Department of Statistics

COURSE STATS 330

Model answers for Assignment 1, 2006

Question 1.

Note: in these solutions I have used some quite advanced R features. Most of you will not have done this in your solutions, but it is not necessary to use advanced features in order to get a good mark. I have included these features in my solutions so that you can see how R can be customized.

Load the data into R.

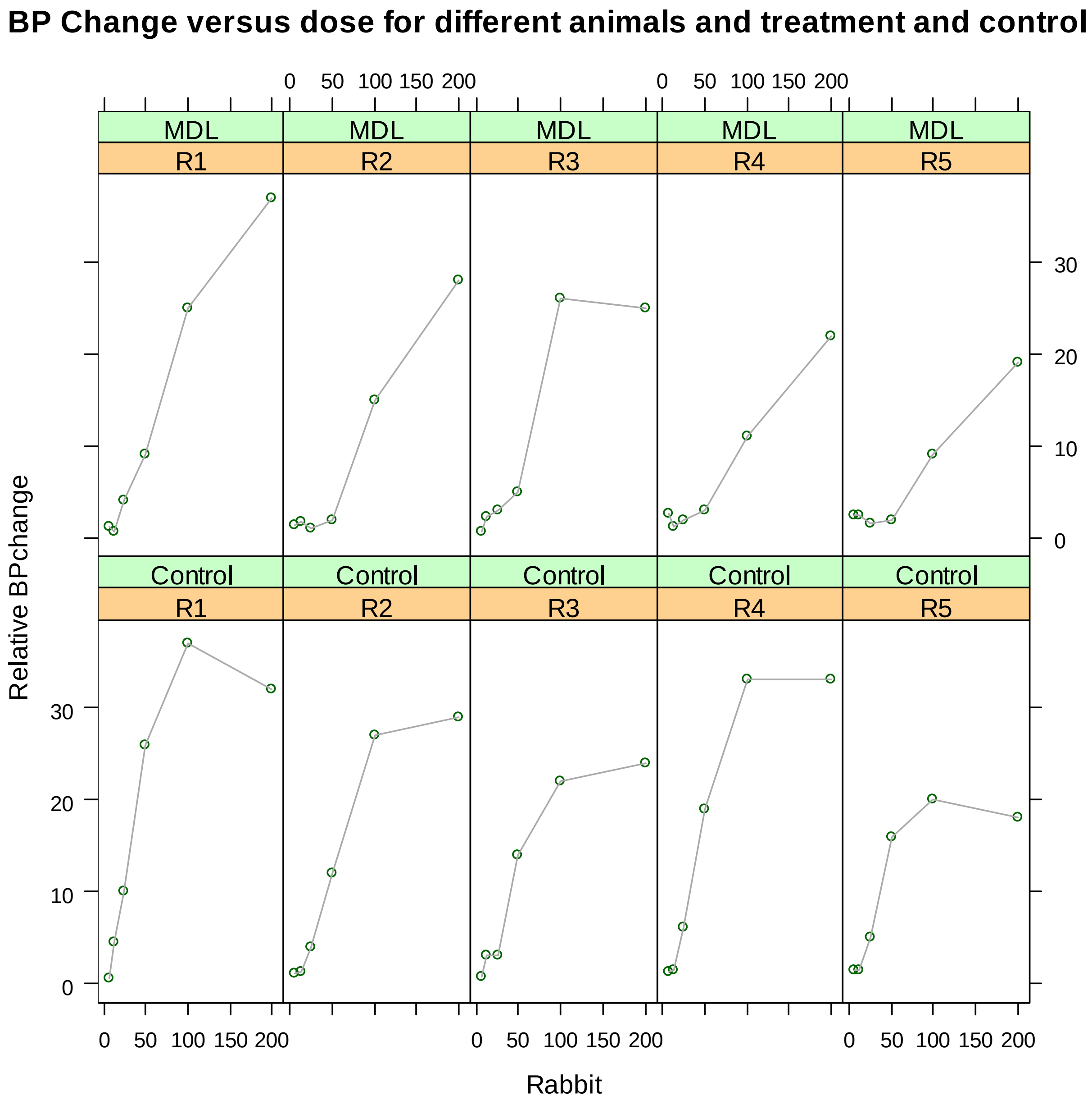
```
rabbit.df = read.csv(file.choose(), header=T)
```

[5 marks]

Draw a suitable graph or graphs that will allow you to examine the effect, if any, of the treatment on the relationship between dose and blood pressure.

```
library(lattice)  # need to load the lattice library

trellis.par.set(background = list(col = "white"),
plot.symbol=list(col="darkgreen"))

# trellis.par.set is the method for changing the background
# colours and plot symbol colours (this is optional)

xyplot(BPchange~Dose|Animal*Treatment, data=rabbit.df,
panel=function(x,y){panel.xyplot(x,y)
panel.linejoin(x,y, horizontal=F)
},
xlab = "Rabbit",
ylab = "Relative BPchange",
main = "BP Change versus dose for different animals and
treatment and control")

# note the use of the argument “panel” to draw connecting
# lines on the plots, this is often quite useful.
```

This code produces the trellis graph overleaf.



