April 21, 2016
Erik Parner

PhD Course in Basic Biostatistics
Exam (J.nr.: 1050/25)

Practical information

Submission and deadline

Individual solutions should be handed in as a single pdf-file by email to the following email-address:

BBEKSAMEN@BIOSTAT.AU.DK

The pdf-file should be named: fullname.pdf. The file must contain both your written answers and the appendix (see below).

Your solution has to be submitted no later than Friday May 13, 2016, at 9 AM..

Guidelines, requirements, and hints for preparing solutions

- Answer all questions. In particular, be aware that some questions comprise several sub-questions that all must be addressed.
- Plot the data whenever reasonable using scatter plots, histograms, Q-Q plots, etc.
- Always specify the statistical model used in the analysis.
- Any quantification of the findings of a statistical analysis in terms of estimates should be accompanied by confidence intervals. Any comparative statement should be backed up by a test and a p-value.
- Model validation is an integral part of any statistical analysis. It is not necessary to ask for relevant model validation to be performed, this should always be done.
- Include the Stata commands used for the analysis (the do/syntax-files) and log/output-files in appendices.
  For the log file it is most convenient to use a monospaced font like Courier, Courier New, Lucida Console, Monaco and Consolas.
- No Stata code or output (except graphs) outside the appendix!
- Formulate the conclusions using relevant terms from the context of the study (it is important to be able to translate the findings from the statistical analysis into conclusions regarding the initial scientific question).

Background information on the data

The high birth weight and long duration of pregnancy found on the Faroe Island have led to the hypothesis that high intake of marine-fat-derived n-3 acids might prolong pregnancy. In a study by Olsen et al (Lancet, 1992) healthy women in week 30 of pregnancy were randomly assigned in a ratio 2:1:1 to daily intake of "fish oil", "olive oil" or "no supplement".

For the present purposes the groups "olive oil" and "no supplement" are combined to form a single "control" group, and we only consider women who gave birth in week 37—42. The data therefore consist of 244 women allocated to the "fish oil" group and 242 women in the "control" group.

The results are contained in the dataset faroe.dta which has the following 3 variables

```
Contains data from faroe.dta
  obs:           486
 vars:             3                          25 Oct 2011 15:03
 size:         4,860 (99.9% of memory free)
-------------------------------------------------------------------------------
              storage  display     value
variable name   type   format      label      variable label
-------------------------------------------------------------------------------
group           byte   %8.0g       trlab      treatment group
ga              byte   %8.0g                  gestational age
weight          float  %9.0g                  birth weight
-------------------------------------------------------------------------------
```

1. Compare the birth weights in the two groups without correction for gestational age. In particular, compute an estimate and a 95%-confidence interval for the difference in expected birth weights.

2. Compare the gestational ages in the two groups.

3. In the following the association between birth weight and gestational age is studied separately for the two groups, so the following should be done separately for each group.

    3.1. Establish a linear regression model of birth weight as dependent variable and gestational age as independent variable.

    3.2. In the regression model, let the intercept correspond to gestational age 37 weeks, and compute estimates and 95%-confidence intervals for all model parameters. What is the interpretation of the regression model and each of its parameters?

    3.3. Is the estimated slope significantly different from zero?

3.4. Based on the estimated model, calculate an interval that contains the birth weight of 95% of future newborns born at 40 weeks.

4. In the following we compare the regression models for the two groups.

4.1. Can the slopes in the two treatment groups be assumed equal?

4.2. Compare the birth weights in the two groups when correcting for gestational age.

5. We will here define post term delivery as a gestational age equal to 42 weeks.

5.1. Compare the risk of post term delivery in the two groups.

5.2. Compare the risk of post term delivery in the two groups when correcting for birth weight in two groups; ≤ 3 kg, >3 kg.

6. Write a brief summary of your analyses and conclusions.