Sweave Tutorial

A.I. McLeod

Abstract

Our purpose is to provide a template and illustrative examples using Sweave.

Keywords: BibTeX, LaTeX, R, Sweave, WinEdt, wrapfig.

1. Introduction

This document and other related files may be downloaded from `http://www.stats.uwo.ca/faculty/aim/2010/SweaveTutorial/`. We found in class that Internet Explorer does not work — you need to use *Mozilla Firefox.*

to get the tutorial files.

Sweave combines LaTeX and R so that beautiful documents can be produced.

Our goal in these notes is to provide a simple template of the type of document that is expected for Assignment 1. The important things to learn are:

1. How to typeset equations in LaTeX
2. Combining R and Sweave in the simplest fashion
3. Use BibTeX for references

The steps in obtaining the final documents are:

1. Write the source for the document in a file, say `eg.Rnw`. Many authors also use the extension `Snw`.
2. Write the file `bib` file. Use `WinEdt`. Use 'Insert' to add entries to your bib-file. Be sure to save this file before trying to compile!
3. Compile using the command `R CMD Sweave eg.Rnw` at the command prompt.[1] This produces a file with extension `tex` that has the LaTeX source. In the present case, the complete file name is `eg.tex`.[2]
4. Load the file into `WinEdt`. Here we use `WinEdt` as the front end to LaTeX. The final pdf can be produced by clicking on the icon at the top showing the lion-pdf icon (PDF TeXify).

[1] On Windows computers, the command prompt is found in the Assessories programs. You also need change directory to the directory where the Rnw-file is located.

[2] You can also use the function codeSweave from inside R to do this. But you need to change the R working directory.

5. Make sure the directory containing the files `eg.Rnw` etc. also contain the file `Sweave`.

After the `eg.Rnw`-file has been created, you can use `R CMD Stangle eg.Rnw` to extract the R scripts used.

There are many books on LaTeX. I recommend that you read Jiang (2009) and Downes (2002-03-22). Both of these references are available as pdf downloads from our WebCT site. If you wish to purchase a book, a standard reference is Mittelbach and Goossens (2004).

There is a recent *CRAN Task View: Reproducible Research* that is of interest to **9864 Students**. See `http://cran.r-project.org/web/views/ReproducibleResearch.html`.

2. R Workspace Management

I wrote an R package McLeod (2010) for this. This package is not required for this course but it may be of interest to you. The main purpose of this section is to show how R packages should be referenced — please see bib file.

3. Normal Distribution and Plot

The normal pdf for with mean μ and standard deviation σ may be written

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{(z-\mu)^2/\sigma^2}$$

where $-\infty < z < \infty$. Notice that $\phi(z) = \phi(-z)$.

The corresponding cdf is

$$\Phi(z) = \int_{-\infty}^{z} phi(t) dt$$

.

In R, the function `dnorm` computes this density function. A quick way of getting basic syntax help for an R function is to use the function `arg`. For example,

```
> x <- seq(70, 130, 15)
> args(dnorm)

function (x, mean = 0, sd = 1, log = FALSE)
NULL

> dnorm(x, 100, 15)

[1] 0.003599398 0.016131382 0.026596152 0.016131382 0.003599398
```

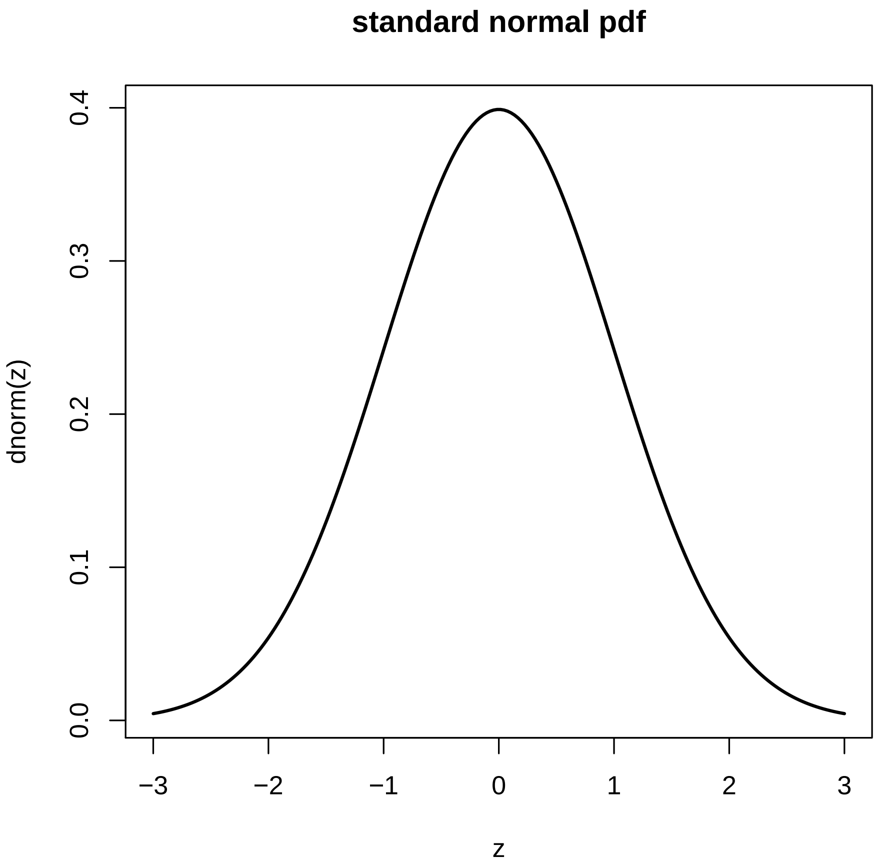
Taking $\mu = 100$ and $\sigma = 15$, the probability density corresponding to 70, 85, 100, 115, 130 may be computed:

```
> x <- seq(70, 130, 15)
> dnorm(x, 100, 15)

[1] 0.003599398 0.016131382 0.026596152 0.016131382 0.003599398
```

