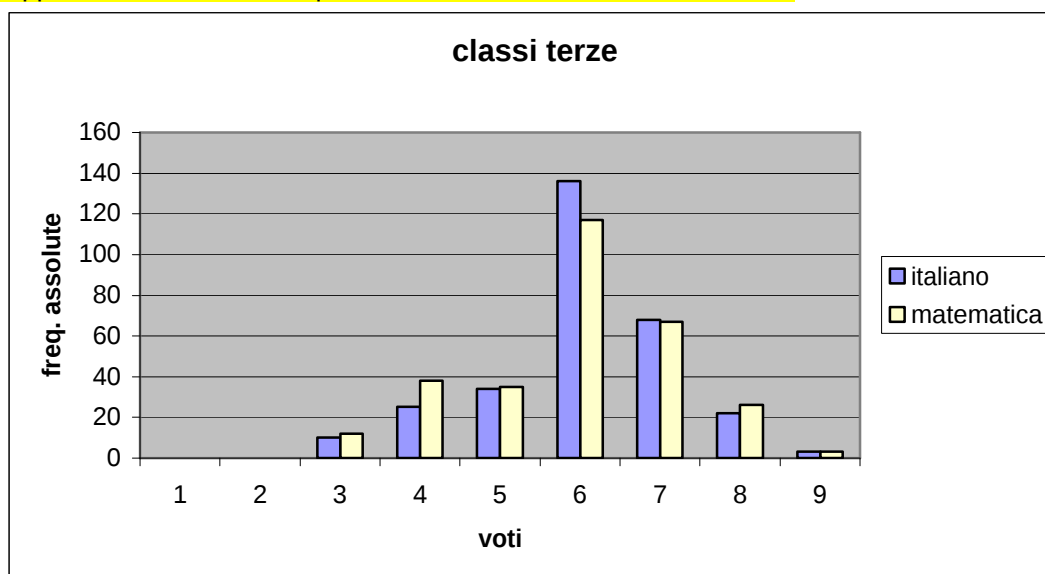
materia	Voti dello scrutinio finale									
	1	2	3	4	5	6	7	8	9	10
italiano	0	0	10	25	34	136	68	22	3	0
matematica	0	0	12	38	35	117	67	26	3	0

Nella tabella presenta due distribuzioni di frequenze assolute, relative ai caratteri “voto in italiano” e “voto in matematica”.

Considerando solo le modalità realmente assunte da ciascun carattere, si ottiene la seguente tabella:

voto dello scrutinio finale	Studenti che hanno riportato tale voto			
	Italiano		matematica	
	Freq. assoluta	Freq. perc.	Freq. assoluta	Freq. perc.
3	10	3.36	12	4.03
4	25	8.39	38	12.75
5	34	11.41	35	11.74
6	136	45.64	117	39.26
7	68	22.82	67	22.48
8	22	7.38	26	8.72
9	3	1.01	3	1.01
Totali	298	100.01	298	99.99

Per questo tipo di distribuzione il diagramma più adeguato è l'**istogramma o diagramma a barre** che permette la rappresentazione delle frequenze di entrambe le variabili statistiche.



L'area di ogni barra rettangolare è proporzionale al valore da rappresentare

Per una distribuzione di frequenza ordinata, come nel nostro esempio, potrebbe essere utile sapere quante siano le unità statistiche con modalità che non superano un determinato valore.

Ad esempio possiamo chiederci quanti studenti (unità statistiche) hanno preso in matematica un voto (variabile statistica) che non supera il 5 (data modalità).

Si può quindi introdurre, per distribuzioni statistiche ordinate rispetto alle modalità di un carattere, la seguente definizione

DEFINIZIONE

Si chiama **frequenza cumulata** di una modalità x_i di un carattere X , il numero di unità statistiche che presentano modalità minore o uguale al valore fissato x_i .

La frequenza cumulata è solitamente indicata con $F(x_i)$ o con $f(x \leq x_i)$.

Dunque si ha $F(x_i) = \sum_{j=1}^i a_j$

DEFINIZIONE

Si definisce inoltre la **frequenza cumulata %** (frequenza cumulata percentuale) associata ad una modalità o a una classe di modalità come il rapporto tra la sua frequenza cumulata e la somma di tutte le frequenze assolute presenti nel campione.

Il computo delle frequenze cumulate è utile nel calcolo di un indice che vedremo nel seguito, la mediana.

Un sinonimo (forse più utilizzato) di frequenza cumulata % è il così detto **quantile**.

Concetto inverso rispetto a quello di frequenza cumulata è quello di **frequenza retrocumulata** ove il processo di sommatoria ha inizio con la modalità più alta.

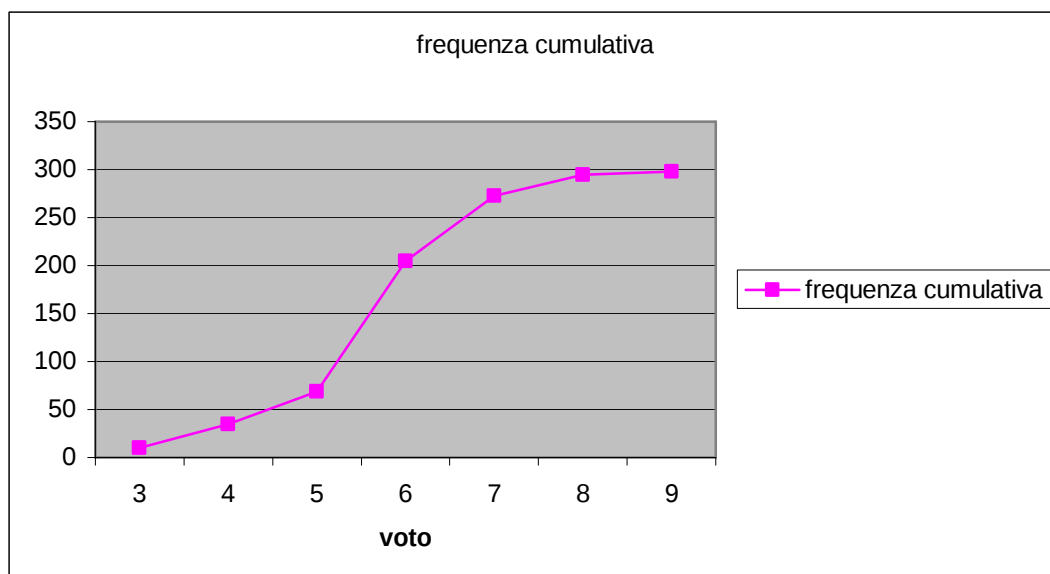
Esempio

In relazione all'esempio precedente, indicare il numero di studenti che:

1. Non hanno la sufficienza in matematica.
2. Hanno voto uguale o inferiore a 6 in italiano
3. Hanno la sufficienza in matematica

1. Frequenza cumulata relativa alla modalità "5" per matematica: $F(5) = f(x \leq 5) = 12 + 38 + 35 = 85$
2. Frequenza cumulata relativa alla modalità 6 per italiano: $F(6) = f(x \leq 6) = 10 + 25 + 34 + 136 = 205$
3. $f(x \geq 6) = N - f(x \leq 5) = 298 - 85 = 213$

Per rappresentare graficamente la frequenza cumulata può essere utile una rappresentazione mediante un poligono di frequenza cumulativa



RAGGRUPPAMENTO DEI DATI IN CLASSI

Supponiamo che da un'indagine effettuata tra gli iscritti di una società sportiva giovanile per conoscere la loro distribuzione in base all'età risultino i seguenti dati:

Età	14	15	16	17	18	19	20	21	22	23	24	25
n.ro iscritti	7	5	6	5	8	9	6	4	7	4	3	2

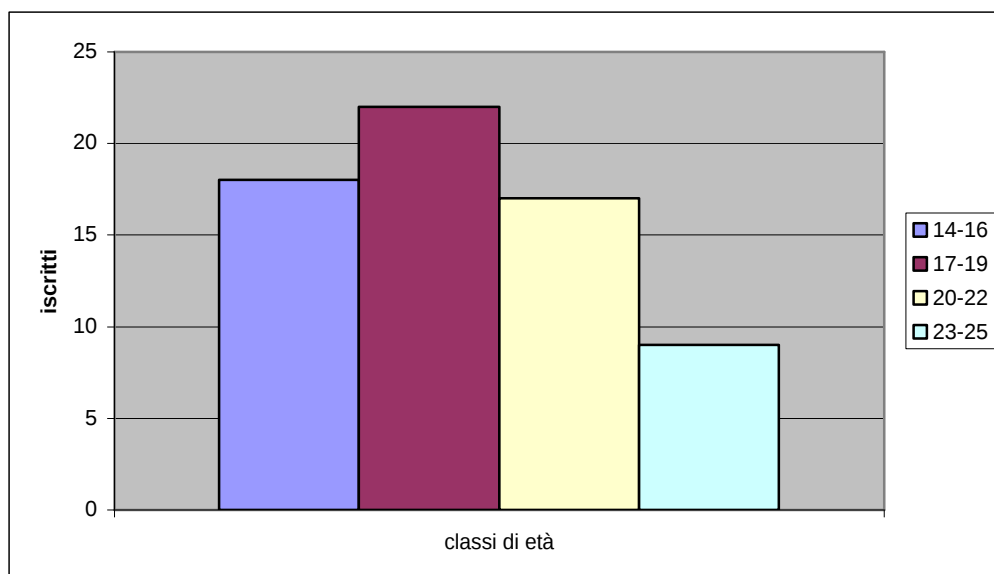
Considerando che la società vuole diversificare i suoi progetti in base alle varie fasce d'età degli iscritti, come può rappresentare i dati in modo da evidenziare la loro distribuzione (secondo questo criterio)?

Raggruppando i dati in classi che comprendano le differenti fasce d'età possiamo ottenere una prima classe che comprende le età 14, 15 e 16, quale sarà la frequenza di questa classe?

Si sommano semplicemente le frequenze delle tre modalità relative al carattere età, quindi $7+5+6 = 18$; allo stesso modo si possono ottenere altre tre classi, infine si otterrà una nuova tabella:

classi di età	14-16	17-19	20-22	23-25
n.ro iscritti	18	22	17	9

Rappresentando i dati mediante un istogramma otteniamo:



Una organizzazione dei dati in classi permette di avere una visione più sintetica dei dati anche se meno dettagliata.

Osservazione: in un istogramma le altezze devono essere direttamente proporzionali al rapporto tra frequenza ed ampiezza della classe.

Esempio (DENSITA' DI FREQUENZA)

E' possibile che in alcune situazioni le classi abbiano ampiezze diverse come nella seguente tabella:

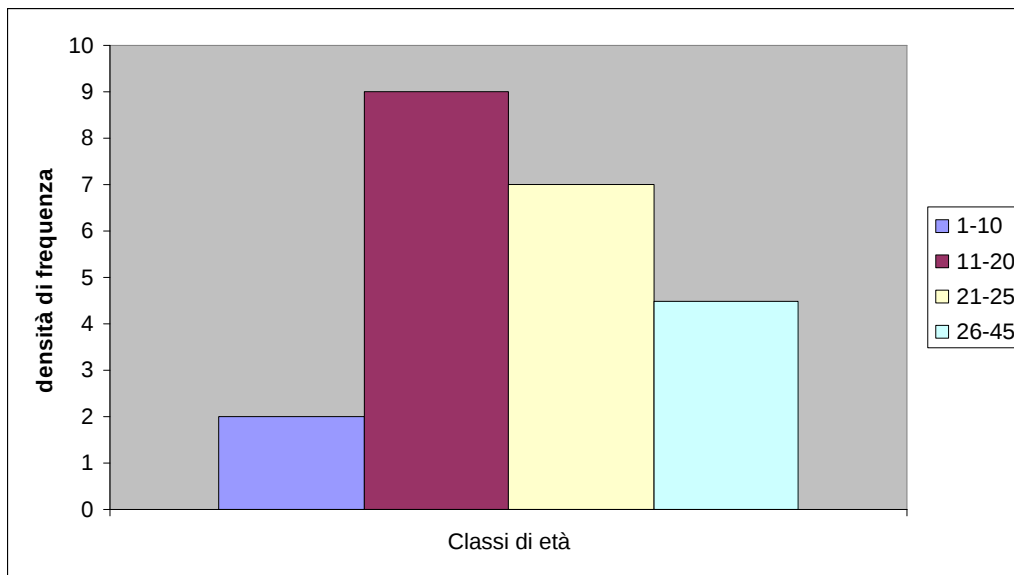
classi di età	1-10	11-20	21-25	26-45
n.ro di persone partecipanti a un viaggio	20	90	35	90

Si può osservare che anche se le frequenze delle classi 11-20 e 26-45 sono entrambe 90, forniscono informazioni diverse. Infatti, le 90 persone distribuite su una fascia di età di ampiezza 10, classe 11-20, sono percentualmente più numerose delle 90 della fascia 26-45, che ha ampiezza 20. Per poterle confrontare, e rappresentare, dobbiamo rendere le frequenze omogenee; si introduce così la densità di frequenza come il rapporto tra la frequenza assoluta della classe e l'ampiezza della classe stessa.

Otteniamo quindi la nuova tabella:

Classi di età	1-10	11-20	21-25	26-45
Frequenza	20	90	35	90
Ampiezza	10	10	5	20
Densità di frequenza	2	9	7	4.5

Costruiamo un istogramma in cui ogni rettangolo ha come base l'ampiezza della classe e come altezza la sua densità di frequenza:



Calcoliamo le aree dei rettangoli rappresentati: sono, rispettivamente, 20, 90, 35, 90. L'area è il prodotto tra ampiezza e densità:

$$\text{ampiezza} \cdot \frac{\text{frequenza}}{\text{ampiezza}} = \text{frequenza}$$

L'area di ciascun rettangolo dà la frequenza della corrispondente classe.

INDICATORI STATISTICI

L'ultima fase di un'indagine statistica è la cosiddetta elaborazione matematica dei dati disponibili. Lo scopo è di ricavare dai dati raccolti tutte le informazioni che possono fornire.

Per comprendere caratteristiche e natura di un fenomeno è utile disporre di indicatori in grado di fornire, con immediatezza e sinteticità, informazioni fondamentali sulle proprietà della sua distribuzione statistica. Le funzioni più importanti che devono essere assolve da tali indicatori sono:

- Limitare gli effetti anomali dovuti agli errori nelle misurazioni e nelle rilevazioni
- Consentire di confrontare distribuzioni diverse di uno stesso fenomeno, riferite a popolazioni o campioni differenti
- Consentire di seguire come varia il fenomeno nel tempo.

INDICI DI POSIZIONE CENTRALE

Il problema di limitare gli effetti dell'errore è di fondamentale importanza nell'elaborazione di dati sperimentali. Uno dei modi per ottenere tale risultato è quello di effettuare più misurazioni della stessa grandezza ed assumere come stima del valore di quella grandezza la media aritmetica delle misure effettuate

A seconda che una modalità di un carattere si presenti una o più volte nella distribuzione, si definisce una media aritmetica semplice o ponderata.

DEFINIZIONI

Si chiama **media aritmetica semplice** di una distribuzione statistica quantitativa, il rapporto tra la somma di tutti i valori assunti dalla variabile X e il numero totale n delle unità statistiche prese in considerazione:

$$M(X) = \bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{\sum_{i=1}^n x_i}{n}$$

Si chiama **media aritmetica ponderata** di una distribuzione statistica quantitativa, il rapporto tra la somma dei prodotti di ciascun valore della variabile X per la rispettiva frequenza e la somma totale delle frequenze:

$$M(X) = \bar{x} = \frac{x_1 \cdot a_1 + x_2 \cdot a_2 + \dots + x_k \cdot a_k}{a_1 + a_2 + \dots + a_k} = \frac{\sum_{i=1}^k x_i \cdot a_i}{\sum_{i=1}^k a_i}$$

Dove k è il numero delle diverse modalità della variabile X.

Osservazioni:

- la media aritmetica è un indice di sintesi che riassume, in un valore unico, le diverse modalità di una variabile statistica. **Rappresenta il valore costante che dovrebbe avere ciascuno dei numeri x_i perché la somma complessiva dei valori resti invariata.**

La paga mensile media di uno studente, ad esempio, è la somma che ciascuno studente riceverebbe dai propri genitori se tutti gli studenti avessero la stessa paga. Ci sono però alcuni studenti che ricevono più della paga media ed altri che ricevono meno.

- La media aritmetica viene usata in tutti quei casi in cui deve rimanere costante la somma dei numeri $x_1, x_2, \dots, x_n$ indipendentemente dal modo con cui questa somma viene ripartita in n parti.

- Il **numero M è effettivamente un valore medio**, cioè compreso fra il minimo ed il massimo di $x_1, x_2, \dots, x_n$. Infatti se disponiamo i valori sopra indicati in ordine non decrescente si ha:

$$x_1 \leq x_2 \leq \dots \leq x_n$$

Ovviamente si ha

$$\underbrace{x_1 + x_1 + \dots + x_1}_{n \text{ volte}} \leq x_1 + x_2 + \dots + x_n \leq \underbrace{x_n + x_n + \dots + x_n}_{n \text{ volte}} \text{ ovvero}$$

$$n \cdot x_1 \leq n \cdot M \leq n \cdot x_n \text{ e quindi } x_1 \leq M \leq x_n$$

- La media aritmetica è un valore sintetico che **permette di effettuare confronti con medie di altre distribuzioni espresse nella stessa unità di misura**. Pertanto sono possibili affermazioni del tipo: "il salario medio della ditta A è maggiore di quello della ditta B", "la statura media degli abitanti della città X è minore di quella degli abitanti della città Y"
- L'aggettivo aritmetica è giustificato dal fatto che la media aritmetica di un numero dispari di valori in **progressione aritmetica è proprio il numero che occupa la posizione centrale**.

Ad esempio la media aritmetica dei numeri 1,3,5,7,9,11,13 è $M = \frac{1+3+5+7+9+11+13}{7} = 7$

- La media aritmetica è influenzata dai valori estremi.**

Ad esempio:

la media aritmetica dei valori 10,12,14 è 12

la media aritmetica dei valori 10,12,50 è 24

la media aritmetica dei valori 1,12,14 è 9

Pertanto un valore molto grande o molto piccolo rispetto agli altri cambia notevolmente il valore della media aritmetica

- Molto spesso invece del termine media aritmetica si utilizza solo media.

Esempio

La seguente tabella riporta i voti in matematica degli studenti della classe IV E, calcolare il voto medio della classe

Voto	1	2	3	4	5	6	7	8	9	10
Numero studenti	0	0	3	5	2	8	5	1	1	0

$$M(X) = \frac{3 \cdot 3 + 4 \cdot 5 + 5 \cdot 2 + 6 \cdot 8 + 7 \cdot 5 + 8 \cdot 1 + 9 \cdot 1}{3 + 5 + 2 + 8 + 5 + 1 + 1} = 5.56$$

I voti finali di una disciplina sono una distribuzione di frequenze, in cui alcuni voti sono generalmente attribuiti a più studenti. Il voto medio è perciò dato dalla media ponderata.

Osservazione:

La media ci dà un'informazione generale sulla classe, infatti la divide in due sottoinsiemi di diversa numerosità, l'insieme degli studenti che hanno un voto minore o uguale alla media (10 studenti) e quello degli studenti che hanno invece un punteggio maggiore (15 studenti).

TEOREMA (PROPRIETA' DI LINEARITA')

Se X è una variabile statistica e $p, q \in \mathbb{R}$, allora si ha:

$$M(p \cdot X + q) = p \cdot M(X) + q$$

Dimostrazione

Dette x_i le modalità della variabile statistica X , a_i le frequenze di ciascuna modalità e k il numero totale delle modalità assunte da X , abbiamo:

$$M(pX) = \frac{\sum_{i=1}^k p x_i a_i}{\sum_{i=1}^k a_i} = p \frac{\sum_{i=1}^k x_i a_i}{\sum_{i=1}^k a_i} = p \cdot M(X)$$

$$M(X + q) = \frac{\sum_{i=1}^k (x_i + q) a_i}{\sum_{i=1}^k a_i} = \frac{\sum_{i=1}^k x_i a_i}{\sum_{i=1}^k a_i} + \frac{\sum_{i=1}^k q a_i}{\sum_{i=1}^k a_i} = M(X) + q$$

Da queste relazioni si ricava la proprietà di linearità della media. ♦

DEFINIZIONE

Data la distribuzione di n numeri $x_1, x_2, \dots, x_n$ e detto M il valore della loro media aritmetica, si chiamano **scarti o scostamenti dei numeri dati dal loro valor medio M** le differenze (positive, nulle o negative) fra ciascuno dei detti numeri ed il loro valor medio M .

TEOREMA (PROPRIETA' DEGLI SCARTI)

La somma algebrica degli scarti dalla media aritmetica è nulla.

