

PhD Course in Basic Biostatistics

Exercises, day 3

You are supposed to complete exercises 3.1, 3.2, 3.3 and 3.4 at the supervised exercises corresponding to day 3. Exercises 3.5 and 3.6 are homework for the next week.

Exercise 3.1

Consider the data set concerning pre- and post-menstrual energy intake from the lecture, *energy.dta*

1. Make four plots similar to the ones presented on pages 12, 13 and 16.
2. Make the statistical analysis found on page 17.
3. Make the plot found on page 21.

Exercise 3.2

Ten persons killed by car accidents had their alcohol concentration measured in two blood samples: one from the left leg and one from the heart, *alcohol.dta*.

1. Plot the data. What do you see in your plots? Is it reasonable to describe the difference between the concentrations measured in the leg and in the heart by a normal distribution?
2. Estimate the mean and the standard deviation of the difference between the concentrations measured in the leg and in the heart.
3. Calculate 95% confidence intervals for the mean difference between concentrations measured in the leg and in the heart.
4. Calculate 95% prediction intervals for the observed difference between concentrations measured in the leg and in the heart.
5. Write a short conclusion on the difference between the alcohol concentration measure in left leg and from the heart. The conclusion should contain information on size of the possible difference and a discussion on the validity of the assumptions behind the statistical analysis.

Exercise 3.3

The table below contains the systolic blood pressure of 16 patients before and after one week's treatment with captopril or placebo. Data are from the randomized controlled trial: *Effect of captopril on kidney function in insulin-dependent diabetic patients with nephropathy* (Hommel *et al.*, *BMJ*, 1986).

Systolic blood pressure of 16 patients before and after treatment with Captopril or Placebo.

Captopril		Placebo	
Baseline	After 1 week	Baseline	After 1 week
147	137	133	139
129	120	129	134
158	141	152	136
164	137	161	151
134	140	154	147
155	144	141	137
151	134	156	149
141	123		
153	142		

The data are found in the file *captopril.dta*.

The authors made a separate analysis for each treatment by means of a paired t-test. They found a significant decrease of the blood pressure in the Captopril group, however the analysis of the placebo group did not show any significant change. The authors concluded: “*Captopril represents a valuable new drug for treating hypertension in diabetics dependent on insulin with nephropathy*”.

1. Reconstruct the original analysis of the authors. What is wrong in their analysis of the data and their interpretation of the result?
2. Make a more correct analysis of the data. What is your conclusion?

Exercise 3.4

We return to the data set with heart period for physically active and passive persons, *hp.dta*. It is well-established that the heart rate is lower at night compared to during the day and that the physically active persons have a lower pulse at rest (compared to persons that are not physically active).

Here we want to compare the “day-night change” for the two types of persons.

1. Make a statistical analysis that compares the *difference* between the day and night heart periods for the two groups.
Write a short conclusion containing information on size of the possible difference and a discussion on the validity of the assumptions behind the statistical analysis.

Another choice is to consider the ratio, i.e. analysing $\log(\text{night/day})$.

2. Make a statistical analysis that uses this measure to compare the two groups. Write a short conclusion containing information on size of the possible difference (both on the log scale and on the original scale) and a discussion on the validity of the assumptions behind the statistical analysis.
3. Compare what you found in 1 and 2 above. Which analysis gives the “best” answer to the question: “Does size of the day to night change depend on whether or not you are physically active?”

Exercise 3.5

Here we will return to the birth weight of sib pairs. In the analyses we have considered the difference in birth weigh, i.e.

$$\text{“birth weight second child} - \text{birth weight first child”}, \quad (\text{A})$$

when we compared birth weight of the two sibs.

Another choice would be to compare the sibs by considering the birth weight ratio:

$$\text{“birth weight second child} / \text{birth weight first child”}. \quad (\text{B})$$

1. Consider sibs pairs of two brothers and make diagnostic plots that examine the two choices (A) and (B). What do you think is the best choice and why?

Exercise 3.6

The Stata-file *temp3.dta* contains measurement of body temperature in 96 patients using rectal Hg thermometer, oral Hg Thermometer, and an electronic Craft thermometer.

1. Compare the two oral temperature measurements.
Is there a systematic difference?
Calculate a 95% prediction interval to illustrate the size of the random difference between the two measurements.
2. Write a short conclusion containing information on size of the possible difference and a discussion on the validity of the assumptions behind the statistical analysis.
3. For a few patients the measurements differ substantially. For these patients, try to identify the measurement which may be in error.