

Probabilità e Statistica

UniVR - Dipartimento di Informatica

Fabio Irimie

2° Semestre 2023/2024

Indice

1	Cos'è la probabilità e la statistica?	2
1.1	Popolazione, variabili e campione	2
1.2	Parametro e Stima	2
2	Variabili	2
3	Statistica descrittiva	2
3.1	Strumenti di sintesi	3
3.1.1	Tabelle di frequenza	3
3.1.2	Distribuzioni	3
3.1.3	Distribuzioni cumulative	3
3.1.4	Grafici	3
3.1.5	Indici statistici-stime	3
4	Frequenze campionarie	3
5	Indici statistici	4
5.1	Indici di posizione	4
5.2	Indici di dispersione	4
5.3	Valori usuali	4

1 Cos'è la probabilità e la statistica?

La statistica è una scienza che si occupa di raccogliere, organizzare, analizzare e interpretare i dati. Nella statistica si cerca di estrapolare informazioni da esperimenti **aleatori** (esperimenti che non si possono ripetere esattamente allo stesso modo) e di prendere decisioni basate su queste informazioni. Ogni esperimento aleatorio ha bisogno di un **modello probabilistico** che ne descriva le caratteristiche principali.

1.1 Popolazione, variabili e campione

- **Popolazione:** tutti i possibili oggetti di un'indagine statistica
- **Individuo:** un singolo oggetto della popolazione
- **Variabile:** una qualsiasi caratteristica di un individuo della popolazione soggetta a possibili variazioni (es. altezza, peso, sesso, ecc.)
- **Range della variabile:** R_x è l'insieme di tutti i possibili valori che la variabile x può assumere
- **Campione:** un sottoinsieme rappresentativo della popolazione composto dalle variabili relative ad un sottoinsieme di individui
- **TODO**

1.2 Parametro e Stima

- **Parametro:** una misura che descrive una proprietà dell'intera popolazione
- **Stima:** una misura che descrive una proprietà del campione che fornisce informazioni sul parametro

2 Variabili

Le variabili possono essere di diverso tipo:

- **Variabili qualitative nominali:**
 - **Ordinali:** possono essere ordinate
 - **Non ordinali:** non possono essere ordinate
- **Variabili quantitative:**
 - **Aleatorie continue:** derivano da processi di misura e assumono i loro range. Sono sottoinsiemi reali
 - **Aleatorie discrete:** assumono valori interi

3 Statistica descrittiva

Consiste nella raccolta, organizzazione, rappresentazione e analisi dei **dati**.

3.1 Strumenti di sintesi

3.1.1 Tabelle di frequenza

Sono tabelle di individui con una certa caratteristica o aventi una caratteristica appartenuta ad un certo intervallo.

- **Frequenza assoluta:** conteggio del numero di individui
- **Frequenza relativa:** percentuale del numero di individui
- **Frequenza cumulativa:** conteggio o percentuale del numero di individui fino ad un certo punto

3.1.2 Distribuzioni

Sono rappresentazioni grafiche delle frequenze di una variabile. Possono essere:

- **Caso discreto:** valore variabile \rightarrow frequenza relativa
- **Caso continuo o numerabile:** intervallo di valori variabile \rightarrow frequenza relativa

3.1.3 Distribuzioni cumulative

Sono distribuzioni che rappresentano la frequenza cumulativa di una variabile. Possono essere:

- **Caso discreto:** valore variabile \rightarrow frequenza cumulativa relativa
- **Caso continuo o numerabile:** intervallo \rightarrow frequenza cumulativa relativa

3.1.4 Grafici

Sono rappresentazioni grafiche delle distribuzioni. Possono essere:

- **Istogrammi:** è costituito da rettangoli, insistenti sulle classi della partizione, attigui le cui aree sono confrontabili con le probabilità.
- **Poligoni di frequenza:** rappresentano le frequenze di una variabile
- **Diagrammi a torta:** rappresentano le frequenze relative di una variabile
- **Boxplot:** rappresentano le frequenze di una variabile

3.1.5 Indici statistici-stime

Sono misure quantitative che forniscono informazioni sulla distribuzione di una certa caratteristica.

4 Frequenze campionarie

Siano $\underline{x} = (\tilde{x}_1, \dots, \tilde{x}_n)$ i valori assunti da una variabile

5 Indici statistici

5.1 Indici di posizione

Forniscono informazioni del valore attorno al quale si posizionano i dati. Sono:

- **Media campionaria:** valore medio dei dati
- **Moda campionaria:** valore che si ripete più frequentemente. Ci possono essere più valori modali. Sia $\underline{y} = (y_1, \dots, y_n)$ il campione ordinato ($y_i \in \{\tilde{x}_1, \dots, \tilde{x}_n\}$ e $y_i \leq y_{i+1}$)
- **Mediana campionaria:** è il valore centrale del campione, una volta ordinato.

$$M = \begin{cases} y_{\frac{n+1}{2}} & \text{se } n \text{ è dispari} \\ \frac{y_{\frac{n}{2}} + y_{\frac{n}{2}+1}}{2} & \text{se } n \text{ è pari} \end{cases}$$

5.2 Indici di dispersione

Forniscono informazioni su quanto i dati si disperdono attorno ad un valore centrale. Sono:

- **Scarto Quadratico Medio:** misura la dispersione dei dati attorno alla media

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

- **Varianza campionaria:** misura la dispersione dei dati attorno alla media

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

- **Deviazione standard campionaria:** misura la distanza dei dati attorno alla media

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

- **Range:** Sia $\underline{x} = (\tilde{x}_1, \dots, \tilde{x}_n)$ un campione di dimensione n . Il range è definito come:

$$R = \max(\underline{x}) - \min(\underline{x})$$

5.3 Valori usuali

Si possono definire **valori usuali** di una variabile i valori del campione compresi tra:

- **Minimo valore "usuale":** media campionaria - 2 deviazioni standard
- **Massimo valore "usuale":** media campionaria + 2 deviazioni standard