Dimostrazione

Si ha

$$\begin{aligned} (x_1 - M) + (x_2 - M) + \dots + (x_n - M) &= (x_1 + x_2 + \dots + x_n) - nM = \\ &= (x_1 + x_2 + \dots + x_n) - n \frac{x_1 + x_2 + \dots + x_n}{n} = (x_1 + x_2 + \dots + x_n) - (x_1 + x_2 + \dots + x_n) = 0 \end{aligned}$$

♦

DEFINIZIONE

Data la distribuzione di n valori positivi: $x_1, x_2, \dots, x_n$ di un carattere quantitativo, si dice **media geometrica semplice G** (oppure M_g) di tali valori, la radice n^{ma} aritmetica del loro prodotto.

$$G = \sqrt[n]{x_1 \cdot x_2 \cdot \dots \cdot x_n}$$

In particolare la media geometrica di due numeri positivi a e b è data dalla radice quadrata del loro prodotto:

$$G = \sqrt{a \cdot b}$$

Se poi degli n numeri x_1 è ripetuto f_1 volte, x_2 è ripetuto f_2 volte, ..., x_k è ripetuto f_k volte con $\sum f_i = n$, allora la media geometrica della variabile statistica può mettersi sotto la forma:

$G = \sqrt[n]{x_1^{f_1} \cdot x_2^{f_2} \cdot \dots \cdot x_k^{f_k}}$ che si chiama anche **media geometrica ponderata** dei k valori $x_1, x_2, \dots, x_k$ di pesi rispettivi $f_1, f_2, \dots, f_k$

Osservazioni:

- G rappresenta il valore che dovrebbe possedere ciascuno degli n numeri $x_1, x_2, \dots, x_n$ affinché rimanga immutato il loro prodotto
- La media geometrica viene quindi usata soprattutto quando i diversi valori vengono per loro natura moltiplicati tra di loro e non sommati. **Esempio tipico sono i tassi di crescita** (anche i tassi d'interesse o i tassi d'inflazione), adeguatamente modificati. In questi casi è più corretto usare questo tipo di media al posto di quella aritmetica, perché ha caratteristiche utili in quelle situazioni.
- Il principale limite è che non si possono usare valori negativi. Una caratteristica è che **valori piccoli (rispetto alla media aritmetica) sono molto più importanti che valori grandi**. In particolare, è sufficiente la presenza di un unico valore nullo, per rendere nulla la media, sia quella semplice che quella ponderata.
- **L'aggettivo geometrica è giustificato dal fatto che, se calcolata per un numero dispari di termini di una progressione geometrica, è uguale al termine della progressione che occupa la posizione centrale.**

Ad esempio la media dei seguenti termini in progressione geometrica : 32,16,8,4,2 è
 $G = \sqrt[5]{32 \cdot 16 \cdot 8 \cdot 4 \cdot 2} = \sqrt[5]{32768} = 8$

- Come la media aritmetica anche quella geometrica **è effettivamente un valore medio**, compreso cioè fra il minimo ed il massimo valore della distribuzione data.

Infatti supposto $x_1 \leq x_2 \leq \dots \leq x_n$ si ha $\underbrace{x_1 \cdot x_1 \cdot \dots \cdot x_1}_{n \text{ volte}} \leq x_1 \cdot x_2 \cdot \dots \cdot x_n \leq \underbrace{x_n \cdot x_n \cdot \dots \cdot x_n}_{n \text{ volte}}$ da cui per definizione di

media geometrica $x_1^n \leq G^n \leq x_n^n$ ed estraendo le radici ennesime $x_1 \leq G \leq x_n$

- **Il reciproco della media geometrica di più numeri è uguale alla media geometrica dei loro reciproci.**

$$\text{Infatti } \sqrt[n]{\frac{1}{x_1} \cdot \frac{1}{x_2} \cdot \dots \cdot \frac{1}{x_n}} = \frac{1}{\sqrt[n]{x_1 \cdot x_2 \cdot \dots \cdot x_n}}$$

n.b.: questa proprietà non vale per la media aritmetica.

Esempio

Tra il 1997 e il 2001 sono stati rilevati i seguenti tassi d'inflazione: 3,2% per il 1997, 2,7% (1998), 2,8% (1999), 2,2% (2000) e 3,2% (2001).

Anno	1997	1998	1999	2000	2001
Tasso	3,2%	2,7%	2,8%	2,2%	3,2%

Per trovare il tasso medio tra il 1997 ed il 2001 calcoliamo la media geometrica

Trattandosi di valori relativi e percentuali, li trasformiamo anzitutto dividendo con 100 e poi sommando loro 1. Otteniamo così per gli ultimi cinque anni dei fattori di moltiplicazione pari a: 1,032 1,027 1,028 1,022 1,032.

Moltiplicando tra di loro questi cinque valori otteniamo

$$\prod x_i = 1,149142$$

Estraendo la radice quinta, si ottiene

$$G = \sqrt[5]{1,149142} = 1,028193$$

DEFINIZIONE

Data la distribuzione di n valori positivi: $x_1, x_2, \dots, x_n$ di un carattere quantitativo, si dice **media armonica semplice** H di tali valori, il reciproco della media aritmetica dei reciproci dei valori dati

$$H = \frac{n}{\frac{1}{x_1} + \dots + \frac{1}{x_n}} \quad \text{cioè} \quad H = \frac{n}{\sum_{i=1}^n \frac{1}{x_i}}$$

Si chiama media armonica ponderata dei k numeri $H = \frac{n}{\frac{f_1}{x_1} + \dots + \frac{f_k}{x_k}}$, dove $f_1, \dots, f_k$ sono le frequenze di

$x_1, x_2, \dots, x_k$

Esempio

Un'automobile percorre un primo silometro a 40km/h e un secondo a 60 km/h . Calcolare la velocità media dell'intero percorso.

L'automobile ha percorso complessivamente 2 km dei quali il primo ha richiesto 1/40 di ora, ossia 60/40=1,5 minuti, mentre il secondo silometro ha richiesto 1/60 di ora cioè 1 minuto. In totale quindi l'automobile ha impiegato 1,5+1=2,1 minuti, che riferiti ad un percorso di 2km danno una velocità media

$$v_m = \frac{s}{t} = \frac{2}{2,5/60} = 48 \text{ km/h}$$

Questa v media è data in funzione delle due velocità parziali da $v_m = \frac{2}{\frac{1}{v_1} + \frac{1}{v_2}}$ che è la media armonica delle due

velocità parziali.

Osservazioni

- H rappresenta il valore che dovrebbe possedere ciascuno degli n numeri affinché resti immutata la somma dei loro reciproci.
- Anche la media armonica è un valore medio, compreso fra minimo e massimo.
- Si dice armonica perché quando viene applicata a una progressione armonica (costituita da un numero dispari di termini) fornisce il valore centrale della progressione stessa .

Abbiamo osservato relativamente alla media aritmetica che questa può essere fortemente influenzata dai valori estremi. Se vi è una grande differenza tra il valore minimo ed il valore massimo della distribuzione può essere più conveniente chiedersi quale sia il valore con la frequenza maggiore piuttosto che quale sia il valore medio.

DEFINIZIONI

Si chiama **moda** (o valore normale) di una distribuzione quantitativa la modalità di maggiore frequenza di un carattere.

Si chiama **mediana** di una distribuzione statistica quantitativa ordinata, la modalità che separa gli elementi della distribuzione in due insiemi di uguale numerosità.

Se il numero delle modalità è dispari la mediana corrisponde alla modalità che occupa la posizione centrale; altrimenti si assume come mediana la media aritmetica delle modalità centrali.

Esempio

Riprendendo un esempio precedente relativo ai voti degli studenti di una classe calcoliamone moda e mediana.

Voto	1	2	3	4	5	6	7	8	9	10
Numero studenti	0	0	3	5	2	8	5	1	1	0

La moda della distribuzione precedente è il voto 6 che ha frequenza 8.

La mediana è 6, poiché il numero di studenti è 25 e la mediana è la modalità relativa al tredicesimo studente.

Esempio

La distribuzione delle abitazioni, costruite in una certa regione nel 2004 secondo il numero delle stanze è la seguente:

Numero delle stanze	Numero delle abitazioni (frequenza)
1	6.005
2	17.437
3	36.243
4	65.623
5	54.324
6	35.233
7	14.247

La moda è 4. Possiamo quindi dire che in quella regione le abitazioni costruite in quell'anno hanno "normalmente" 4 stanze

Osservazioni:

- In molti fenomeni collettivi può essere più espressivo il "valore normale" che altri valori medi.

Così ad esempio è più espressivo il salario normale dei lavoratori di una ditta che non il salario medio, perché quest'ultimo può essere pesantemente influenzato da retribuzioni molto alte di un piccolo numero di lavoratori della ditta. In altre parole il salario normale dà un'idea immediata della retribuzione che normalmente percepiscono quei lavoratori.

- La moda può non esistere e, anche se esiste, può non essere unica.

Per esempio l'a distribuzione 3,5,7,9,11,13 non ha moda

La distribuzione 2,3,4,4,4,5,5,7,7,7,9 ha due mode : 4 e 7. In questi casi la distribuzione viene detta **bimodale**. Una distribuzione dotata di una sola moda si dice invece **unimodale**.

La mediana come abbiamo visto rappresenta il valore centrale di una distribuzione di modalità di un carattere quantitativo perché lascia il 50% dei dati alla sua sinistra e l'altro 50% alla sua destra. Oltre alla mediana si usano in statistica talora altri valori medi posizionali che ripartiscono il numero dei dati in della distribuzione in un determinato numero di parti uguali. Per esempio i valori Q_1 , Q_2 , Q_3 che dividono la distribuzione in 4 parti uguali sono detti primo, secondo e terzo **quartile**, il valore Q_2 è la mediana. Analogamente i valori che dividono i dati in 10 parti uguali sono detti **decili** e si indicano con D_1 , .. D_9 . I valori che dividono in 100 parti uguali si dicono **percentili** e si indicano con P_1 , ..., P_{99} .

Esempio (calcolo della mediana per distribuzioni di dati raggruppati in classi)

Nel caso in cui i dati forniti siano aggregati in classi, la determinazione delle mediane non è immediata. Consideriamo la seguente distribuzione di persone in base all'altezza (espressa in centimetri):

classi di altezze	n.ro di persone	Frequenze cumulate
150-160	15	15
160-170	40	55
170-180	25	80
180-190	14	94
190-200	6	100

Poiché le persone esaminate sono complessivamente 100, per determinare l'altezza mediana dobbiamo contare le prime $100/2=50$ persone disposte per altezze crescenti.

Leggiamo dalla tabella che la 50-esima persona si trova nella classe 160-170, quindi questa è la classe mediana. A noi interessa sapere qual'è l'altezza mediana che sarà un valore del tipo $160+x$, con x una parte dell'ampiezza 10 della classe mediana.

Per calcolare x , supponiamo che le persone si distribuiscano nella classe 160-170 in maniera uniforme, vale la proporzione $x:10=35:40$, dove 40 è la differenza tra la frequenza cumulata della classe mediana (55) e quella della classe precedente (15), mentre 35 è la differenza tra il valore mediano (50) e la frequenza cumulata della classe precedente la classe mediana (15).

Dalla proporzione abbiamo:

$$x = \frac{10 \cdot 35}{40} = 8.75$$

quindi l'altezza mediana è $160+8.75=168.75$.

INDICI DI VARIABILITA'

Dopo aver introdotto gli indici di posizione osserviamo che però essi non sempre danno un'informazione esauriente per esprimere sinteticamente il fenomeno oggetto di studio.

Il poeta romano Carlo Alberto Salustri (1871 - 1950), noto con lo pseudonimo - anagramma del proprio cognome - di Trilussa, è spesso ricordato per l'aforisma del pollo: *Dalli conti che se fanno, risulta che ce tocca 'n pollo a testa. Ma si nun centra nelle spese tue, c'entra nella statistica lo stesso perché c'è 'nartro che ne magna due*

Distribuzioni che assumono lo stesso valore medio in realtà possono essere molto diverse, consideriamo le seguenti distribuzioni che hanno tutte media uguale a 10:

A={10 10 10 10 10}

B={2 5 10 15 18}

C={8 9 10 11 12}

Per una descrizione più completa è necessario studiare come variano, come si disperdono, cioè come si sparpagliano i singoli valori attorno alla media considerata.

A questo scopo sono stati introdotti gli indici di variabilità.

I principali indici di variabilità sono:

il campo di variazione

lo scarto semplice medio

lo scarto quadratico medio e la varianza

DEFINIZIONE

Si chiama **campo di variabilità** (o escursione) di un insieme di n valori, la differenza tra il valore massimo ed il valore minimo.

Il campo di variabilità è una **misura grossolana della dispersione** perché su di esso non influiscono i valori intermedi della distribuzione. Ha però il pregio di essere di **facile calcolo** e particolarmente **utile nei piccoli campioni**.

Questa misura della variabilità trova vaste applicazioni nel controllo statistico di qualità, in meteorologia, in campo economico.

DEFINIZIONE

Dato un insieme di valori $x_1, x_2, \dots, x_n$ e detta M la loro media aritmetica si chiama **scarto semplice medio** dalla media aritmetica la media dei valori assoluti degli scarti dalla loro media aritmetica.

Detto S_M lo scarto semplice medio si ha per definizione

$$S_M = \frac{|x_1 - M| + \dots + |x_n - M|}{n} \text{ se i dati sono distribuiti con frequenze } f_1, f_2, \dots, f_k \text{ si ha}$$

$$S_M = \frac{|x_1 - M| f_1 + \dots + |x_k - M| f_k}{f_1 + \dots + f_k}, \text{ dove } M \text{ è la media ponderata.}$$

Osservazioni:

- Mentre, come abbiamo già visto, la somma degli scarti relativi è sempre 0, la **somma degli scarti assoluti è 0 se e solo se i dati sono tutti uguali tra loro**
- **tanto più lo scarto medio è piccolo tanto più i valori osservati si addensano attorno alla loro media aritmetica.**
- Questo indice è poco usato per via della **scomodità** di trattare i valori assoluti dal punto di vista matematico

DEFINIZIONI

Si chiama **varianza** di una distribuzione, la media aritmetica dei quadrati degli scarti dalla media, e si indica con σ^2 oppure con $\text{VAR}(X)$.

Si chiama **scarto quadratico medio** di una distribuzione, la radice quadrata della varianza e si indica con σ

Si ha quindi $\sigma^2 = \frac{(x_1 - M)^2 + (x_2 - M)^2 + \dots + (x_n - M)^2}{n}$ oppure considerando la media ponderata

$$VAR(X) = \sigma^2 = \frac{\sum_{i=1}^k (x_i - M)^2 \cdot f_i}{\sum_{i=1}^k f_i}$$

Per lo scarto quadratico:

$$\sigma = \sqrt{\frac{(x_1 - M)^2 + (x_2 - M)^2 + \dots + (x_n - M)^2}{n}} \text{ per dati semplici,}$$

$$\sigma = \sqrt{\frac{\sum_{i=1}^k (x_i - M)^2 \cdot f_i}{\sum_{i=1}^k f_i}} \text{ per dati ponderati}$$

Osservazioni:

- La varianza è un numero positivo essendo la media di numeri al quadrato.
- Quando si usa la varianza si opera sui quadrati degli scarti e questo può alterare certi aspetti. Ciò risulta evidente se i dati hanno una certa unità di misura, allora la varianza risulta espressa nel quadrato della unità di misura. E' per questo che si è introdotto anche lo scarto quadratico medio.
- La varianza può essere anche espressa in funzione dei dati originari e non degli scarti, dalla proprietà di linearità della media abbiamo:

$$\sigma^2 = {}_2M - M^2 \text{ dove abbiamo indicato con } {}_2M \text{ la media aritmetica dei quadrati dei dati}$$

Lo scarto quadratico medio sarà quindi:

$$\sigma = \sqrt{{}_2M - M^2}$$

- Anche lo scarto quadratico medio è tanto minore quanto è maggiore l'addensamento dei dati attorno alla loro media aritmetica ed è nullo se e solo se i dati sono tutti uguali.
- Lo scarto quadratico medio è un indice di variabilità assoluta e viene espresso nella medesima unità di misura rispetto alla quale si esprimono i dati statistici. Il rapporto $\frac{\sigma}{|M|}$, fra lo scarto quadratico medio ed il valore assoluto della media aritmetica, rappresenta un indice di variabilità relativa, ed è chiamato **coefficiente di variabilità**.

Tale coefficiente è un numero puro e viene generalmente espresso in percentuale.

Non può essere utilizzato se $M=0$ o se lo scarto quadratico medio è maggiore del modulo di M .

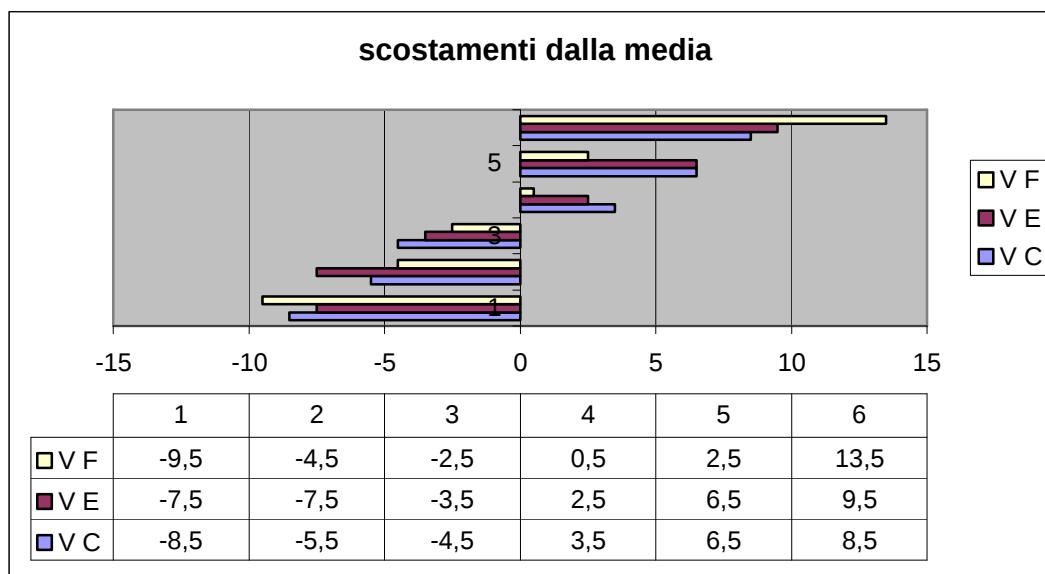
Esempio

Per la partecipazione ad una gara di matematica le scuole debbono inviare i nominativi di una squadra di 6 studenti di una stessa classe. Si attua perciò un test preliminare (punteggio max 100 punti) di selezione e sulla base della media dei 6 migliori punteggi risultano 3 squadre a pari merito. Sulla base di quale criterio può essere scelta la squadra da inviare alla gara?

Classe	Punteggi degli studenti					
V C	73	76	77	85	88	90
V E	74	74	78	84	88	91
V F	72	77	79	82	84	95

La media aritmetica è la stessa per le 3 squadre: 81,5.

Possiamo utilizzare il foglio elettronico anche per rappresentare graficamente gli scarti dalla media per ogni classe



Elaborando i dati si ottiene

Classe	Campo di variazione	Scarto quadratico medio
V C	17	6.45
V E	17	6.63
V F	23	7.14

L'analisi del solo campo di variazione indica risultati più omogenei per gli studenti delle classi V C e V E, anche se in V F esiste un risultato di rilievo (95/100). Lo scarto quadratico medio della V C è il minimo, utilizzando il criterio dello scarto quadratico medio la classe da inviare allora è la V C.

Esempio

I risultati dell'ultimo test di matematica degli studenti di una classe terza sono di seguito riportati (in trentesimi). Individua quali studenti si discostano dal punteggio medio per più di una oppure per due volte lo scarto quadratico medio.

Numero studente	1	2	3	4	5	6	7	8	9	10	11	12	13
punteggio	15	17	24	25	29	28	16	15	19	25	21	20	28

Numero studente	14	15	16	17	18	19	20	21	22	23	24	25
punteggio	21	20	19	18	13	18	22	26	27	13	15	18

Elaborando i dati con il foglio elettronico otteniamo:

n.ro	x	x_i^2	$x_i - M$
1	15	225	-5.48
2	17	289	-3.48
3	24	576	3.52
4	25	625	4.52
5	29	841	8.52
6	28	784	7.52
7	16	256	-4.48
8	15	225	-5.48
9	19	361	-1.48
10	25	625	4.52
11	21	441	0.52
12	20	400	-0.48

13	28	784	7.52
14	21	441	0.52
15	20	400	-0.48
16	19	361	-1.48
17	18	324	-2.48
18	13	169	-7.48
19	18	324	-2.48
20	22	484	1.52
21	26	676	5.52
22	27	729	6.52
23	13	169	-7.48
24	15	225	-5.48
25	18	324	-2.48
totale	512	11058	
medie	20.48	442.32	

L'elaborazione dei dati con il foglio elettronico

fornisce i seguenti valori:

$$M(X)=20.48 \quad \sigma=4.78$$

$$M(X)-\sigma=15.70 \quad M(X)+\sigma=25.26$$

$$M(X)-2\sigma=10.92 \quad M(X)+2\sigma=30.04$$

Tutti i punteggi appartengono all'intervallo

$$[M - 2\sigma, M + 2\sigma] = [10.98, 30.04], \text{ cioè non vi è}$$

alcun punteggio che si discosta dalla media per più di due volte lo scarto quadratico medio.