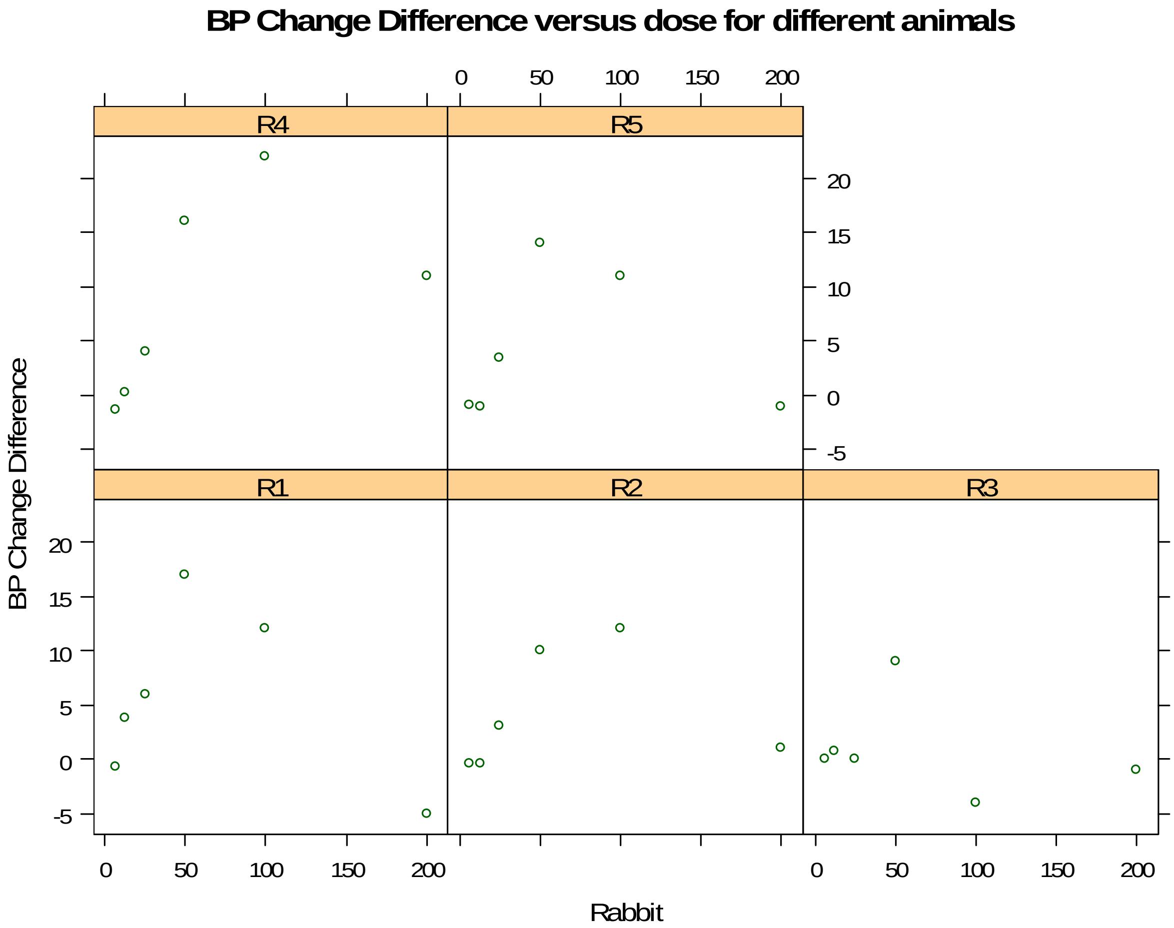
What conclusions do you reach? Do all rabbits respond in the same way?
At higher doses, the BPchange flattens out for the control observations, but not to the same extent for the MDL observations (except for rabbit R3)

