Swiss Federal Institute of Technology Zurich

Seminar for
Statistics

Departement of Mathematics

Master Thesis

Spring 2015

Nicolas Bennett

The title of my thesis which should be split on several lines if it is too long

Submission Date: August 25th 2015

Co-Adviser: Anna Drewek
Adviser: Prof. Dr. Peter Bühlmann

[B]ecause we tend to reward others when they do well and punish them when they do badly, and because there is regression to the mean, it is part of the human condition that we are statistically punished for rewarding others and rewarded for punishing them — Kahneman (2012).

Preface

First words and acknowledgements. And we add a lot of text to make sure that it spans more than one line, as otherwise it may not show up.

Abstract

This example thesis briefly shows the main features of our thesis style, and how to use it for your purposes.

Contents

List of Figures

List of Tables

Notation

Explain your symbols and abbreviations.

Chapter 1

Introduction

Infectious diseases have played an undeniably important role in human history. With human populations becoming sufficiently aggregated to sustain direct life cycle viral and bacterial infectants around 2000 BC, devastating invasions of a growing number of pathogens started to occur (Dobson and Carper, 1996).

One of the earliest well documented incidence of a large-scale epidemic is known as the Plague of Athens. Starting in 430 BC and lasting roughly three years, a highly infectious disease killed 75'000 to 100'000 people or 25% of Athen's population. This catastrophic event is attributed either to smallpox, a viral infection with *Variola major* of typhus, caused by *Rickettsia* bacteria (Littman, 2009).

The bacterium *Yersinia pestis* caused three major plague pandemics in the early and late middle ages, as well as in the late 19th century. Originating in northern Africa in 523 AD and spreading around the Mediterranean basin throughout the years 541–546, the Plague of Justinian is assumed to have killed up to half of the population of affected areas. The effect on cities was disproportionately severe. In Constantinople, for example, an estimated 230'000 people out of 375'000 lost their lives to the disease (Treadgold, 1997). Returning in the years 1347–1351, known today as the Black Death, a plague pandemic again wiped out around half of Europe's population. Death toll estimates range from 15 to 23.5 million (Zietz and Dunkelberg, 2004). Leaving behind a grim cultural heritage, this catastrophe had a lasting effect on economic and social structures in Europe. The third large-scale outbreak started around 1855 in southern China and quickly spread to Japan, Taiwan and India again wreaking havoc on the affected population.

Bringing diseases such as smallpox, measles (an infection with the *Measles virus*) and typhus to the Americas during the European invasion of the New World had grave repercussions for the indigenous population, carrying no natural resistance towards the newly introduced pathogens. It is estimated that the population of present day Mexico fell from 20 million to 1.6 million over the course of the 16th century due to multiple disease epidemics, critically contributing to the successful colonization of the new continents (Dobson and Carper, 1996).

Cholera and influenza are further contagious diseases with high mortality rates, responsible for global epidemics. *Vibrio cholerae*, a bacterium which causes infections of the intestine, became widespread in the early 19th century and caused seven pandemics since, the last of which only started in 1961. Antibacterial treatment of sewage and purification of drinking water greatly help to prevent and contain spreading of the disease but in areas with inadequate sanitation, such as Haiti after the 2010 earthquake, it remains a pathogen difficult to control. The influenza virus causes seasonal epidemics characterized by low lethality rates among people with intact immune systems[1]. Irregularly occurring influenza pandemics, initiated by zoonosis of new virus strains, against which no natural immunity exists, however, are accompanied by much higher lethality rates. The most significant such event is known today as the Spanish flu pandemic of 1918, costing the lives of 50–100 million, nearly half of which were young, healthy adults (Taubenberger and Morens, 2006).

With better knowledge of these diseases, effective countermeasures could be developed. Identifying vectors and natural reservoirs, as well as understanding how transmission between infected individuals occur greatly helps to stymie burgeoning outbreaks of infective pathogens and prevent their spreading. In the case of plague, insecticides killing fleas were successfully used as a prophylactic measure, as well as controlling rat populations. Improvements in sanitary conditions and general population health are further important contributing factors to the decline of certain infectious diseases. Most important of all are advancements in medicine such as the development of vaccines and antibiotics. Among the great successes of widespread vaccination efforts is the global eradication of smallpox through a coordinated initiative lead by the World Health Organization in the 1970's.