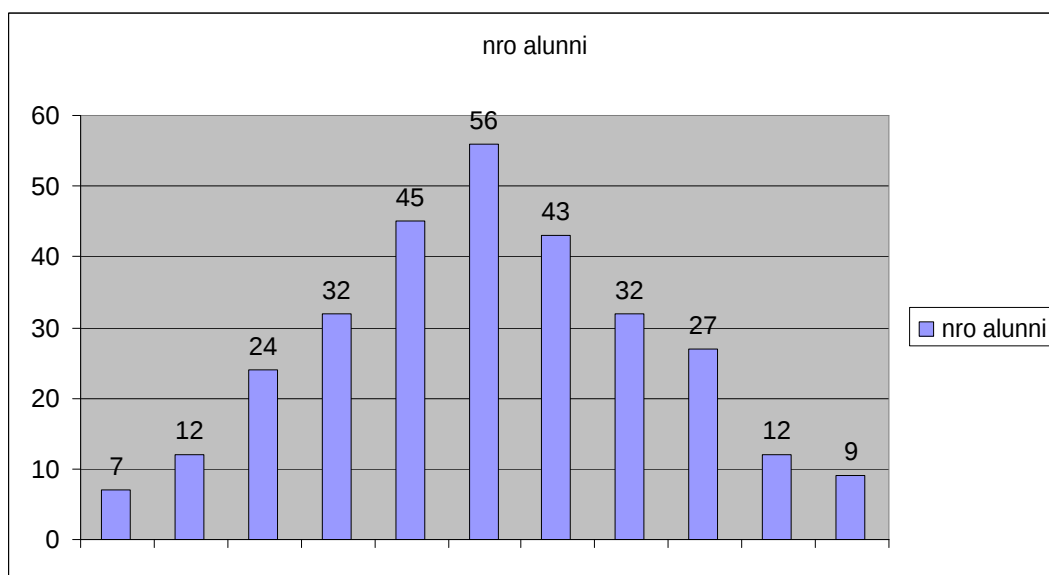
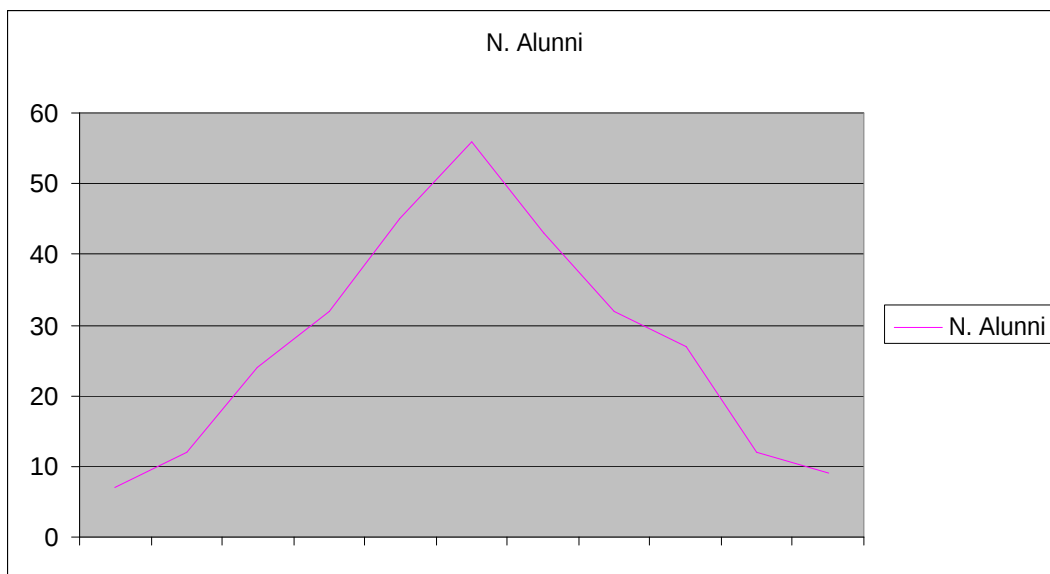
LA DISTRIBUZIONE NORMALE

Lo scarto quadratico medio risulta importante per lo studio delle distribuzioni che si traducono graficamente in diagrammi con forma campanulare e che prendono il nome di **distribuzioni normali**.

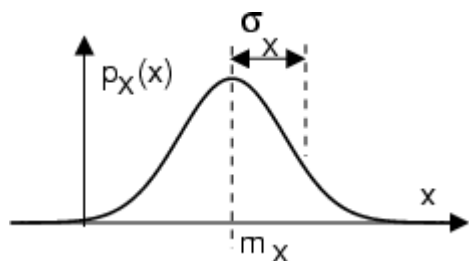
Esempio

Si esegue un'indagine sugli studenti di una scuola per rilevare il tempo impiegato da ognuno per raggiungere da casa la scuola. Si raccoglie un considerevole numero di dati intervistando, in maniera casuale, gli alunni. Si effettua lo spoglio e si raggruppano le rilevazioni in intervalli di tempo di 5 minuti ottenendo la tabella seguente:

Tempo	[0 – 5[	[5 – 10[	[10 – 15[	[15 – 20[	[20 – 25[	[25 – 30[	[30 – 35[	[35 – 40[	[40 – 45[	[45 – 50[	[50 – 55[
N. Alunni	7	12	24	32	45	56	43	32	27	12	9



Per facilitare la comprensione della curva si è supposto che i dati rilevati comportino una distribuzione quasi simmetrica, una situazione ideale che in alcuni casi è possibile raggiungere aumentando considerevolmente il numero degli intervistati. Se anziché 299 alunni si fossero intervistati tutti gli alunni è molto probabile che si sarebbe ottenuta una curva del tipo



La curva è detta **curva normale o Gaussiana**.

Questa curva risulta un insostituibile aiuto nella ricerca e nello studio di molti problemi in campo sociologico, economico, biologico, ecc.. Molte indagini portano alla scoperta che la popolazione oggetto della rilevazione si distribuisce secondo una curva normale. Invece di eseguire i calcoli sui valori della distribuzione ottenuti con la

rilevazione si effettuano allora sulla curva normale teorica. Questo permette di determinare rapidamente molti indici quali mediana, media e gli indici di variabilità. Consente inoltre di formulare leggi statistiche per l'intero universo d'indagine.

Calcoliamo la media e lo scarto quadratico medio per l'esempio precedente
Completiamo la tabella con i calcoli parziali per la ricerca degli indici

Tempo	Valore centrale a_i	Frequenze f_i	$a_i f_i$	$a_i - M$	$(a_i - M)^2$	$(a_i - M)^2 f_i$
[0-5[	2,5	8	20	-25	625	5000
[5-10[	7,5	12	90	-20	400	4800
[10-15[	12,5	26	325	-15	225	5850
[15-20[	17,5	32	560	-10	100	3200
[20-25[	22,5	44	990	-5	25	1100
[25-30[	27,5	56	1540	0	0	0
[30-35[	32,5	44	1430	5	25	1100
[35-40[	37,5	32	1200	10	100	3200
[40-45[	42,5	26	1105	15	225	5850
[45-50[	47,5	12	570	20	400	4800
[50-55[	52,5	8	420	25	625	5000
Totali		300	8250	0	2750	39900

La media vale $M=27,5$

Osservazioni

- In una distribuzione normale la media aritmetica, la mediana e la moda coincidono
- Le maggiori frequenze si addensano attorno alla media. Nell'esempio 144 alunni su 300 impiegano da casa a scuola tra i 20 e i 35 minuti.
- Scarto quadratico medio:

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (a_i - M)^2 f_i}{\sum_{i=1}^n f_i}} = \sqrt{\frac{39900}{300}} = 11,53$$

Mediamente i tempi si discostano dalla media di un valore pari a 11,53 minuti in più o in meno.

Caratteristiche di una distribuzione normale

- La curva gaussiana risulta nel primo tratto crescente, la concavità è rivolta verso l'alto, poi presenta una concavità verso il basso fino a raggiungere un punto di massimo. Il tratto decrescente è simmetrico a quello crescente.
- Il punto di massimo coincide con la media, la moda e la mediana. La probabilità che un ragazzo impieghi tra i 25 e i 30 minuti per andare da casa a scuola è maggiore di tutte le altre.
- Se una distribuzione è gaussiana si dimostra che nell'intervallo $M \pm \sigma$ sono distribuite il 65% delle unità statistiche; nell'intervallo $M \pm 2\sigma$ circa il 95% delle unità statistiche e quasi la totalità dei casi appartengono all'intervallo $M \pm 3\sigma$. Lo scarto quadratico medio è utile per giudicare se un elemento della popolazione è normale cioè se cade nell'intervallo. $M \pm \sigma$.

Nel nostro esempio si dice che è nella norma un tempo di percorrenza nell'intervallo $27,5 \pm 11,53$, cioè compreso tra 16 e 39 minuti.

Si potrebbe invece considerare fuori dalla norma un alunno che impieghi per andare da casa a scuola meno di 16 o più di 39 minuti.

- Entro l'intervallo $M - 2\sigma$ $M + 2\sigma$ cadono il 95% circa delle unità statistiche, in termini probabilistici diremo che la probabilità che una unità statistica scelta a caso appartenga a questo intervallo è del 95%.
- La curva normale è particolarmente importante per le grandezze che vengono rilevate attraverso prove ripetute. Quando si compiono rilevazioni per misurare una profondità marina, l'altezza di una montagna, la lunghezza di una strada... si commettono errori casuali che dipendono dal misuratore e da altre cause imponderabili, sicché misurando più volte lo stesso oggetto si ottengono ogni volta valori diversi. Tali "errori" non sono eliminabili. Riportando su un diagramma i risultati ottenuti dalle prove ripetute si nota che questi si distribuiscono secondo la curva gaussiana. La media individua la misura tipica più attendibile del valore cercato.

DISTRIBUZIONE STATISTICA CONGIUNTA

Per una stessa collettività è spesso interessante studiare il caso in cui ogni unità statistica è caratterizzata simultaneamente da più caratteri. Quando si considerano due soli caratteri si parla di **distribuzione statistica congiunta o doppia**.

Queste distribuzioni possono essere rappresentate da una tabella a doppia entrata.

In un esempio precedente abbiamo considerato due caratteri diversi, della stessa collettività statistica con stesse modalità, cioè la collettività studenti e le variabili statistiche “voto in matematica” e “voto in italiano”.

Vediamo come esempio la **tabella a doppia entrata** che rappresenta i dati congiunti dei voti finali in italiano e matematica della collettività studenti

Voti in matematica	Voti in italiano							
	3	4	5	6	7	8	9	tot
3	7	3	2					12
4	1	13	10	13	1			38
5	2	5	12	9	5	2		35
6		3	7	90	14	3		117
7		1	2	15	42	7		67
8			1	9	5	9	2	26
9					1	1	1	3
tot	10	25	34	136	68	22	3	298

In ciascuna cella di una tabella di distribuzione congiunta appare una frequenza, assoluta o relativa. La frequenza assoluta indica il numero delle unità che presentano contemporaneamente la modalità x_i del carattere X e la modalità y_j del carattere Y, ci sono ad esempio 3 studenti che hanno 4 in italiano e contemporaneamente 3 in matematica, incrocio linea 3 e colonna 4.

Ci si può chiedere:

Qual è la distribuzione dei voti in matematica degli studenti che hanno 6 in italiano?

Si introduce così la **distribuzione condizionata**.

DEFINIZIONE

Si chiama **distribuzione condizionata** la distribuzione che si ottiene fissando una particolare modalità per uno dei due caratteri e considerando le frequenze associate a tutte le modalità dell'altro carattere

Si chiama **distribuzione marginale** la distribuzione che si ottiene considerando le frequenze associate a uno solo dei due caratteri, indipendentemente dall'altro

Osservazioni

- La distribuzione condizionata è quella che si ottiene considerando le frequenze di una particolare riga o colonna.
- Le distribuzioni marginali sono dunque quelle dei totali per righe e per colonne: ogni tabella a doppia entrata ha così due distribuzioni marginali.

Esempio

La seguente tabella indica la distribuzione congiunta relativa alla lingua straniera studiata e alla classe di iscrizione in un liceo

Classe di iscrizione	Lingua straniera studiata			
	Inglese	Francese	Tedesco	Spagnolo
I	102	45	37	15
II	114	53	19	10
III	98	41	18	13
IV	85	44	26	9
V	88	37	21	11

Determinare le distribuzioni marginali e quella condizionata all'ipotesi: studenti iscritti alla classe terza.

La distribuzione dei totali per riga è quella marginale relativa al carattere “classe di iscrizione”. La distribuzione dei totali per colonna è quella relativa al carattere “lingua straniera studiata”.

	Inglese	Francese	Tedesco	Spagnolo	totale
I	102	45	37	15	199
II	114	53	19	10	196
III	98	41	18	13	170
IV	85	44	26	9	164
V	88	37	21	11	157
totale	487	220	121	58	886

La distribuzione condizionata ai soli studenti iscritti alla classe terza è quella corrispondente alla terza riga della tabella.

Da una tabella di frequenze assolute si possono derivare 3 diverse tabelle di frequenze relative.

Se indichiamo con $A(i,j)$ il generico elemento della tabella corrispondente alla riga i alla colonna j e con s una qualsiasi somma per riga o per colonna, abbiamo 3 possibilità:

$$\text{frequenze relative per riga} \quad \frac{A(i, j)}{s(i)}$$

$$\text{frequenze relative per colonna} \quad \frac{A(i, j)}{s(j)}$$

$$\text{frequenze relative globali} \quad \frac{A(i, j)}{N}$$

Dove N è il numero totale delle unità considerate

Se le frequenze relative sono moltiplicate per 100 allora i valori esprimono frequenze percentuali.

Dall'esempio precedente, dividendo ogni numero della tabella per il totale degli alunni, si ottiene la tabella delle frequenze relative globali.

	Tabella delle frequenze percentuali (globali)				
	Inglese	Francese	Tedesco	spagnolo	Frequenze marginali
I	11.5%	1.7%	4.2%	1.7%	22.5%
II	12.9%	6.0%	2.1%	1.1%	19.2%
III	11.1%	4.6%	2.0%	1.5%	19.2%
IV	9.6%	5.0%	2.9%	1.0%	18.5%
V	9.9%	4.2%	2.4%	1.2%	17.7%
Frequenze marginali	55.0%	24.9%	13.6%	6.5%	100.0%

Ciascuna riga della tabella rappresenta una distribuzione condizionata in cui i valori sono espressi in forma percentuale. L'ultima riga è la distribuzione marginale relativa al carattere "lingua straniera".

Una ulteriore tabella si può costruire dividendo ogni dato per il totale della propria riga.

	Tabella delle frequenze percentuali per riga				
	Inglese	Francese	Tedesco	spagnolo	Frequenze marginali
I	51.3%	22.6%	18.6%	7.5%	100%

II	58.2%	27.0%	9.7%	5.1%	100%
III	57.6%	24.1%	10.6%	7.6%	99.9%
IV	51.8%	26.8%	15.9%	5.5%	100%
V	56.1%	23.6%	13.4%	7.0%	100.1%
Frequenze marginali	55.0%	24.9%	13.6%	6.5%	100.0%

Ciascuna riga della tabella rappresenta una distribuzione condizionata in cui i valori sono espressi in forma percentuale. L'ultima riga, che peraltro coincide con l'ultima riga dell'altra tabella, è la distribuzione marginale relativa al carattere "lingua straniera studiata".

Un'altra tabella può essere costruita dividendo ciascun dato per il totale di colonna.

	Tabella delle frequenze percentuali per colonna (in percentuale)				
	Inglese	Francese	Tedesco	spagnolo	Frequenze marginali
I	20.9%	20.5%	30.6%	25.9%	22.5%
II	23.4%	24.1%	15.7%	17.2%	22.1%
III	20.1%	18.6%	14.9%	22.4%	19.2%
IV	17.5%	20.00%	21.5%	15.5%	18.5%
V	18.1%	16.8%	17.4%	19.00%	17.7%
Totali	100%	100%	100.1%	100%	100%

La distribuzione della tabella a doppia entrata rappresenta una distribuzione congiunta di due caratteri X e Y, questi possono essere tra loro indipendenti oppure avere una qualche forma di relazione.

DEFINIZIONE

Una variabile statistica X è in **assoluto indipendente** dalla variabile Y se l'insieme dei valori assunti da X non varia al variare della modalità y_i .

Quindi se tutte le distribuzioni condizionate a una specifica modalità y_i hanno la stessa distribuzione di frequenze relative, coincidente con quella marginale.

Esempio

	y_1	y_2	y_3	tot
x_1	0.25	0.25	0.25	0.25
x_2	0.35	0.35	0.35	0.35
x_3	0.40	0.40	0.40	0.40
tot	1.00	1.00	1.00	1.00

	y_1	y_2	y_3	tot
x_1	0.20	0.25	0.30	0.28
x_2	0.45	0.40	0.30	0.45
x_3	0.35	0.35	0.40	0.27
tot	1.00	1.00	1.00	1.00

Nella prima tabella il carattere X è indipendente in maniera assoluta dal carattere Y, mentre nella seconda tabella vi è una certa dipendenza tra i due caratteri in quanto le distribuzioni per colonna sono diverse le une dalle altre. La dipendenza tra i caratteri può essere più o meno accentuata, vediamo con un esempio come si può introdurre un opportuno indice che misura tale grado di dipendenza statistica.

Esempio

La tabella seguente riporta i voti finali in matematica in quattro classi terze di una stessa scuola. Vogliamo analizzare se il carattere "voto in matematica" è indipendente dal carattere "classe frequentata".

Tabella delle distribuzioni dei voti finali in matematica					
O_{ij}	IIIA	IIIB	IIIC	IIID	totali
3	1	3	4	1	9
4	6	4	5	3	18
5	3	2	3	4	12
6	10	13	9	14	46
7	5	4	5	2	16
8	1	1	0	2	4
Totali	26	27	26	26	105

Si parte dalla cosiddetta *ipotesi nulla*, per la quale si suppone che vi sia indipendenza tra i caratteri "voto finale in matematica" e sezione di iscrizione".

Costruiamo quindi una tabella delle frequenze teoriche attese a_{ij} , nel caso in cui la distribuzione dei voti fosse del tutto indipendente dalla classe frequentata.

Se i valori a_{ij} non dipendessero dalla particolare colonna allora avremmo:

$$\frac{a_{ij}}{\text{totale di riga } i} = \frac{\text{totale di colonna } j}{\text{totale generale}}$$

$$a_{ij} = \frac{(\text{totale di riga } i) \cdot (\text{totale di colonna } j)}{\text{totale generale}}$$

$$\text{frequenza teorica nella casella (3;IIIA)} = \frac{9 \cdot 26}{105} \cong 2,2$$

$$\text{frequenza teorica nella casella (3;IIIB)} = \frac{9 \cdot 27}{105} \cong 2,3$$

otteniamo così

Tabella delle frequenze teoriche attese					
a_{ij}	IIIA	IIIB	IIIC	IIID	totali
3	2.2	2.3	2.2	2.2	9
4	4.5	4.6	4.5	4.5	18
5	3.0	3.1	3.0	3.0	12
6	11.4	11.8	11.4	11.4	46
7	4.0	4.1	4.0	4.0	16
8	1.0	1.0	1.0	1.0	4
Totali	26	27	26	26	105

Adesso consideriamo le differenze tra le frequenze osservate e quelle teoriche attese.

Tali differenze sono chiamate **contingenze** e sono indicate con c_{ij} :

$$c_{ij} = O_{ij} - a_{ij}$$

Tabella delle contingenze

Tabella delle contingenze					
c_{ij}	IIIA	IIIB	IIIC	IIID	totali
3	-1.2	0.7	1.8	-1.2	0
4	1.5	-0.6	0.5	-1.5	0
5	0.0	-1.1	0.0	1.0	0
6	-1.4	1.2	-2.4	2.6	0
7	1.0	-1.0	1.0	-2.0	0
8	0.0	0.0	-1.0	1.0	0
Totali	0	0	0	0	

Per eliminare l'effetto di compensazione dei voti positivi e negativi senza comunque modificare la dimensione delle contingenze, si calcola, per ogni casella, il valore

$$\frac{c_{ij}^2}{a_{ij}} = \frac{(o_{ij} - a_{ij})^2}{a_{ij}}$$

Costruiamo dunque un'altra tabella

Tabella dei valori c_{ij}^2/a_{ij}				
	IIIA	IIIB	IIIC	IIID
3	0.68	0.20	1.41	0.68
4	0.53	0.09	0.07	0.48
5	0.0	0.38	0.0	0.36
6	0.17	0.12	0.50	0.60
7	0.27	0.00	0.27	0.97
8	0.0	0.0	0.99	1.03

La somma di tutti questi valori ci dà la distanza della distribuzione congiunta da quella teorica nel caso di indipendenza assoluta. Tale somma si chiama indice “**chi quadrato**” di Pizzetti-Pearson e si scrive χ^2 .

$$\chi^2 = \sum_{i,j} \frac{(o_{ij} - a_{ij})^2}{a_{ij}}$$

Nel nostro esempio $\chi^2 \cong 9,8$

Nel caso di indipendenza assoluta dei due caratteri l'indice chi assume il suo minimo che è uguale a 0, mentre il suo massimo dipende dal numero n delle unità statistiche e dal numero delle modalità assunte dai caratteri.

