October 27, 2015
Erik Parner

PhD Course in Basic Biostatistics
Exam (J.nr.: 1050/24)

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 November 20, 2015, 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 data consist of a random sample of 1000 families containing data on the first and second born child in the family. The included variables are

id	Family id
sex1	Sex of first born
bweight1	Birth weight of first born
mage1	Mothers age at birth of the first born
sex2	Sex of second born
bweight2	Birth weight of second born
mage2	Mothers age at birth of the second born

1. Consider the chance of getting a boy the second time. Estimate this if the older sib was a boy and if she was a girl.
   Test the hypothesis of no association between the sex of the first and the second child.

2. Is the time between the births of the two children depending on the sex of the first born?

3. Compare the birth weight of the first and second born.
   Compare the chance of a birth weight below 3600g of the first and second born.

4. Predict the birth weight of the second born (hint: create a prediction interval. Make an interpretation of the prediction interval).
   Predict the birth weight of the second born using the sex of the second born.

5. Predict the birth weight of the second born if the birth weight of the first born is 3000g (hint: establish a linear regression with the birth weight of the second born as outcome and birth weight of the first born as explanatory variable. Interpret the parameters in the regression model).
   Present a prediction formula for the birth weight of the second born for various values of birth weights of the first born.

6. Establish a multiple linear regression model with the birth weight of the second born as outcome and
   A. birth weight of the first born
   B. sex of the second born
   as explanatory variables.
   Test the hypothesis of equal slope for boys and girls.
   Based on the linear regression model predict the birth weight of the second born boy if the birth weight of the first born is 3000g.
   Present a prediction formula for the birth weight of the second born for various values of birth weights of the first born and sex of the second born.

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