Despite development of means to treat and prevent many previously devastating diseases, infectious pathogens remain a serious threat to global health. In 2012, an estimated total of 58.3 million people died (20.1% in high, 29.4% in upper-middle, 36.5% in lower-middle and 14% in low income countries). Figure 1.1 partitions the total death count into World Bank income groups and causes. In low income countries, infective diseases are the most prevalent cause of death (39.6%), followed by maternal and perinatal complica-

[1] In spite of low lethality, these seasonal epidemics still incur significant economic damages. The World Health Organization (2003) estimates annual health care costs and loss of productivity due to influenza at US $71–176 billion for the United States of America alone.

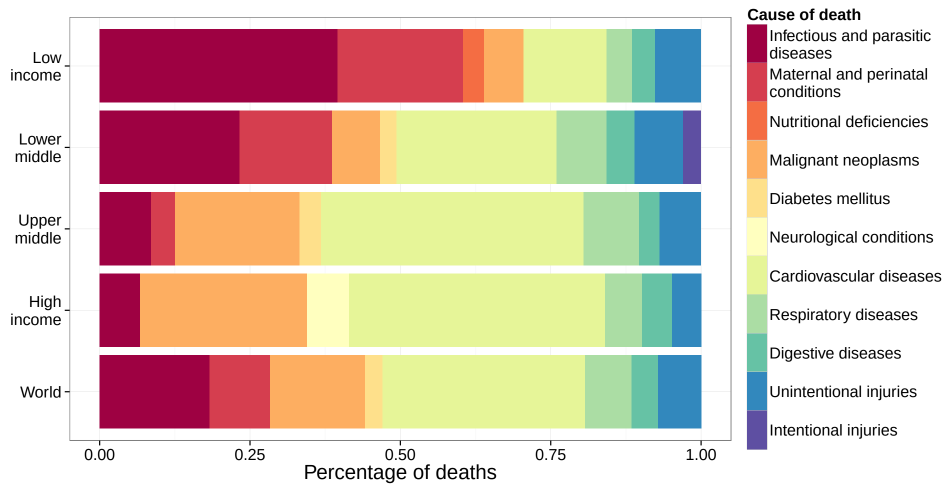

Figure 1.1: Relative frequencies of death causes in 2012 by World Bank income groups. Binning is based on Gross National Income (GNI) per capita and the thresholds are \$1'045 or less for low income, \$1046 to \$4125 for lower-middle, \$4126 to \$12745 for upper-middle and \$12746 or more for high income economies. The data was obtained from the World Health Organization (2012).

tions with substantial margin (20.8%). In lower middle income countries, cardiovascular conditions catch up (26.5%), almost matched in frequency by infectious diseases (23.3%). In upper middle (8.5%) and high income countries (6.7%), the importance of infectious disease is weakened but is still accountable for a significant number of deaths. Globally, infectious diseases are the second most frequent cause of death (18.3%), only preceded by cardiovascular diseases (33.7%).

Focusing only on deaths caused by infectious disease, lower respiratory infections are most frequent (for each income region individually, low to high: 28.7%, 30.8%, 43.5% and 57.7% as well as worldwide: 34.5%; cf. figure 1.2). Diarrhoeal diseases and HIV/AIDS are the next most common worldwide (16.9% and 17.3%, respectively) where diarrhea is more prevalent in lower income regions (16.6% and 21.4% versus 7% and 5.6%), while AIDS plays a major role irrespective of income region (low to high: 20.4%, 13.3%, 26.2% and 11.3%).

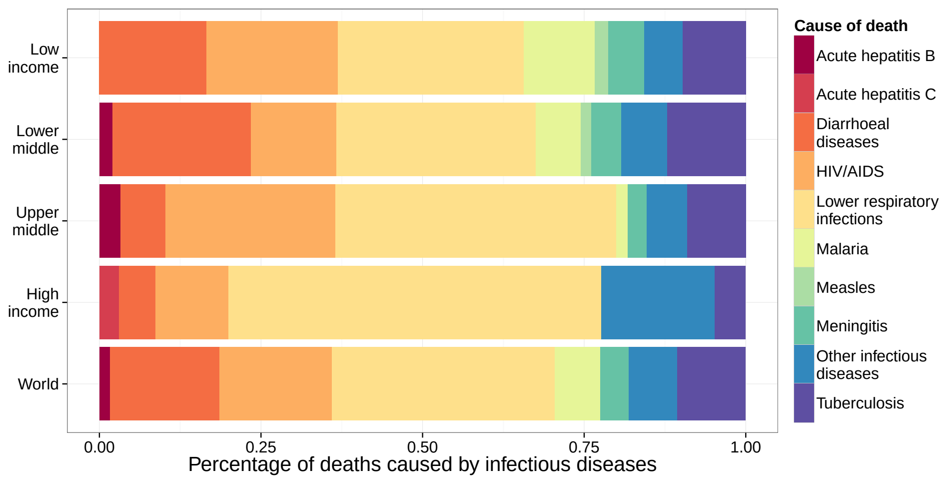

Figure 1.2: Relative frequencies of death causes in 2012 by World Bank income groups. Binning is based on Gross National Income (GNI; see figure 1.1). The data was obtained from the World Health Organization (2012).

