

Dataset and variables

Lung Capacity as a function of age, height, smoking status, gender and caesarian (dataset: **LungCapData.xlsx**).

In this study, we have the explanatory variables:

Age (years),

Height (inches)

Smoke (no/yes)

Gender (female/male)

Caesarian (no/yes)

and the Lung Capacity (response variable LungCap in liters).

Questions

- a) Run a simple linear regression model for each explanatory variable with response variable LungCap (remember to convert character variables to factors). Interpret the results.
- b) Run a multiple linear regression model with response variable LungCap and all explanatory variables with $p < 0.2$ in the univariable analysis. Interpret the results.
- c) How much of the variability in the response variable is explained by the model?
- d) Check the diagnostics for the multiple regression model graphically.
- e) Run Backward elimination and Forward selection. Is the same final model?
- f) Present your results in a table with all the necessary information.