Il rapporto $\Phi = \frac{\chi^2}{N}$ si chiama **contingenza quadratica media**. Anche questo indice è nullo in caso di indipendenza assoluta e tanto più vicino a zero quanto più la connessione tra i caratteri è debole.

Introduciamo un altro indice di dipendenza tra due caratteri, indice di **contingenza di Cramer**, così definito:

$$C = \frac{\Phi}{h-1} = \frac{\chi^2}{N(h-1)}$$

dove h è il minore tra il numero di righe e quello di colonne meno 1 e N è il numero delle unità statistiche.

Il valore C è più indicativo per stabilire la distanza della distribuzione congiunta da quella teorica rispetto al chi quadrato, infatti assume un valore minimo 0 (indipendenza assoluta) e massimo 1 (dipendenza massima).

Nell'esempio precedente ci sono 6 righe e 4 colonne allora $h=4-1=3$, il numero n delle unità statistiche è 105, allora $C=0.03$, il risultato è un numero molto vicino allo zero, quindi si può sostenere che tra il voto di matematica e la classe frequentata non ci sia dipendenza.

Possiamo quindi riassumere i passi da effettuare per analizzare la dipendenza statistica tra due caratteri:

1. si costruisce preliminarmente la tabella a doppia entrata delle frequenze dei due caratteri, completa di totali
2. si costruisce la tabella delle frequenze teoriche
3. si costruisce la tabella delle contingenze
4. si costruisce la tabella dei quadrati delle contingenze divisi per le frequenze teoriche
5. si sommano i valori della tabella precedente per ottenere χ^2
6. dividendo χ^2 per N si ottiene Φ

$$7. \text{ si può poi calcolare l'indice di Cramer } C = \frac{\Phi}{h-1} = \frac{\chi^2}{N(h-1)}$$

Retta di regressione e correlazione lineare

Abbiamo già accennato prima che la ricerca sperimentale utilizza la statistica per stabilire la misura della dipendenza di un carattere Y da un altro carattere X. Il problema che si pone è vedere se esiste una funzione matematica f che sia espressa da una formula algebrica tale che $f(X)=Y$

Se i due caratteri sono entrambi di tipo quantitativo costituiscono una distribuzione doppia e il processo di ricerca di un legame di dipendenza viene detto **regressione**.

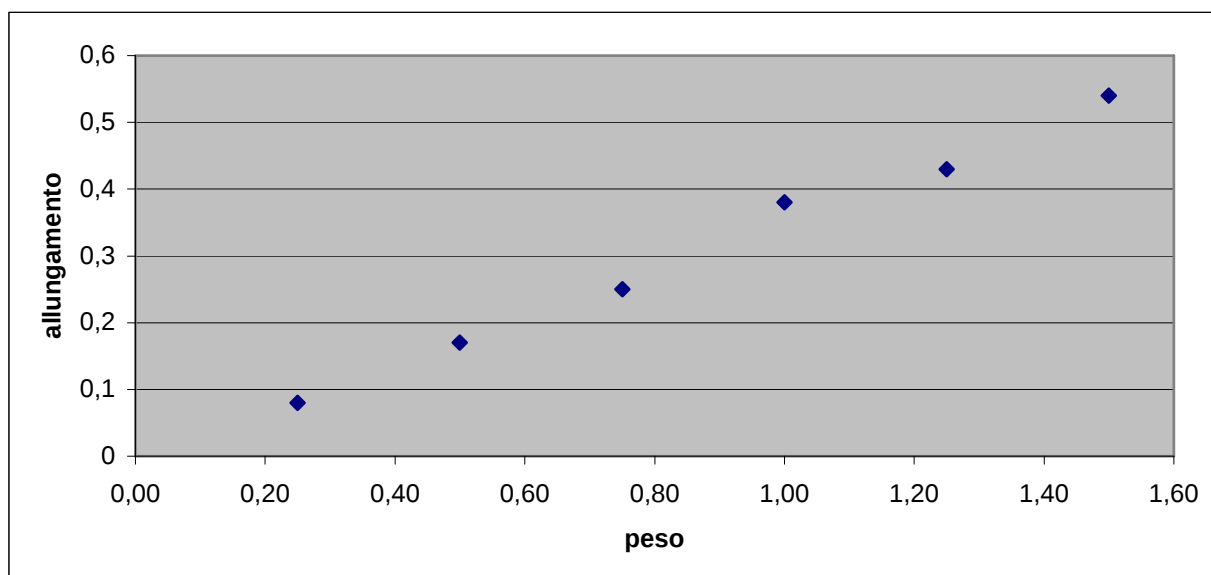
Per effettuare una regressione si fa riferimento a modelli teorici di vario tipo: lineare, parabolico, esponenziale, logaritmico. Vediamo ora di occuparci del **modello lineare**.

Esempio

Nella tabella che segue è registrato l'allungamento di una molla Y rispetto alla posizione di riposo in corrispondenza del peso X applicato alla molla:

peso X in N	0.25	0.50	0.75	1.00	1.25	1.50
Allungamento Y in m	0.08	0.17	0.25	0.38	0.43	0.54

Questa tabella descrive una funzione statistica, il carattere Y è trovato in corrispondenza del carattere X, è possibile rappresentare tale funzione in un riferimento cartesiano



è facile osservare **dal grafico che i punti sono quasi in linea retta**, dobbiamo stabilire se esiste realmente tra i valori X e Y una legge matematica di dipendenza lineare, cioè: $Y=mX+q$.

Un modo per determinare una legge di dipendenza che approssimi una funzione statistica è quello dei minimi quadrati, cioè rendere minima la somma dei quadrati delle differenze tra dati reali e dati teorici.

Questo metodo è detto **metodo dei minimi quadrati**.

Una volta stabilito che i dati reali rappresentati graficamente possono essere approssimati da una funzione lineare, come nel nostro esempio, bisogna stabilire quali sono i migliori valori da assegnare a m e q.

La retta che meglio approssima matematicamente la relazione tra due distribuzioni X e Y è detta **retta di regressione**.

Indicati con y_i il valore effettivo del carattere Y e con $\hat{y}_i$ il valore teorico corrispondente atteso, il metodo dei minimi quadrati consiste nel trovare quella funzione che nel caso lineare è $Y=mX+q$, che renda minima la seguente misura di distanza:

$$\sum_{i=1}^n (y_i - f(x_i))^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

Nel caso di dipendenza perfetta questa somma risulta nulla.

Tornando al nostro esempio calcoliamo la retta dei minimi quadrati:

DEFINIZIONE

Si definisce **baricentro** di una distribuzione doppia (X,Y) il punto che ha per coordinate $(\bar{x}; \bar{y})$ le medie aritmetiche di ognuna delle distribuzioni

Si può dimostrare che la retta di regressione ottenuta con il metodo dei minimi quadrati deve passare per il baricentro dei dati.

Nel nostro caso si ha

$$\bar{x} = 0,875 \text{ e } \bar{y} = 0,3008\bar{3}$$

Il fascio di rette passanti per il punto $(\bar{x}; \bar{y})$ ha equazione $y - \bar{y} = m(x - \bar{x})$
allora, sostituendo, otteniamo

$$y - 0.308\bar{3} = m(x - 0.875)$$

allora

$$y = mx + 0.308\bar{3} - m \cdot 0.875$$

La retta di regressione appartiene a tale fascio.

Dobbiamo adesso calcolare il coefficiente angolare m, applichiamo quindi una traslazione degli assi cartesiani in modo che il baricentro dei dati coincida con l'origine degli assi.

$$\begin{cases} x' = x - \bar{x} \\ y' = y - \bar{y} \end{cases}$$

L'altra cosa che dobbiamo considerare è che il metodo dei minimi quadrati rende minimo il valore della somma

$$\sum_{i=1}^n (y'_i - \hat{y}'_i)^2 = \sum_{i=1}^n (y'_i - mx'_i)^2 = \sum_{i=1}^n (y'^2_i - 2mx'_i y'_i + m^2 x'^2_i) = \left(\sum x'^2_i\right)m^2 - \left(2\sum x'_i y'_i\right)m + \left(\sum y'^2_i\right)$$

L'espressione così ottenuta è una funzione polinomiale di secondo grado che ha come unica variabile m, l'equazione rappresenta una parabola con la concavità rivolta verso l'alto. Il valore di m per cui è minima tale funzione è dato dall'ascissa del vertice della parabola.

$$m = \frac{-2\sum x'_i y'_i}{2\sum x'^2_i} = \frac{\sum x'_i y'_i}{\sum x'^2_i}$$

Facendo gli opportuni calcoli nel nostro esempio:

$$m \cong 0.3669$$

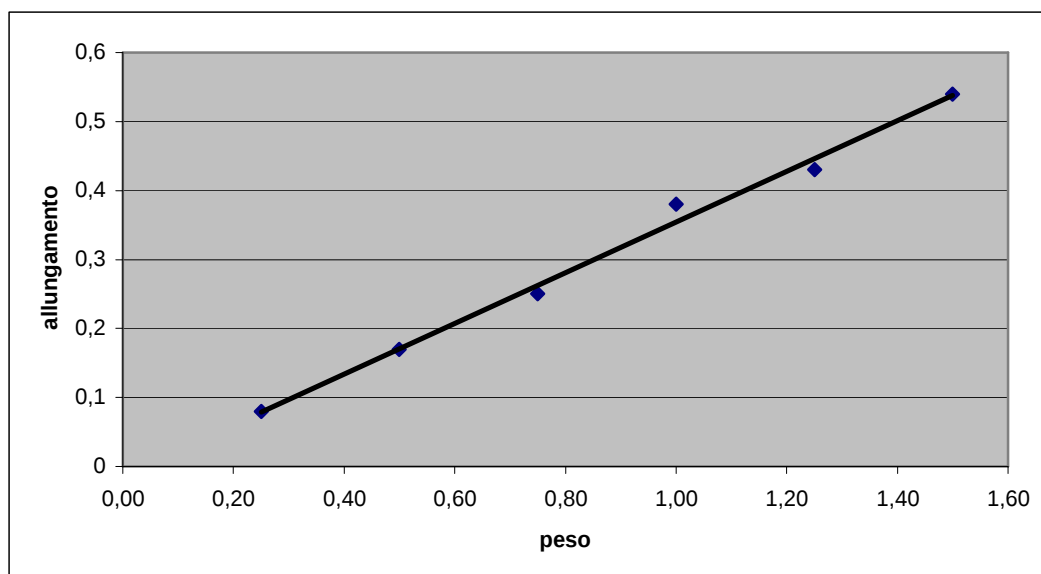
la retta di regressione è dunque

$$y = 0.3669x - 0.013$$

Con un errore trascurabile tale retta può essere così scritta:

$$y = 0.37x$$

La retta riportata di seguito è la retta di regressione relativa all'esempio:



Generalizzando, la retta di regressione con il metodo dei minimi quadrati è:

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

Ricordando che la varianza è

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

e detta **covarianza** di due distribuzioni statistiche

$$COVAR(X, Y) = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n},$$

possiamo generalizzare la formula per la ricerca del coefficiente angolare della retta di regressione in questo modo:

$$m = \frac{COVAR(X, Y)}{VAR(X)}$$

Quindi la retta di regressione sarà:

$$y - \bar{y} = \frac{COVAR(X, Y)}{VAR(X)} (x - \bar{x})$$

Esempio

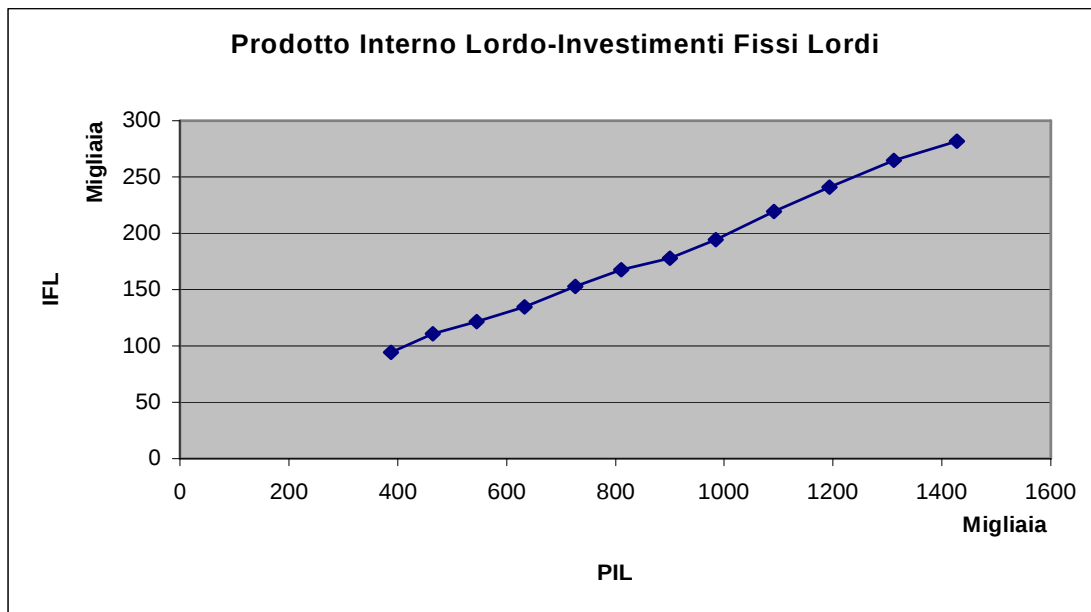
Di seguito sono riportati i dati delle serie storiche relative ai valori correnti degli investimenti fissi lordi e del prodotto interno lordo (dati in miliardi di lire. Fonte ISTAT).

Elabora i dati e analizza se esiste dipendenza di un carattere rispetto all'altro determinando la retta di regressione:

anno	PIL	IFL
1980	387669	94062
1981	464030	110683
1982	545124	121734
1983	633436	134842
1984	725760	152603

1985	810580	167593
1986	899903	177654
1987	983803	194102
1988	1091837	219252
1989	1193462	241023
1990	1311638	264889
1991	1427342	281984

Si rappresentano i dati graficamente:



Osservando il diagramma si capisce che è possibile approssimare la funzione statistica con una funzione matematica lineare.

Calcoliamo i vari elementi che ci servono per disegnare la retta di regressione lineare (i dati sono considerati in migliaia di lire):

$$\bar{x} = 873 \quad \bar{y} = 180 \quad \text{COVAR}(X, Y) = 18988 \quad \text{VAR}(X, Y) = 104683$$

L'equazione della retta di regressione è pertanto:

$$y = \frac{18988}{104683}(x - 873) + 180$$

$$\text{cioè: } y = 0.181x + 21.834$$

Nel caso appena considerato della retta di regressione, il ruolo delle due variabili è diverso, una è considerata come variabile indipendente, mentre l'altra come variabile dipendente. Spesso i ruoli delle due variabili non sono esattamente definiti o definibili a priori, cioè i ruoli possono essere facilmente scambiabili.

Allora si avranno due rette di regressione

di y su x, $y = mx + q$

e di x su y, $x = m'y + q'$

anche la seconda retta di regressione passa per il baricentro dei dati e il coefficiente angolare m' si ottiene

scambiando tutte le x_i con le y_i : $m' = \frac{\text{COVAR}(X, Y)}{\text{VAR}(Y)}$, le due rette ovviamente si intersecano nel punto di

coordinate $(\bar{x}; \bar{y})$, a questo punto è possibile definire un nuovo indice che misura la correlazione tra i caratteri X e Y.

DEFINIZIONE

Si chiama **indice di correlazione lineare di Bravais Pearson** il valore dell'espressione seguente:

$$r = \frac{\text{COVAR}(X, Y)}{\sigma(X) \cdot \sigma(Y)}$$

L'indice di correlazione lineare è compreso tra -1 e 1 e assume questi valori estremi in caso di dipendenza perfetta, mentre assume il valore 0 in caso di indipendenza assoluta. Valori positivi di r indicano una correlazione positiva, cioè la pendenza della retta che descrive l'andamento della dipendenza è positivo (retta crescente), mentre valori negativi indicano che la pendenza della retta che descrive l'andamento della dipendenza è negativo (retta decrescente).

Per quanto riguarda l'esempio precedente, l'indice r di correlazione è positivo, il coefficiente angolare della retta di regressione di X su Y è:

$$m' = \frac{18988}{3457} \cong 5.49$$

L'equazione di questa retta è:

$$x = 5.49y - 115.2$$

Esempio di esercitazione in laboratorio

Vediamo un esempio tratto da un'esercitazione del Corso di Analisi dei Dati del corso di Laurea in Scienze Motorie (Facoltà di medicina-Università di Ferrara).

Analizziamo un esperimento di memorizzazione a breve termine dei primi 10 nomi propri che compaiono in una pagina scelta a caso dall'elenco telefonico di Ferrara. Ogni prova dell'esperimento consiste nella scelta di una nuova pagina, nella lettura dei primi 10 nomi, nella trascrizione su un foglio dei nomi memorizzati. Al termine di ogni prova vengono annotati data, ora, esito, fattori contingenti (parametri che descrivono la situazione in cui la prova si è svolta e che potrebbero influenzare la singola prestazione come le condizioni atmosferiche, stanchezza, ecc.) e fattori sistematici (parametri che descrivono e quantificano azioni sistematiche che si ritiene possano influenzare a medio termine la prestazione media del soggetto).

Impostazione iniziale. Impostiamo una tabella in cui inserire i dati relativi alle prove. Ad ogni prova sarà dedicata una riga della tabella, le cui colonne saranno: **Data**, **Ora**, **E1** (numero di nomi scritti), **E2** (numero di nomi corretti), **C1** (ore di sonno nelle ultime 24 ore), **C2** (numero di nomi diversi), **S1** (ore di studio), **S2** (ore di lettura). Creiamo una riga con l'intestazione delle colonne: scriviamo i titoli delle 8 colonne nelle caselle dalla D9 alla K9. I dati relativi alle singole prove verranno riportati nelle corrispondenti caselle delle righe sottostanti (dalla 10 in poi).

Compilazione. Supponiamo di aver compilato 40 righe della tabella, dalla 10 alla 49 incluse, con i dati relativi ad altrettante prove ripetute dell'esperimento.

Elaborazione: formule. Utilizziamo alcune delle colonne a destra della tabella per svolgere semplici elaborazioni sui dati relativi a singoli esperimenti. Ad esempio, in colonna M, calcoliamo il numero di nomi scritti, ma sbagliati. Evidentemente, per ogni esperimento questo dato può essere ottenuto come differenza tra il numero di nomi scritti e il numero di nomi corretti. Pertanto non era necessario annotarlo, ma può essere calcolato a posteriori sulla base di quanto annotato in tabella. Per ottenere in posizione M10 il dato relativo alla prima prova, scriviamo nella cella la formula =F10-G10. Premendo invio la formula viene valutata e il risultato visualizzato nella cella M10. Si noti che qualsiasi modifica del contenuto delle celle F10 o G10 provoca la rivalutazione della formula di cui F10 e G10 sono operandi e l'aggiornamento del risultato visualizzato in posizione M10.

Volendo ripetere l'elaborazione per tutte le righe della tabella si può (ovviamente) ripetere l'inserimento della formula per ogni riga, oppure copiare la formula. Questo avviene selezionando con il mouse la cella M10, scegliendo il comando Modifica-Copia, selezionando con il mouse le celle dalla M11 alla M49 e scegliendo il comando Modifica-Incolla. L'effetto del comando è la replica della formula con i riferimenti traslati rigidamente. Ad esempio, poichè tra M10 e M12 ci sono due posizioni verticali di differenza, la formula copiata in posizione M12 conterrà riferimenti con numeri di riga incrementati di 2. In tal modo in ogni cella della colonna M viene riportata la formula riferita ai dati sulla stessa riga.

Elaborazione: funzioni. Le funzioni predefinite utilizzabili per comporre formule ed effettuare elaborazioni sono disponibili attraverso il menu Inserisci-Funzione.

Elaborazione: riferimenti fissi. Supponiamo di voler annotare in colonna N il numero di giorni trascorsi dall'inizio degli esperimenti. Poichè la data del primo esperimento è stata annotata in posizione D10, e la data di ogni esperimento è stata annotata in colonna D, alla riga 20, il numero di giorni trascorsi dal primo esperimento è restituito dalla formula $=D20-D10$, inserita in posizione N20. Copiando la formula in posizione N21, in base a quanto detto nel paragrafo precedente, la formula diventa $=D21-D11$, mentre ciò che si vorrebbe ottenere è $=D21-D10$ (cioè si vorrebbe che il primo riferimento venisse aggiornato e il secondo no). Per ottenere questo effetto è possibile bloccare il numero di riga del secondo riferimento facendolo precedere dal simbolo \$. Se la formula scritta in posizione N20 è $=D20-D\$10$, copiandola in posizione N21 diventa $=D21-D\$10$. In generale, il simbolo \$ deve precedere qualsiasi indice di riga o colonna che non deve essere modificato durante le operazioni di copiando o spostamento della formula. La formula può essere copiata in tutte le celle dalla N10 alla N49.

