Statistics & R

Exercises

1) *Look at the websites for **R** to look at the various notes on usage of **R** referred to on page 2. This is useful if you want to install **R** on your own machine, (10 minutes maximum).*

Trust you've done this by now

2) *Call up the R programme.*

a) *Find out what libraries are available on your system by typing* `library()`.

I get:

```
>
> library()
>
```

```
Packages in library 'C:/Users/nick/Documents/R/win-library/2.13':

anacor                  Simple and Canonical Correspondence Analysis.
ca                      Simple, Multiple and Joint Correspondence
                        Analysis
car                     Companion to Applied Regression
coda                    Output analysis and diagnostics for MCMC
colorspace              Color Space Manipulation
deldir                  Delaunay Triangulation and Dirichlet (Voronoi)
                        Tessellation.
fda                     Functional Data Analysis
fossil                  Palaeoecological and Palaeogeographical
                        Analysis Tools
ISwR                    Introductory Statistics with R
maps                    Draw Geographical Maps
maptools                Tools for reading and handling spatial objects
rgl                     3D visualization device system (OpenGL)
scatterplot3d           3D Scatter Plot
sp                      classes and methods for spatial data
spdep                   Spatial dependence: weighting schemes,
                        statistics and models
vegan                   Community Ecology Package
zoo                     S3 Infrastructure for Regular and Irregular
                        Time Series (Z's ordered observations)

Packages in library 'C:/Program Files/R/R-2.13.1/library':

base                    The R Base Package
boot                    Bootstrap Functions (originally by Angelo Canty
                        for S)
class                   Functions for Classification
cluster                 Cluster Analysis Extended Rousseeuw et al.
codetools               Code Analysis Tools for R
compiler                The R Compiler Package
```


```
datasets                The R Datasets Package
foreign                 Read Data Stored by Minitab, S, SAS, SPSS,
                        Stata, Systat, dBase, ...
graphics                The R Graphics Package
grDevices               The R Graphics Devices and Support for Colours
                        and Fonts
grid                    The Grid Graphics Package
KernSmooth              Functions for kernel smoothing for Wand & Jones
                        (1995)
lattice                 Lattice Graphics
MASS                    Support Functions and Datasets for Venables and
                        Ripley's MASS
Matrix                  Sparse and Dense Matrix Classes and Methods
methods                 Formal Methods and Classes
mgcv                    GAMs with GCV/AIC/REML smoothness estimation
                        and GAMMs by PQL
nlme                    Linear and Nonlinear Mixed Effects Models
nnet                    Feed-forward Neural Networks and Multinomial
                        Log-Linear Models
rpart                   Recursive Partitioning
spatial                 Functions for Kriging and Point Pattern
                        Analysis
splines                 Regression Spline Functions and Classes
stats                   The R Stats Package
stats4                  Statistical Functions using S4 Classes
survival                Survival analysis, including penalised
                        likelihood.
tcltk                   Tcl/Tk Interface
tools                   Tools for Package Development
utils                   The R Utils Package
```

The first set are libraries I have downloaded from CRAN and installed myself. The others are libraries that are bundled with the package **R** and installed automatically. Of the second set some (such as `stats`) are automatically loaded into every **R** session and others such as `MASS` must be loaded when you need to use the functions or datasets that they contain.

b) *If the MASS library (**M**odern **A**pplied **S**tatistics with **S**) is available then find out what is inside it (i.e. what statistical facilities or commands it has and what data sets are provided with it) with `library(help=MASS)`. Note that **R** thinks that upper case LETTERS and lower case letters are different so you must type `MASS` and not `Mass` or `mass`. Note that the list of commands and data sets are all given in one single alphabetic list.*

c) *Open the MASS library with `library(MASS)`and find out what data sets are available to you with `data()`. You will see that they are listed first for those from the library MASS and then the standard ones that come with the base package. You*

will need to open the MASS library for almost all exercises in this course and checking what data sets are available is always a good idea before you try to open a particular data set since some have UPPER CASE letters in their names and others do not, e.g. data set `abbey` *is all lower case but* `Aids2` *has a capital letter.* ***R*** *will not recognise* `aids2` *as the name of a data set.*

d) try pressing the up arrow key ↑ and notice that you can retrieve commands that you issued — these can be edited and this saves a lot of time if you want to correct a small typing error in a complicated command.

All of these are important to do yourself

3) *Repeat all of the commands (**including the mistakes!**) given in the notes on pages 8 to 12, looking at the comments on these given on pages 13 – 15.*

All of these are useful to do if you are new to **R**.

