

Elementi di Matematica e di Statistica

Distribuzioni Statistiche e Indici di Posizione

Docente: Riccardo Ievoli
`riccardo.ievoli@unife.it`

Corso di Laurea in Biotecnologie
a.a 2025-2026

08/10/2025

Outline

- 1 La Matrice dei dati
- 2 Distribuzioni di Frequenza
- 3 Misure di Sintesi

La Matrice dei dati

I dati vengono riportati in una tabella con le seguenti caratteristiche:

- ogni **riga** rappresenta una unità statistica e scorrendola è possibile leggere gli attributi di ogni unità statistica (soggetti)
- ogni **colonna** contiene i valori osservati per una determinata variabile in ogni unità statistica

La Matrice dei dati

Esempio (1)

- Matrice dei dati: si tratta di tabella unità \times variabili
- Esempio fittizio: dataset aziende italiane

ID	Geo	n dip	Settore	Fatturato ($\times 10^3 \text{€}$)
Id ₁	Nord	12	Edilizia	1250
Id ₂	Sud	100	Edilizia	100
Id ₃	Centro	15	Editoria	2000
\vdots	\vdots	\vdots	\vdots	\vdots
Id _n	Nord	30	Tessile	500

- Unità statistiche: ??
- Variabili **quantitative**: ??
- Variabili **qualitative**: ??

La Matrice dei dati

Esempio (2)

- Esempio fittizio: dataset pallacanestro

Giocatore	Gare	Minuti (media)	Ruolo	Punti (media)
G_1	32	16,2	Playmaker	4,8
G_2	17	13,7	Guardia	6,2
G_3	25	16,5	Centro	5,0
\vdots	\vdots	\vdots	\vdots	\vdots
G_n	28	27,5	Ala(G)	15,6

- Unità statistiche:** ??
- Variabili quantitative:** ??
- Variabili qualitative:** ??

La Matrice dei dati

Esempio

- Il primo approccio con la matrice di dati è di tipo **descrittivo**
- L'obiettivo è **esplorare** preliminarmente le variabili a disposizione.

Analisi univariate

Tecniche esplorative che permettono di sintetizzare l'informazione in una colonna (variabile) contenuta nel dataset.

Distribuzioni di Frequenza

Definizione

Sintesi descrittiva che permette di valutare la frequenza con cui la variabile studiata assume determinati valori

Dipende dalla tipologia di variabile (qualitativa/quantitativa):

- Variabili **qualitative**: tabelle, diagramma a barre (o a torta);
- Variabili **quantitative**: tabelle, istogramma, boxplot.

Distribuzioni di Frequenza

Focus: Variabili qualitative (e quantitative)

La variabile di interesse X presenta k diverse modalità nella matrice dei dati:

$$x_1, \dots, x_i, \dots, x_k.$$

Questa scrittura permette di indicare una sequenza di valori, con i che assume la funzione di indice.

Distribuzioni di Frequenza

Frequenze Assolute

- Sono definite come il conteggio di unità statistiche del campione che mostrano la i -esima modalità.
- La frequenza assoluta della modalità x_i è indicabile con n_i .
- Semplice proprietà, utile per definire l'operatore **sommatoria**:

$$\sum_{i=1}^k n_i = n_1 + \dots + n_k = n.$$

Distribuzioni di Frequenza

Frequenze Relative e Percentuali

- Non sempre le frequenze assolute risultano facilmente leggibili ed interpretabili
- É utile affiancarle alle **frequenze relative**, che forniscono immediatamente l'idea del peso che una modalità ha all'interno della distribuzione
- Sono definite come:

$$f_i = \frac{n_i}{n}.$$

Moltiplicando f_i per 100 si ottengono le frequenze **relative percentuali**

Distribuzioni di Frequenza

Tipologie di Frequenza

- Frequenze assolute, indicate con n_i
- Frequenze relative o percentuali, ossia f_i , p_i
- Frequenze cumulate assolute, relative o percentuali (N_i ; F_i ; P_i)

Analisi grafica delle frequenze

- Variabili quantitative: istogramma
- Variabili qualitative: diagramma a barre (grafico a “torta”)

Distribuzioni di Frequenza

Riepilogo

Semplici			Cumulate		
Absolute	Relative	Percentuali	Absolute	Relative	Percentuali
n_i	f_i	p_i	N_i	F_i	P_i
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
Tot.	n	1	100%	/	/

