October 10, 2016
Erik Parner

PhD. Course in Basic Biostatistics
Exercises, day 6

You are supposed to complete exercises 6.1, 6.2 and 6.3 at the supervised exercises corresponding to day 6. The rest of the time you should catch up on the first five days of exercises.

Exercise 6.1
The file *haemoglob.dta* contains the haemoglobin data used at the lectures day 6.
Here we will go through the analysis of these data in Stata.

1. Make a scatter plot of the haemoglobin levels against the type.

2. Make the three QQ-plots.

3. Run the Stata command `oneway haemo type, tab`. Find in the output:
   a) The count, mean and standard deviation in each group.
   b) Bartlett's test for equal standard deviations (variances).
   c) The estimate of the common variance (in the line within groups).
   d) The F-test of the identical means in three groups.

The `oneway` command will not give estimates of the differences between the groups or the confidence intervals for the means assuming equal standard deviations .
This is most easily done by the `regress` with or without the `nocons` option

4. Run the commands
   ```
   regress haemo ib1.type
   ```

   (Stata 10 and earlier:
   ```
   char type[omit] 1
   xi: regress haemo i.type )
   ```

   Find in the output:
   a) The estimate of the standard deviation.
   b) The F-test of the identical means in three groups.
   c) The mean for type 1, the difference in mean for type 2 compared to type 1 and the difference in mean for type 3 compared to type 1.
   Run:
   ```
   lincom 2.type – 3.type
   ```
   (Stata 10 or earlier: `lincom _Itype_2 - _Itype_3`)
   in order to get the difference between type 2 and 3.

5. Run the command

```
regress haemo ibn.type, nocons
```

(Stata 10 and earlier: `xi, noomit: regress haemo i.type, nocons`) to get the confidence intervals for the means assuming equal standard deviations

6. Generate a new variable containing the residuals (here we call this variable `res`)

```
predict res if e(sample), residual
```

Make a QQ-plot of the residuals under the model.

Note, the `lincom` and `predict` commands are so-called post-estimation commands. That is, they use the last `anova` or `regress` command as input.

7. Run

```
kwallis haemo, by(type)
```

to get the Kruskal-Wallis test.

Exercise 6.2

Twenty-two patients undergoing cardiac bypass surgery were randomized to one out of three ventilation groups (Amess *et al., Lancet* 1978). The red cell folate levels (μg/l) after 24 hours in three groups of cardiac bypass patients given different levels of nitrous oxide ventilation are found in *folate.dta.*

1. Compare the three methods of ventilation.

Exercise 6.3

In a randomized trial involving terminally ill cancer patients (*Cameron & Pauling, Proc. Natl Acad. Sci. USA* 1978) looked at the time to death of patients suffering from different types of cancer. The data set *cancer.dta* is considering a small subset of the study. **Note, all patients died during the follow-up period, i.e. no data were censored.**

One way to compare the time to death in different groups is to assume a so-called *accelerated waiting time model*, which is given by:

time to death in group 2 = const * time to death in group 1

Use what you have learned today together with a suitable transformation to analyze the data and answer the following questions:

1. Is there a statistical significant difference in the median survival for the five types of cancer?

2. Estimate the median survival in breast and stomach cancer.

3. Estimate the ratio between the median survival in breast and stomach cancer.

4. Discuss the validity of the accelerated waiting time model for these data.