

Regression models in R

Exercises

Antonio Gasparrini

01 November 2023

Sections 1-2

1. Using the `MTCARS` dataset in the package `DATASETS`, perform a linear regression between consumption (miles per gallon) and the following explanatory variables: gross horse power, weight, type of engine, and number of carburettors. The last variable must be categorical, with categories higher than two or not. Save the results in a new object `mylm`.
2. Determine the class of the object `mylm` and describe its structure. Extract the coefficients and their variance-covariance matrix, and then the value of the coefficient of determination R^2 .
3. Using the `ESOPH` dataset from the package `DATASET`, perform a logistic regression modelling the odds of being a case in terms of age and consumption of alcohol and tobacco (*hint*: use `cbind(ncases, ncontrols)` as the response variable), as suggested in the help page of `glm()`. Save the results in a new object `myglm`.

Sections 3-5

4. Derive the model frame of the regression model saved in the object `mylm`. Compare the first rows with those of the data frame `mtcars`. What do you notice?
5. Plot the residuals of the linear regression model vs the explanatory variable weight. The obtain the predictions for the observed weights, the average gross horsepower, a straight engine and two carburettors, and plot them against weights
6. Summarize the results of the two regression models stored in the objects `mylm` and `myglm`. What do you notice in the output of the logistic model?
7. Perform a significance test on the effect tobacco consumption on the risk of esophageal cancer in the logistic model, using a likelihood ratio test (see `help(anova)`).