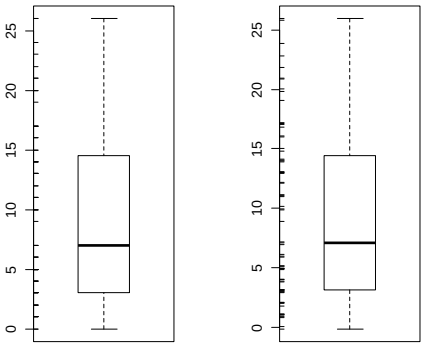
4) *If you have time then use the help system to find out more about some commands and then read ahead in section 2.2 Graphical Summaries and try out some of the worked examples there on pages 23 & 24, and maybe further….*

5) *Look again at the data on InsectSprays: Try the following:*

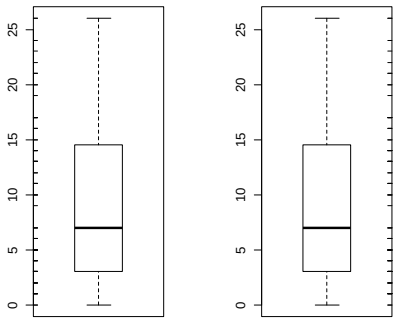
```
data(InsectSprays)
attach(InsectSprays)
xcount<- jitter(count)
par(mfrow=c(1,2))
boxplot(count)
rug(count,side=2)
boxplot(xcount)
rug(xcount,side=2)
```

The effect of `jitter(count)` *is to separate out the individual observations by a very small amount so that you can see (for example) whether they are evenly spaced our or clustered together in small groups. Now produce a display with the rug plot of the jittered data on the boxplot of the actual counts.*

```
>
> library()
> library(help=MASS)
> library(MASS)
> data()
> attach(InsectSprays)
> xcount<- jitter(count)
> par(mfrow=c(1,2))
> boxplot(count)
> rug(count,side=2)
> boxplot(xcount)
> rug(xcount,side=2)
>
```




```
> boxplot(count)
> rug(count,side=2)
>
> boxplot(count)
> rug(count,side=4)
>
```

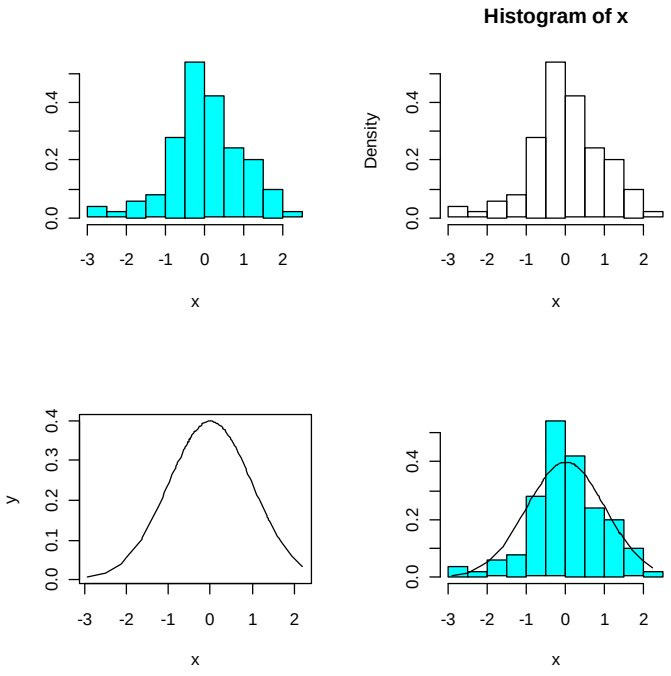


Note that putting the rug on the right hand side (with side=4) avoids confusion with tick marks on the vertical scale.

6) *Following the example on P32, generate 100 random numbers from a standard Normal distribution.*

a) Display them in a histogram with a plot of the density of N(0,1) superimposed (as in P32/33).

```
> par(mfrow=c(2,2))
> x<-rnorm(100)
> x<-sort(x)
> y<-exp(-x*x/2)/sqrt(2*pi)
> truehist(x)
> hist(x,probability=TRUE)
> plot(x,y,type='l')
> truehist(x)
> lines(x,y,type='l')
>
```



b) Repeat the histogram with a kernel density estimate instead of the 'true' density, using

i) the default bandwidth

ii) bandwidth adjusted to a smaller value

iii) bandwidth adjusted to a larger value.

```
> par(mfrow=c(2,2))
> x<-rnorm(100)
> kdedefault<-density(x)
> kdeadjdown<-density(x,adjust=0.3)
> kdeadjup<-density(x,adjust=1.5)
> kdeadjupup<-density(x,adjust=4.0)
> truehist(x)
> lines(kdedefault)
>
> truehist(x)
> lines(kdeadjdown)
>
> truehist(x)
> lines(kdeadjup)
>
> truehist(x)
> lines(kdeadjupup)
```