

Data Analytics: Homework 2

2025-05-23

Homework 2

Consegna 3 giugno 2025

Si considerino i dati contenuti nel file `fev.csv`. Si tratta di informazioni raccolte su 654 bambini e adolescenti. La variabile chiave è `FEV`, la capacità polmonare (in litri). Inoltre si dispone dell'età (`AGE`), della statura (`HEIGHT`, in pollici) e di due variabili dicotomiche, genere (`SEX` - 1 = M) e fumo (`SMOKE` - 1 = Fumatore).

1. Si svolga l'analisi univariata di ciascuna variabile, proponendo visualizzazioni dei dati adeguate.
2. Si usi uno strumento grafico per evidenziare le eventuali differenze nella capacità polmonare per i maschi e le femmine.
3. Si conduca una analisi di regressione per prevedere la capacità polmonare utilizzando come predittore la statura e si fornisca la rappresentazione grafica e gli indici per valutare la qualità di tale rappresentazione. Cosa succede se la statura viene trasformata in centimetri?
4. Si calcolino i residui della regressione di cui al punto 3 e si produca un grafico con sull'asse orizzontale il valore della variabile dipendente e su quello verticale i residui. Si aggiunga al grafico due linee orizzontali in corrispondenza di più o meno la radice quadrata della varianza dei residui.
5. Si ipotizzi che la relazione fra le due variabili non sia lineare, ma lineare a tratti. Cioè si immagini che vi sia una relazione lineare per le stature fino a 63 pollici e un'altra per le stature oltre 63. Si svolga l'analisi e si rappresenti questa su un grafico.
6. Si valuti la qualità della regressione a tratti con un opportuno misura in cui la qualità dipende dalla distanza del valore osservato da quello previsto al quadrato. Si confronti con quanto ottenuto al punto 3 e si commenti.
7. Si svolga una analisi come al punto 5 ma imponendo che fra la seconda retta di regressione (quella per i soggetti più alti) e la prima (per i soggetti bassi non vi sia discontinuità). Ovvero la retta di regressione fino a 63 ha in quel punto lo stesso valore della retta che parte da 63 in poi. Anche in questo caso si misuri la qualità della analisi e si confronti.