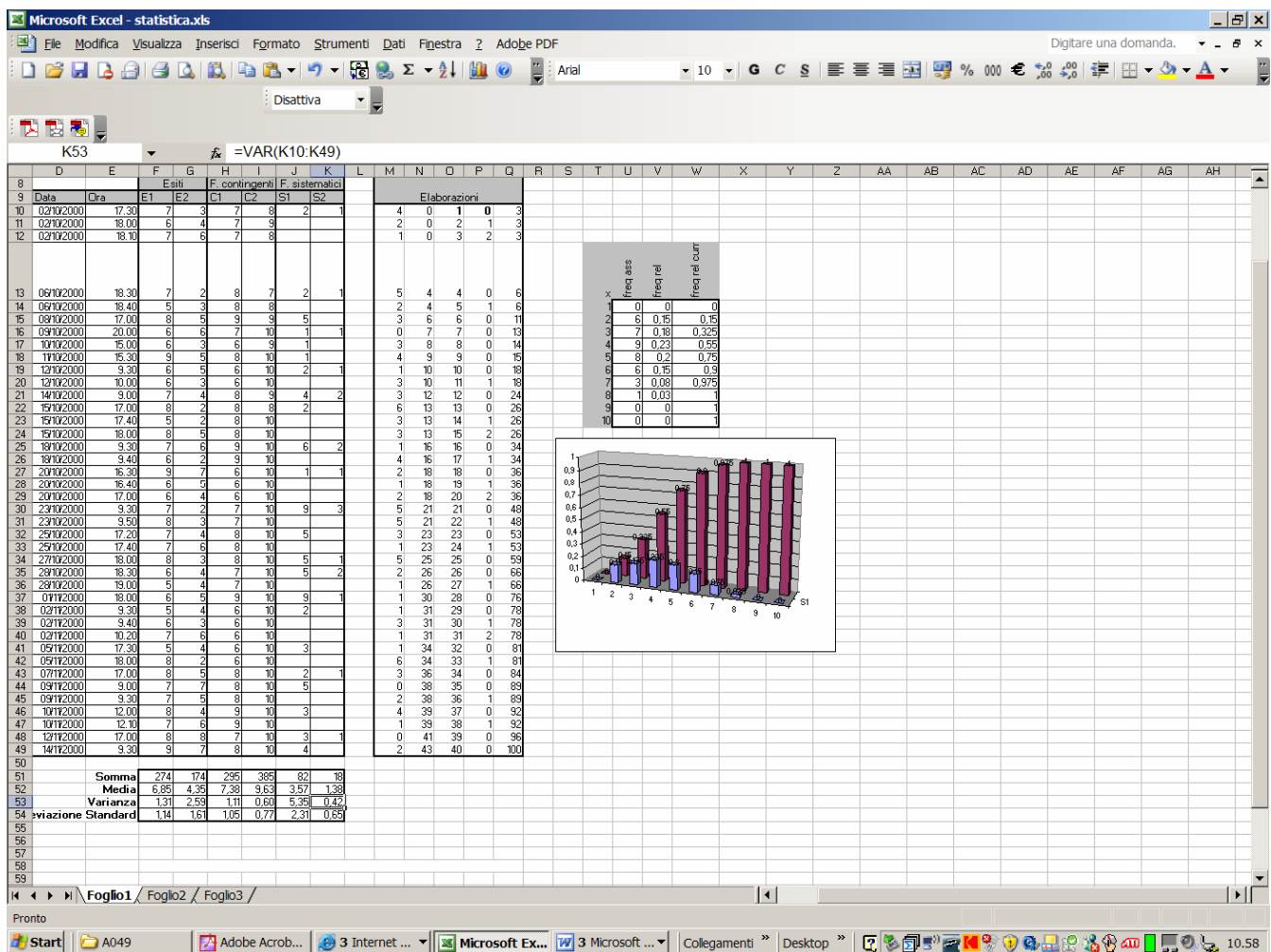
Elaborazione: dati incrementali. In colonna O riportiamo il numero totale di prove effettuate (cioè il numero d'ordine della prova a cui la riga è dedicata). Scriviamo 1 in posizione O10 in quanto la riga 10 è la prima in cui sia stato annotato un esperimento. Poichè in ogni riga è stato annotato un esperimento il numero d'ordine di un qualsiasi esperimento può essere ottenuto sommando 1 al numero d'ordine dell'esperimento precedente. Pertanto scriviamo $=O10+1$ in posizione O11 e copiamo la formula in tutte le celle dalla O12 alla O49. Un procedimento simile è stato utilizzato per calcolare, in colonna Q, le ore totali di studio e lettura dall'inizio degli esperimenti.

Elaborazione: condizioni. In colonna P scriviamo il numero di prove già effettuate nel corso dello stesso giorno. La regola per determinare tale numero è la seguente: se la data della prova corrente è uguale a quella della prova precedente, allora il numero di prove già effettuate è uguale a quello calcolato alla riga precedente più 1, altrimenti il numero di prove già effettuate nello stesso giorno è 0. Per specificare tale regola è possibile utilizzare la funzione SE. Ad esempio, in posizione P11 scriviamo l'espressione $=SE(D11=D10;P10+1;0)$. I tre argomenti separati da punto e virgola rappresentano rispettivamente: la condizione, la formula da utilizzare se la condizione è verificata, la formula da utilizzare se la condizione non è verificata. La formula può essere copiata su tutte le celle dalla P12 alla P49, mentre in posizione P10 va inserito il valore iniziale 0.

Elaborazione: dati d'insieme. Ai piedi di ogni colonna della tabella calcoliamo alcune proprietà d'insieme dei dati raccolti nella colonna. Utilizziamo la formula $=SOMMA(F10:F49)$ in posizione F51 per calcolare il numero totale di nomi trascritti. Analogamente utilizziamo le formule $=MEDIA(F10:F49)$, $=VARIANZA(F10:F49)$ e $=DEV.STD(F10:F49)$ in posizione F52, F53 e F54 per calcolare media, varianza e deviazione standard dell'insieme di dati raccolti in colonna F. Per applicare le stesse formule ad altre colonne di dati è sufficiente copiarle nelle rispettive colonne alle righe 51, 52, 53 e 54.

Elaborazione: frequenze relative. Si supponga di voler valutare la frequenza relativa di un evento (ad esempio la corretta memorizzazione di 6 nomi). Per farlo occorre contare quante volte in colonna G compare il numero 6 e dividere per il numero di prove. La funzione CONTA.SE può essere utilizzata a questo scopo. Il primo argomento è l'insieme di celle su cui effettuare il conteggio, il secondo argomento è il valore da cercare. Nel nostro caso: $=CONTA.SE(G10:G49;6)$. Nel file di esempio, le occorrenze di tutti i possibili esiti sono calcolate e annotate in posizione U14:U23. Dividendo i conteggi per il numero di prove si ottengono le frequenze relative riportate in colonna V. Infine, le frequenze cumulate, sono riportate in colonna W. Il calcolo diretto delle frequenze relative poteva essere effettuato utilizzando la funzione FREQUENZA.

Grafici: Selezioniamo la regione V14:W23 e scegliamo Inserisci-Grafico. Tra i tipi standard scegliamo Istogramma con visione tridimensionale e andiamo avanti fino alla comparsa di un grafico che mostra l'andamento della frequenza relativa e relativa cumulata



(file statistica.xls)

Verifica sommativa di fine modulo (2h di tempo)

Esercizio 1

La tabella rappresenta la distribuzione relativa a una simulazione di 1000 lanci di 3 dadi non truccati.

Punteggio	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
frequenza	3	14	22	37	72	98	128	134	110	114	93	74	34	34	28	5

Calcola la frequenza di volte in cui:

è uscito il 13

è uscito un punteggio minore o uguale a 10

è uscito un punteggio superiore a 11

è uscito un punteggio compreso tra 6 e 14, estremi esclusi

è uscito un numero dispari

è uscito un numero primo

Esercizio 2

Il prezzo di una polizza assicurativa ha subito negli ultimi 5 anni gli aumenti percentuali indicati nella tabella seguente

anno	1	2	3	4	5
Incremento perc.	12%	9%	10,5%	4,3%	6,5%

Calcolare l'aumento medio percentuale annuo.

Esercizio 3

In 5 diversi supermercati di una stessa zona si sono rilevati i prezzi al chilogrammo di alcuni formaggi (stessa marca e qualità)

	Negozi A	Negozi B	Negozi C	Negozi D	Negozi E
Grana padano	9,90	10,60	10,90	10,20	10,30

Pecorino romano	8,00	5,50	8,20	8,10	8,10
fontina	7,20	7,40	7,50	7,80	7,60
Provolone dolce	8,90	8,10	9,00	8,70	8,50

Calcola il prezzo medio per ciascun tipo di formaggio ed individua il tipo di formaggio rispetto al quale vi è la maggior variazione di prezzo rispetto alla media.

Esercizio 4

In questa tabella sono riportate le misure del volume di due diverse quantità di aria a differenti temperature:

temperature	10	20	30	40	50	60	70
Volume A	10.41	10.92	11.25	11.65	11.94	12.35	12.85
Volume B	20.87	21.58	22.15	22.95	23.82	24.35	25.92

Verifica se esiste dipendenza lineare del carattere volume dal carattere temperatura.

Determina l'equazione della retta di regressione del volume B sul volume A

Esercizio 5

ad un campione di 118 studenti sono stati somministrati due test, di cui il secondo dopo un opportuno intervento didattico. I risultati dei test sono stati classificati in quattro categorie di valutazione, ottenendo la seguente tabella

Risultati I test	Risultati II test			
	insufficiente	sufficiente	buono	ottimo
insufficiente	7	8	3	1
Sufficiente	1	12	13	3
Buono	0	5	19	12
ottimo	1	3	11	19

Calcola l'indice di contingenza quadratica media e discuti la dipendenza tra i due caratteri.

Punteggio attribuito

Esercizio numero	Punteggio grezzo
1	6
2	2
3	5
4	10
5	8
totale	31

Griglia di Misurazione

riferita alla verifica sommativa dell'Unità Didattica:

Punteggio Grezzo (Totale 31)	Voto in Decimi (ottenuto con la proporzione)	Voto in decimi (una proposta)
0	0-1	3
1		
2		
3		
4	1-2	
5		
6		
7		
8	2-3	
9		
10		
11	3-4	4
12		
13		
14	4-5	

15		
16	5-6	
17		
18		
19	6-7	6
20		
21		
22	7-8	7
23		
24		
25	8-9	8
26		
27		
28	9-10	9
29		
30		
31	10	10

U.D. 3 ELEMENTI DI CALCOLO DELLE PROBABILITA'

Obbiettivi specifici U.D.3

Conoscenze

- Conoscere il concetto degli eventi.
- Conoscere le varie definizioni di probabilità.
- Conoscere l'impostazione assiomatica della probabilità.
- Conoscere il concetto di probabilità della somma logica di eventi.
- Conoscere il concetto di probabilità del prodotto logico di eventi.
- Conoscere il concetto di probabilità condizionata.
- Conoscere il teorema di Bayes.
- Svolgere esperienze interessanti con un foglio elettronico a sostegno, chiarimento, dei concetti relativi e degli esercizi proposti nel testo.

Competenze e Capacità

- Saper costruire ed operare nell'ambito dei modelli probabilistici.
- Saper calcolare la probabilità in semplici casi.
- Saper calcolare la probabilità di un evento condizionato al verificarsi di un altro evento.
- Saper utilizzare la formula di Bayes.
- Utilizzare le conoscenze e le competenze acquisite per risolvere esercizi.
- Individuare in problemi la necessità di giungere alla soluzione mediante l'uso del calcolo delle probabilità.

Prerequisiti U.D. 3

- Le quattro operazioni aritmetiche e relative proprietà.
- L'operazione di elevazione a potenza e relative proprietà.
- Elementi di base del calcolo letterale.
- Nozioni generali sulle frazioni algebriche.
- Il calcolo combinatorio.
- Conoscenza basilari sulla teoria degli insiemi.
- Diagrammi di Venn
- Conoscenza del foglio elettronico

Contenuti U.D.3

- La storia della probabilità.
- La concezione classica della probabilità.
- La concezione statistica della probabilità.
- La concezione soggettiva della probabilità.
- L'impostazione assiomatica della probabilità.
- La probabilità della somma logica di eventi.
- La probabilità condizionata.
- La probabilità del prodotto logico di eventi.
- Il Teorema di Bayes.
- Il calcolo delle probabilità con un foglio elettronico

Tempi U.D.3

Lezioni frontali e svolgimento degli esercizi	6h
Laboratorio	4
Verifica formativa + discussione	2h
Verifica sommativa + discussione	3h
TOTALE	15h

Per un totale di 15 ore che, tenuto conto delle 5 ore settimanali di matematica, equivalgono a circa tre settimane di lavoro. La previsione è da intendersi elastica, perché occorre tener conto dell'andamento e dei processi di apprendimento della classe.

Per questa unità didattica si prevede di svolgere quasi tutte le lezioni con l'ausilio del computer. Si analizzeranno infatti i dati relativi ai vari esempi utilizzati mediante un software foglio elettronico ad esempio excel che permetterà anche la rappresentazione grafica dei dati.

Sviluppo dei contenuti U.D.3

Introduzione storica al calcolo delle probabilità

Ben prima del Medio Evo, il gioco era diffusissimo in Europa. Ad un certo momento, ignoto, i dadi alla fine rimpiazzarono gli astragali come strumenti di gioco e, dal momento che le carte da gioco non apparvero sin verso il 1350, i giochi per un migliaio d'anni dovettero essere condotti prevalentemente con i dadi. Gli sforzi della Chiesa e dello Stato per controllare i mali associati al gioco furono infruttuosi.

Il gioco d'azzardo più antico di cui si abbia notizia, è una specie di lancio del dado, chiamato gioco della pietra egiziana. Ed è proprio in seguito a questo tipo di gioco che Cicerone, nella sua opera *De Divinatione* (da previsione del futuro, 44 a.C.), cerca di approfondire i termini di sorte, caso e prove ripetute, avvicinandosi così alle problematiche attuali del calcolo delle probabilità.

I primi studi che portarono successivamente a concetti legati alla probabilità possono essere trovati però solo a metà del XVI secolo in *Liber de ludo aleæ* di Girolamo Cardano (scritto nel 1526, ma pubblicato solo un secolo e mezzo dopo, nel 1663) e in *Sulla scoperta dei dadi* di Galileo Galilei (pubblicato nel 1656) nei quali i due autori ottengono degli elenchi di numeri facendo ricorso alle permutazioni.

Il problema della ripartizione della posta in gioco nel caso che un gioco d'azzardo debba essere interrotto, venne affrontato da Luca Pacioli, noto anche come Fra Luca dal Borgo, nella sua *Summa de arithmetica, geometria, proportioni et proportionalita* (pubblicata nel 1494) e successivamente da Tartaglia, per poi essere risolto da Pascal e Fermat.

L'interesse di Pascal fu risvegliato dal cavaliere De Méré, spirito vivace, matematico almeno discreto, ed accanito giocatore d'azzardo.

De Méré riportò a Pascal vari problemi, tra cui la ripartizione della posta tra due giocatori che debbono interrompere il gioco, ed un altro problema, meno rilevante per gli sviluppi matematici, ma molto interessante per altri versi: egli si lamentò che la matematica lo faceva perdere al gioco, perché aveva calcolato per una combinazione ai dadi una probabilità maggiore di $1/2$, aveva scommesso a lungo su tale combinazione, ma invece di vincere, perdeva. Pascal rifece i calcoli e trovò che in realtà la probabilità dell'evento considerato era minore di $1/2$, cosicché le perdite conseguite concordavano con quanto ci si poteva aspettare.

La nascita del concetto moderno di probabilità viene proprio attribuita ai due grandi scienziati Blaise Pascal (1623-1662) e Pierre de Fermat (1601-1665), in particolar modo nella corrispondenza che si scambiavano discutendo dei problemi posti da De Méré.

Nello stesso periodo Christiaan Huygens (1629-1695) scrive *de ratiociniis in aleæ ludo* nel quale utilizza il concetto di valore atteso e il campionamento statistico con e senza riposizione.

I suoi lavori influenzano tra l'altro Pierre de Montmort (1678-1719) che scrive nel 1708 *Essai d'analyse sur le jeux de hasard*, ma anche Jakob Bernoulli e Abraham de Moivre.

Pascal annuncia nel 1654 all'Accademia di Parigi che sta lavorando sul problema della ripartizione della posta in gioco. E in una lettera del 29 luglio dello stesso anno a Fermat propone la soluzione del problema, affrontato con il metodo per ricorrenza, mentre Fermat utilizzava metodi basati sulle combinazioni.

Nel 1713 viene pubblicato postumo *Ars conjectandi* di Jakob Bernoulli dove viene formulato il primo teorema limite, ovvero la legge dei grandi numeri.

Nel 1812, il francese Laplace introduce la definizione classica di probabilità di un evento casuale, intesa come il rapporto tra il numero dei casi favorevoli al realizzarsi dell'evento e il numero dei casi possibili, a condizione che quest'ultimi si possano ritenere ugualmente possibili. Sulla base di questa definizione nasce la teoria classica della probabilità, il cui campo di applicabilità è, però, molto ristretto. Rimane comunque l'unica teoria fino all'inizio del XX secolo.

Solo nel '900, negli anni '30, si viene a creare pure una moderna teoria della probabilità grazie soprattutto a Andrey Nikolaevich Kolmogorov che nel 1933 sviluppa la teoria assiomatica in *Grundbegriffe der Wahrscheinlichkeitsrechnung* ispirandosi alla teoria della misurazione o delle scale di misura il cui dibattito - in quel periodo - era particolarmente acceso tra le psico-discipline.

Nella prima metà del '900 si imposta anche la teoria soggettivista, la cui formulazione è dovuta a Bruno de Finetti.

Gli eventi

Nel mondo che ci circonda troviamo una quantità innumerevole di situazioni probabilistiche.

L'idea di probabilità viene spesso introdotta anche nel linguaggio comune pur senza che ne sia stata data una precisa definizione o un contenuto numerico. Si parla, ad esempio di probabilità di vittoria di una certa squadra di calcio, di probabilità che domani piova o no, di probabilità che l'attuale Governo cada prima della fine dell'anno.

Per introdurre l'argomento, da un punto di vista matematico cominciamo a considerare i seguenti avvenimenti:

1. nel gioco della tombola, durante una estrazione, si estrae un numero compreso tra 1 e 90;
2. lanciando un dado esce il numero 7;
3. nel gioco della tombola, durante una estrazione, si estrae il numero 70.

Tali avvenimenti possono verificarsi oppure no e, in questo senso, vengono detti **eventi**.

DEFINIZIONE

Un **evento** è un avvenimento, descritto da una proposizione, che può accadere o non accadere.

Gli eventi che accadono con certezza (come il primo) si dicono **eventi certi**.

Gli eventi che non possono mai verificarsi si dicono **eventi impossibili** (come l'evento 2).

Ci sono eventi che dipendono dal caso e quindi possono accadere, ma senza certezza (come nel caso dell'esempio 3). Questi ultimi si dicono **eventi aleatori**.

Osservazione: Aleatorio deriva dal latino alea, che significa "dado", ricordiamo la famosa frase "Alea iacta est" di Cesare dopo il passaggio del Rubicone.

Tali eventi possono essere visti come possibili risultati di un esperimento il cui esito è casuale e pertanto è detto esperimento aleatorio.

Ogni singolo risultato possibile viene detto **evento elementare o campione**.

Dalla casualità di un evento aleatorio, diventa necessario stimare la possibilità che esso ha di accadere. Per determinare questa stima si introduce il concetto di **probabilità**.

La concezione classica della probabilità

Iniziamo sempre da un esempio.

Abbiamo un sacchetto contenente 5 palline numerate da 1 a 5 che differiscono solo per il numero riportato, impercettibile al tatto. Estraiamo una pallina. L'esito dell'esperimento dà luogo ai seguenti possibili risultati che costituiscono tutti gli eventi elementari relativi a questo esperimento:

- E_1 = "Esce il numero 1"
- E_2 = "Esce il numero 2"
- E_3 = "Esce il numero 3"
- E_4 = "Esce il numero 4"
- E_5 = "Esce il numero 5".

L'insieme di tutti gli eventi elementari si chiama **universo degli eventi** o anche **spazio campionario**.

Nel nostro esempio possiamo rappresentare l'universo come

$$U = \{1, 2, 3, 4, 5\}$$

L'insieme U rappresenta l'insieme dei **casi possibili**. Avendo considerato palline tutte uguali, possiamo dire che tutti i casi sono ugualmente possibili. Osserviamo inoltre che i casi possibili si escludono a vicenda.

Consideriamo l'evento aleatorio E = "Esce una pallina con un numero dispari"

Il sottoinsieme $E = \{1, 3, 5\}$ rappresenta l'insieme dei **casi favorevoli**, cioè quello dei casi in cui E è verificato.

