STAT3008 Exercise 1 Solutions
(2011 – 2012 2nd Semester)

Q1.

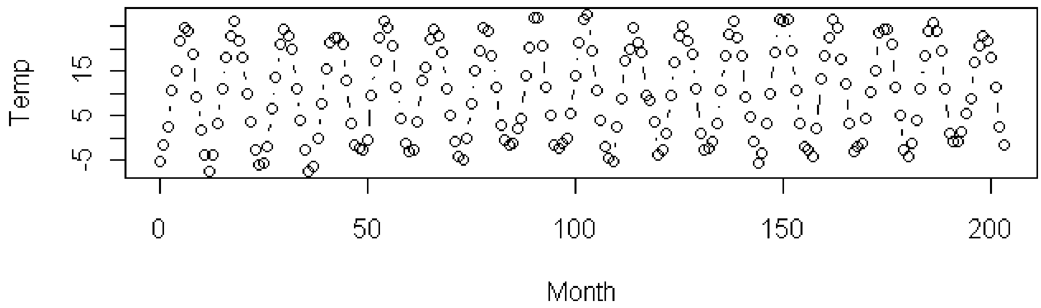
Problem 1.2

1.2.1 From the graph, it appears to be a null plot with no particularly interesting characteristics.

1.2.2

R codes:

```
library(alr3)
plot(Mitchell,type="b")
```



Scaling does matter. The points have been joined with lines to emphasize the temporal pattern in the data: temperature is high in the summer but low in the winter.

Q2.

Problem 1.3

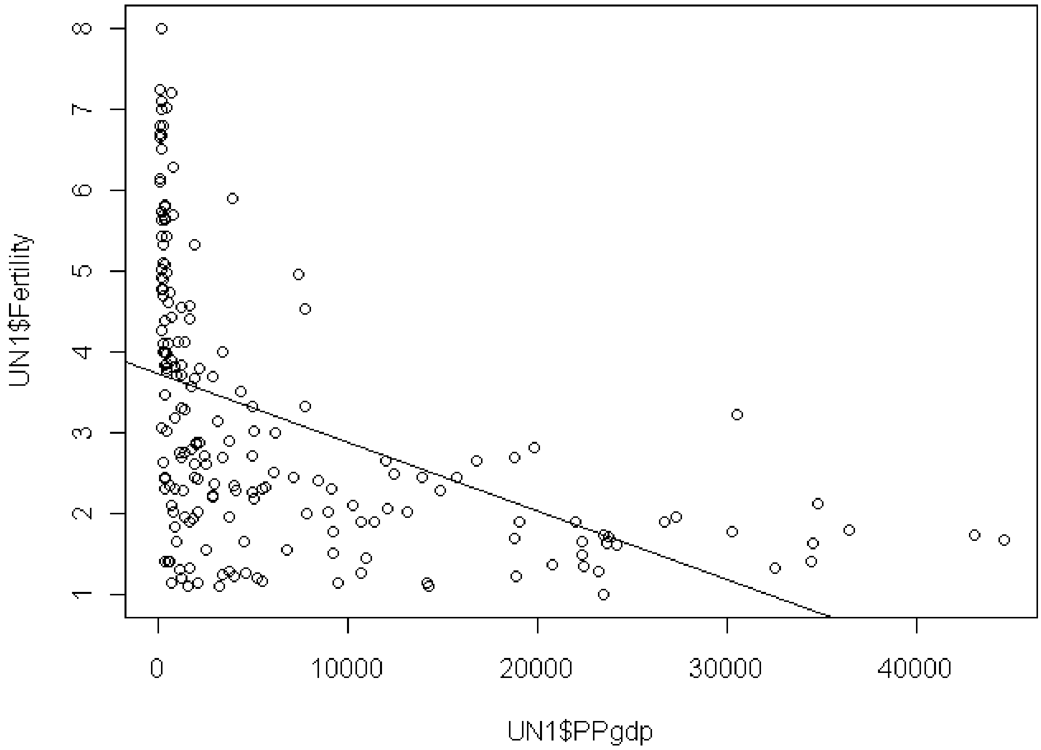
1.3.1

The predictor is a function of *PPgdp* while the response is a function of *Fertility*.