Chapter 2

Mathematical Background

Modeling the relationship among variables is one of the most important applications of statistical theory. The study of regression analysis (and the closely related notion of correlation) started to form towards the end of the 19th century with Sir Francis Galton's study of height heredity in humans and his observation of regression towards the mean. Over the next few years, Udny Yule and Karl Pearson cast the developed concepts into precise mathematical formulation, in turn building on work performed by Adrien-Marie Legendre and Carl Friedrich Gauss who developed the method of least squares almost a century earlier (Allen, 1997).

A multiple linear regression model can be written in matrix-vector form as

$$y = X\beta + \varepsilon \tag{2.1}$$

where $y \in \mathbb{R}^n$ is the vector of observations on the dependent variable, the design matrix $X \in \mathbb{R}^{n \times p}$ contains data on the independent variables, $\beta \in \mathbb{R}^p$ is the p-dimensional parameter vector and the error term $\varepsilon \in \mathbb{R}^n$ captures effects not modeled by the regressors. Without loss of generality, all variables are assumed to be expressed as deviations from their means and measured on the same scale.

In order to find unknown coefficients β_i, the ordinary least squares estimator minimizes the residual sum of squares, the squared differences between observed responses and their predictions according to the linear model.

$$\widehat{\beta} = \arg\min_{\beta} \|y - X\beta\|^2 \tag{2.2a}$$

$$= (X^T X)^{-1} X^T y \tag{2.2b}$$

Some assumptions are typically associated with linear regression models that yield desirable attributes for the estimates. None of these restrictions are imposed on the explanatory variables; they can be continuous or discrete and combined as well as transformed arbitrarily. Furthermore, in practice, it is irrelevant whether the covariates are treated as random variables or as deterministic constants. With exception of the field of econometrics it appears that the majority of literature adheres to the latter interpretation and therefore, statements will not explicitly be conditional on covariate values.

Linearity. The relationship between dependent and independent variables should be linear (after suitable transformations) and individual effects additive. If this cannot be satisfied, a linear model is not suitable.

Full rank. For the matrix $X^T X$ to be invertible, it has to have full rank p. Therefore $n \leq p$ and all covariates must be linearly independent.

Exogeneity. All independent variables should be known exactly i.e. contain no measurement or observation errors as only the mean squared error of the dependent variable is minimized. Additionally, all important causal factors have to be included in the model. Exogeneity implies $\mathbf{E}[\varepsilon_i] = 0 \;\; \forall\, i$, as well no correlation between regressors and error terms (Hayashi, 2000).

Spherical errors. This includes both homoscedasticity or constant error variance: $\mathbf{E}[\varepsilon_i^2] = \sigma^2 \;\; \forall\, i$ and uncorrelated errors $\mathbf{E}[\varepsilon_i \varepsilon_j] = 0 \;\; \forall\, i \neq j$. These two conditions can be written more compactly as $\mathrm{Var}[\varepsilon] = \sigma^2 I_{n \times n}$.

Normality. To have some additional desirable characteristics of the estimated coefficients, it can be required that the errors ε_i be jointly normally distributed. With the above restrictions on expectation and variance, this yields $\varepsilon \sim \mathcal{N}_n(0, \sigma^2 I_{n \times n})$.

Violations of these assumptions have varying consequences. In case of perfect multicollinearity, the ordinary least squares estimator $\widehat{\beta}$ as defined in (2.2b) does not exist. Recovering such a situation is possible by using a generalized matrix inverse (for example the Moore–Penrose pseudoinverse) or employing a regularization scheme such as ridge regression.

Omitting a variable that is both correlated with dependent variables and has an effect on the response (a nonzero true coefficient) will introduce bias in the parameters. The method of instrumental variables can help to produce an unbiased estimator.