Posiamo allora generalizzare il concetto di evento aleatorio definendolo come ogni possibile sottoinsieme di U . Gli eventi aleatori sono quindi tutti i possibili sottoinsiemi di U , l'insieme dei quali è l'insieme delle parti di U che chiamiamo **spazio degli eventi**. In particolare:

- i sottoinsiemi di un solo elemento sono gli eventi elementari
- il sottoinsieme $\emptyset$ è l'evento impossibile
- l'insieme U è l'evento certo

Consideriamo il quoziente

$$\frac{\text{numero dei casi favorevoli}}{\text{numero dei casi possibili}} = \frac{3}{5}$$

otteniamo una stima sulla possibilità che l'evento E si verifichi che viene chiamato **probabilità di E**.

Osservazione

Bisogna far notare ai ragazzi che la formula precedente vale solo nel caso in cui i singoli eventi abbiano tutti identiche possibilità di verificarsi.

Possiamo quindi meglio precisare la formula mediante la definizione seguente, dovuta al matematico **Laplace** (1749-1827).

DEFINIZIONE

La **probabilità** di un evento **E** è il quoziente fra il numero dei casi favorevoli **f** e quello dei casi possibili **u**, quando sono tutti ugualmente possibili.

$$p(E) = \frac{f}{u}$$

La definizione appena data non è esente da possibili critiche. Infatti in essa il concetto di probabilità viene introdotto sfruttando la nozione di eventi equiprobabili che a sua volta implica la conoscenza dell'idea di probabilità. Ne nasce un evidente giro vizioso che non sfuggì a Laplace. Per superarlo egli ricorse al cosiddetto "**principio di ragione sufficiente**", in base al quale

"due eventi sono da considerarsi **equiprobabili** quando non sussistono validi motivi per pensare che uno di essi possa verificarsi più facilmente dell'altro"

Per esempio nel caso delle palline precedenti non avevamo motivi per dire che le probabilità relative alle loro estrazioni fossero diverse.

Esempio1

Estraiamo una carta da un mazzo di 52 carte. Vogliamo calcolare la probabilità che si verifichino i due venti:

E_1 ="estrazione di una figura rossa"

E_2 ="estrazione di una carta di picche"

In entrambi i casi i casi possibili sono 52. Dunque $u=52$.

Il numero delle figure rosse è 6 (jack,donna,re di quadri e cuori), quindi per E_1 si ha $f = 6$.

Il numero delle carte di picche 13, quindi per E_2 si ha $f = 13$.

Dalla definizione classica di probabilità si ha:

$$p(E_1) = \frac{f}{u} = \frac{6}{52} = \frac{3}{26} \approx 0,11$$

$$p(E_2) = \frac{f}{u} = \frac{13}{52} = \frac{1}{4} = 0,25$$

Osservazione didattica: dalla definizione della probabilità, possiamo osservare che:

- La **probabilità è un numero puro**, cioè non ha dimensione.
- Se un evento è impossibile, il numero dei casi favorevoli è 0 e quindi

$$p = \frac{f}{u} = \frac{0}{u} = 0.$$

Pertanto **la probabilità di un evento impossibile è 0.**

- Se un evento è certo, il numero dei casi favorevoli è uguale a quello dei casi possibili e quindi

$$p = \frac{f}{u} = \frac{u}{u} = 1.$$

Pertanto **la probabilità di un evento certo è 1.**

- Per gli eventi aleatori, il numero f dei casi favorevoli è compreso fra 0 e u :
 $0 < f < u.$

Dividendo tutti i termini della doppia disuguaglianza per u , si ottiene:

$$\frac{0}{u} < \frac{f}{u} < \frac{u}{u}, \quad \text{ossia} \quad 0 < p < 1.$$

Pertanto **la probabilità di un evento aleatorio è compreso fra 0 e 1.**

In generale, considerando assieme i tre casi, possiamo dire che la probabilità di un evento è compresa fra 0 e 1, estremi inclusi: $0 \leq p \leq 1$.

Spesso, il valore della probabilità viene espresso in termini percentuali.

E' comune l'affermazione, per un evento che si reputa certo, che esso si verificherà al 100%

Consideriamo ora nel caso del sacchetto con le 5 palline l'evento
 $\bar{E}$ = "non esce una pallina con un numero pari"

Questo evento si verifica se e solo se non si verifica l'evento

E = "esce una pallina con un numero pari"

$\bar{E}$ è dunque la negazione di E . le palline che verificano E sono 2, le restanti $5-2=3$ palline verificano $\bar{E}$.

$$\text{Dunque } p(\bar{E}) = \frac{3}{5} = \frac{5-2}{5} = 1 - \frac{2}{5} = 1 - p(E).$$

In generale, dato un evento E , il suo evento contrario indicato con $\bar{E}$ è quello che si verifica solo quando non si verifica E e risulta $p(\bar{E}) = 1 - p(E)$

La probabilità e il calcolo combinatorio.

Prima di introdurre un nuovo esempio ricordiamo alcune formule incontrate nello studio del calcolo combinatorio:

$$D'_{n,k} = n^k \quad \text{disposizioni con ripetizione}$$

$$P_n^{(r,s,\dots)} = \frac{n!}{r!s!\dots} \quad \text{permutazioni con ripetizione}$$

$$D_{n,k} = n(n-1) \cdot \dots \cdot (n-k+1) \quad \text{disposizioni semplici}$$

$$C_{n,k} = \frac{D_{n,k}}{P_k} = \frac{D_{n,k}}{k!} \quad \text{combinazioni semplici}$$

$$\binom{n}{k} = \binom{n}{n-k} \quad \text{proprietà dei coefficienti binomiali (Legge delle classi complementari)}$$

Esempio 1

Lanciamo contemporaneamente tre monete. Consideriamo l'evento:

E = "le facce presentano una testa (T) e due croci (C)".

I casi possibili sono tutti i modi con cui T e C possono presentarsi:

TTT TTC TCT CTT TCC CTC CCT CCC cioè $D'_{2,3} = 2^3 = 8$.

I casi favorevoli sono TCC CTC CCT, cioè tutti i possibili modi con cui si possono presentare una testa e due croci. Per calcolarli utilizziamo le permutazioni di 3 elementi di cui 2 ripetuti.

Ne segue che $P_3^{(2)} = \frac{3!}{2!} = 3$. La probabilità è $p(E) = \frac{3}{8}$.

Rileviamo che l'impostazione, il procedimento e il risultato non cambiano se lanciamo successivamente tre volte la stessa moneta.

Esempio 2

Da un'urna contenente 4 palline bianche e 6 nere estraiamo *successivamente* cinque palline, senza rimettere ogni volta la pallina estratta nell'urna. Consideriamo l'evento:

E_1 = "escono successivamente nell'ordine 2 bianche e 3 nere".

Dobbiamo supporre di distinguere per l'ordine di uscita ogni possibile raggruppamento, anche se identico ad altri per composizione. I casi possibili si possono quindi calcolare con le **disposizioni semplici**:

$$D_{10,5} = 10 \cdot 9 \cdot 8 \cdot 7 \cdot 6 = 30240$$

I casi favorevoli sono tutti i gruppi formati da 2 palline bianche delle quattro contenute nell'urna e dai gruppi formati da tre palline nere delle sei contenute nell'urna:

$$D_{4,2} \cdot D_{6,3} = (4 \cdot 3) \cdot (6 \cdot 5 \cdot 4) = 1440.$$

Si ha che $p(E_1) = \frac{D_{4,2} \cdot D_{6,3}}{D_{10,5}} = \frac{1440}{30240} = \frac{1}{21}$.

Consideriamo ora l'evento:

E_2 = "escono 2 bianche e 3 nere".

Questa volta non interessa l'ordine. L'evento è verificato non solo quando la successione è:

bianca, bianca, nera, nera, nera;

ma anche quando è:

bianca, nera, bianca, nera, nera;

nera, bianca, bianca, nera, nera;

....

nera, nera, nera, bianca, bianca.

I casi favorevoli sono quindi quelli precedenti moltiplicati per le permutazioni di 5 elementi di cui 2 e 3 ripetuti:

$$(D_{4,2} \cdot D_{6,3}) \cdot P_5^{(2,3)} = 1440 \cdot \frac{5!}{2!3!} = 14400.$$

Pertanto $p(E_2) = \frac{14400}{30240} = \frac{10}{21}.$

Possiamo concludere che nelle estrazioni successive senza reinserimento si utilizzano le disposizioni semplici se è essenziale l'ordine di uscita, le combinazioni semplici se l'ordine non interessa.

La concezione frequentista della probabilità

L'impostazione classica della probabilità non è però esauriente e ammette alcune alternative radicalmente diverse. Applichiamo la definizione classica al caso del lancio di un dado: la probabilità che esca 6 è uguale al rapporto tra il numero di facce con sei punti e il numero di facce totali. Esistono tuttavia alcuni problemi: che cosa accade se una faccia del dado è stata alleggerita? Oppure appesantita? Conseguenze:

- la previsione del calcolo non è più adeguata al risultato.
- il denominatore non è più la somma dei casi ugualmente possibili.

La definizione appare chiusa su se stessa; con ciò non va scartata ma solo applicata con attenzione, essendo insostituibile in molti casi. Tutte le volte che non si è certi di poter applicare la definizione classica di probabilità, occorre verificare il risultato di quest'ultima mediante l'esperimento, ossia eseguendo un numero sufficientemente grande di prove, dalle quali si ottiene una nuova valutazione della probabilità che fa uso dell'esperienza, quindi empirica. Il calcolo delle probabilità diventa così una branca delle scienze naturali, della stessa natura della geometria. Come lo scopo della geometria è lo studio dei fenomeni spaziali, così lo scopo della teoria della probabilità è lo studio degli eventi ripetibili e la frequenza relativa della ripetizione di un evento osservato diventa la sua misura di probabilità. Questa concezione frequentistica o empirica che si contrappone alla concezione classica, la quale concerne una probabilità a priori, è dovuta a Richard von Mises che è conosciuta come impostazione frequentistica o concezione statistica.

Supponiamo di avere un'urna contenente palline colorate, ma non sappiamo né quali sono i colori né quante sono le palline. Ovviamente non possiamo aprire l'urna per esaminarne il contenuto.

L'unica cosa che ci rimane da fare è quella di estrarre a sorte un gran numero di volte delle palline, rimettendo ogni volta la pallina estratta nell'urna, in modo che ogni estrazione sia effettuata nelle stesse condizioni.

Effettuiamo 80 estrazioni e otteniamo la seguente tabella:

Colore	Numero palline
rosso	5
giallo	18
nero	22
verde	35

Utilizzando il valore ottenuto calcoliamo il rapporto fra il numero delle volte in cui è uscito un determinato colore e il numero delle prove effettuate tutte nelle stesse condizioni.

Il rapporto tra il numero delle prove in cui si è verificato l'evento ed il numero totale delle prove si definisce **frequenza relativa**.

DEFINIZIONE

La **frequenza relativa $f(E)$** di un evento sottoposto a n esperimenti effettuati tutti nelle stesse condizioni, è il rapporto fra il numero delle volte m che si è verificato e il numero n delle prove effettuate.

Colore	Numero palline	Frequenza relativa
rosso	5	$\frac{5}{80} = \frac{1}{16}$

giallo	18	$\frac{18}{80} = \frac{9}{40}$
nero	22	$\frac{22}{80} = \frac{11}{40}$
verde	35	$\frac{35}{80} = \frac{7}{16}$
totale	80	1

Osservazione: osserviamo che il valore della frequenza di un evento è compreso tra 0 e 1. Dobbiamo inoltre sottolineare che la frequenza 0 non significa che l'evento è impossibile, ma soltanto che non si è mai verificato. Analogamente, frequenza 1 non significa che l'evento è certo, ma soltanto che in quella serie di esperimenti è stato sempre osservato. Osserviamo anche che nelle prove effettuate non è uscita nessuna blu, questo però non significa che non ve ne sia alcuna.

Se ripetiamo l'esperimento, senz'altro otterremo valori diversi. Inoltre, se aumentiamo il numero delle prove si rileva un fatto molto interessante e cioè che il valore della frequenza tende a un valore costante che si può ritenere come la probabilità dell'evento.

Dalla definizione della frequenza relativa si giunge a formulare la seguente legge sperimentale

Legge empirica del caso

Dato un evento E , sottoposto a n prove tutte nelle stesse condizioni, il valore della frequenza relativa $f(E) = \frac{m}{n}$ tende al valore della probabilità $p(E)$, all'aumentare del numero n di prove effettuate.

Per esempio se sappiamo che la probabilità di ottenere testa dal lancio di una moneta è, secondo la definizione classica, $p=1/2$, allo stesso valore tende la frequenza se sperimentalmente eseguiamo un numero altissimo di lanci.

La *legge empirica del caso* è alla base della **definizione statistica** o **frequentistica** della probabilità.

DEFINIZIONE

La **probabilità statistica** di un evento E è la frequenza relativa del suo verificarsi quando il numero di prove effettuato è da ritenersi "sufficientemente alto".

Nell'impostazione classica il valore della probabilità è calcolato **a priori**, ossia prima che l'esperimento avvenga, mentre il valore della frequenza è un valore **a posteriori**.

Ci sono moltissimi eventi per i quali è impossibile calcolare la probabilità applicando l'impostazione classica. Per eventi di questo tipo si è costretti ad applicare l'impostazione frequentistica.

Per esempio, nel campo delle assicurazioni si calcola con questo metodo:

- la probabilità di incidenti automobilistici;
- la probabilità di vita e di morte;
- la probabilità di furti.

Per tutti questi eventi occorre fondare il calcolo su quanto è avvenuto in passato e cercare statisticamente le relative frequenze che costituiranno le probabilità degli eventi.

Altri esempi in cui viene utilizzata la probabilità statistica sono:

- la probabilità che hanno dei macchinari di produrre pezzi difettosi;
- la probabilità di contrarre una determinata malattia;
- la probabilità che un farmaco sia efficace.

Esempio

Per il calcolo delle probabilità di vita e di morte sono state costruite delle tabelle demografiche apposite che, partendo da una collettività chiusa di 100000 nati vivi, riporta anno dopo anno le persone viventi a una certa età.

Sulla tavola di mortalità per il 1987-maschi leggiamo che per l'età 50 i viventi sono 93300 e per l'età 85 i viventi sono 19384.

La probabilità che ha un 50-enne di raggiungere gli 85 anni è:

$$p_1 = \frac{19384}{93300} \approx 0,2074$$

Mentre la probabilità di morire entro gli 85 anni è: $p_1 \approx 1 - 0,2074 \approx 0,7926$

La concezione soggettiva della probabilità

A questo punto è opportuno far notare che anche la teoria frequentistica della probabilità, proprio perchè a *posteriori*, ha dei limiti di applicazioni. Si richiede, infatti, che le prove sperimentali possano essere ripetute un gran numero di volte, tutte nelle *stesse condizioni*. Ciò, ovviamente, non è possibile per tutti i tipi di eventi. Infatti se un evento non si è mai realizzato, ma si potrebbe realizzare, non è possibile stabilirne la probabilità; come ad esempio: qual è la probabilità che un terremoto distrugga Roma? Oppure che ci sia un extraterrestre rinchiuso nella vostra cantina? La definizione frequentistica della probabilità trova larghe applicazioni nella statistica; con il campionamento dà ottimi risultati, ma non è universale. La matematica non deterministica è in difficoltà proprio sulla definizione di probabilità. Per risolvere questo problema, e quindi anche quelli legati alla impostazione classica, è stato proposto in questo secolo una nuova impostazione del concetto, basato sulla cosiddetta definizione *soggettiva* dovuta al grande matematico italiano **Bruno De Finetti**.

Una persona sta valutando se partecipare o non partecipare a un gioco d'azzardo nel quale si vince se, lanciando contemporaneamente due dadi, escono due numeri pari.

Secondo l'impostazione classica, egli sa che la probabilità di vittoria è $\frac{9}{36} = \frac{1}{4} = 25\%$, ma ha osservato che su

21 lanci le due facce pari sono uscite 7 volte, cioè con una frequenza di $\frac{1}{3}$, cioè del 33,3% circa.

Decide allora di proporre a chi tiene il banco di giocare una posta di 10 euro per ricevere in caso di vittoria 34 euro.

La valutazione che ha fatto è del tutto personale. Ha determinato il valore della probabilità secondo una valutazione **soggettiva**.

Possiamo determinare il valore che ha attribuito alla probabilità dell'evento "uscita di due facce pari" mediante il rapporto tra la posta che è disposto a pagare e la somma che dovrebbe ricevere in caso di vincita:

$$p(E) = \frac{10}{34} = \frac{5}{17} \approx 29,4\%$$

La valutazione di questa persona ha tenuto conto, anche se non in modo esplicito, sia dell'impostazione classica sia di quella statistica.

Osservazione: il modo di procedere soggettivo è l'unico utilizzabile per quegli eventi per i quali non è possibile calcolare teoricamente il numero dei casi favorevoli e possibili e non si può sottoporre l'evento a prove sperimentali ripetute nelle stesse condizioni.

Siamo in questa situazione se vogliamo stimare la probabilità di vittoria di una squadra di calcio a un torneo.

DEFINIZIONE

La **probabilità soggettiva** di un evento è la **misura del grado di fiducia** che una persona attribuisce al verificarsi dell'evento, secondo la sua opinione. Il valore si ottiene effettuando il rapporto fra la somma **P** che si è disposti a pagare in una scommessa, e la somma **V** che si riceverà nel caso l'evento si verifichi.

Osservazione: deve sussistere la **condizione di coerenza**: la persona che accetta di pagare **P** per ottenere **V** deve anche essere disposta a ricevere **P** per pagare **V** nel caso l'evento si verifichi.

Esempio

A una corsa di cavalli una persona è disposta a pagare 90 euro per ricevere 120 euro in caso di vittoria di un determinato cavallo.

Soluzione

Calcoliamo la probabilità di vittoria che attribuisce al cavallo:

$$p(E) = \frac{90}{120} = \frac{3}{4}.$$

Per la condizione di coerenza deve essere disposto, scambiando i ruoli, a ricevere 90 euro e pagare 120 euro in caso di vittoria del cavallo. Si dice anche che la vittoria del cavallo è pagata 4 a 3.

In conclusione, la valutazione soggettiva della probabilità porta a considerare il calcolo della probabilità come una scommessa.

Osservazione: il valore della probabilità, con questa concezione, varia da individuo ad individuo, ma in ogni caso esso è compreso tra 0 e 1 ; si suppone infatti che si è disposti a pagare una somma minore o al più uguale di quella della vincita.

L'impostazione assiomatica della probabilità

La necessità di superare le critiche all'impostazione classica e a quella statistica del calcolo delle probabilità ha portato nel 1933 il matematico russo Kolmogorov ad una impostazione assiomatica che sistema in modo rigoroso le conoscenze e le applicazioni che si sono sviluppate nel tempo.

Osservazione didattica: si può a questo punto accennare ai ragazzi cosa sia una teoria assiomatica. Quindi evidenziare cosa sono gli enti primitivi e gli assiomi.

Lanciamo successivamente una moneta due volte. I possibili esiti dell'esperimento costituiscono un insieme chiamato **spazio dei campioni**:

$$U = \{TT, CC, TC, CT\}.$$

Consideriamo i seguenti eventi e le proposizioni che li esprimono:

$$\begin{aligned} E_1 &= \text{"croce esce una volta"}; \\ E_2 &= \text{"croce esce due volte"} \end{aligned}$$

L'evento E_1 è verificato dagli eventi dell'insieme $M = \{CT, TC\}$ mentre l'evento E_2 è verificato dall'unico elemento dell'insieme $N = \{CC\}$.

Possiamo quindi associare a ogni evento un sottoinsieme di U . Allora tutti i possibili eventi che si possono associare all'esperimento sono le parti di U e costituiscono il cosiddetto spazio degli eventi:

$$P(U) = \{\{TT\}, \{CC\}, \{TC\}, \{CT\}, \{TT, CC\}, \{TT, TC\}, \{TT, CT\}, \{CC, TC\}, \{CC, CT\}, \{TC, CT\}, \{TT, CC, TC\}, \{TT, CC, CT\}, \{CC, TC, CT\}, \{TT, TC, CT\}, \{TT, CC, TC, CT\}, \emptyset\}.$$

Questo esempio ci permette di puntualizzare alcuni concetti fondamentali dell'impostazione assiomatica:

- tutti i possibili risultati o **campioni** di uno esperimento aleatorio sono gli elementi di un insieme U chiamato **spazio dei campioni**;
- un **evento** E si identifica con un **sottoinsieme di** U ed è verificato quando l'esito dell'esperimento coincide con uno dei suoi elementi;
- l'evento **impossibile** è l'insieme vuoto $\emptyset$;
- l'evento **certo** coincide con l'insieme U ;
- un evento **elementare** è un sottoinsieme di U costituito da un solo campione;
- l'insieme di tutti i possibili eventi coincide con l'insieme delle parti di U e si dice **spazio degli eventi**.

Considerando gli eventi come insieme, possiamo effettuare le usuali operazioni fra insiemi, alle quali corrispondono operazioni logiche tra le proposizioni che descrivono gli eventi.

Osservazione didattica: possiamo ora ricordare alcune operazioni logiche.