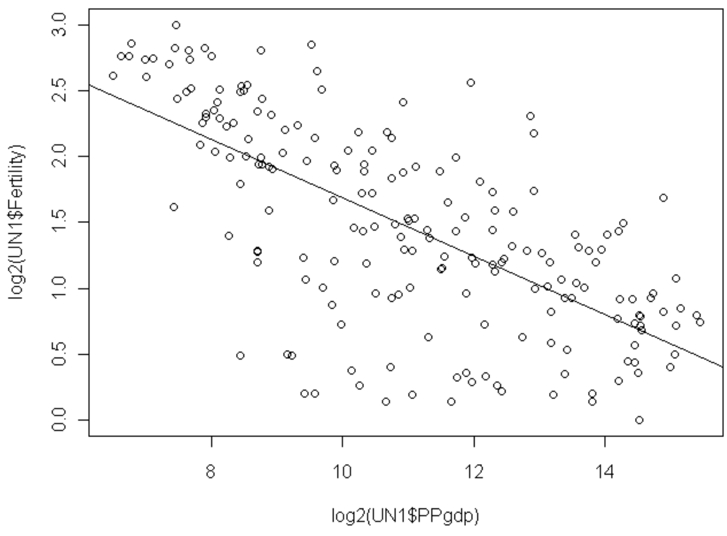
1.3.2

R codes:

```
library(alr3)
plot(UN1$PPgdp,UN1$Fertility)
fittedline=lm(UN1$Fertility~UN1$PPgdp)
abline(fittedline)
```



Simple linear regression is not a good summary of the graph. The mean function does not appear to be linear and also the variance should not be constant.
1.3.3
R Codes:
library(alr3)
plot(log2(UN1$PPgdp),log2(UN1$Fertility))
fittedline=lm(log2(UN1$Fertility)~log2(UN1$PPgdp))
abline(fittedline)

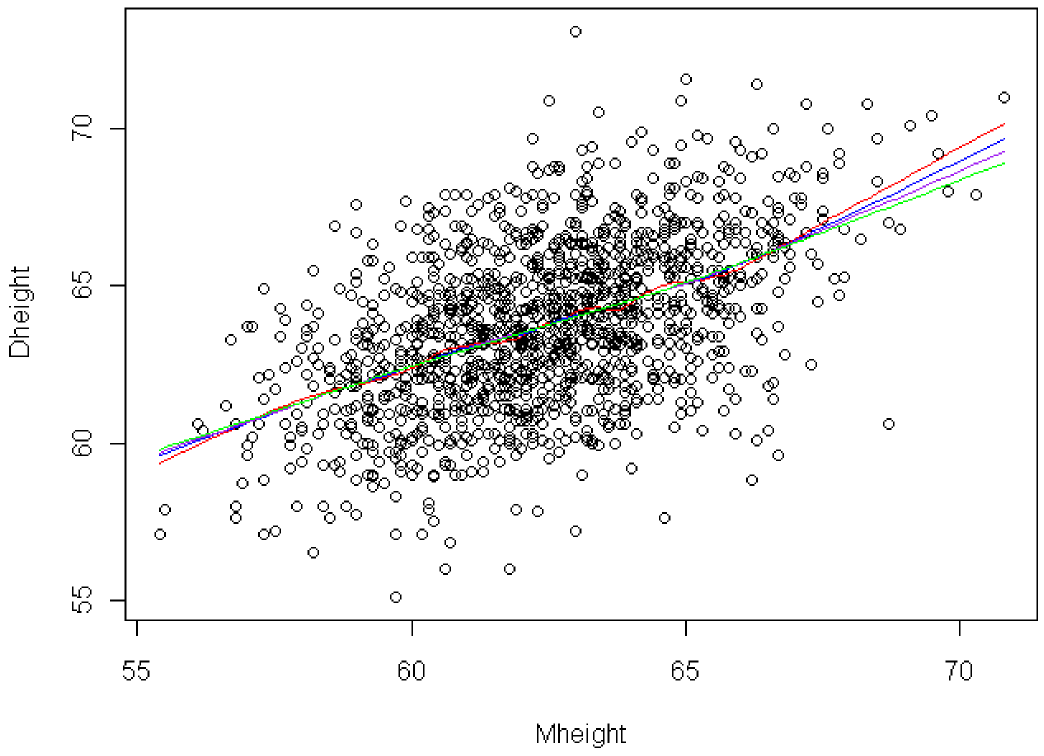


Simple linear regression is much more appropriate after transformation of the response and predictor. The mean function appears to be nearly with fairly constant variance.