The assumption of spherical errors ensures that the least squares estimator is the best linear unbiased estimator in the sense that it has minimal variance among all linear unbiased estimators. Heteroscedasticity and autocorrelation do not cause coefficient estimates to be biased but can introduce bias in OLS estimates of variance, causing inaccurate standard errors. A generalized least squares estimator (for example weighted least squares)

Chapter 3

Writing scientific texts in English

This chapter was originally a separate document written by Reto Spöhel. It is reprinted here so that the template can serve as a quick guide to thesis writing, and to provide some more example material to give you a feeling for good typesetting.

3.1 Basic writing rules

The following rules need little further explanation; they are best understood by looking at the example in the booklet by Knuth et al., §2–§3.

Rule 3.1 Write texts, not chains of formulas.

More specifically, write full sentences that are logically interconnected by phrases like 'Therefore', 'However', 'On the other hand', etc. where appropriate.

Rule 3.2 Displayed formulas should be embedded in your text and punctuated with it.

In other words, your writing should not be divided into 'text parts' and 'formula parts'; instead the formulas should be tied together by your prose such that there is a natural flow to your writing.

3.2 Being nice to the reader

Try to write your text in such a way that a reader enjoys reading it. That's of course a lofty goal, but nevertheless one you should aspire to!

Rule 3.3 Be nice to the reader.

Give some intuition or easy example for definitions and theorems which might be hard to digest. Remind the reader of notations you introduced many pages ago – chances are he has forgotten them. Illustrate your writing with diagrams and pictures where this helps the reader. Etc.

Rule 3.4 Organize your writing.

Think carefully about how you subdivide your thesis into chapters, sections, and possibly subsections. Give overviews at the beginning of your thesis and of each chapter, so the reader knows what to expect. In proofs, outline the main ideas before going into technical details. Give the reader the opportunity to 'catch up with you' by summing up your findings periodically.

Useful phrases: 'So far we have shown that ...', 'It remains to show that ...', 'Recall that we want to prove inequality (7), as this will allow us to deduce that ...', 'Thus we can conclude that .... Next, we would like to find out whether ...', etc.

Rule 3.5 Don't say the same thing twice without telling the reader that you are saying it twice.

Repetition of key ideas is important and helpful. However, if you present the same idea, definition or observation twice (in the same or different words) without telling the reader, he will be looking for something new where there is nothing new.

Useful phrases: 'Recall that [we have seen in Chapter 5 that] ...', 'As argued before / in the proof of Lemma 3, ...', 'As mentioned in the introduction, ...', 'In other words, ...', etc.

Rule 3.6 Don't make statements that you will justify later without telling the reader that you will justify them later.

This rule also applies when the justification is coming right in the next sentence! The reasoning should be clear: if you violate it, the reader will lose valuable time trying to figure out on his own what you were going to explain to him anyway.

Useful phrases: 'Next we argue that ...', 'As we shall see, ...', 'We will see in the next section that ..., etc.

3.3 A few important grammar rules

Rule 3.7 There is (almost) *never* a comma before 'that'.

It's really that simple. Examples:

> We assume that ...
> *Wir nehmen an, dass ...*

It follows that . . .
Daraus folgt, dass . . .

'thrice' is a word that is seldom used.
'thrice' ist ein Wort, das selten verwendet wird.

Exceptions to this rule are rare and usually pretty obvious. For example, you may end up with a comma before 'that' because 'i.e.' is spelled out as 'that is':

For $p(n) = \log n/n$ we have . . . However, if we choose p a little bit higher, that is $p(n) = (1+\varepsilon)\log n/n$ for some $\varepsilon > 0$, we obtain that. . .

Or you may get a comma before 'that' because there is some additional information inserted in the middle of your sentence:

Thus we found a number, namely n_0, that satisfies equation (13).

If the additional information is left out, the sentence has no comma:

Thus we found a number that satisfies equation (13).

(For 'that' as a relative pronoun, see also Rules 3.9 and 3.10 below.)

Rule 3.8 There is usually no comma before 'if'.

Example:

A graph is not 3-colorable if it contains a 4-clique.
Ein Graph ist nicht 3-färbbar, wenn er eine 4-Clique enthält.

However, if the 'if' clause comes first, it is usually separated from the main clause by a comma:

If a graph contains a 4-clique, it is not 3-colorable .
Wenn ein Graph eine 4-Clique enthält, ist er nicht 3-färbbar.

There are more exceptions to these rules than to Rule 3.7, which is why we are not discussing them here. Just keep in mind: don't put a comma before 'if' without good reason.

Rule 3.9 Non-defining relative clauses have commas.

Rule 3.10 Defining relative clauses have no commas.

In English, it is very important to distinguish between two types of relative clauses: defining and non-defining ones. This is a distinction you absolutely need to understand to write scientific texts, because mistakes in this area actually distort the meaning of your text!

