

Esercizio di riepilogo #01

Inserisci il file **S1_Dataset_red** in una cartella **nephrology_study_01**, crea un progetto R associandolo alla cartella e crea uno script per rispondere ai seguenti punti.

1. Importa il dataset
2. Uniforma i nomi con la funzione **clean_names**
3. Ricodifica le variabili qualitative usando le seguenti regole
 - **Sesso**: 0 = female, 1 = male
 - **Diab**: 0 = no, 1 = sì
 - **Cvd**: 0 = no, 1 = sì
 - **Fumo**: 0 = no, 1 = sì
 - **Stadio**: 1 = 3a, 2 = 3b, 3 = 4
 - **Nefrobases**: 1=Polycystic, 2=Diabetic nephropathy, 3=Hypertensive nephropathy, 4=Glomerulonephritis, 5=Other, 6=Unknown
 - **Ace**: 0 = no, 1 = sì
 - **Ara**: 0 = no, 1 = sì
 - **Aceara**: 0 = no, 1 = sì
 - **Epo**: 0 = no, 1 = sì
 - **Morte**: 0 = no, 1 = sì
4. Dividi la variabile **bmi** in 5 classi di uguale ampiezza aggiungendo la corrispondente colonna alla tabella
5. Dividi la variabile **bmi** in 5 classi di uguale frequenza aggiungendo la corrispondente colonna alla tabella
6. Esplora il dataset usando la **glimpse** (disponibile in **tidyverse**)
7. Esplora il dataset usando la funzione **skim** (disponibile in **skimr**)
8. Costruisci una tabella di frequenza per la variabile **nefrobase**
9. Costruisci una tabella di frequenza per la variabile **stadio**
10. Costruisci una tabella di frequenza per la variabile **nefrobase** e **stadio** riportando le frequenze assolute
11. Costruisci una tabella di frequenza per la variabile **nefrobase** e **stadio** riportando le frequenze percentuali (rispetto al gran totale)
12. Costruisci una tabella di frequenza per la variabile **nefrobase** e **stadio** riportando i profili colonna (le frequenze percentuali rispetto ai totali di colonna) e congiuntamente le frequenze assolute
13. Crea una nuova tabella **male_and_smoker** riportando solo i pazienti maschi fumatori
14. Crea una nuova tabella **male_or_smoker** riportando i pazienti maschi e le donne solo se fumatrici
15. Calcola le seguenti statistiche di sintesi per le variabili **Bmi**, **Age** e **Pas**:

- minimo
- primo quartile
- media
- mediana
- terzo quartile
- massimo
- scarto quadratico medio
- campo di variazione (differenza tra il valore minimo e massimo)
- indice di asimmetria (funzione **skewness** disponibile nel package **moments**)
- indice di curtosi (funzione **kurtosis** disponibile nel package **moments**)

16. Ripeti il punto precedente calcolando separatamente le statistiche per i pazienti maschi e femmine

17. Costruisci una tabella a doppia entrata riportando le modalità della variabile **nefrobase** sulle righe, le modalità della variabile **stadio** sulle colonne e la media della variabile **bmi** in corrispondenza di ciascun incrocio

18. Salva lo script in un file **esercizio_01_cognome_nome.R** (inserendo il tuo cognome e nome)

19. Salva il workspace in un file **esercizio_01_cognome_nome.RData**

20. Salva il progetto, comprimi la cartella associata al progetto in un file dal nome **esercizio-01-cognome_nome.zip**