PROJECT 3: OTHER REGRESSIONS
MASM22/FMSN30: LINEAR AND LOGISTIC REGRESSION 7.5 HP, 2022
Oral presentation: **23–25 May 2022**
Submit the slides **before** the presentation

Plasma Beta-carotene concentrations/Number of days in hospital

The objective is to continue either

(A) Project 1 and re-model the concentration of plasma beta-carotene, including the 0, or

(B) Project 2 and model the actual number of days in hospital, `hospdays` added in the complete data set `hospitaldays.txt`,

using some other type of regression. Choose one of the three alternatives below, giving you six different combinations to choose from.

For any of the approaches you should determine the best model (in some sense) using all the tools at your disposal. You can transform some of the x-variables and experiment with adding interactions between variables where this might improve the fit. You can use `cut()` to divide continous variables into categorical versions, if necessary.

Concentrate your presentation on the results and discussion. Use suitable plots and comment on interesting finds. The scope of your work is that you should be able to give a presentation lasting 10–15 minutes.

Alt.1 Poisson and/or Negative binomial

Since (A) the concentrations `betaplasma` or (B) the number of days in hospital `hospdays`, are non-negative integer values, we can treat them as count data. Use `betaplasma` or `hospdays` as Y-variable and determine whether you can use a Poisson regression or if Negative binomial regression fits the data better. Use the usual model selection tools to find suitable x-variables for the model.

Alt.2 Ordinal logistic regression

Divide the concentrations, `betaplasma` or days in hospital, `hospdays`, into 3 (or 4) categories with increasing levels. Model the probabilities for the categories using ordinal logistic regression. Use the usual model selection tools to find a good model.

You can investigate the Goodness-of-fit comparing the predictions to the true outcome. The confusion matrix is now a 3×3-table but you can still calculate sensitivity and specificity (except that you get one for each category).

You could also ignore the fact that the categories are ordered and model the probabilities using a multinomial logistic regression instead and compare the results with the ones from the ordinal logistic model.

Alt.3 Quantile regression

Model some of the quantiles of `betaplasma` or `hospdays`. Start with only one continuous x-variable before you build a more complicated model.

End of Project 3