It's probably easier to explain first what a *non-defining* relative clause is. A non-defining relative clauses simply gives additional information *that could also be left out* (or given in a separate sentence). For example, the sentence

> The `WeirdSort` algorithm, which was found by the famous mathematician John Doe, is theoretically best possible but difficult to implement in practice.

would be fully understandable if the relative clause were left out completely. It could also be rephrased as two separate sentences:

> The `WeirdSort` algorithm is theoretically best possible but difficult to implement in practice. [By the way,] `WeirdSort` was found by the famous mathematician John Doe.

This is what a non-defining relative clause is. *Non-defining relative clauses are always written with commas.* As a corollary we obtain that you cannot use 'that' in non-defining relative clauses (see Rule 3.7!). It would be wrong to write

> The `WeirdSort` algorithm, that was found by the famous mathematician John Doe, is theoretically best possible but difficult to implement in practice.

A special case that warrants its own example is when 'which' is referring to the entire preceding sentence:

> Thus inequality (7) is true, which implies that the Riemann hypothesis holds.

As before, this is a non-defining relative sentence (it could be left out) and therefore needs a comma.

So let's discuss *defining* relative clauses next. A defining relative clause tells the reader *which specific item the main clause is talking about*. Leaving it out either changes the meaning of the sentence or renders it incomprehensible altogether. Consider the following example:

> The `WeirdSort` algorithm is difficult to implement in practice. In contrast, the algorithm that we suggest is very simple.

Here the relative clause 'that we suggest' cannot be left out – the remaining sentence would make no sense since the reader would not know which algorithm it is talking about. This is what a defining relative clause is. *Defining relative clauses are never written with commas.* Usually, you can use both 'that' and 'which' in defining relative clauses, although in many cases 'that' sounds better.

As a final example, consider the following sentence:

> For the elements in $\mathcal{B}$ which satisfy property (A), we know that equation (37) holds.

Table 3.1: Things you (usually) don't say

It holds (that) ...	We have ...	*Es gilt ...*
('Equation (5) holds.' is fine, though.)		
x fulfills property $\mathcal{P}$.	x satisfies property $\mathcal{P}$.	*x erfüllt Eigenschaft $\mathcal{P}$.*
in average	on average	*im Durchschnitt*
estimation	estimate	*Abschätzung*
composed number	composite number	*zusammengesetzte Zahl*
with the help of	using	*mit Hilfe von*
surely	clearly	*sicher, bestimmt*
monotonously increasing	monotonically incr.	*monoton steigend*
(Actually, in most cases 'increasing' is just fine.)		

This sentence does not make a statement about all elements in $\mathcal{B}$, only about those satisfying property (A). The relative clause is *defining*. (Thus we could also use 'that' in place of 'which'.)

In contrast, if we add a comma the sentence reads

> For the elements in $\mathcal{B}$, which satisfy property (A), we know that equation (37) holds.

Now the relative clause is *non-defining* – it just mentions in passing that all elements in $\mathcal{B}$ satisfy property (A). The main clause states that equation (37) holds for *all* elements in $\mathcal{B}$. See the difference?

3.4 Things you (usually) don't say in English – and what to say instead

Table 3.1 lists some common mistakes and alternatives. The entries should not be taken as gospel – they don't necessarily mean that a given word or formulation is wrong under all circumstances (obviously, this depends a lot on the context). However, in nine out of ten instances the suggested alternative is the better word to use.

Chapter 4

Typography

4.1 Punctuation

Rule 4.1 Use opening (‘) and closing (’) quotation marks correctly.

In LaTeX, the closing quotation mark is typed like a normal apostrophe, while the opening quotation mark is typed using the French *accent grave* on your keyboard (the *accent grave* is the one going down, as in *frère*).

Note that any punctuation that *semantically* follows quoted speech goes inside the quotes in American English, but outside in Britain. Also, Americans use double quotes first. Oppose

> “Using ‘lasers,’ we punch a hole in ... the Ozone Layer,” Dr. Evil said.

to

> ‘Using “lasers”, we punch a hole in ... the Ozone Layer’, Dr. Evil said.

Rule 4.2 Use hyphens (-), en-dashes (–) and em-dashes (—) correctly.

A hyphen is only used in words like ‘well-known’, ‘3-colorable’ etc., or to separate words that continue in the next line (which is known as hyphenation). It is entered as a single ASCII hyphen character (`-`).