Do all rabbits respond in the same way?
Rabbit R1 has the biggest relative change, rabbit R5 the smallest. Rabbit R3's response to the high dose is different from the other rabbits.
The differences between the individual rabbits are bigger than the treatment effect, making the treatment effect hard to detect. We could get a better idea of the treatment effect by looking at the differences between the MDL and control observations. First, we make a new data frame for the differences:

```
new.df = data.frame(Dose= rabbit.df[1:30,2],
Animal = rabbit.df[1:30,4],
Difference = rabbit.df[1:30,1] -rabbit.df[31:60,1] )
```

Then do a trellis plot of the differences:

```
xyplot(Difference~Dose|Animal, data=new.df,
xlab = "Rabbit",
ylab = "BP Change Difference",
main = "BP Change Difference versus dose for different
animals")
```



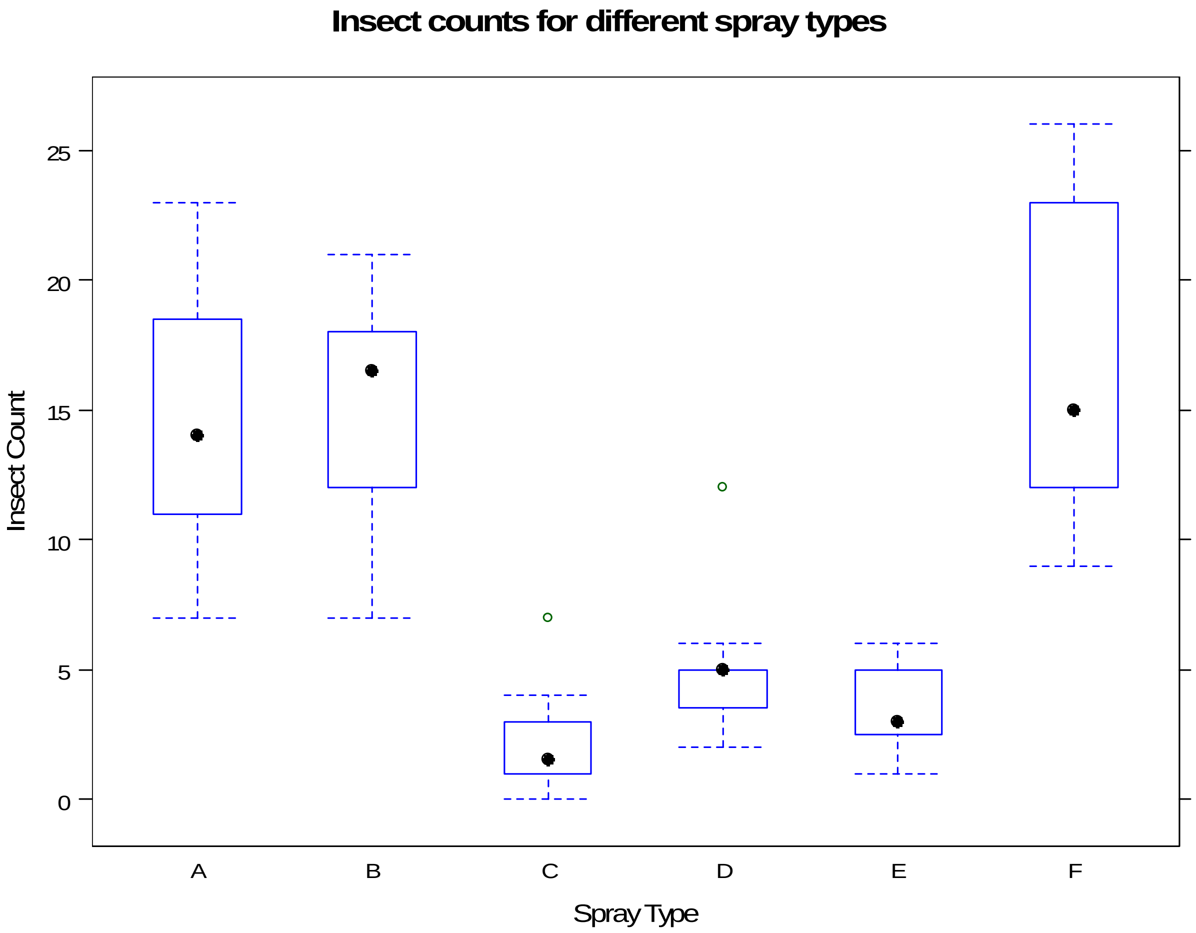
Now the conclusions are clearer: the MDL observations tend to be greater than the control observations, and the differences increase as dose increases up to about 100, then decrease. This holds in general, except for rabbit R3, where the MDL treatment seems to have little effect. *[8 marks for graphs, 7 marks for conclusions]*

Question 2.