To plot the pdf we use,

```
> z <- seq(-3, 3, length.out = 200)
> plot(z, dnorm(z), type = "l", lwd = 2, color = "blue")
> title("standard normal pdf")
```



4. Binomial Distribution and Plot

The binomial distributions may be defined as the sum of n independent Bernouilli trials with outcomes 0 or 1 with probabilities p and $1 - p$ respectively.

The probability function may be written,

$$f(x) = \binom{n}{x} p^x (1 - p)^{n-x},$$

where $x = 0, 1, \ldots, n$. So the cdf is

$$F(x) = \sum_{k=0}^{x} f(k),$$

for $x = 0, \ldots, n$.

We already saw how to plot a probability function.

Let's see how to construct a beautiful table. For this purpose we use the library `xtable`. Since this is not one of the base libraries[3], if you are working at home, you may need to download it. On our department network, it should already be available.

Next you create a matrix.[4] Be sure to use `dimnames` to name the rows and columns. Then use `xtable` to convert to LaTeX.

```
> x <- 0:5
> px <- pbinom(x, 5, 0.5)
> m <- matrix(c(x, px), byrow = TRUE, nrow = 2)
> dimnames(m) <- list(c("x", "f(x)"), NULL)

> library(xtable)
> xtable(m)
```

	1	2	3	4	5	6
x	0.00	1.00	2.00	3.00	4.00	5.00
f(x)	0.03	0.19	0.50	0.81	0.97	1.00

This is not so beautiful because x should be integer valued. To correct this we will need to create a **dataframe**. Basically a dataframe is like a matrix except that each column can be numeric, character or factor. Let's try again, this time using `format`.

```
> m <- matrix(c(format(x), format(round(px, 5))), byrow = TRUE,
+     nrow = 2)
> rownames(m) <- c("x", "f(x)")
> colnames(m) <- rep(" ", length(x))
> xtable(m)
```

x	0	1	2	3	4	5
f(x)	0.03125	0.18750	0.50000	0.81250	0.96875	1.00000

This is better but not great since the spacing of the lines at the top of the table is ugly. Things can be improved using `print` and also giving a caption. To fully understand this you need to realize that `xtable` produces output of class 'xtable' and their is a 'print'-method, `print.xtable` available.

[3] See CRAN FAQ, Section 5. `http://cran.r-project.org/doc/FAQ/R-FAQ.html`.

[4] The input to xtable can be other R objects, including 'data.frame', 'lm' and many others.

```
> m <- matrix(c(format(x), format(round(px, 5))), byrow = TRUE,
+     nrow = 2)
> rownames(m) <- c("x", "f(x)")
> colnames(m) <- rep(" ", length(x))
> print(xtable(m, caption = "Binomial probabilities, n=5, p=0.5"),
+     include.colnames = FALSE, hline.after = c(1))
```

x	0	1	2	3	4	5
f(x)	0.03125	0.18750	0.50000	0.81250	0.96875	1.00000

Table 1: Binomial probabilities, n=5, p=0.5

Finally, we should mention that often if we preparing a statistical analysis to be read by others, we really don't want to distract them with lot's of R code detail. This is easily done as in this example,

