Instructions for Reports

Anastasija Tetereva

0. General Guidelines

What follows here are guidelines for writing your reports (both the group assignments and the individual assigment).

- Length of your report:
  - 4 pages (not having ridiculously small fonts), preferably no appendix.
  - 2 pages for code.
- Clear style of writing.
- Tables and figures are numbered and they must be referred to in the text. Only include a table or figure if it supports the story you want to tell. In the text, capitalize numbered items (thus Table 3, Figure 1, Section 2) but lower case should be used whenever the item is not numbered (thus as can be seen in this figure, the previous table, the next section).
- Figures and tables have a caption that has sufficient information to understand what can be seen in the figure or read from the table.
- Always motivate a choice that you have made. Why would you choose for ridge instead of LASSO regression? Why do you choose to interpret certain weights and other not?
- Use consistent notation: the mathematical symbols used need to be formatted exactly the same everywhere in your text. For example, if $\mathbf{X}$ is the data matrix, then it should always be referred to as $\mathbf{X}$ and not as X, $\mathbf{X}$, or X.
- Equations are a part of the sentence and never appear on their own. If the sentence ends after an equation, then write a period. If the sentences continues, a comma after the equation may be needed.
- No subsections: the report is so short that subsections do not make sense.
- We strongly recommend to use R Markdown for your report. The present document can serve as an example. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. The main advantage is that the text of your report and the code for generating tables and figures are in the same document. As such, it allows to reproduce the results in your report very easily. It is nicely integrated in R-Studio. In the appendix, we give some examples on how R-code, plots, tables, mathematical symbols, and equations are integrated in the text. For more details on using R Markdown see http://rmarkdown.rstudio.com.
- Hand in your assignment as a pdf file. If you have LaTeX installed on the same computer as where you run R Markdown, you may be able to choose the option Knit to PDF to create the pdf file directly from R-Studio. Without LaTeX, you can use Knit to HTML and open the created html file in your browser and print it to a pdf file from your browser.

1. Introduction

- There should be a substantive research question and beforehand you need to introduce some reasoning why the research question is interesting and/or relevant.
- The technique is a tool to answer your research question so (in this course) it is not a part of the research question.
- The introduction is typically one paragraph (usually less than half a page).

2. Data

Provide enough description of the data so that the reader understands where they come from, who gathered them, what the values mean and possibly some simple (univariate statistics). Usually, this section is not more than half a page.

3. Methods

- The methods section should be written for general data, not for the specific data at hand.
- First paragraph should discuss the main feature of the technique in intuitive terms.
- Then a more technical discussion of the technique should be given (and generally this will include some formulas).
- After this diagnostics (could be t-values, R^2, etc.) should be discussed.
- Combine the diagnostics with what you are looking for: e.g., we prefer larger R^2 values as more variance of the dependent variable is explained by the model.
- Whenever you introduce a mathematical symbol, be sure to define explicitly what it is.

4. Results

- No new diagnostics should be introduced here (as you do this in the methods section).
- Only present relevant tables and figures.
- Carefully check if your plot needs to have an aspect ratio of 1 (as for biplots as without it projection does not work anymore).
- We would like to see a serious substantive interpretation (say about half a page or so).

5. Conclusions and Discussion

Usually one or two paragraphs that answer your research question based on the results you found. Possibly discuss how the research could have been done better

6. Code

Code should be clear, have comments, should do what it is supposed to do. An example of the R code of the first lecture is given here.

```
load("Advertising.Rdata")  # Load the Advertsing data set
attach(Advertising)        # Set focus to the Advertising data set
                           # This allows to use variable names as
                           # variables in the model specification
result <- lm(Sales ~ TV)   # Call the linear regression model (lm):
                           # Sales dependent, TV as predictor
summary(result)            # Give a summary of the results object
```

References

References should go here.

