STAT3008 Exercise 5 Solutions

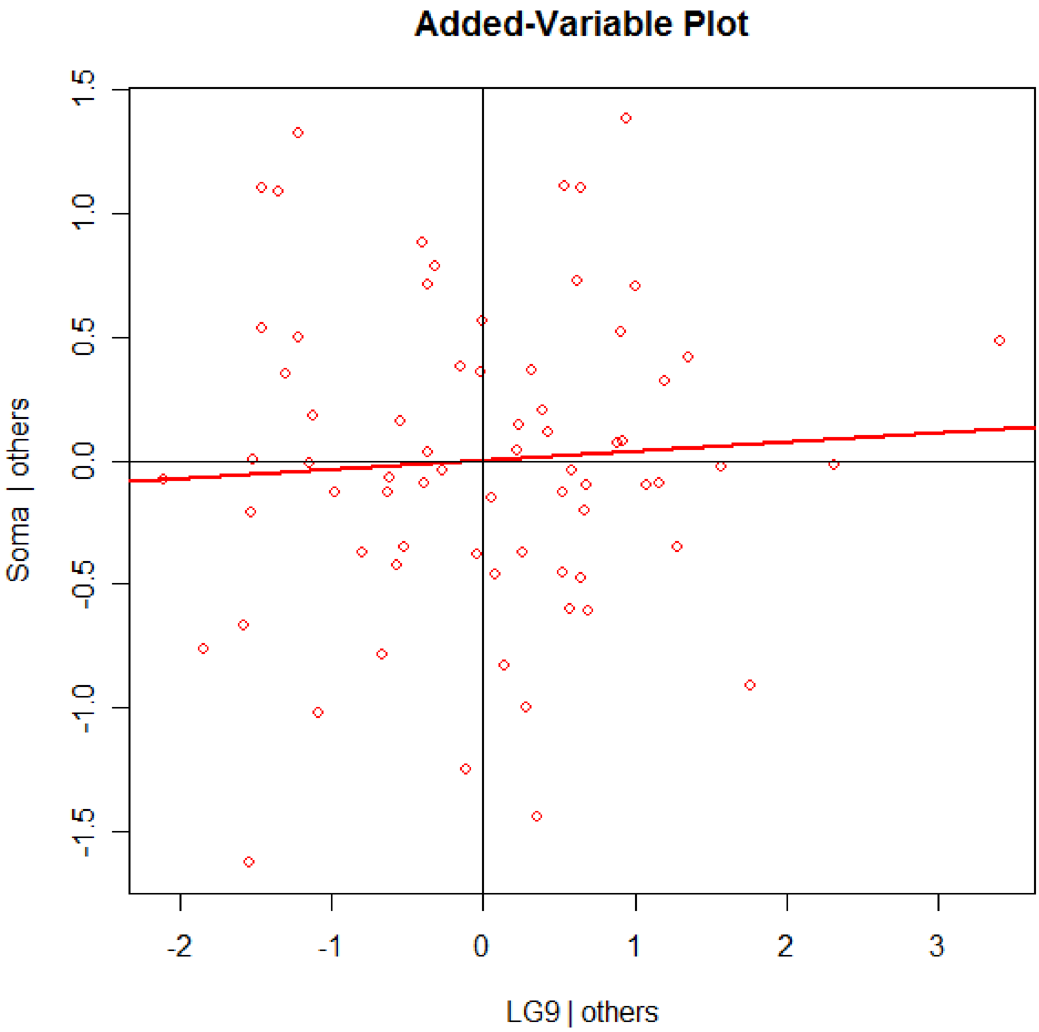
(2011-2012 2nd Semester)

Q1.

Problem 3.1.2

R Codes:

```
library(car)
av.plots(lm(Soma~WT9+LG9,data=BGSgirls))
abline(v=0,h=0,col="black")
```



Problem 3.1.3

R Codes:

```
library(car)
m1=lm(Soma~HT2+WT2+HT9+WT9+ST9,BGSgirls)
```

summary(m1)
Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept) 8.8590417 2.3764431 3.728 0.000411
HT2 -0.0792535 0.0354034 -2.239 0.028668
WT2 -0.0409358 0.0754343 -0.543 0.589244
HT9 -0.0009613 0.0260735 -0.037 0.970704
WT9 0.1280506 0.0203544 6.291 3.2e-08
ST9 -0.0092629 0.0060130 -1.540 0.128373

Residual standard error: 0.5791 on 64 degrees of freedom
Multiple R-Squared: 0.5211,
F-statistic: 13.93 on 5 and 64 DF, p-value: 3.309e-09

$R^2 = 0.5211$, $\widehat{\sigma^2} = 0.5791^2 = 0.3354$

For t-test, $H_0 : \beta_j = 0$ with other β s arbitrary.