Date due proposizioni p , q , la loro disgiunzione $p \vee q$ è falsa quando entrambe sono false e vera negli altri casi. Invece la loro congiunzione $p \wedge q$ è vera quando entrambe sono vere e falsa negli altri casi.

$$\begin{array}{ccc} p & q & p \vee q \\ p & q & p \wedge q \end{array}$$

V	V	V	V	V	V
V	F	V	V	F	F
F	V	V	F	V	F
F	F	F	F	F	F

Esempio

All'evento:

$$E_1 \vee E_2 = \text{"croce esce una o due volte"}$$

disgiunzione delle due proposizioni, viene associata l'unione

$$M \cup N = \{CC, CT, TC\}$$

dei due sottoinsiemi corrispondenti. L'evento viene indicato anche con $E_1 \cup E_2$.

Consideriamo ora l'evento

$$E_3 = \text{"testa esce la prima volta"},$$

al quale corrisponde il sottoinsieme $P = \{TT, TC\}$.

All'evento

$$E_1 \wedge E_3 = \text{"croce esce una volta e testa esce la prima volta"},$$

congiunzione delle due proposizioni, viene associata l'intersezione

$$M \cap P = \{TC\}$$

dei due sottoinsiemi corrispondenti. Indichiamo l'evento anche con $E_1 \cap E_3$. L'evento $E_1 \wedge E_3$ si può esprimere anche con "prima esce testa e poi croce".

In generale, possiamo dare le seguenti definizioni.

Definizione

Dato un evento E , l'evento complementare E^C di E rispetto ad U è detto **evento contrario** di E . Tale evento si verifica se e solo se non si verifica E .

Definizione

Dati due eventi E_1 ed E_2 , entrambi sottoinsiemi di U , l'evento $E_1 \cup E_2$ è detto **evento unione** o **somma logica** di E_1 ed E_2 . Esso si verifica al verificarsi di almeno uno degli eventi dati.

Definizione

Dati due eventi E_1 ed E_2 , entrambi sottoinsiemi di U , l'evento $E_1 \cap E_2$ è detto **evento intersezione** o **prodotto logico** di E_1 ed E_2 . Esso si verifica al verificarsi di entrambi gli eventi dati.

Definizione

Due o più eventi si dicono **incompatibili** o **disgiunti** se essi hanno a due a due intersezione vuota, cioè se il verificarsi di uno di essi esclude il verificarsi di uno qualunque dei rimanenti. Nel caso contrario si diranno **"compatibili"**.

Esempio

Relativamente al lancio di un dado consideriamo gli eventi

$$E_1 = \{e_1, e_3, e_5\} \quad E_2 = \{e_3, e_6\} \quad E_3 = \{e_2\} \quad E_4 = \{e_4, e_6\}.$$

Ebbene, sono incompatibili, ad esempio gli eventi:

E_1, E_3 (dato che $E_1 \cap E_3 = \emptyset$)
 E_1, E_3, E_4 (dato che $E_1 \cap E_3 = E_1 \cap E_4 = E_3 \cap E_4 = \emptyset$).

Sono, invece, compatibili ad esempio gli eventi:

E_1, E_2 (perché $E_1 \cap E_2 \neq \emptyset$),
 E_1, E_2, E_3 (perché $E_1 \cap E_2 \neq \emptyset$),
 E_1, E_2, E_3, E_4 (perché $E_1 \cap E_2 \neq \emptyset$ ed $E_2 \cap E_4 \neq \emptyset$).

Avendo fissato tutti questi elementi si ha la seguente definizione assiomatica di probabilità.

DEFINIZIONE

La **probabilità assiomatica** è una funzione reale **p** che associa a ogni evento **E** dello spazio degli eventi un numero reale, in modo da soddisfare i seguenti assiomi:

- (1) $p(E) \geq 0$;
- (2) $p(U) = 1$;
- (3) se due eventi E_1 ed E_2 sono tali che $E_1 \cap E_2 = \emptyset$ (eventi incompatibili), allora:

$$p(E_1 \cup E_2) = p(E_1) + p(E_2).$$

Osservazione: l'impostazione assiomatica non fornisce alcun procedimento per determinare la probabilità di un evento, ma i valori che vengono assegnati agli eventi devono rispettare gli assiomi.

Esempio

L'insieme **U** è costituito da tre eventi elementari **A, B, C** i quali hanno le seguenti probabilità:

$$p(A) = \frac{3}{8} \quad p(B) = \frac{2}{5} \quad p(C) = \frac{9}{40}.$$

Tali valori soddisfano gli assiomi della probabilità. In particolare essendo:

$$\begin{aligned} A \cup B \cup C &= U \\ A \cap B &= A \cap C = B \cap C = \emptyset \\ p(A) + p(B) + p(C) &= \frac{3}{8} + \frac{2}{5} + \frac{9}{40} = 1 = p(U). \end{aligned}$$

Dalla definizione assiomatica possiamo dedurre le seguenti proprietà:

- A. $p(\emptyset) = 0$;
- B. $0 \leq p(E) \leq 1$;
- C. $p(E^c) = 1 - p(E)$;
- D. se $E_1 \subseteq E_2$, allora $p(E_2 - E_1) = p(E_2) - p(E_1)$;
- E. se $E_1 \cap E_2 \neq \emptyset$ e $E_1 \not\subseteq E_2$, allora $p(E_2 - E_1) = p(E_2) - p(E_1 \cap E_2)$.

Vale il seguente teorema.

TEOREMA Probabilità della somma logica di due eventi.

La probabilità della somma logica di due eventi E_1 ed E_2 è uguale alla somma delle loro probabilità diminuita della probabilità del loro evento intersezione:

$$p(E_1 \cup E_2) = p(E_1) + p(E_2) - p(E_1 \cap E_2).$$

In particolare se gli eventi sono **incompatibili**:

$$p(E_1 \cup E_2) = p(E_1) + p(E_2).$$

Analogamente, si può generalizzare la relazione precedente al caso di n eventi; essa si riduce alla seguente quando sono tutti *incompatibili*.

TEOREMA DELLA PROBABILITA' TOTALE

Dati n eventi a due a due incompatibili $E_1, E_2, \dots, E_n$, la probabilità della loro unione è uguale alla somma delle loro singole probabilità:

$$p(E_1 \cup E_2 \cup \dots \cup E_n) = p(E_1) + p(E_2) + \dots + p(E_n).$$

Esempio

Un'urna contiene 15 palline identiche al tatto numerate da 1 a 15. Estraiamo contemporaneamente due palline e consideriamo i seguenti eventi:

E_1	=	"escono due numeri pari"
E_2	=	"escono due numeri dispari"
E	=	"escono due numeri pari o due numeri dispari"

Gli eventi E_1 ed E_2 si escludono a vicenda e i loro insiemi sono disgiunti, sono eventi incompatibili. L'evento E si verifica quando l'esito dell'estrazione è un elemento dell'insieme unione $E_1 \cup E_2$.

Essendo $E_1 \cap E_2 = \emptyset$, $p(E) = p(E_1 \cup E_2) = p(E_1) + p(E_2)$.

Quindi:

$$p(E) = \frac{\binom{7}{2}}{\binom{15}{2}} + \frac{\binom{8}{2}}{\binom{15}{2}} = \frac{7 \cdot 6}{15 \cdot 14} + \frac{8 \cdot 7}{15 \cdot 14} = \frac{7}{15}$$

Consideriamo poi gli eventi

E_3	=	"escono due numeri dispari"
E_4	=	"escono due multipli di 3"
E	=	"escono due numeri dispari o multipli di 3"

E si verifica quando l'esito dell'estrazione è un elemento dell'unione dei due insiemi $E_3 \cup E_4$, gli eventi relativi però questa volta sono compatibili.

$$p(E) = p(E_3 \cup E_4) = p(E_3) + p(E_4) - p(E_3 \cap E_4).$$

Gli esiti dell'estrazione che sono comuni ai due eventi sono tutte le coppie che si possono formare con i numeri 3, 9 e 15.

Quindi:

$$p(E) = \frac{\binom{8}{2}}{\binom{15}{2}} + \frac{\binom{5}{2}}{\binom{15}{2}} - \frac{\binom{3}{2}}{\binom{15}{2}} = \frac{7 \cdot 8}{15 \cdot 14} + \frac{5 \cdot 4}{15 \cdot 14} - \frac{3 \cdot 2}{15 \cdot 14} = \frac{70}{15 \cdot 14} = \frac{1}{3}.$$

La probabilità condizionata

Un'urna contiene 12 palline identiche numerate da 1 a 12. Sappiamo che lo spazio dei campioni è

$$U = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$$

e l'evento

E = "estrazione di una pallina con un numero divisibile per 3"

ha probabilità $p(E) = \frac{4}{12} = \frac{1}{3}$.

Eventi correlati positivamente

Consideriamo ora l'evento:

E_1 = "estrazione di una pallina con un numero maggiore di 7"

per il quale $p(E_1) = \frac{5}{12}$.

Valutiamo ora la probabilità di **E** quando sappiamo che l'evento **E₁** si è verificato.

Indichiamo questa probabilità con il simbolo

$p(E|E_1)$ **probabilità di E condizionata (o subordinata) a E_1**

A questo punto gli esiti possibili non sono più 12 ma 5, in quanto l'insieme universo si è ridotto a:

$$U' = E_1 = \{8, 9, 10, 11, 12\}$$

I casi favorevoli sono i 2 elementi dell'insieme $E \cap E_1 = \{9, 12\}$.

La probabilità è: $p(E|E_1) = \frac{2}{5}$, che è un valore maggiore di $\frac{1}{3}$.

L'informazione supplementare ha aumentato la probabilità di **E**. I due eventi **E** ed **E₁** si dicono **correlati positivamente**.

Eventi stocasticamente indipendenti

Consideriamo l'evento:

E₂ = " estrazione di una pallina con un numero pari"

per il quale $p(E_2) = \frac{6}{12} = \frac{1}{2}$.

Valutiamo la probabilità di **E** supponendo che sia verificato l'evento **E₂**. Indichiamo la probabilità di **E** condizionata a **E₂** con il simbolo $p(E|E_2)$. Anche in questa situazione abbiamo un'informazione in più. Gli esiti possibili non sono più 12 ma sono diventati 6, in quanto l'insieme universo si è ridotto a:

$$U' = E_2 = \{2, 4, 6, 8, 10, 12\}$$

e i casi favorevoli sono i 2 elementi dell'insieme $E \cap E_2 = \{6, 12\}$.

La probabilità è: $p(E|E_2) = \frac{2}{6} = \frac{1}{3}$, che è uguale al valore di $p(E) = \frac{1}{3}$.

In questo caso l'informazione supplementare non ha mutato il valore della probabilità dell'evento **E**. I due eventi **E** ed **E₂** si dicono **stocasticamente indipendenti**.

Eventi correlati negativamente

Consideriamo l'evento:

E₃ = " estrazione di una pallina con un numero minore di 8"

per il quale $p(E_3) = \frac{7}{12}$.

Valutiamo la probabilità di **E** supponendo che sia verificato l'evento **E₃**. Indichiamo la probabilità di **E** condizionata a **E₃** con il simbolo $p(E|E_3)$. Gli esiti possibili non sono più 12 ma sono 7, in quanto l'insieme universo si è ridotto a:

$$U' = E_3 = \{1, 2, 3, 4, 5, 6, 7\}$$

e i casi favorevoli sono i 2 elementi dell'insieme $E \cap E_3 = \{3, 6\}$.

La probabilità è: $p(E|E_3) = \frac{2}{7}$, che è minore del valore di $p(E) = \frac{1}{3}$.

L'informazione supplementare in questo caso ha portato a un valore minore della probabilità. I due eventi **E** ed **E₃** si dicono **correlati negativamente**.

Generalizzando enunciamo il seguente teorema:

TEOREMA

La probabilità condizionata di un evento E_1 rispetto a un evento E_2 , non impossibile, è determinata dalla formula:

$$p(E_1|E_2) = \frac{p(E_1 \cap E_2)}{p(E_2)}, \quad \text{con } p(E_2) \neq 0.$$

Analogamente :

$$p(E_2|E_1) = \frac{p(E_1 \cap E_2)}{p(E_1)}, \quad \text{con } p(E_1) \neq 0.$$

Dimostrazione

Chiamiamo k il numero degli esiti favorevoli al verificarsi di E_2 ed r quello degli esiti favorevoli al verificarsi di E_1 nell'ipotesi che E_2 si sia verificato, ossia il numero di elementi di $E_1 \cap E_2$. Si ha:

$$p(E_1|E_2) = \frac{r}{k}.$$

Dividendo numeratore e denominatore per n , numero dei campioni di U , otteniamo:

$$p(E_1|E_2) = \frac{r}{k} = \frac{\frac{r}{n}}{\frac{k}{n}} = \frac{p(E_1 \cap E_2)}{p(E_2)}.$$

La probabilità del prodotto logico di eventi

TEOREMA DELLA PROBABILITA' COMPOSTA

La probabilità dell'evento composto o prodotto logico degli eventi E_1 ed E_2 è uguale al prodotto della probabilità dell'evento E_1 per la probabilità dell'evento E_2 nell'ipotesi che E_1 sia verificato:

$$p(E_1 \cap E_2) = p(E_1) \cdot p(E_2|E_1).$$

In particolare nel caso di eventi **stocasticamente indipendenti**:

$$p(E_1 \cap E_2) = p(E_1) \cdot p(E_2).$$

Esempio

Estraiamo una carta da un mazzo di 52 carte. L'evento:

$$E = \text{"esce un re nero"}$$

è formato dai due eventi:

$$E_1 = \text{"esce una carta con un seme nero"}$$

$$E_2 = \text{"esce un re"}.$$

Essendo $E = \{\text{re picche, re fiori}\}$, la probabilità del prodotto logico di E_1 e E_2 è:

$$p(E) = p(E_1 \cap E_2) = \frac{2}{52} = \frac{1}{26}.$$

Questo risultato si può ottenere anche in un altro modo.
Dalla relazione della probabilità condizionata:

$$p(E_2|E_1) = \frac{p(E_1 \cap E_2)}{p(E_1)}$$

otteniamo:

$$p(E_1 \cap E_2) = p(E_1) \cdot p(E_2|E_1).$$

Applichiamo questa relazione nel nostro esempio.

$$\text{Abbiamo } p(E_1) = \frac{26}{52} = \frac{1}{2}.$$

Per l'evento E_2 condizionato a E_1 , essendo uscita una carta nera, i casi possibili sono 26, mentre i casi favorevoli sono 2:

$$p(E_2|E_1) = \frac{2}{26} = \frac{1}{13}.$$

$$\text{Pertanto } p(E_1 \cap E_2) = p(E_1) \cdot p(E_2|E_1) = \frac{1}{2} \cdot \frac{1}{13} = \frac{1}{26}.$$

Con il teorema della probabilità composta possiamo calcolare la probabilità degli eventi composti anche quando non è possibile effettuare il calcolo diretto.

Esempio

Da una rilevazione statistica è risultato che, al primo controllo dopo 2000 Km, su 100 automobili 8 hanno rilevato difetti solo alla alimentazione e 3 solo ai freni. Calcoliamo la probabilità che una macchina presenti questi due difetti che risultano indipendenti:

$$E_1 = \text{"avere difettosa l'alimentazione"}, \quad p(E_1) = \frac{8}{100} = 8\%;$$

$$E_2 = \text{"avere difettosi i freni"}, \quad p(E_2) = \frac{3}{100} = 3\%;$$

$$p(E_1 \cap E_2) = \frac{8}{100} \cdot \frac{3}{100} = \frac{24}{10000} = \frac{0,24}{100} = 0,24\%$$

Se invece fra le 8 macchine che hanno difetti alla alimentazione 2 presentano anche difetti ai freni, non potremo più considerare i due eventi come indipendenti. In questo caso la probabilità di avere difetti all'alimentazione e ai freni è:

$$p(E_1 \cap E_2) = p(E_1) \cdot p(E_2|E_1) = \frac{8}{100} \cdot \frac{2}{8} = \frac{2}{100} = 2\%.$$

Il teorema della probabilità composta si può estendere a più eventi che si devono verificare uno dopo l'altro, considerando sempre quello precedente come verificato. Nel caso di tre eventi la formulazione è la seguente:

$$p(E) = p(E_1 \cap E_2 \cap E_3) = p(E_1) \cdot p(E_2|E_1) \cdot p(E_3|(E_1 \cap E_2)),$$

che per eventi indipendenti si semplifica:

$$p(E) = p(E_1 \cap E_2 \cap E_3) = p(E_1) \cdot p(E_2) \cdot p(E_3).$$

Il Teorema di Bayes

Abbiamo le seguenti due urne:

- ♦ Urna 1: 3 palline bianche e 2 nere;

- ♦ Urna 2: 4 palline bianche e 5 nere.

Calcoliamo la probabilità che scegliendo a caso un'urna ed effettuando l'estrazione di una pallina questa sia bianca. Per la scelta dell'urna ci affidiamo al lancio di un dado: se viene un numero minore di tre effettueremo l'estrazione dalla prima urna, altrimenti dalla seconda.

Gli eventi relativi alla scelta dell'urna sono:

E_1 = "faccia del dado con un numero minore di tre";

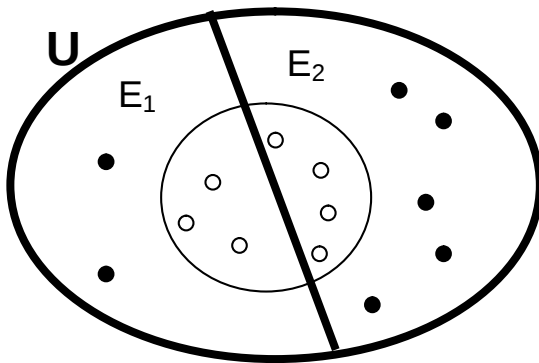
E_2 = "faccia del dado con un numero maggiore o uguale a tre".

Questi due eventi sono incompatibili ed esauriscono tutte le possibilità del lancio del dado.

Essi hanno probabilità:

$$p(E_1) = \frac{2}{6} = \frac{1}{3}, \quad p(E_2) = \frac{4}{6} = \frac{2}{3}.$$

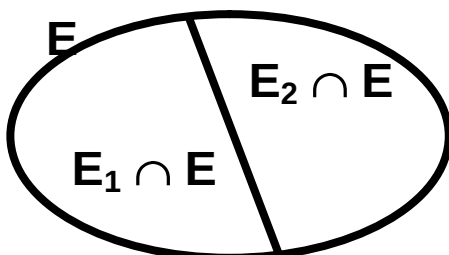
Collegiamo gli eventi E_1 ed E_2 alle due urne con le palline bianche e nere considerando la rappresentazione



Allora l'evento:

E = "estrazione di una pallina bianca"

è un sottoinsieme dell'evento $U = E_1 \cup E_2$ ed è l'unione di due eventi incompatibili.



$E_1 \cap E$ = "uscita della faccia del dado con un numero minore di tre ed estrazione di una pallina bianca dalla prima urna";

$E_2 \cap E$ = "uscita della faccia del dado con un numero maggiore o uguale a tre ed estrazione di una pallina bianca dalla seconda urna".

Gli eventi condizionati relativi all'estrazione della pallina bianca avendo scelto l'urna sono:

$E|E_1$ = "estrazione pallina bianca avendo scelto la prima urna";

$E|E_2$ = "estrazione pallina bianca avendo scelto la seconda urna".

Essi hanno probabilità: $p(E|E_1) = \frac{3}{5}, \quad p(E|E_2) = \frac{4}{9}.$

Applicando il teorema della probabilità composta, abbiamo che gli eventi composti $E_1 \cap E$ ed $E_2 \cap E$ hanno rispettivamente probabilità:

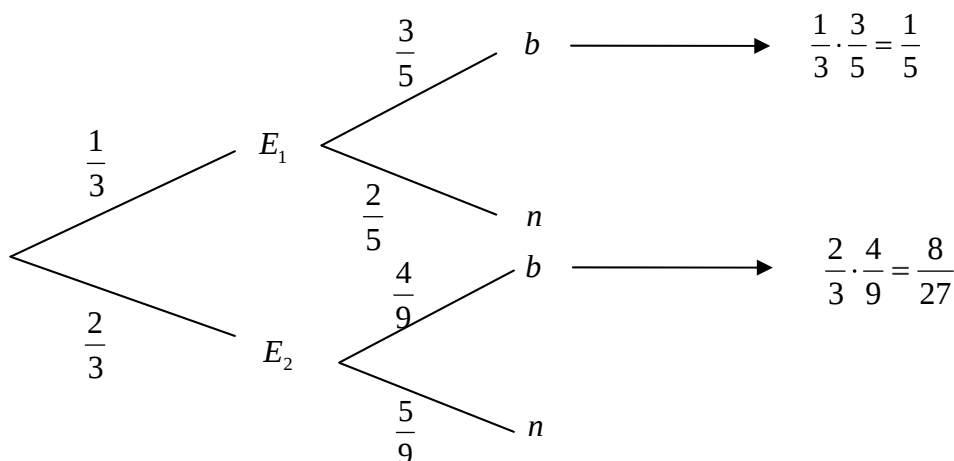
$$p(E_1 \cap E) = p(E_1) \cdot p(E|E_1) = \frac{1}{3} \cdot \frac{3}{5} = \frac{1}{5};$$

$$p(E_2 \cap E) = p(E_2) \cdot p(E|E_2) = \frac{2}{3} \cdot \frac{4}{9} = \frac{8}{27}.$$

Possiamo ora calcolare la probabilità dell'evento E , che è unione dei due eventi composti: $E = (E_1 \cap E) \cup (E_2 \cap E)$:

$$p(E) = p(E_1 \cap E) + p(E_2 \cap E) = p(E_1) \cdot p(E|E_1) + p(E_2) \cdot p(E|E_2) = \frac{1}{5} + \frac{8}{27} = \frac{67}{135}.$$

Possiamo sintetizzare tutto il ragionamento nel modo seguente:



Partendo da sinistra, percorrendo i rami del diagramma ad albero, leggiamo la successione degli eventi che formano l'evento composto ed effettuiamo il prodotto delle probabilità. I rami uscenti da un nodo rappresentano eventi incompatibili e sommando le probabilità segnate su essi otteniamo il valore uno.