Schema riassuntivo delle frequenze

- **Relative:** $f_i = n_i / \sum_i n_i$; si ricordi che $\sum_i f_i = 1$
- **Percentuali:** $p_i = f_i \cdot 100 = n_i / \sum_i n_i \cdot 100$; $\sum_i p_i = 100\%$
- **Cumulate assolute:** $N_i = n_i + \sum_i n_{(i-1)}$
- Cumulate relative e percentuali

$$F_i = \frac{(n_i + \sum_i n_{(i-1)})}{n}; \quad P_i = F_i \cdot 100$$

Distribuzioni di Frequenza

Riepilogo (Esempio)

- Esempio: esito dell'esame di statistica
- Due 18, Tre 24, Due 27, Tre Trenta
- Numero di osservazioni: $n = \sum_i n_i = 10$

Voto	Semplici			Cumulate		
	n_i	f_i	p_i	N_i	F_i	p_i
18						
24						
27						
30						
Tot.				/	/	/

Misure di Sintesi

Obiettivo: Sintetizzare informazioni del campione o della popolazione

Tre tipologie di misure (indici)

- 1 Indici *assoluti*: espressi nella stessa unità di misura della variabile oggetto di studio. Esempio: media, mediana.
- 2 Indici *relativi*: Non sono vincolati ad alcuna unità di misura. Esempio: Coefficiente di Variazione
- 3 Indici *normalizzati*: assumono valori in un intervallo finito, $\in [0; 1]$ oppure $\in [-1; 1]$. Esempio: indice di correlazione.

Categorie di indici

- Indici di posizione (e tendenza centrale): medie, mediana, moda
- Indici di variabilità: Varianza, Range

Misure di Sintesi

Misure di tendenza centrale

In statistica è necessario individuare delle misure di sintesi utili per caratterizzare l'ordine di grandezza della variabile studiata.

La domanda guida è del tipo “*dove siamo?*” (si parla di *location*)

Famiglia delle misure di **tendenza centrale**. Alcuni esempi sono:

- La media aritmetica
- La mediana e altre misure di posizione (quartili e percentili)
- La moda

Misure di Sintesi

La Moda

È definita come la modalità del carattere che si osserva con la frequenza più elevata nel collettivo studiato.

Può essere calcolata per variabili qualitative e quantitative

Nel caso di variabili qualitative (o anche in qualche caso quantitative discrete):

- Per determinare la moda basta osservare la distribuzione delle frequenze
- La moda è la modalità che presenta il valore più elevato nella colonna delle frequenze.

Misure di Sintesi

La Moda (2)

Rapporto tra moda e distribuzione:

- Distribuzione **unimodale**: è presente una sola modalità relativa alla frequenza più elevata
- Distribuzione **multimodale**: se due (bimodale) o più modalità sono associate alla frequenza più elevata.
- *Assenza* di moda: tutte le modalità presentano la stessa frequenza.

Misure di Sintesi

La Moda: Esercitazione

Recap: La moda è la modalità che presenta la maggior frequenza.

- Esempio 1: risultati esame di statistica, $n = 10$

Matr.	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10
Voto	18	18	30	28	25	26	25	25	24	25

- Esempio 2: soddisfazione per un servizio ($n = 100$)

Modalità	Frequenze (n_i)
Per nulla soddisfatta/o	13
Poco soddisfatta/o	22
Abbastanza soddisfatta/o	48
Del tutto soddisfatta/o	17

Misure di Sintesi

Media Aritmetica

La media aritmetica può essere definita come il valore associato alle unità statistiche se ognuna di esse *ricevesse lo stesso ammontare della variabile oggetto di studio*.

Se consideriamo la sequenza degli n valori osservati come:

$$x_1, \dots, x_i, \dots, x_n;$$

si definisce media aritmetica:

$$\bar{x} = \frac{\sum_i^n x_i}{n} = \frac{x_1 + x_2 + \dots + x_i + \dots + x_n}{n}$$

Per calcolare la media aritmetica abbiamo bisogno che la variabile sia **quantitativa** (numerica).

Misure di Sintesi

Media Aritmetica: Proprietà

- 1 Identità di somma:** è il valore che, se sostituito al valore assunto da ogni unità, lascia inalterato l'ammontare complessivo del carattere. In simboli:

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

- 2 La somma algebrica degli scarti (differenze) dalla media è pari a 0:** definendo con scarto la differenza tra ogni valore osservato e la media \bar{x} , la somma algebrica di tutti gli scarti è pari a 0. In formule:

$$\sum_{i=1}^n (x_i - \bar{x}) = 0$$

Misure di Sintesi

Media Aritmetica: Proprietà (2)

- 3 Minimo della somma dei quadrati degli scarti:** se si considera ora la somma degli scarti al quadrato (devianza):

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

la media è quel valore che rende minima questa somma. In pratica, sostituendo qualunque altro valore al posto di \bar{x} , si ottiene un valore più elevato.

