March 25th 2009
Morten Frydenberg

PhD Course in Basic Biostatistics
Exercises, day 7

Exercise 7.1
The file *postterm.dta* contains the post term delivery data used at the lectures day 7.
Below we will go through the analysis of these data in Stata.
We will use parity=0 as reference (`char parity [omit] 0`).

1. Make a two by two table showing the association between parity and post term delivery. Calculate, by hand, the odds for post term delivery in each of the two parity groups (no CI's).

2. Run the Stata command `cs ptd parity, or woolf`.
   Find all the numbers on page 4.

3. Run the Stata command `xi: logit ptd i.parity`.
   Find the log odds for the reference group.
   Find the log odds ratio with standard error and confidence intervals on page 11.
   The output also contains two z-tests, what do we test here?

4. Run the Stata command `xi: logit ptd i.parity, or`.
   Find the odds ratio with confidence interval and the z-test.
   ln(0.9) =- 0.10536 so run `lincom _Iparity_1 +0.10536`
   to get the test for OR=0.9 on page 12.

Generate a new variable `age30= age-30`.
We will now look at the association between age and the risk of post term delivery among women with parity ==0.

5. Run the Stata commands
   ```
   logit ptd age30  if parity==0
   logit, or
   lincom _cons,or.
   ```
   Find all estimates, confidence intervals, and the tests shown on page 19.

6. Calculate, based on the output, the odds ratio (with CI) for post term delivery comparing two women (both given birth to their first child), who differ 10 years in

age.
Run the Stata commands `lincom 10*age30,or` and find the result .

Back to comparing the parity groups.

7. Run the Stata command `xi: logit ptd i.parity*age30.`
   Find the estimates with confidence interval and the z-test from page 27.

8. Run the Stata command `xi: logit ptd i.parity age30.`
   Find the estimates with confidence interval from page 29.
   Run the Stata command `xi: logit, or.`
   Find the age-adjusted OR with confidence interval from page 29.
   Comment on the z-test.

Exercise 7.2

The file *tatsoib.dta* contains data concerning a clinical trial of a new drug. Seventy patients were randomized to one of two groups: **I** receiving the new drug and **II** receiving a placebo 'drug'. The effect of the treatment was measured as the change (after-before) in the concentration of a specific substance *Tatsoib* in the blood. The objective of the treatment was to increase the level of *Tatsoib*. An increase of 1mg/l was considered clinically important.
As the age was suspected to influence the change in *Tatsoib,* the age of the patient was registered. At the start of the data analysis the investigators compared the age distributions in the two groups. They concluded, not surprisingly, that there was no statistically significant difference between the age distributions.

1. Make a statistical analysis that confirms this result.

The investigators concluded: *"As there was no statistically significant difference between the age distributions in the two groups, age was not adjusted for in the statistical analysis of the effect of the drug."*

2. Describe the change in *Tatsoib* in each of the two groups.

3. Estimate the effect of the new drug adjusted for a possible placebo effect.
   Write a conclusion on the possible effect of the new drug.

Another way of measuring the effect of the treatment is to see how many patients experienced an increase in *Tatsoib.* (Twenty-six patients in group **I** and seventeen in group **II** had an increase in *Tatsoib.*)

4. Perform a relevant statistical analysis of these data.

5. Comment on the similarities and differences between the analyses and results in questions 3 and 4.

Now, let us return to the possible association between age and the change in *Tatsoib.*
First consider the placebo group **II**.

6. Argue that it is reasonable to describe the relation between age and the change in *Tatsoib* as approximately linear.
   Estimate parameters describing the linear relationship. Comment on the interpretation of these estimates.

7. Repeat this for group **I**.

Based on this we will now compare the two groups:

8. Show that the slopes in the two groups can be assumed to be identical..
   What is the interpretation of this?

Assuming that the slopes are the same, the two groups can be compared by an *analysis of covariance*, i.e. a linear regression including "group" and "age".

9 Do this and write a conclusion on the age-adjusted difference between the two groups.

10. Comment on the similarities and differences between the analyses and results in questions 3 and 9.
    Discuss the citation on the previous page.

An "age-adjusted" analysis of the binary outcome (question 4) can be made by a logistic regression.

11 Do this and write a conclusion on the age-adjusted difference between the two groups.
   Comment on the similarities and differences between this analysis and the result in question 10 and what you found in question 4.