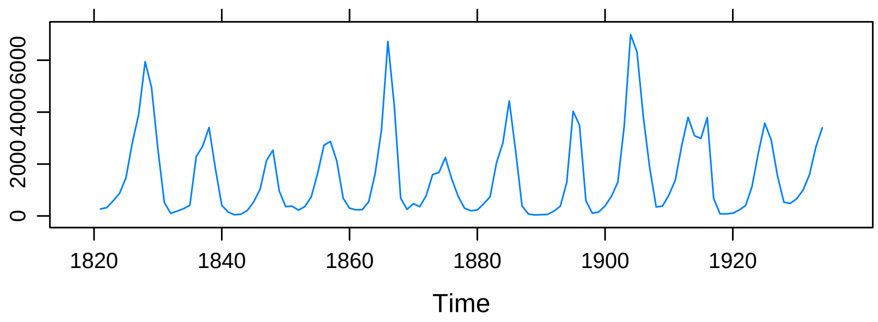
x	0	1	2	3	4	5
f(x)	0.03125	0.18750	0.50000	0.81250	0.96875	1.00000

Table 2: Binomial probabilities, n=5, p=0.5

5. Lattice Graphics

Lattice graphics, unlike the standard graphics function `plot()`, does not produce the plot as a side effect. Instead the lattice function returns an object with class "trellis". This object is then display using the function `print.trellis`. All this happens by default and in normal simple usage the casual user does not notice this difference between standard graphics and lattice graphics. In order to use lattice graphics with Sweave we will need to "print" the object. Here is an example of how it works.

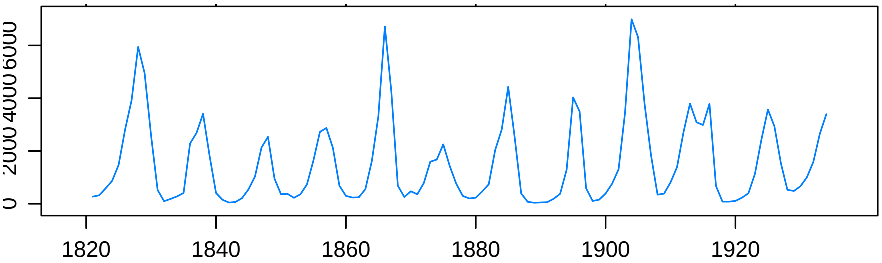
```
> library(lattice)
> print(xyplot(lynx, aspect = 0.25))
```



The above example is ugly when included in your report because of all the extra space it leaves!

6. Cropping

We need to remove the white space. This can be done but it is a little complicated. In the above example, we can crop to obtain:



The above cropped version fits nicely in our text.

This is done by editing the `tex` file. We need to replace the `\includegraphics` directive. In the current example we see the filename is `AIMSweave-lynxPlot-repeat`.

```
\includegraphics[viewport=25 160 450 275, clip]{AIMSweave-lynxPlot-repeat}
```

The coordinates used in `viewport` are briefly described below.

```
viewport=
    llx lly urx ury;
    specify the bounding box relative to the lower left corner of the existing one;
    the units are obtained from the graphic file itself.
```

Fortunately Sweave is *very smart*, after you the above change to your `tex` file, running Sweave again produces a new `tex` file but it does not overwrite this change!

To start a new page, use the LaTeXmarkup `\ vfill\ newpage`.

