Regression

Master Data Science
Summer 2023

Prof. Dr. Marlene Müller
marlene.mueller@bht-berlin.de

Exercises 5

Exercise 1

Load the dataset `Affairs` from the R package `AER` (`library(AER); data(Affairs)`). Check the data documentation: `?Affairs`.

To estimate a logit model, we need a dependent variable Y with only to values 1 and 0. A useful approach is to generate Y from the variable `affairs` (say $Y = 1$ if the number of of affairs is positive and $Y = 0$ otherwise). Do the following analyses:

(a) Explore graphically the effect of single explanatory variables on Y (use for example: `spineplot`, `barplot`, or `mosaicplot`).

(b) Fit at least three different logit models (some of them should be nested) and do interpret the estimated coefficients.

(c) Use one of your models to predict. The link here is logit, so the inverse link for prediction is $F(u) = \dfrac{1}{1 + e^{-u}}$. Derive the formula for doing it using a pocket calculator in an exam situation.

Exercise 2

Redo the analysis from Exercise 1 using probit models (use `link='probit'` in the `family` parameter of `glm`). In particular: How could you do the prediction here without using R?

Exercise 3

Compare your estimated models from Exercise 1. Instead of the F test for linear models we do now use χ^2 tests. The syntax is similar: `anova(glm1,glm2,test="Chisq")` Additionally, also compare AIC values and apply `stepAIC`.

Exercise 4

Load again the dataset `Affairs` from the R package `AER`. We aim to estimate a poisson regression now, so that we use $Y =$ `affairs`. Do the following analyses:

(a) Explore graphically the effect of single explanatory variables on Y (cf. Exercise 1(a)).

(b) Fit at least three different poisson models (some of them should be nested) and do interpret the estimated coefficients. Use the canonical link function.

(c) Use one of your models to predict. The canonical link here is $\log$, so the inverse link for prediction is $\exp$. Derive the formula for doing it using a pocket calculator in an exam situation.

Exercise 5

Compare your estimated models from Exercise 4. Instead of the F test for linear models we do now use χ^2 tests. The syntax is similar: `anova(glm1,glm2,test="Chisq")` Additionally, also compare AIC values and apply `stepAIC`.