To denote ranges of numbers, chapters, etc., use an en-dash (entered as two ASCII hyphens `--`) with no spaces on either side. For example, using Equations (1)–(3), we see...

As the equivalent of the German *Gedankenstrich*, use an en-dash with spaces on both sides – in the title of Section 3.4, it would be wrong to use a hyphen instead of the dash. (Some English authors use the even longer emdash (—)

instead, which is typed as three subsequent hyphens in LaTeX. This emdash is used without spaces around it—like so.)

4.2 Spacing

Rule 4.3 Do not add spacing manually.

You should never use the commands `\\` (except within tabulars and arrays), `\␣` (except to prevent a sentence-ending space after Dr. and such), `\vspace`, `\hspace`, etc. The choices programmed into LaTeX and this style should cover almost all cases. Doing it manually quickly leads to inconsistent spacing, which looks terrible. Note that this list of commands is by no means conclusive.

Rule 4.4 Judiciously insert spacing in maths where it helps.

This directly contradicts Rule 4.3, but in some cases TeX fails to correctly decide how much spacing is required. For example, consider

$$f(a,b) = f(a+b,a-b).$$

In such cases, inserting a thin math space `\,` greatly increases readability:

$$f(a,b) = f(a+b,\,a-b).$$

Along similar lines, there are variations of some symbols with different spacing. For example, Lagrange's Theorem states that $|G| = [G : H]|H|$, but the proof uses a bijection $f\colon aH \to bH$. (Note how the first colon is symmetrically spaced, but the second is not.)

Rule 4.5 Learn when to use `\␣` and `\@`.

Unless you use 'french spacing', the space at the end of a sentence is slightly larger than the normal interword space.

The rule used by TeX is that any space following a period, exclamation mark or question mark is sentence-ending, except for periods preceded by an upper-case letter. Inserting `\` before a space turns it into an interword space, and inserting `\@` before a period makes it sentence-ending. This means you should write

```
Prof.\ Dr.\ A. Steger is a member of CADMO\@.
If you want to write a thesis with her, you
should use this template.
```

which turns into

> Prof. Dr. A. Steger is a member of CADMO. If you want to write a thesis with her, you should use this template.

The effect becomes more dramatic in lines that are stretched slightly during justification:

> Prof. Dr. A. Steger is a member of CADMO. If you

Rule 4.6 Place a non-breaking space (~) right before references.

This is actually a slight simplification of the real rule, which should invoke common sense. Place non-breaking spaces where a line break would look 'funny' because it occurs right in the middle of a construction, especially between a reference type (Chapter) and its number.

4.3 Choice of 'fonts'

Professional typography distinguishes many font attributes, such as family, size, shape, and weight. The choice for sectional divisions and layout elements has been made, but you will still occasionally want to switch to something else to get the reader's attention. The most important rule is very simple.

Rule 4.7 When emphasising a short bit of text, use `\emph`.

In particular, *never* use bold text (`\textbf`). Italics (or Roman type if used within italics) avoids distracting the eye with the huge blobs of ink in the middle of the text that bold text so quickly introduces.

Occasionally you will need more notation, for example, a consistent typeface used to identify algorithms.

Rule 4.8 Vary one attribute at a time.

For example, for WEIRDSORT we only changed the shape to small caps. Changing two attributes, say, to bold small caps would be excessive (LaTeX does not even have this particular variation). The same holds for mathematical notation: the reader can easily distinguish g_n, $G(x)$, $\mathcal{G}$ and G.

Rule 4.9 Never underline or uppercase.

No exceptions to this one, unless you are writing your thesis on a typewriter. Manually. Uphill both ways. In a blizzard.

4.4 Displayed equations

Rule 4.10 Insert paragraph breaks *after* displays only where they belong. Never insert paragraph breaks *before* displays.

LaTeX translates sequences of more than one linebreak (i.e., what looks like an empty line in the source code) into a paragraph break in almost all contexts. This also happens before and after displays, where extra spacing is inserted to give a visual indication of the structure. Adding a blank line in these places may look nice in the sources, but compare the resulting display

$$a = b$$

to the following:

$$a = b$$

The first display is surrounded by blank lines, but the second is not. It is bad style to start a paragraph with a display (you should always tell the reader what the display means first), so the rule follows.

Rule 4.11 Never use `eqnarray`.