`a)` *Use trellis graphics to draw a series of side-by side boxplots to illustrate the relationship between a continuous variable and a categorical variable (use the data frame* ***`InsectSprays,`*** *treating the count variable as continuous). Provide a title and informative axis labels. [5 marks, 4 for the plot and 1 for comments]*

The following code does the trick:

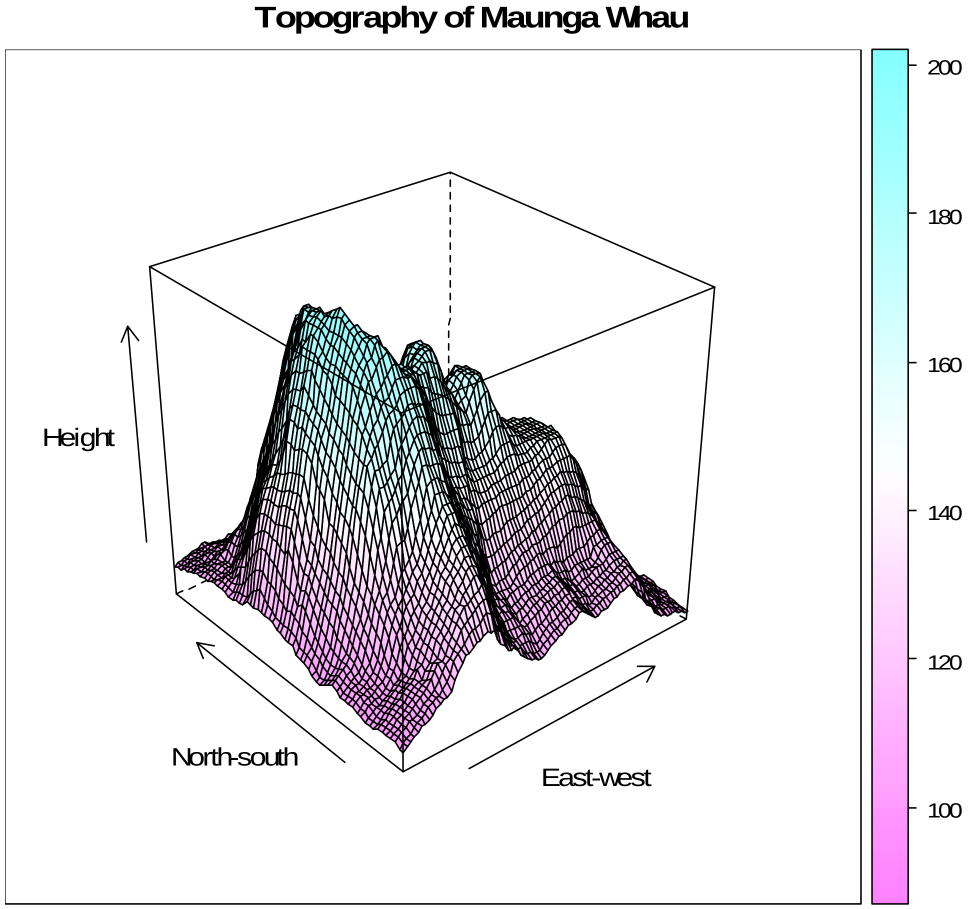
```
trellis.par.set(background = list(col = "white"),
box.rectangle=list(col="blue"),
box.umbrella=list(col="blue"))

# change the default colours for the boxplot

bwplot(count~spray, data=InsectSprays, xlab="Spray Type",
ylab="Insect Count", main = "Insect counts for different
spray types")
```



The insect counts for insecticides A, B, F are greater than the other types. *[5 marks, 4 for the plot and 1 for comments]*

b) A wireframe plot, using the data frame **`volcano.`** Use informative labels for the axes**.** *[5 marks, 4 for the plot and 1 for comments]*

```
wireframe(volcano, xlab="East-west",ylab = "North-south",
zlab="Height", main="Topography of Maunga Whau",drape=T)
```



