PROJECT 3: OTHER REGRESSIONS

MASM22/FMSN30: LINEAR AND LOGISTIC REGRESSION, 2019

Oral presentations

Monday 27 May, Tuesday 28 May, Wednesday 29 May

Instructions

Choose presentation time and form groups of (approximately) two students at

`https://matstat.sam.cs.lth.se/LabsSelectSession?occasionId=475`

The presentations take place in MH:Sigma which is located in the South end of the Mathematics library on the second floor. If the library is closed, go to the entrance by the South stairs and knock and I will let you in. After the presentation, and **before your oral exam**, or that of your group partner, you must e-mail the slides to `fmsn30@matstat.lu.se`. Subject: `Project3 by studid1 and studid2`.

If you do not give a presentation at any of these dates, you will have to wait until next year!

You may attend via Skype if you cannot be physically present.

Project 3 *does not* require a written report, you will instead produce a presentation (with slides). Each group gives an oral presentation (in English) lasting approximately 10 minutes (15 if you are three people in the group). The presentation should be split about evenly between the group members. It will be followed by a 5 minute discussion with at least one another group attending.

Focus on results and conclusions, and rehearse the presentation to keep the duration within the allowed 10 min or you might experience the following `http://goo.gl/Q1y3s` ;-) You should manage to say everything you planned to say without rushing through the slides.

Create presentation slides (in English) with the tool you are most comfortable with (e.g. Powerpoint or similar, Beamer-LaTeX,...). You can use my laptop (Windows 10) for the presentation, or bring your own. If you are using mine, please bring your presentation on a USB stick or email it to `fmsn30@matstat.lu.se`, subject: `Project3 by studid1 and studid2`, before you are due to present.

Important: if your presentation is not a pdf file (e.g. is a Powerpoint file) I strongly recommend to also bring a pdf version as a plan-B (missing fonts with computers different from the one used to create the presentation is not a rare event with Powerpoint files. This should not happen with pdf files).

1 Atmospheric particles in Olso — revisited

We will use the data from Project 1 again, this time trying to model the PM_{10} concentration directly. Since the loglin model seemed to be the best one we could suspect that a Generalized linear model with a log-link function, $\ln \mu_i = \mathbf{x}_i\boldsymbol{\beta}$, would be a good idea. Since the concentration is measured as integers a Poisson or a Negative binomial regression model might work. Or we could use Quantile regression and ignore the distribution type. Some things that you could do:

- Use each of the three model types (Poisson, Negative binomial and Quantile) to model how the PM_{10} concentration depends on the number of cars. Use your models to calculate 95 % prediction intervals for the concentration as a function on the number of cars. Compare the results using the different models and also with the loglin model in Project 1.
- Use one of the model types to model how the PM_{10} concentration depends on the number of cars, the windspeed, the temperature, the (categorized) temperature difference and the (categorized) wind direction. Fit the best model you can, using these variables.
- Other things you can think of. . .

2 Useful R-commands (see also Lecture 10)

- Fit a poisson regression:
  `mymodel <- glm(y ~ x, data = mydata, family = "poisson")`
  Fit a negative binomial regression. Requires `library(MASS)`:
  `mymodel <- glm.nb(y ~ x, data = mydata)`
  Fit a quantile regression for several $\alpha =$ `tau`. Requires `library(quantreg)`:
  `qmodel <- rq(y ~ x, data = mydata, tau = c(...))`
- Predicted means and their confidence interval
  ```
  x0 <- data.frame(x = c(...))
  x0$mu <- predict(mymodel, x0, type = "response")
  xb <- predict(mymodel, x0, se.fit = TRUE)
  x0$lo <- exp(xb$fit - 1.96*xb$se.fit)
  x0$hi <- exp(xb$fit + 1.96*xb$se.fit)
  ```
- Probability density function for a Po(*mu*)-distribution:
  `dpois(y, mu)`
  Probability density function for a NB(*mu*, *theta*)-distribution:
  `dnbinom(y, mu = mu, size = theta)`
- Likelihood ratio test for comparing a Poisson model (reduced) and a corresponding Negative binomial model (full).
  `-2*(logLik(mymodel.pois)[1] - logLik(mymodel.nb)[1])`
- Poisson and Negative binonial: Confidence intervals for regression parameter, likelihood ratio test, AIC/BIC, deviance residuals, Leverage and Cook's distance as before (see Project 2).
- For Quantile regression

  Confidence intervals calculated in
  `summary(qmodel)`
  Plotting β-estimates for several α-values:
  `plot(summary(qmodel))`
  Comparing full and reduced models:
  `anova(qmodel.red, qmodel.full)`