

An Example KnitR Document

Gregory Macfarlane

September 5, 2012

1 Introduction

The purpose of this document, and its associated folder, is to provide a testing ground for your initial work using `knitr`, `R`, and `LATEX`. This document should be a thorough, if not comprehensive, treatment of the tools you will use. Feel free to contribute to the project by adding your own sections, trying new things, and committing to the repository.

1.1 Academic Software

`R` is a widely-used general purpose statistical program, based on the `S` programming language. It is available for all major computing platforms, and is freely distributed under the General Public License. It has advantages and disadvantages relative to `Stata`, `SAS`, or other programs, but I personally feel it is an extremely powerful program that is worth far more than its cost.

`LATEX` is a free and open-source document processing system that is highly programmable and produces very attractive output (such as this document). Most of the top academic journals provide `LATEX` files so that authors do not need to worry about formatting rules, and some (for instance, all Elsevier publications) even accept `.tex` files directly.

`knitr` is a free and open-source program that processes `R` script for inclusion in a `LATEX`-compiled document. The `RStudio` integrated development environment (IDE) includes `knitr` processing as a compilation option in its most recent available version. Using `knitr` (or its predecessor, `Sweave`), a highly productive academic workflow is as follows:

1. Create an `R`-noweb (`.rnw`) file with `LATEX` document elements and `R` statistical chunks.
2. Run `knitr` on that file to create a `.tex` file.
3. Run `pdflatex` on the `.tex` file to create a PDF that combines statistical analysis and discussion.

The advantage of this method is that your entire project, from data to conclusions, takes place in a single file. Every assumption in your project is documented in code that directly accompanies the discussion in your research. Assuming that your code is documented

properly, you could allow another researcher or reviewer to replicate your findings and identify potential weaknesses.

2 An Example

In order to give an example of how R chunks look in an `.rnw` document, and how to use other tools in `knitr`, I am going to run through a simple linear model. To include R chunks in the document, we use the following notation:

```
<<chunk_name, options>>=  
R code  
@
```

So, to load an R package that I want to use in my analysis, I use the following command (you may need to install the “apsrtable” package):

```
<<load_packages>>=  
library(apsrtable) # pretty model output tables  
@
```

```
library(apsrtable) # pretty model output tables
```

If I just use the default options, it will print out a very pretty code chunk, as you see above. You can suppress this with the `echo=FALSE` option, globally or just in a particular chunk. In this example, I copy the the example given for the `lm` command when you call the help file in R.

```
## Annette Dobson (1990) 'An Introduction to  
# Generalized Linear Models'.  
## Page 9: Plant Weight Data.  
ctl <- c(4.17, 5.58, 5.18, 6.11, 4.5, 4.61, 5.17,  
        4.53, 5.33, 5.14)  
trt <- c(4.81, 4.17, 4.41, 3.59, 5.87, 3.83, 6.03,  
        4.89, 4.32, 4.69)  
group <- gl(2, 10, 20, labels = c("Ctl", "Trt"))  
weight <- c(ctl, trt)  
lm.D9 <- lm(weight ~ group)  
lm.D90 <- lm(weight ~ group - 1)
```

Once the two models are estimated, I can use their objects in a number of ways. For example I can make a figure of the regression diagnostics, which is given in Figure 1. I can also print out a model comparison table, which is given in Table 1. Note that I do not have to specify reference numbers directly, but can simply call the object’s label.

To compile this document, you can simply push the “Compile PDF” button at the top of the editor console in RStudio.

```
par(mfrow = c(2, 2)) #multiple plots
plot(lm.D9)
```