7. Resizing

```
> qqnorm(rt(100, 5))
> title(sub = expression(paste("100 ", t[5], " RVS")))
```

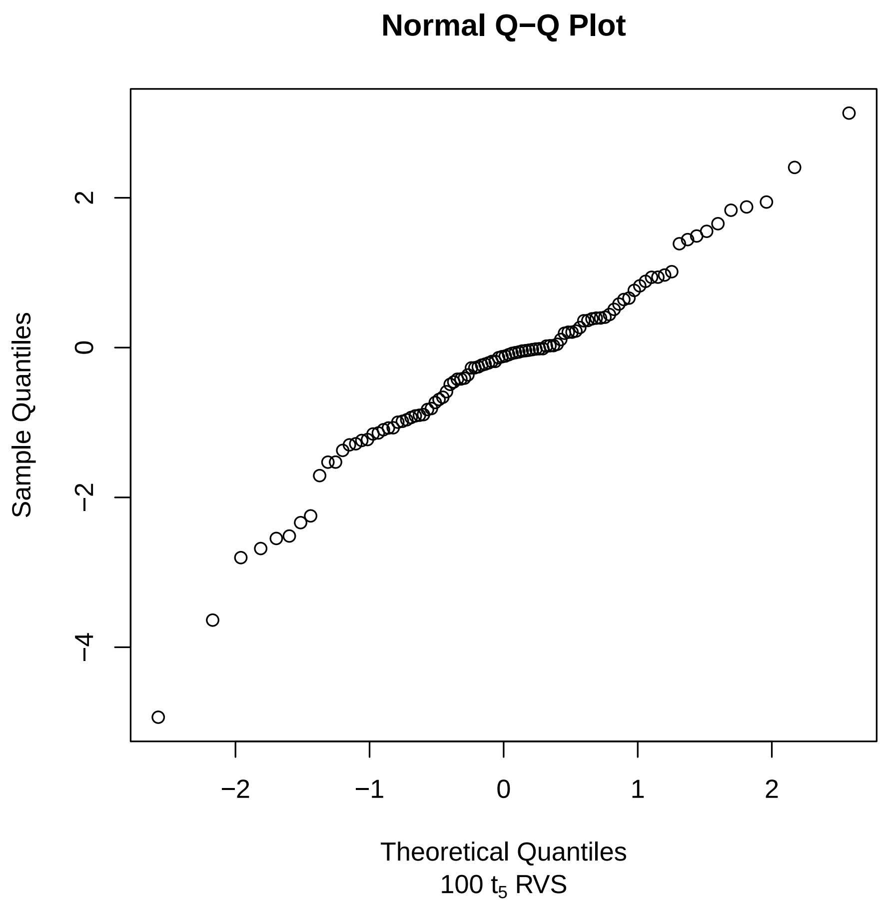


Figure 1: Normal probability plot of 100 RVS from t_5 distribution

The proper way to resize a graphic in Sweave:

```
\setkeys{Gin}{width=0.5\textwidth}
```

In Figures 1 and 2, I used the LaTeX `figure` environment that provides options for captions and for symbollically referring to figure by its number — like I just did! See `.Rnw` file.

Also in Figure 1, I used LaTeX `center` environment to center the figure.

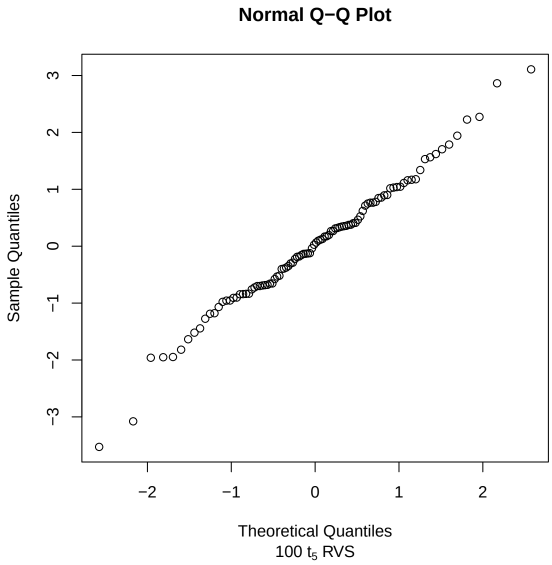


Figure 2: Above plot resized using Sweave markup.

8. Literate Statistical Reports

Using Sweave is still a very good idea even if we don't show any R code since we it allows us and others to easily reproduce our work. This is the essence of *reproducible research*. Sometimes this has also been called "Literate Statistical Practice" Rossini and Leisch (2003).

Further articles on Sweave are avaiable Professor Friedrich Leisch homepage `http://www.stat.uni-muenchen.de/~leisch/Sweave/`.

Professor Murdoch has a tutorial page on Sweave: `http://www.stats.uwo.ca/faculty/murdoch/computing/`.

Comments or suggestions welcome, please email `aimcleod@uwo.ca`.

References

Downes M (2002-03-22). *Short Math Guide for LATEX*. American Mathematical Society. URL `http://www.ams.org/tex/short-math-guide.html`.

Jiang W (2009). *An Introduction to LATEX*. Department of Statistical and Actuarial Science, The University of Western Ontario, London, Ontario. Available on Course WebCT site.

McLeod AI (2010). *rwm: R Workspace Management*. URL `http://CRAN.R-project.org/package=rwm`.

Mittelbach F, Goossens M (2004). *The LATEXCompanion*. Addison-Wesly, Boston.

Rossini A, Leisch F (2003). *Literate Statistical Practice*. URL `http://www.bepress.com/cgi/viewcontent.cgi?article=1017&context=uwbiostat`.

Affiliation:

A. Ian McLeod

Department of Statistical and Actuarial Sciences
The University of Western Ontario
London, Ontario, Canada
E-mail: aimcleod@uwo.ca
URL: http://www.stats.uwo.ca/faculty/~aim/