Addizionando le probabilità degli eventi prodotto dei percorsi otteniamo la probabilità dell'evento considerato:

$$p(E) = \frac{1}{5} + \frac{8}{27} = \frac{67}{135}$$

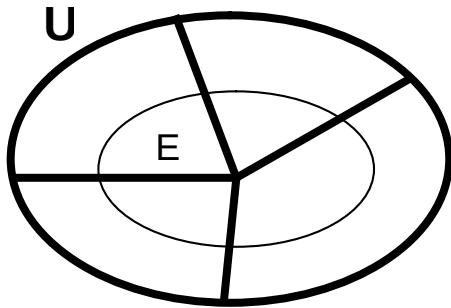
Generalizziamo dicendo che:

“ Un evento si può esprimere come unione di eventi composti a due a due incompatibili nel seguente modo:

$$E = (E_1 \cap E) \cup (E_2 \cap E) \cup \dots \cup (E_n \cap E)$$

dove $E_1, E_2, \dots, E_n$ costituiscono una partizione dello spazio dei campioni U , cioè sono n eventi:

- non impossibili: $E_i \neq \emptyset$ per $i = 1, 2, 3, \dots, n$;
- incompatibili a due a due: $E_i \cap E_j = \emptyset$ per $i, j = 1, 2, 3, \dots, n$ e $i \neq j$;
- tali che la loro unione $E_1 \cup E_2 \cup \dots \cup E_n$ sia un evento certo.



Applicando il teorema della probabilità totale, si ha:

$$p(E) = p(E \cap E_1) + p(E \cap E_2) + \dots + p(E \cap E_n)$$

e applicando il teorema del prodotto logico di eventi si ha la seguente formula:

$$p(E) = p(E_1) \cdot p(E|E_1) + p(E_2) \cdot p(E|E_2) + \dots + p(E_n) \cdot p(E|E_n)$$

formula di disintegrazione

la cui applicazione risulta facilitata utilizzando i diagrammi ad albero.

Consideriamo ora l'esperimento relativo all'estrazione di una pallina bianca di due urne, la cui scelta è stabilita dal lancio di un dado. Supponiamo che l'evento:

E = "estrazione di una pallina bianca"

si sia verificato. Proviamo a rispondere alla seguente domanda:

"qual è la probabilità che la pallina estratta provenga dalla prima urna?".

Siamo in una situazione completamente diversa da quella precedente. Infatti abbiamo sempre calcolato la probabilità di un evento che potrebbe accadere conoscendo le cause che stanno alla base del suo verificarsi. Ora **siamo di fronte a un evento che si è verificato e vogliamo conoscere la probabilità da assegnare alla causa che può averlo prodotto**.

Trascriviamo le seguenti nozioni già adottate:

E₁ = "faccia del dado con un numero minore di tre";

E₁ ∩ E = "uscita della faccia del dado con un numero minore di tre ed estrazione di una pallina bianca dalla prima urna";

E|E₁ = "estrazione di una pallina bianca essendo uscita la faccia del dado con un numero minore di tre".

I valori della probabilità che abbiamo calcolato precedentemente sono:

$$p(E_1 \cap E) = p(E_1) \cdot p(E|E_1) = \frac{1}{5} \quad \text{e} \quad p(E) = \frac{67}{135}.$$

Dalla relazione della probabilità composta:

$$p(E \cap E_1) = p(E) \cdot p(E_1|E),$$

Ricaviamo $p(E_1|E)$ e, dopo aver notato che $p(E \cap E_1) = p(E_1 \cap E)$, sostituiamo i valori calcolati precedentemente, ottenendo che, avendo riscontrato l'uscita di una pallina bianca, la probabilità che essa provenga dalla prima urna è:

$$p(E_1|E) = \frac{p(E \cap E_1)}{p(E)} = \frac{p(E_1 \cap E)}{p(E)} = \frac{\frac{1}{5}}{\frac{67}{135}} = \frac{27}{67}.$$

Quindi la probabilità della causa dell'evento che si è verificato si ottiene facendo il rapporto tra la probabilità dell'evento, verificata la causa, e la probabilità totale dell'evento.

Generalizziamo il problema.

Sia U uno spazio di campioni ed $E \subset U$ un evento che supponiamo sia verificato. Consideriamo una partizione di U in n eventi $E_1, E_2, \dots, E_n$. Possiamo scrivere:

$$E = (E \cap E_1) \cup (E \cap E_2) \cup \dots \cup (E \cap E_n)$$

e la probabilità che l'evento E_i sia stato la **causa** di E si ottiene come rapporto fra la probabilità di $E_i \cap E$ e la probabilità dell'evento totale E . Ciò conduce alla seguente formula, nota come **teorema di Bayes**.

TEOREMA di BAYES

La probabilità che, essendosi verificato un evento E , la causa che sta alla sua origine sia l'evento E_i , con $i = 1, 2, \dots, n$, è:

$$p(E_i|E) = \frac{p(E_i) \cdot p(E|E_i)}{p(E)}$$

dove $p(E)$ è la probabilità dell'evento totale:

$$p(E) = p(E_1) \cdot p(E|E_1) + p(E_2) \cdot p(E|E_2) + \dots + p(E_n) \cdot p(E|E_n).$$

Il teorema di Bayes trova applicazioni nel campo del controllo della qualità, in medicina, in farmacia e ogni qualvolta è necessario valutare il "peso" di una causa di fronte al verificarsi di un effetto.

La speranza matematica e i giochi equi

Una persona decide di partecipare al seguente gioco.

Viene lanciato un dado. Se esce il numero 1 essa riceve 7 euro, se esce un numero pari 4 euro, mentre in caso contrario deve pagare 2 euro. Vengono effettuati 8 lanci e alla fine si rileva che il numero 1 è uscito una volta, un numero pari 2 volte e per 5 volte la persona ha dovuto effettuare il pagamento pattuito.

Vediamo se è stato conveniente partecipare, il bilancio finale è dato da:

$$EUR(7+4 \times 2 - 2 \times 5) = EUR 5$$

Il guadagno totale è stato positivo e il guadagno medio è stato di $5/8 = 0,625$ euro. Possiamo ottenere lo stesso valore scrivendo i calcoli nella forma:

$$7 \cdot \frac{1}{8} + 4 \cdot \frac{2}{8} - 2 \cdot \frac{1}{3} = 0,625$$

Se al posto delle frequenze empiriche sostituiamo le probabilità abbiamo:

$$7 \cdot \frac{1}{6} + 4 \cdot \frac{1}{2} - 2 \cdot \frac{1}{3} = 2,5$$

Questo valore rappresenta non la vincita media riscontrata ma quella che il giocatore avrebbe potuto sperare di realizzare in media. Questo valore si chiama proprio **speranza matematica**.

Se la speranza matematica è nulla il gioco è equo, se è positiva è favorevole, altrimenti sfavorevole.

Nel nostro esempio la speranza è 2,5 quindi il gioco è favorevole.

Consideriamo il caso in cui il gioco preveda una posta da pagarsi prima dell'inizio.

Ad esempio in caso di esito favorevole il giocatore incasserà una vincita lorda di 16,5 euro se esce la faccia 1 e 13,5 se esce un numero pari.

La posta P che il giocatore deve pagare, affinché il gioco sia equo, è la somma dei prodotti delle vincite lorde per le rispettive probabilità di vincita:

$$P = 16,5 \cdot \frac{1}{6} + 13,5 \cdot \frac{1}{2} = 9,5$$

Si chiama speranza matematica M di una somma S il prodotto dell'importo della somma per la probabilità p di ottenerla. $M=Sp$.

Se la posta P è uguale a M il gioco è equo. Quindi il rapporto P/M caratterizza l'equità del gioco.

Facciamo degli esempi:

Roulette, giocando al "rosso e nero" (ad esempio $E = \text{"rosso"}$):

$$p(E) = 18/37$$

$$S = 2P$$

$$M/P = 36/37 = 0,973 \text{ (97,3\%)}$$

Dunque il gioco non è equo.

Roulette, puntando su un numero (es. $E = \text{"esce il 24"}$)

$$P(E) = 1/37$$

$$S = 36P$$

$$M/P = 36/37 = 0,973$$

In effetti tutte le vincite della roulette sono calcolate in modo da mantenere un rapporto costante fra puntata e speranza matematica.

Lotto, puntando su un numero su una ruota (es. $E = \text{"esce il 25 sulla ruota di Bari"}$)

$$P(E) = 5/90 = 1/18$$

$$S = 10,9 P$$

$$M/P = 10,9/18 = 0,605 \text{ (60,5\%)}$$

Da questi esempi si vede, ad esempio, che il gioco della roulette è molto più equo del lotto. Anzi si potrebbe dire per inciso che sembra anche "troppo equo", nel senso che la speranza matematica differisce soltanto del 2.7% dalla puntata. Presumibilmente bisogna tener conto che, anche se il gioco è quasi equo dal punto di vista probabilistico, non lo è dal punto di vista psicologico, e infatti quasi sempre i giocatori seguitano a giocare anche dopo che hanno vinto, e spesso finiscono per perdere tutta la cifra che erano disposti a giocare e sono costretti a smettere (esiste un classico problema del calcolo delle probabilità detto della "rovina del giocatore").

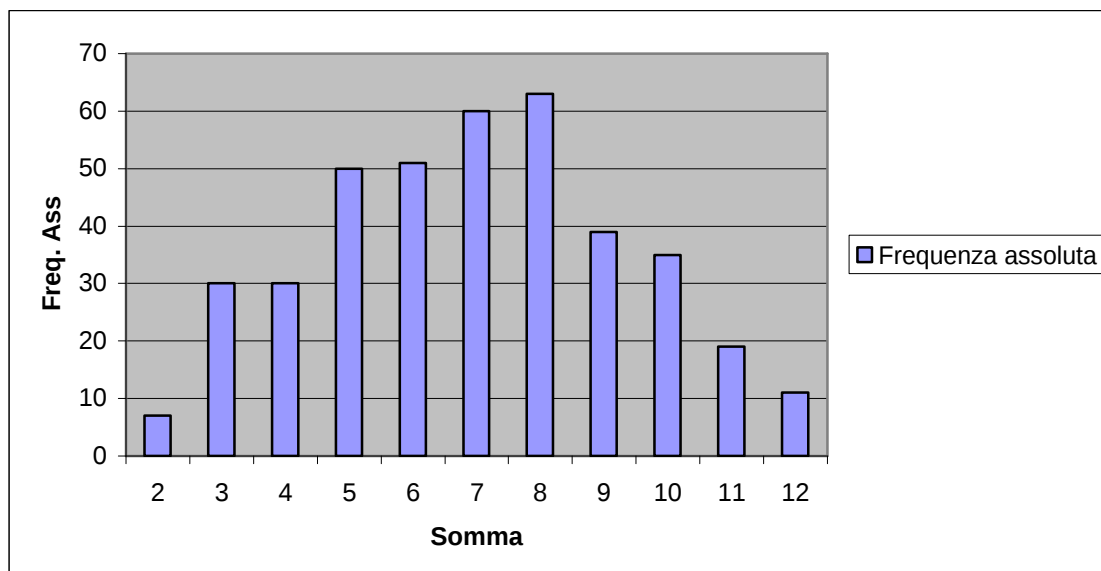
Esempio di esercitazione di laboratorio

1 Simulazione del lancio di due dadi con Excel (vedi file lancio_dadi.xls)

Costruiamo un foglio elettronico che riporti nelle colonne B e D le uscite dei due dadi (usando in maniera opportuna la funzione casuale). Calcolare la somma e utilizzando la funzione conta.se() calcolare frequenze assolute e relative

Microsoft Excel - lancio_dadi.xls									
File Modifica Visualizza Inserisci Formato Strumenti Dati Finestra ? Adobe PDF									
Digitare una domanda.									
Arial 10 G C S									
Disattiva									
H29									
	A	B	C	D	E	F	G	H	I
1		Dado 1		Dado 2	Somma	Nro estrazioni	Somma	Frequenza ass	Frequenza rel
2									
3		2		3	5	400	2	10	0,025
4		6		1	7		3	20	0,05
5		4		2	6		4	25	0,0625
6		3		1	4		5	41	0,1025
7		4		2	6		6	54	0,135
8		6		2	8		7	80	0,2
9		4		1	5		8	63	0,1575
10		3		6	9		9	42	0,105
11		4		1	5		10	35	0,0875
12		2		3	5		11	16	0,04
13		4		5	9		12	8	0,02
14		4		2	6				
15		6		2	8				
16		2		6	8				
17		4		2	6				
18		5		4	9				
19		4		4	8				
20		6		3	9				
21		1		6	7				
22		2		2	4				
23		3		5	8				
24		5		4	9				
25		6		6	12				
26		3		4	7				
27		1		6	7				
28		3		4	7				
29		1		3	4				
30		5		3	8				
31		2		2	4				
32		5		3	8				
33		2		5	7				
34		2		5	7				
35		2		3	5				

Simuliamo un numero elevato di lanci, ad esempio 400, e osserviamo il grafico delle frequenze che si viene ad ottenere.

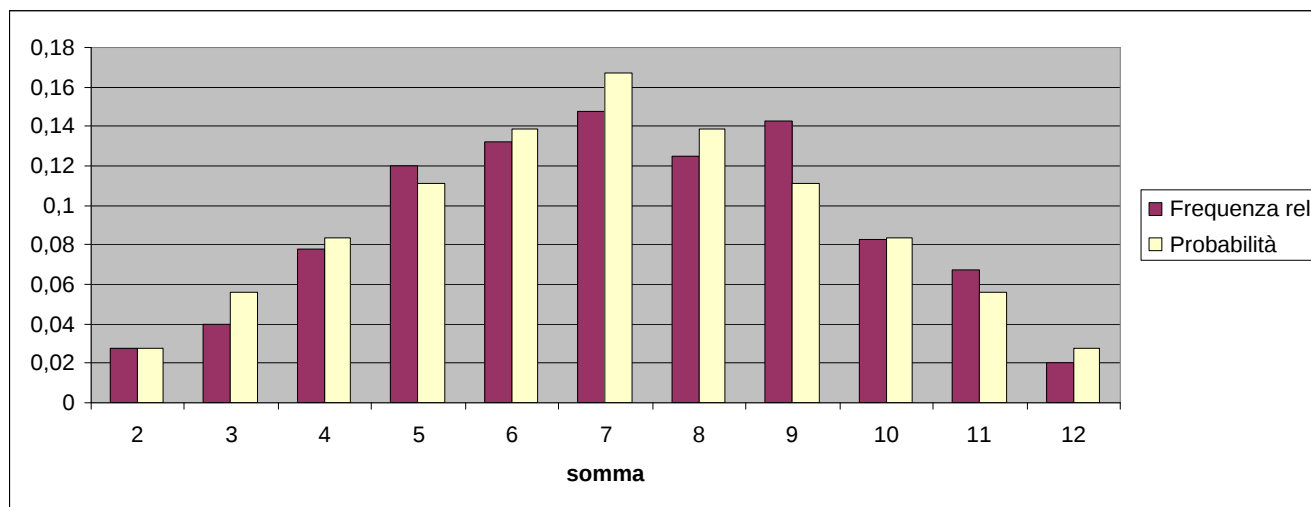


I casi centrali sono i più frequenti. Perché?

Valutiamo le probabilità dei diversi eventi. I casi possibili nel lancio di due dadi sono tutte le coppie che si possono ottenere associando una faccia del primo dado con una qualunque faccia del secondo dado. Quanti sono quindi?

Completiamo una tabella per calcolare le probabilità secondo la definizione classica

Modalità presentazione	Nro casi favorevoli	Nro casi possibili	Probabilità
1+1	1	36	0,027778
1+2;2+1	2	36	0,055556
1+3;3+1;2+2	3	36	0,083333
1+4;4+1;2+3;3+2	4	36	0,111111
1+5;5+1;2+4;4+2;3+3	5	36	0,138889
1+6;6+1;2+5;5+2;3+4;4+3	6	36	0,166667
2+6;6+2;3+5;5+3;4+4	5	36	0,138889
3+6;6+3;4+5;5+4	4	36	0,111111
4+6;6+4;5+5	3	36	0,083333
5+6;6+5	2	36	0,055556
6+6	1	36	0,027778



Confrontando il grafico delle probabilità con quello sperimentale ci possiamo chiedere come mai quest'ultimo non sia simmetrico come ci si aspetterebbe dalla valutazione delle probabilità.

Il grafico avrebbe un andamento più vicino alle previsioni se il numero di lanci fosse maggiore?

Si possono fare diverse simulazioni aumentando via via il numero di lanci.

Verifica sommativa di fine modulo (2h di tempo)

Esercizio 1

Un'urna contiene 10 palline numerate da 1 a 10 indistinguibili. Calcolare la probabilità che:

- estraendo una pallina essa abbia il numero 5
- estraendo una pallina essa abbia un numero divisibile per 4
- estraendo una pallina essa non abbia un numero divisibile per 4

Esercizio 2

Prima dell'inizio di una partita di calcio un tifoso sarebbe disposto a scommettere 14 euro per ricevere 20 euro sulla vincita della sua squadra del cuore, 2 euro per riceverne 10 in caso di pareggio e infine 0,5 euro per riceverne 5 in caso di perdita. Calcolare le probabilità che attribuisce ai tre eventi: vincita, pareggio, sconfitta.

Esercizio 3

Un'urna contiene 4 palline rosse numerate da 1 a 4 e 6 palline nere numerate da 1 a 6. Si estraggono successivamente due palline, senza rimettere la pallina estratta nell'urna. Calcolare le probabilità che:

- le palline estratte siano di colore uguale
- le palline estratte siano rosse o rechino un numero pari
- almeno una pallina estratta sia rossa

Esercizio 4

Un'urna contiene 40 palline numerate da 1 a 40. consideriamo gli eventi: A="esce un numero divisibile per 5", B="esce un numero maggiore di 20".

Calcolare la probabilità condizionata di A rispetto a B

Esercizio 5

Un'urna contiene 4 palline bianche e 5 nere ed una seconda urna 6 bianche e 4 nere. Si è scelta un'urna a caso e sono state estratte contemporaneamente una pallina bianca e una nera. Calcoliamo la probabilità che esse provengano dalla prima urna

Esercizio 6

Tre persone sono candidate a una carica. A ha la probabilità del 40% di essere eletto, B del 35% e C del 25%. C ritira la propria candidatura. Calcolare le nuove probabilità di vittoria di A e B.

Punteggio attribuito

Esercizio numero	Punteggio grezzo
1	4
2	5
3	5
4	5
5	5
6	6
totale	31

Griglia di Misurazione

riferita alla verifica sommativa dell'Unità Didattica:

Punteggio Grezzo (Totale 31)	Voto in Decimi (ottenuto con la proporzione)	Voto in decimi (una proposta)
0	0-1	3
1		
2		
3		
4	1-2	
5		
6		
7		
8	2-3	
9		
10		
11	3-4	4
12		
13		
14	4-5	5
15		
16	5-6	

17		6
18		
19	6-7	7
20		
21	7-8	8
22		
23	8-9	9
24		
25	9-10	10
26		
27	10	
28		
29		
30		
31		

BIBLIOGRAFIA

- Bergamini-Trifone-Dieghi : “Elementi di matematica- Modulo α giallo rosso verde- Il calcolo combinatorio, la probabilità e le distribuzioni di probabilità” Zanichelli 2001
- Frascini-Grazzi : “MATEMATICA per i Licei Scientifici-3 Analisi Calcolo combinatorio” Atlas 2007
- Maraschini-Palma: “Format,COM la formazione matematica per il triennio” Paravia 1996
- Zwirner-Scaglianti: “Il metodo matematico orientato all’economia 4” Cedam 1995
- Bucchini-Magi-Gambardella “Elementi di matematica-teoria” Calderoni 1995
- 4° corso MPI-UMI in Didattica della Matematica per Docenti di Scuole Medie, Un itinerario didattico per la Statistica (1997)
- Appunti del modulo di statistica descrittiva e probabilità del corso SSIS di Laboratorio didattico di attività trasversali. Prof. Fucci
- <http://www.sti.uniurb.it/bogliolo/didattica/AnalisiDati/>
- <http://www.liceofoscarini.it/studenti/probabilita/storia.html>

This document was created with Win2PDF available at <http://www.win2pdf.com>.
The unregistered version of Win2PDF is for evaluation or non-commercial use only.
This page will not be added after purchasing Win2PDF.