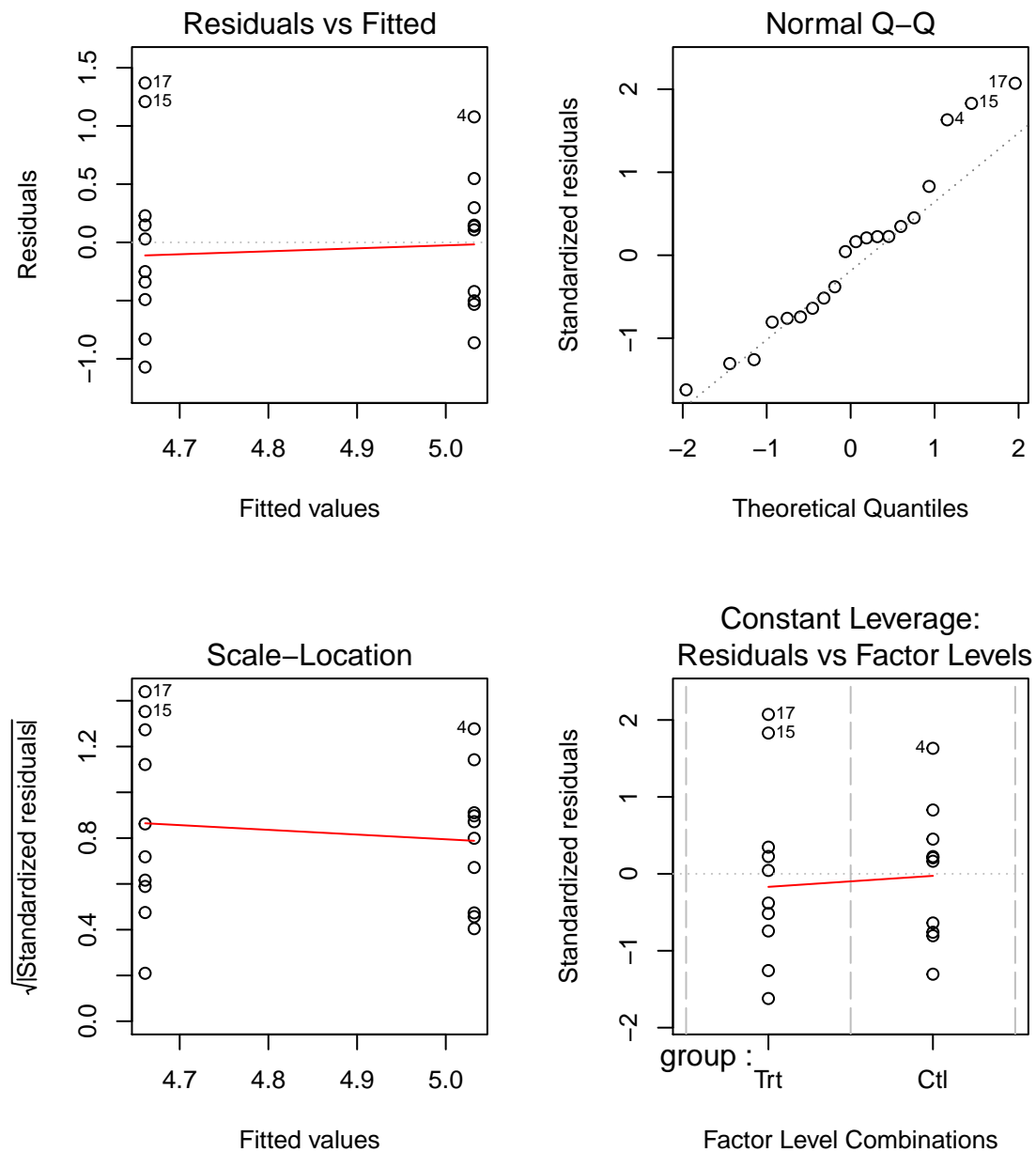


Figure 1: Standard regression diagnostic plots, for the effect of treatment group on plant weight.

Table 1: Model Comparison

	Default		No Intercept	
(Intercept)	5.03*	(0.22)		
groupTrt	−0.37	(0.31)	4.66*	(0.22)
groupCtl			5.03*	(0.22)
N	20		20	
R^2	0.07		0.98	
adj. R^2	0.02		0.98	
Resid. sd	0.70		0.70	

Standard errors in parentheses

* indicates significance at $p < 0.05$