Practical Exercises for **Day 6**

Sonja Hartnack, Terence Odoch & Muriel Buri

October 2017

Exercise 17

- Load the `ToothGrowth` data set and run the following four linear regression models.

```
mod1 <- lm(len ~ dose.fac, data = ToothGrowth)
mod2 <- lm(len ~ supp, data = ToothGrowth)
mod3 <- lm(len ~ dose.fac + supp, data = ToothGrowth)
```

- Have a look at the summary of these models.

- How do you interpret the model coefficients?

- Which model is best?

Exercise 18

- Load the `water` data set and fit a multiple linear regression model. Use `mortality` as your response variable and add `hardness` and `location` as an explanatory variable.

- Check the underlying model assumptions.

- Add an interaction term between `hardness` and `location` to the above estimated multiple linear regression model.

- Interpret the interaction coefficient `hardness:locationSouth`.

- Check the underlying model assumptions.

- Which one is the better model? With or without the interaction term?

- How to derive confidence intervals for the regression coefficient of `hardness` and `location`?

Exercise 19

Hypothetical example - from Kirkwood and Sterne, Medical Statistics, 2nd ed., p. 177

- Read in the data set `lepto`. This study presents a serology survey of leptospira sero-prevalence in rural and urban areas of the west indies.

- Encode the numeric variable `antibodies` as a factor with levels 0 and 1.

- Make a crosstable with the risk factor `exposure` and `antibodies`.

- Run a Chi-squared test, a Fisher's exact test and a logistic regression (`glm`) to assess if the `exposure` (living in rural vs. urban areas) is a risk factor.

- Create a subset for `male` and `female` based on the variable `gender`.

- Repeat the crosstable, Chi-squared test, Fisher's exact test and a logistic regression (`glm`) for the subsets **separately**.

- Does the conclusion of your research question change with the analysis of the subsets? (Research question: Is the `exposure` (rural and urban areas) a risk factor?)

- Fit a logistic regression model (`glm`) with `exposure` and `gender` as explanatory variables.

- **SPECIAL FOR GUMA**: Is `exposure` being from an urban area a risk factor?