It is at the root of most ill-spaced multiline displays. The *amsmath* package provides better alternatives, such as the `align` family

$$\begin{aligned} f(x) &= \sin x, \\ g(x) &= \cos x, \end{aligned}$$

and `multline` which copes with excessively long equations:

$$\begin{multline} \mathrm{P}\big[X_{t_0} \in (z_0, z_0 + dz_0], \ldots, X_{t_n} \in (z_n, z_n + dz_n]\big] \\ = \nu(dz_0) K_{t_1}(z_0, dz_1) K_{t_2 - t_1}(z_1, dz_2) \cdots K_{t_n - t_{n-1}}(z_{n-1}, dz_n). \end{multline}$$

4.5 Floats

By default this style provides floating environments for tables and figures. The general structure should be as follows:

```
\begin{figure}
  \centering
  % content goes here
  \caption{A short caption}
  \label{some-short-label}
\end{figure}
```

Note that the label must follow the caption, otherwise the label will refer to the surrounding section instead. Also note that figures should be captioned at the bottom, and tables at the top.

The whole point of floats is that they, well, *float* to a place where they fit without interrupting the text body. This is a frequent source of confusion and changes; please leave it as is.

Rule 4.12 Do not restrict float movement to only 'here' (`h`).

If you are still tempted, you should avoid the float altogether and just show the figure or table inline, similar to a displayed equation.

Chapter 5

First Chapter SfS Template

5.1 To include a picture

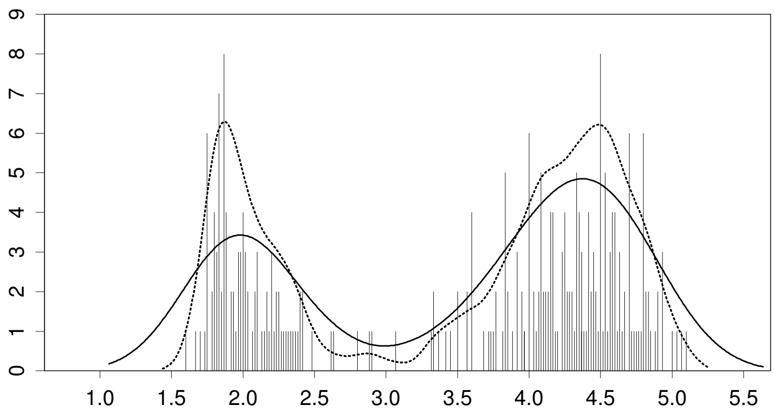

Figure 5.1: Old Faithful Geyser eruption lengths, $n = 272$; binned data and two (Gaussian) kernel density estimates ($\times 10$) with $h = h^* = .3348$ and $h = .1$ (dotted).

Or also with `includegraphics`:

5.2 To make a proof

Proof $1 + 1 = 2$ $\square$

5.3 To include R code

See information in Appendix A.

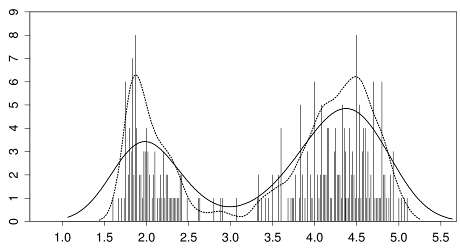

Figure 5.2: Old Faithful Geyser eruption lengths, $n = 272$; binned data and two (Gaussian) kernel density estimates ($\times 10$) with $h = h^* = .3348$ and $h = .1$ (dotted).

5.4 Other information

Put a text between quotes: make sure to use nice quotes, such as “quote”.

Cite a document in the bibliography (an example here): Gelman, Jakulin, Pittau, and Su (2008). Or mention that Konis (a person) or Hastie, Tibshirani, and Friedman (multiple persons) have already done quite a bit work.

Referencing a different part of your work: please refer to Appendix A.

Appendix A

Complementary information

Additional material. For example long mathematical derivations could be given in the appendix. Or you could include part of your code that is needed in printed form. You can add several Appendices to your thesis (as you can include several chapters in the main part of your work).

A.1 Including R code with verbatim

A simple (rather too simple, see A.2) way to include code or *R* output is to use `verbatim`. It just prints the text however it is (including all spaces, "strange" symbols,...) in a slightly different font.

```
## loading packages
library(RBGL)
library(Rgraphviz)
library(boot)

## global variables
X_MAX <- 150

   This allows me to put as many s  p a   c es   as I want.
I can also use \ and ‘ and & and all the rest that is usually only
accepted in the math mode.

I can also make as
                  many
             line
    breaks as
I want... and
             where I want.
```

