Esercizio per “Data Analysis modulo 1”

Usando i dati nel file “heigth_weight” allegato a questo testo svolgere i seguenti punti:

1. caricare tutti i dati in un unico dataframe detto “TUTTI”.
2. creare due dataframe indipendenti uno MASCHI ed un FEMMINE che contengano rispettivamente i record degli individui di genere maschile e quelli di genere femminile.
3. far stampare a console il sommario statistico dei tre dataframe
4. creare tre figure, ciascuna con tre subplot (uno per ciascun dataframe) che mostrino rispettivamente gli istogrammi del BMI, dell'height e del weight su 20 bin
5. usando il dataframe TUTTI calcolare gli indici di correlazione di PEARSON tra le variabili numeriche (trascurare il genere)
6. Individuata la coppia di variabili in TUTTI che risultano maggiormente correlate calcolare un modello di regressione lineare per calcolare una variabile a partire dall'altra.
7. Formare un nuovo dataframe chiamato Z con due colonne: weight e height estratti dal dataframe TUTTI ma trasformati con il metodo Z-score.
8. Far stampare a console il sommario statistico di Z.
9. Calcolare un modello di regressione lienare che leghi tra loro le variabili di Z
10. Formare un nuovo dataframe Z2 che contenga solo i record per cui weight z-normalizzato si trovi tra il primo e il terzo quartile
11. 10 produrre un plot dei record in Z2
12. Calcolare un modello di regressione lineare che leghi tra lor le due variabili di Z2: è migliore di quello ottenuto usando l'intero Z?.