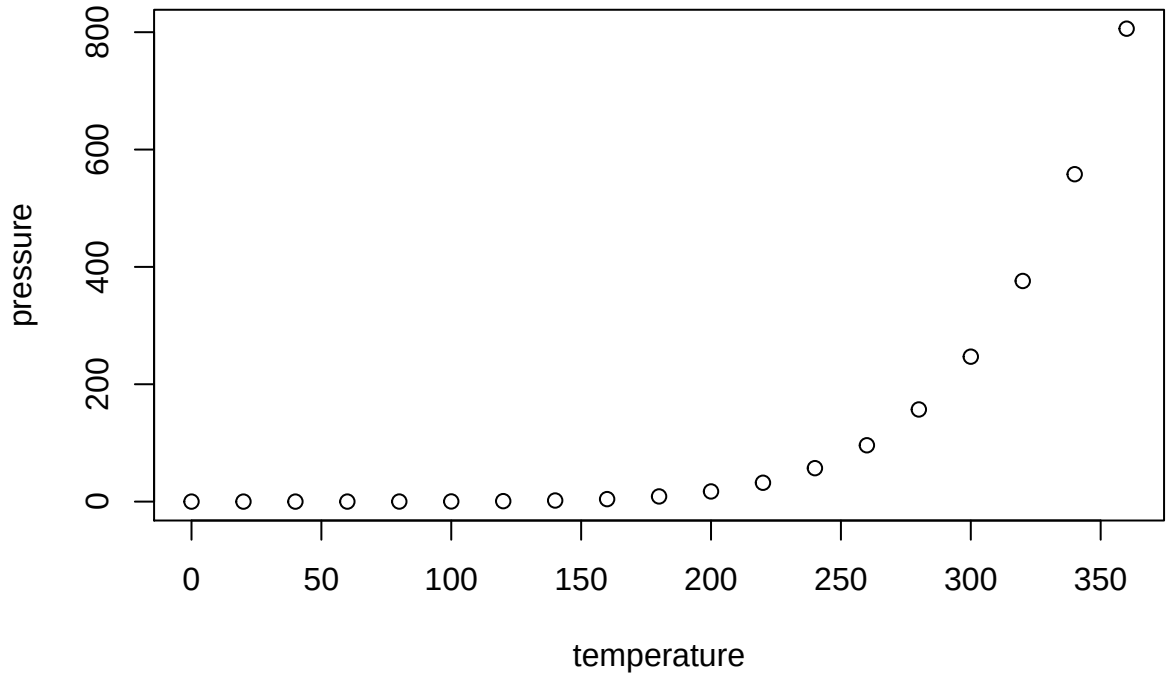

Figure 1: **Figure 1**: This is the caption of a scatterplot showing pressure as a function of temperature.

Appendix A. R Markdown

The present document is an example of an R Markdown document. When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
summary(cars)
```

```
##      speed           dist
##  Min.   : 4.0   Min.   :  2.00
##  1st Qu.:12.0   1st Qu.: 26.00
##  Median :15.0   Median : 36.00
##  Mean   :15.4   Mean   : 42.98
##  3rd Qu.:19.0   3rd Qu.: 56.00
##  Max.   :25.0   Max.   :120.00
```

You can also embed plots, for example:

Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot. Other options can be specified between the { } characters and affect how the R output is being presented. For more details, see https://yihui.name/knitr/options/. For some reason, the captions of the figures and tables in the present document produces a single figure or table number in the html version, but a double one in the pdf version. Do not worry about that.

An example of a table is given in Table 1.

Table 1: **Table 1**: The first six rows of the pressure data.

temperature	pressure
0	0.000
20	0.001

temperature	pressure
40	0.006
60	0.030
80	0.090
100	0.270

A.1. Getting Started with Equations in Markdown

Equations and mathematical symbols can be easily added to a markdown file. A fraction of two-thirds can be written as $\frac{2}{3}$. Other expressions are, e.g., $\hat{\lambda} = 1.02$ and $\sqrt{4} = 2$. The formulas so far were inline formulas. Formulas can appear also appear on their own line (in so-called display math mode)

$$\alpha, \beta, \gamma, \Gamma,$$

$$a \pm b,$$

$$x \geq 15,$$

and

$$a_i \geq 0 \quad \forall i.$$

An $n \times m$ matrix $\mathbf{A}$ can be shown as

$${}_n\mathbf{A}_m = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{pmatrix}.$$

A.2. Other Examples for Formulas and Equations

The binomial probability:

$$f(y|N,p) = \frac{N!}{y!(N-y)!} \cdot p^y \cdot (1-p)^{N-y} = \binom{N}{y} \cdot p^y \cdot (1-p)^{N-y}.$$

To calculate the **mean** of n observations of variable x, you can use:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i.$$

Note that this equation looks quite nice above where it's in display math mode. It is more compact but not quite as nice looking if we present it using inline mode, e.g., $\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$.

Let us do the same with the equation for **variance**. First the inline version, which is $\sigma^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}$. And then the display mode version:

$$\sigma^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}.$$

Next, it is good to look at the equation for **covariance** to see how it is just a generalization of variance to two variables. An inline version of the equation is $cov_{x,y} = \sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y})/(n-1)$. And, the display mode is:

$$cov_{x,y} = \frac{\sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y})}{n-1}.$$

And, finally, we will end with the **standard deviation**. Here is the inline version, $\sigma = \sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2 / (n-1)}$.

And here is the display version.

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}}.$$

These equations are coded in LaTeX. There helpful online editors to build up the equation, such as http://visualmatheditor.equatheque.net/VisualMathEditor.html. Make your equation there and copy the code into the RMarkdown document in between dollar signs (1 or 2 on either end depending on whether you want the equation in line or in display mode).

Equation numbering can be done manually as follows:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}}, \quad (1)$$

$$cov_{x,y} = \frac{\sum_{i=1}^{n} (x_i - \overline{x}) \cdot (y_i - \overline{y})}{n-1}. \quad (2)$$