13 September 2016
Erik Parner

PhD Course in Basic Biostatistics
Exercises, day 2

You are supposed to complete exercises 2.1, 2.2, 2.3 and 2.4 at the supervised exercises corresponding to day 2. Exercises 2.5 and 2.6 are *homework* for the next week. Before you do the exercises, you should review the "day2.do"-file which shows the analyses presented in the lectures of Day 2 – this do-file uses the two dataset *normtemp.dta* and *bwsmoking.dta*.

Exercise 2.1
The fish oil supplement data, *fishoil.dta,* contains the difference in systolic blood pressure. In this exercise we will consider the possible effect of fish oil supplement on the increase in *systolic blood pressure* and go through an analysis similar to the one you saw at the lecture.

1. Make Q-Q plots of the difference in systolic blood pressure for each of the two groups. What are your comments to the plots?

2. Make a short description of the difference in systolic blood pressure using
```
bysort group:summarize systol, detail
```
and
```
ttest systol, by(group).
```

Comment on the descriptives (not the test!).

3. Test the hypothesis of the common standard deviation in the two groups.

4. Find the common standard deviation. You have to do the calculations by hand!

5. Return to the output from the t-test command and write a conclusion on the possible effect of fish oil on the change in systolic blood pressure during pregnancy. The conclusion should contain information on size of the possible effect, whether or not it is statistical significant and a discussion on the validity of the assumptions behind the statistical analysis.

Exercise 2.2
In the experiment above two women in the control group had a decrease in the systolic blood pressure of more than 50 mmHg.
1. Exclude these two women and repeat the analysis above. How does this affect your conclusions?

Exercise 2.3

The data set *hp.dta* contains data on the heart period (the average time in ms between two consecutive heart beats) during night and day for a group of healthy persons divided into physically active and passive persons.

1. Describe the heart period during day for the 'active' and the 'passive' persons.

2. Calculate a 95% prediction interval for the heart period for a passive person during day. Do the same for the active group. Compare the intervals and comment on whether these are valid 95%-prediction intervals.

3. Compare the heart period during day in the two groups. The comparison should (among other things) include a non-parametric test of no difference between the two groups.

4. Write a conclusion on the possible difference between the heart period during the day for physical active and passive persons. The conclusion should contain information on size of the possible difference, whether or not it is statistical significant and a discussion on the validity of the assumptions behind the statistical analysis.

Exercise 2.4
This exercise focuses on power and sample size calculations.

1. Look at the plot on page 44 in today's lecture. Suppose that you want to design a study comparing two groups, with the same standard deviation (sd=8) and a hypothesised difference in means of 5:
How many participants should you have in each group to obtain a power of 90%.

2. Answer the same question using the formula on page 50 and using Stata.

3. Now, suppose that you have planned a study with one intervention group and one control group both with 200 participants. Your plan is to compare the systolic blood pressure in the two groups and you expect the mean SBP in the control group to be 130 mmHg and in the intervention group to be 125 mmHg. Furthermore, your best guess is that the standard deviation in both groups will be 25 mmHg. Determine the statistical power of this study.

Exercise 2.5
We continue with heart period data, *hp.dta*. Generate a new variable containing the natural logarithm of the heart period during day (i.e. log(day)).

1. Repeat what you did in exercise 2.3 on this variable.

2. Comment on the similarities and differences on the two set of analyses.

Exercise 2.6

Here we will return to the birth of sib pairs we considered last week.
Again we will restrict the analysis to sib pairs **where the oldest is a boy**.

1. Compare the birth weight of the youngest for girls and boys.
   Write a conclusion of the analysis. (The conclusion should contain information on size of the possible difference, whether or not it is statistical significant and a discussion on the validity of the assumptions behind the statistical analysis.)

2. Compare the difference in birth weight for the two sibs if the youngest is a girl to the difference if the youngest is a boy. Write a conclusion of the analysis. (The conclusion should contain information on size of the possible effect, whether or not it is statistical significant and a discussion on the validity of the assumptions behind the statistical analysis.)

3. Comment on the similarities and differences between the two analyses and results.