

Esercizio di riepilogo #02

Inserisci il file **esercizio_01_cognome_nome.RData** che hai salvato al punto 19 dell'esercizio 1 in una cartella **nephrology_study_02**, crea un progetto R associandolo alla cartella e crea uno script per rispondere ai seguenti punti.

1. Carica il workspace contenuto nel file **esercizio_01_cognome_nome.RData** in memoria
2. Rappresenta la variabile **bmi** usando 15 classi di uguale ampiezza
3. Ripeti il punto precedente ma dividendo i dati in quattro pannelli in funzione delle tre modalità della variabile **stadio** (distribuzioni condizionate) e per tutti i dati (distribuzione marginale)
4. Rappresenta la variabile **bmi** usando un boxplot
5. Rappresenta la variabile **bmi** per i differenti livelli della variabile **stadio** usando boxplot paralleli
6. Rappresenta la variabile **nefrobase** usando un diagramma a barre
7. Costruisci la tabella di frequenza incrociando le variabili **nefrobase** (sulle righe) e **stadio** (colonne)
8. Rappresenta la variabile **nefrobase** usando un diagramma a barre impilato (`position = "fill"`) usando il colore per rappresentare le differenti modalità della variabile **stadio**
9. Rappresenta la variabile **nefrobase** usando un diagramma a barre accostato (`position = "dodge"`) usando il colore per rappresentare le differenti modalità della variabile **stadio**
10. Costruisci la tabella di frequenza incrociando le variabili **nefrobase** (sulle righe) e **stadio** (colonne) calcolando i profili colonna (distribuzioni di **nefrobase** condizionate su **stadio**)
11. Rappresenta la tabella ottenuta al punto precedente usando un diagramma a barre (suggerimento: `position = "dodge"`)
12. Rappresenta la variabile **stadio** usando un diagramma a barre
13. Costruisci un diagramma di dispersione delle variabili **eta** e **bmi**
14. Costruisci un diagramma di dispersione delle variabili **eta** e **bmi** aggiungendo sul grafico una retta di regressione lineare (suggerimento: funzione **geom_smooth**)
15. Costruisci un diagramma di dispersione delle variabili **eta** e **bmi** distinguendo i punti in base alle due modalità della variabile **morte**, usando il colore e la forma dei punti
16. Ripeti il punto precedente aggiungendo al grafico due rette di regressione lineare, una per ciascun gruppo di pazienti definiti dalle due modalità della variabile **morte** (suggerimento: funzione **geom_smooth**)
17. Crea una tabella con le sole variabili: **eta**, **e_gfr**, **bmi**, **pas**, **pad**, **u_prot** e **hb**

18. Calcola la matrice di correlazione a partire dalla tabella ottenuta al punto precedente (utilizza la funzione **cor** disponibile nella libreria base di R e la funzione **correlate** disponibile nel package **corr**)
19. Costruisci la matrice dei diagrammi di dispersione sulla tabella ottenuta al punto 17 (suggerimento: funzione **ggpairs** disponibile nel package **GGally**)
20. Salva lo script in un file **esercizio_02_cognome_nome.R**, il workspace in un file **esercizio_02_cognome_nome.RData** (inserendo il tuo cognome e nome) e comprimi il progetto in un file dal nome **esercizio-02-cognome_nome.zip**