

PES - Probabilità e Statistica per l'informatica

Elia Ronchetti

Marzo 2022

Indice

1	Introduzione	3
2	Analisi Descrittiva	4
2.1	Descrivere i dati	4
2.1.1	Rappresentazione dei dati	4
2.1.2	Dati Bivariati	5
2.2	Riassumere i dati	5
2.2.1	Indici di posizione	5
3	Esame	6

Capitolo 1

Introduzione

Il corso di probabilità e statistica per l'informatica è diviso in 2 parti

1. Statistica Descrittiva - Descrivere e riassumere i dati
 - (a) Probabilità - Descrivere matematicamente i fenomeni casuali
2. Statistica inferenziale - Trarre conclusioni dai dati

Capitolo 2

Analisi Descrittiva

2.1 Descrivere i dati

Per descrivere una raccolta dati in maniera chiara e immediata è utile utilizzare una **tabella delle frequenze** all'interno della quale sono contenuti:

- Valori
- Frequenze Assolute - Numero di volte in cui compare "i" nell'insieme di dati
- Frequenze Relative - Frazione di volte in cui compare i nell'insieme di dati
- Percentuali - (Frequenza relativa x 100)

I dati possono essere

- Qualitativi
- Quantitativi

Noi useremo i dati **quantitativi**

2.1.1 Rappresentazione dei dati

Per rappresentare le frequenze (assolute o relative) risulta efficace e immediato l'utilizzo di un grafico a barre detto istogramma, esso rappresenta in graficamente la tabella, chiaramente da esso è possibile risalire alla tabella stessa. Capita di avere degli insiemi di dati che assumono un valore elevato di valori distinti, per questo conviene suddividerli in classi e determinare la

frequenza di ciascuna classe. In questo modo c'è una perdita d'informazioni (sui valori specifici), ma così facendo possiamo calcolare le frequenze delle classi e avere un'idea migliore della distribuzione dei dati.

2.1.2 Dati Bivariati

Quando per ciascun individuo vengono misurate due variabili ci troviamo un insieme di N dati a coppie detti **dati bivariati**. Anche in questo caso è possibile calcolare le frequenze, in questo caso detto **frequenze congiunte**.

è possibile, inoltre, misurare la correlazione tra le due variabili attraverso per esempio un diagramma di dispersione (detto anche scatterplot).

Correlazione non significa causalità! Non è detto che l'aumento di una variabile causi la diminuzione dell'altra o viceversa, potrebbe esserci una causa comune.

2.2 Riassumere i dati

Dopo aver rappresentato i dati vogliamo ora riassumerli mediante quantità numeriche, dette **Statistiche Campionarie**, al fine di sintetizzare le proprietà salienti dei dati.

2.2.1 Indici di posizione

Per definire il centro dell'insieme dei dati definiamo la

Media Campionaria
$$\frac{x_1 + x_2 + \dots + x_n}{N}$$

Per misurare il valore in posizione centrale (considerando l'insieme di dati ordinato), utilizziamo la

Mediana

- Se N dispari - $X_{\frac{N+1}{2}}$
- Se N pari - (manca da scrivere la formula in latex)

La mediana è insensibile alle code

Capitolo 3

Esame

L'esame sarà strutturato nella seguente maniera

Parte 1 - Teoria 8 Domande a risposta multipla - Punteggio 10/30

Parte 2 - Pratica 4 Esercizi a risposta aperta - Punteggio 20/30

Progetto (facoltativo) Progetto R, da consegnare prima dell'esame, può fornire un massimo di 2/30