

Statistica descrittiva



Analisi esplorativa e descrittiva dei dati



Primo approccio con i dati

- **Analisi esplorativa**

- Identificare possibili errori grossolani (es. codifica “99” per dati mancanti)
- Inquadramento della scala di misura delle variabili (molto rilevante per il software di analisi)

- **Analisi descrittiva**

- Sintetizzare l'informazione contenuta nei dati (indici di posizione, dispersione...)

- **Buona pratica:** resistere alla tentazione di partire «lancia in resta» e buttarsi ad analizzare qualsiasi insieme di dati...



Perchè descrivere i dati?

- Raramente i dati possono essere condivisi, specialmente nell'ambito medico (regole per il trattamento dei dati sensibili) → necessità di descriverli a chi non può vederli
- Potremmo essere interessati a confrontare i dati del nostro campione con quelli di un **altro campione** → per il motivo sopra, il dataset esterno non sarà a nostra disposizione e dovremo limitarci a **confrontare** informazioni riassuntive
- I grafici sono un ottimo punto di partenza, ma un confronto rigoroso deve essere basato su **quantità numeriche** → indici di posizione e dispersione



Terapia riabilitativa motoria

Il dataset **arm** contiene dati raccolti per uno studio sulla **riabilitazione motoria** a seguito di un infarto. Un gruppo di pazienti è stato sottoposto ad una terapia riabilitativa innovativa basata su una tecnologia di realtà virtuale, l'altro gruppo invece ha seguito una terapia convenzionale.

La principale variabile di interesse è la *Functional Independence Measure* (**FIM**), uno standard internazionale di misura della disabilità (è una scala da 0 a 126 punti che si calcola in base alla risposta di un questionario di 18 domande).



Terapia riabilitativa motoria

- 1) Esplorare i dati e descrivere il campione
- 2) Confrontare a livello descrittivo i due gruppi di pazienti in base al trattamento
- 3) Sintetizzare i risultati in una tabella

Il dataset “**Visite_pz_SCC**” contiene i valori di numerosi parametri clinici rilevanti per lo stato di salute di un gruppo di pazienti cardiologici.

Ottenere una tabella descrittiva per le variabili elencate.

*Cosa si può notare rispetto alla variabile BNP ? Il BNP è un peptide natriuretico cerebrale (o peptide natriuretico di tipo B) ossia è una sostanza prodotta dai ventricoli del cuore in risposta agli sbalzi di pressione che si verificano in caso di scompenso cardiaco.

Dal punto di vista della statistica descrittiva, la sua distribuzione ha delle caratteristiche particolari.

Come mai c'è tanta differenza tra la media e la mediana ? Che tipo di distribuzione presenta ?

Opzionale: che tipo di trasformazione della scala potrebbe essere opportuna per rendere la distribuzione meno «asimmetrica» ?

SESSO MASCHILE (N, %)	
Età (media ± dev.std)	
CMP_PERICARDIO (N, %)	
CPT_IPERTENSIVA (N, %)	
CPT_ISCHEMIA (N, %)	
DM_D (N, %) [diabete]	
EMBOLIA_POLMONARE (N, %)	
FIBRILLAZIONE_ATRIALE (N, %)	
IPERTENSIONE_ARTERIOSA (N, %)	
PAS (media ± dev.std)	
PAD (media ± dev.std)	
PAS (mediana e IQR)	
PAD (mediana e IQR)	
BNP (media ± dev.std)	
BNP (mediana e IQR)	

Creare una tabella descrittiva del dataset “Visite_pz_SCC” dove si riportano i seguenti dati confrontando gli uomini e le donne. Commentare quali variabili sembrano distribuite in modo diverso tra uomini e donne.

	UOMINI	DONNE
Età (media $\pm$ dev.std)		
CMP_PERICARDIO (N, %)		
CPT_IPERTENSIVA (N, %)		
CPT_ISCHEMIA (N, %)		
DM_D (N, %) [diabete]		
EMBOLIA_POLMONARE (N, %)		
FIBRILLAZIONE_ATRIALE (N, %)		
IPERTENSIONE_ARTERIOSA (N, %)		
PAS (media $\pm$ dev.std)		
PAD (media $\pm$ dev.std)		
PAS (mediana e IQR)		
PAD (mediana e IQR)		
BNP (media $\pm$ dev.std)		
BNP (mediana e IQR)		



Rappresentare graficamente le variabili riportate nella tabella precedente, sempre confrontando le loro distribuzioni negli uomini e nelle donne. Utilizzare i grafici opportuni, in base alla scala di misura della variabile considerata.