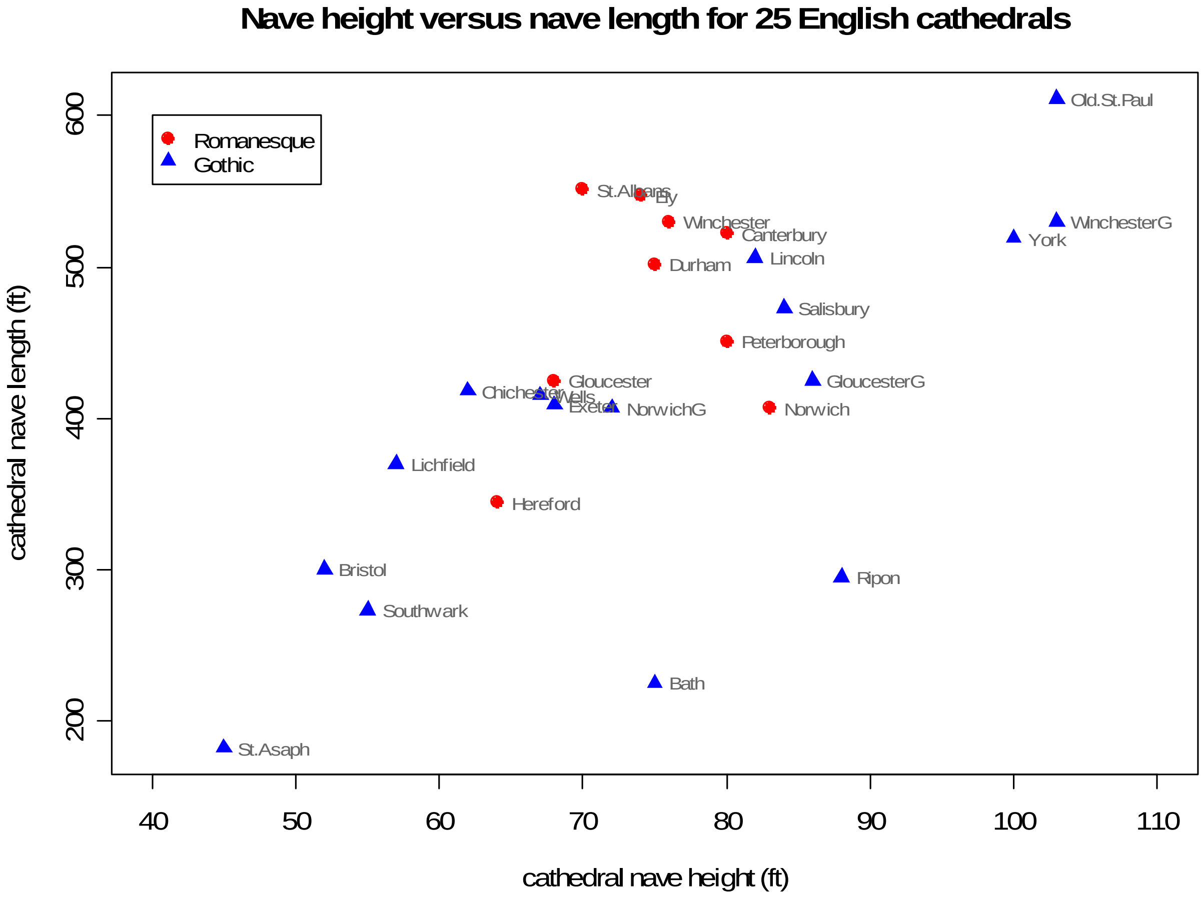
The plot gives a good idea of the surface (in this case the topography of Mt Eden!) Note that the input to the wireframe function is a data frame with a particular structure, with rows representing the x coordinates (longitude in this example), and columns the y coordinates (latitude in this example). The actual data values are the z-values (height in this case) at the x-y point defined by a particular row and column. See the documentation for the volcano data set (type **?volcano**)

[5 marks, 4 for the plot and 1 for comments]

c) The following code draws the picture:

```
plot(c(40,110), range(cathedral$y),
xlab="cathedral nave height (ft)",
ylab = "cathedral nave length (ft)",
main="Nave height versus nave length for 25 English
cathedrals",
type="n")

# type=”n”   suppresses the plotting of the points,
# we use points for this to control the colours (using
col)
# and plotting symbols (using pch)

points(cathedral$x,cathedral$y, pch = rep(c(16, 17),
c(9, 16)), col = rep(c("red", "blue"), c(9,16)), cex=1.2)

# cex = 1.2  increases the size of the points. The first 9
# observations are Romanesque, the next 16 Gothic

text(cathedral$x,cathedral$y, cathedral.names,
col="grey40", cex=0.7, pos=4)

# col=”grey40” writes the text in a slightly lighter
manner
# to allow the points to stand out better

legend(40,600, legend = c("Romanesque","Gothic"),
col = c("red", "blue"),
pch = c(16,17), cex=0.8, pt.cex=1.1)

# legend  writes the legend on the plot. pt.cex enlarges
# the symbol in the legend
```

PTO for graph



There is quite a strong relationship between nave height and nave length for the Gothic cathedrals, (except for Ripon and Bath) but the relationship is weaker for Romanesque cathedrals.

[10 marks, 8 for the plot and 2 for comments]

[Total: 40 marks]