Misure di Sintesi

Media Aritmetica: Alcuni Commenti

- Quando la distribuzione è **simmetrica**, media aritmetica e mediana coincidono. Inoltre, se la distribuzione è anche **unimodale**, oltre che simmetrica, anche la moda coincide con mediana e media.
- Quando la distribuzione è **asimmetrica**, le tre misure iniziano a differenziarsi tra loro.
- La media aritmetica è sensibile ai valori estremi. Si dice che non è un indicatore **robusto**, al contrario della mediana (che vedremo tra qualche slide)

Misure di Sintesi

Media Aritmetica: Esempio

- Risultati esame di statistica, $n = 10$

Matr.	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10
Voto	18	30	28	30	25	26	26	25	24	25

- Qual è la media voto all'esame?

Misure di Sintesi

Misure di Posizione

- Le misure di posizione sono definite basandosi sull'idea di **ordinamento** delle osservazioni.
- Esempio: variabile altezza registrata per 10 cavalli. Si ordinano in maniera **crescente** le osservazioni.

141.5	165.9	161.1	145.0	161.1	165.8	154.4	155.5	151.2	160.2
				↓	↓				
				↓	↓				
141.5	145.0	151.2	154.4	155.5	160.2	161.1	161.0	165.0	165.0

Tali misure non possono quindi essere calcolate per le
variabili qualitative

Misure di Sintesi

La Mediana

Si tratta del valore della variabile che divide in due parti di eguale numerosità le unità statistiche (il collettivo):

- se n è dispari: la mediana è il valore assunto dall'unità statistica con posizione

$$Pos(Me) = \frac{(n+1)}{2}$$

ovvero $x_{[Pos(Me)]}$

- se n è pari: si hanno due posizioni centrali:

$$\left(\frac{n}{2}; \frac{n}{2} + 1\right)$$

- Se la variabile è numerica la mediana è la **semisomma** dei valori che occupano quelle posizioni $\frac{x_{[n/2]} + x_{[n/2+1]}}{2}$
- Se la variabile è qualitativa ordinabile si hanno **due modalità mediane**

Passi per il calcolo della mediana

- 1 Si ordina la distribuzione
- 2 Si calcola la posizione centrale (o le posizioni centrali)
- 3 La mediana è quel valore che occupa la posizione centrale (o la semisomma dei due valori che occupano le 2 posizioni centrali)

Misure di Sintesi

La Mediana (2)

Esercitazione: esempio Beverly Hills

- Dati: sette abitazioni con metrature diverse
- ...ma una di queste è una villa californiana
- Metri quadri: 80; 75; 110; 100; 85; 740; 70

Domande

Qual è la media?

Qual è la mediana?

Commentare i risultati ottenuti

Riassunto: cosa abbiamo imparato oggi?

- 1 La matrice dei dati
- 2 Distribuzioni di frequenza
- 3 Frequenze assolute, relative, percentuali
- 4 Misure di Tendenza centrale: media e moda
- 5 Misure di Posizione: mediana

Materiale supplementare

Esercizi per casa (1)

I $n = 7$ ospedali di una città metropolitana hanno le seguenti disponibilità in termini di posti letto:

18; 49; 524; 67; 67; 53; 1370

- 1 Calcolare la media aritmetica dei posti letto
- 2 Calcolare la mediana dei posti letto
- 3 Commentare i risultati ottenuti

Materiale supplementare

Esercizi per casa (2)

Si dispone di informazione sulla velocità massima di 16 calciatori.

Giocatore	G1	G2	G3	G4	G5	G6	G7	G8
Speed (km/h)	33,9	33,2	33,2	32,8	37,2	31,5	31,2	31,1
Giocatore	G9	G10	G11	G12	G13	G14	G15	G16
Speed (km/h)	31,0	30,3	29,8	29,7	29,3	25,6	20,5	20,3

- 1 Calcolare la media aritmetica della velocità massima
- 2 Calcolare la mediana della velocità massima
- 3 Commentare i risultati ottenuti

Materiale supplementare

Esercizi per casa (3)

Alla fine dell'anno accademico si registra il numero di esami sostenuto da una coorte di 160 iscritte/i al primo anno di un corso di laurea.

N. Esami	Semplici			Cumulate		
	n_i	f_i	p_i	N_i	F_i	p_i
0	30					
1	23					
2	25					
3	18					
4	19					
5	20					
6	35					
Tot.				/	/	/

- Completare la tabella delle frequenze
- Qual è la moda? Riuscite ad individuare la media e la mediana del numero di esami (facoltativo)?