A.2 Including R code with the listings package

However, it is much nicer to use the *listings* package to include R code in your report. It allows you to number the lines, color the comments differently than the code, and so on.

```
## example to generate an .eps file with the function ps.latex()
## Author: Sarah Gerster and Martin Machler
## Last revision: 16 Aug 2011

require("sfsmisc") # pdf.latex(), pdf.end(), etc

pdf.latex(file='test_plot.pdf') #, main=TRUE)
## no main=TRUE is needed to leave enough space for the plot title
## but see below

## make sure the legends are large enough
par(cex=1.5)

## Make sure your lines are "visible" enough. Otherwise your plot
## won't look very nicely in your text.
plot(-10:10, (-10:10)**2, type="l", lty=5,
     xlab="my_x", ylab="my_y",
     ## no main title: NOT recommended for figures in text which
     ## have a \caption{..}
     lwd=4, col='blue')
lines(-10:10, 0:20, type="p", lwd=4, pch=23,col='red')
legend(-3, 90, c("func1","func2"),lwd=4,col=c('blue', 'red'),
       lty=c(1,1),cex=1)
pdf.end() # starts the previewer (which refreshes itself;
          # at least on Linux at SfS
```

A.3 Using Sweave to include R code (and more) in your report

The easiest (and most elegant) way to include R code and its output (and have all your figures up to date with your report) is to use Sweave. You can find an introduction Sweave in `/u/sfs/StatSoftDoc/Sweave/Sweave-tutorial.pdf`.

Appendix B

Yet another appendix....

B.1 Description

Something details.

Something else other definition.

B.2 Tables

Refer to Table B.1 to see a left justified table with caption on top.

Table B.1: Results.

Student	**Grade**
Marie	6
Alain	5.5
Josette	4.5
Pierre	5

Bibliography

Allen, M. P. (1997). The origins and uses of regression analysis. In *Understanding Regression Analysis*, pp. 1–5. Springer US.

Dobson, A. P. and E. R. Carper (1996, feb,). Infectious Diseases and Human Population History. *BioScience 46*(2), 115–126 CR – Copyright © 1996 Oxford Universi.

Gelman, A., A. Jakulin, M. G. Pittau, and Y.-S. Su (2008). A weakly informative default prior distribution for logistic and other regression models. *Ann. Appl. Stat. 2*(4), 1360–1383.

Hastie, T., R. Tibshirani, and J. Friedman (2009). *The Elements of Statistical Learning*, Volume 1.

Hayashi, F. (2000). *Econometrics*. Princeton, NJ [u.a.]: Princeton Univ. Press.

Kahneman, D. (2012, mar,). Quotation: Kahneman on Contingencies. *Journal of the Experimental Analysis of Behavior 97*(2), 182.

Konis, K. P. (2007). *Linear programming algorithms for detecting separated data in binary logistic regression models*. Doctoral, University of Oxford.

Littman, R. J. (2009, oct,). The Plague of Athens: Epidemiology and Paleopathology. *Mount Sinai Journal of Medicine: A Journal of Translational and Personalized Medicine 76*(5), 456–467.

Taubenberger, J. and D. M. Morens (2006). 1918 Influenza: the Mother of All Pandemics. *Emerging Infectious Diseases Volume 12*(1).

Treadgold, W. (1997, jun,). *A History of the Byzantine State and Society*. Stanford: Stanford University Press.

World Health Organization (2003). Fact sheet No. 211: Influenza.

World Health Organization (2012). Mortality and global health estimates: Causes of death.

Zietz, B. P. and H. Dunkelberg (2004). The history of the plague and the research on the causative agent Yersinia pestis. *International Journal of Hygiene and Environmental Health 207*(2), 165–178.

Epilogue

A few final words. Test 2

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Declaration of originality

The signed declaration of originality is a component of every semester paper, Bachelor's thesis, Master's thesis and any other degree paper undertaken during the course of studies, including the respective electronic versions.

Lecturers may also require a declaration of originality for other written papers compiled for their courses.

I hereby confirm that I am the sole author of the written work here enclosed and that I have compiled it in my own words. Parts excepted are corrections of form and content by the supervisor.

Title of work (in block letters):

Authored by (in block letters):
For papers written by groups the names of all authors are required.

Name(s): **First name(s):**

With my signature I confirm that
- I have committed none of the forms of plagiarism described in the 'Citation etiquette' information sheet.
- I have documented all methods, data and processes truthfully.
- I have not manipulated any data.
- I have mentioned all persons who were significant facilitators of the work.

I am aware that the work may be screened electronically for plagiarism.

Place, date **Signature(s)**

For papers written by groups the names of all authors are required. Their signatures collectively guarantee the entire content of the written paper.