Q3

R Codes:

```
library(alr3)
plot(heights)
lines(stats::lowess(heights,f=0.2),col="red")
lines(stats::lowess(heights,f=0.4),col="blue")
lines(stats::lowess(heights,f=0.6),col="purple")
lines(stats::lowess(heights,f=0.8),col="green")
```

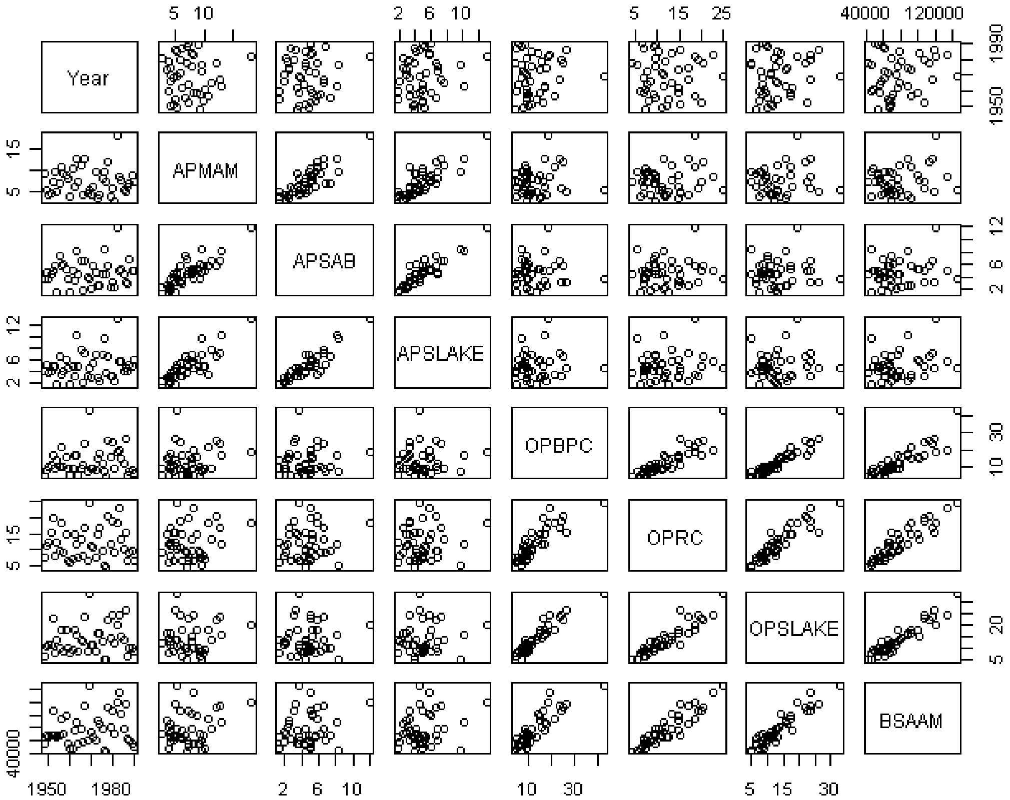


Q4

R Codes:

library(alr3)

plot(water)



According to the above graph,

(1) Two pairs of variables that are not correlated: YEAR and APMAM, YEAR and APSAB.

(2) Two pairs of variables that are correlated: APMAM and APSAB, OPBPC and OPRC.

(3) One pair of variables that do not have a constant variance function: APSAB and OPRC.

(4) One pair of variables that an outlier exists: APSAB and OPBPC.

(5) One pair of variables that an influential point exists: OPBPC and BSAAM

Note that there are not just only the above correct answers. There are also many other answers that are correct.

Q5

$$\sum_i (x_i - \overline{x})(y_i - \overline{y})$$

$$= \sum_i (x_i - \overline{x}) y_i - \sum_i (x_i - \overline{x}) \overline{y}$$

$$= \sum_i (x_i - \overline{x}) y_i - \overline{y} \sum_i (x_i - \overline{x})$$

$$= \sum_i (x_i - \overline{x}) y_i - \overline{y}[\sum_i x_i - \sum_i \overline{x}]$$

$$= \sum_i (x_i - \overline{x}) y_i - \overline{y}(n \times \overline{x} - n \times \overline{x})$$

$$= \sum_i (x_i - \overline{x}) y_i$$

For $\sum_i (x_i - \overline{x})(y_i - \overline{y}) = \sum_i x_i (y_i - \overline{y})$, it is similar to what we did in the above steps.