

DEPARTMENT OF STATISTICS
Course STATS 330: Advanced Statistical Modelling
Tutorial Sheet I: July 29, 2010

This tutorial is designed to give you practice in the following:

- Logging on to the system
- Download a data file from the class web site
- Create an R data frame from the file
- Turn a numeric variable into a factor
- Draw some plots
- Get data into Ggobi and spin it
- Make a Microsoft Word document containing text and graphics copied from R

In this tutorial we will be using the **petrol vapour data** on the website. It will also be discussed in class in Lecture 7. After completing the tutorial, you should be able to complete Assignment 1.

Task 1: Logging on

Log on to the system. Open the My Documents folder on the desktop. Make a folder entitled 330 Data Sets

Task 2: Download a data file

Start Internet Explorer, and go to the course web page at

www.stat.auckland.ac.nz/~lee/330

Click on **Data Sets**, and then click on **Petrol Vapour Data**

Save the data as a text file **vapour.txt** in the folder you made in Task 1.

Task 3: Create a data frame

Start R. Make a data frame **vapour.df** from the data in the text file you downloaded in **Task 2**, by typing

```
> vapour.df<-read.table(file.choose( ),header=T)
```

and selecting the file **vapour.txt**. Type **vapour.df** to inspect the contents of the data frame.

Task 4: Create a factor

Suppose we want to divide the data into 2 groups: Group1 with values of **hc** less than 30, and group 2 the remaining data.

We can make a variable group having values 1 and 2 according as **hc** is less than 30 or not in various ways. One way is to create a vector **group** having values all 2, and then modify it so that the observations for which **hc<30** have **group** reset to 1. The code to do this is

```
group = rep(2, 125) # this makes 125 repeats of the
value 2
# (there are 125 observations in the data set)
group[vapour.df$hc<30] = 1 # this is subsetting
```

The variable group is not a factor, but rather is a numeric vector. Suppose we want to turn it into a factor with “levels” Group1 and Group2. We type

```
group.factor = factor(group, labels=c("Group1", "Group2"))
```

To add this variable to the data frame (although this is not really necessary) you can type

```
vapour.df = data.frame(vapour.df, group.factor)
```

Task 5: Make some plots

Draw the following plots:

- boxplots of **group.factor** versus **t.temp** (what happens if we plot **group** against **t.temp**?)
- a pairs plot
- a coplot of **t.temp** and **t.vp**, conditioning on **p.temp**
- using the (non-trellis) function **coplot**
- using the trellis function **xyplot**
- a 3-dimensional scatter plot of **t.temp**, **t.vp** and **p.temp**

Points to note:

1. You will have to attach the data frame for some of the plots (otherwise the function can't find the data), better, use e.g. **vapour.df\$t.temp**
2. You will have to load the trellis library (actually called the lattice library in R) to get the **xyplot** function.
3. You will have to load the **scatterplot3d** library to get the **scatterplot3d** function.

Sample code:

```
# for the boxplots
plot(group.factor, vapour.df$t.temp, ylab = "Tank
Temperature", xlab = "Group")

# or

plot(t.temp~group.factor, data=vapour.df, ylab = "Tank
Temperature", xlab= "Group")
```

```

# for the pairs plot
pairs(vapour.df)

# for the coplot, using the coplot function
coplot(t.temp~t.vp|p.temp, data=vapour.df)

# same using the trellis function xyplot
library(lattice)
xyplot(t.temp~t.vp|equal.count(p.temp), data=vapour.df)

# for the 3d scatterplot
library(scatterplot3d)
scatterplot3d(vapour.df$t.temp, vapour.df$t.vp,
vapour.df$p.temp, type="h")
cloud(t.temp~t.vp*p.temp, data=vapour.df)

```

Task 6: Enter the data into Ggobi, and spin the plot

To get data into Ggobi, you must create a CSV file somewhere on the computer you are using. We will make a suitable CSV file from the R data frame **vapour.df**.

We need to put the file somewhere. R has a “default folder”, which depends on where R was started from. You can reset this using the File menu in R: first click on the console window, then select “Change dir...” from the file menu. This will bring up a dialog box which you can use to navigate your way to a suitable place to store the CSV file. Then type

```
write.table(vapour.df, "vapour.csv", row.names=FALSE,
col.names=TRUE, quote=FALSE, sep=",")
```

This will create a suitable CSV file called `vapour.csv` in the place you have nominated.

Now fire up Ggobi. Choose “Open” from the file menu, navigate to the place you have stored your file `vapour.csv`, and click “Open”. Then select “Rotate” from the View menu. You can stop the rotating plot by clicking the pause box. You can then rotate the point cloud manually with the mouse. You should be able to see three clusters of points in the data.

Task 7: Make a Word Document containing the data and the 3-d scatterplot

Open MS Word. In R, type **vapour.df** to get a listing of the data frame. Cut and past the first 10 observations of this data frame into Word. Then redraw the 3d scatter plot.

Cut and paste this into Word. (Use Ctrl-W to copy and Ctrl-V to paste.) Print out the Word document.