$H_1 : \beta_j \neq 0$ and other β s arbitrary

Therefore, β_0, β_1 and β_4 are significant at 5% of significance level.

(i) $H_0 : E(Soma|X) = \beta_0 + \beta_1 HT2 + \beta_2 WT2 + \beta_4 WT9 + \beta_5 ST9$

$H_1 : E(Soma|X) = \beta_0 + \beta_1 HT2 + \beta_2 WT2 + \beta_3 HT9 + \beta_4 WT9 + \beta_5 ST9$

R Codes:
library(car)
m1=lm(Soma~HT2+WT2+HT9+WT9+ST9,BGSgirls)
m2=lm(Soma~HT2+WT2+WT9+ST9,BGSgirls)
anova(m2,m1)
F-statistic=0.00014=(-0.037)*(-0.037)=t-statistic^2 ,
p-value=0.9707.
t-test is equivalent to this F-test.
As a result, H_0 is not rejected at 5% significance level. i.e. Soma is independent of HT9.

(ii) $H_0 : E(Soma|X) = \beta_0 + \beta_1 HT2 + \beta_2 WT2 + \beta_5 ST9$

$$H_1: E(Soma|X) = \beta_0 + \beta_1 HT2 + \beta_2 WT2 + \beta_3 HT9 + \beta_4 WT9 + \beta_5 ST9$$

R Codes:

m3=lm(Soma~HT2+WT2+ST9,BGSgirls)

anova(m3,m1)

F-statistic=27.78

p-value=2.065e-09

Therefore, H_0 is rejected at 5% significance level.

(iii) (a) **R Codes:**

predict(m1,data.frame(HT2=50,WT2=15,HT9=100,WT9=30,ST9=10),interval="confidence",level=.99)

The confidence interval is [4.9166, 10.9536]

(b) **R Codes:**

predict(m1,data.frame(HT2=50,WT2=15,HT9=100,WT9=30,ST9=10),interval="confidence",level=.99)

The prediction interval is [4.5476, 11.3226]

(c) **R Codes:**

i=diag(rep(1,6))

x=cbind(matrix(1,nr=70),BGSgirls$HT2,BGSgirls$WT2,BGSgirls$HT9,BGSgirls$WT9,BGSgirls$ST9)

predict(m1,data.frame(HT2=50,WT2=15,HT9=100,WT9=30,ST9=10),interval="confidence",level=.99)[1]-sqrt(6*qf(0.95,6,64)*matrix(c(1,50,15,100,30,10),nc=6)%*%solve(t(x)%*%x,i)%*%t(matrix(c(1,50,15,100,30,10),nc=6)))*0.5791

The confidence band is [3.7630, 12.1072]

Q2.

Problem 2.7

2.7.1

$$RSS = \sum (y_i - \beta_1 x_i)^2$$

$$\frac{\partial RSS}{\partial \beta_1} = -2\sum x_i (y_i - \beta_1 x_i)$$

$$\left.\frac{\partial RSS}{\partial \beta_1}\right|_{\beta_1 = \widehat{\beta}_1} = -2\sum x_i (y_i - \widehat{\beta}_1 x_i) = 0$$

$$\text{Therefore, } \widehat{\beta}_1 = \frac{\sum x_i y_i}{\sum x_i^2}$$

$$E(\widehat{\beta}_1) = E(\frac{\sum x_i y_i}{\sum x_i^2}) = \frac{\sum x_i E(y_i)}{\sum x_i^2}$$

$$= \frac{\sum x_i(\beta_1 x_i)}{\sum x_i^2} = \beta_1 \frac{\sum x_i^2}{\sum x_i^2} = \beta_1$$

Thus, $\widehat{\beta}_1$ is unbiased for β_1.

$$Var(\widehat{\beta}_1) = Var(\frac{\sum x_i y_i}{\sum x_i^2}) = \frac{\sum x_i^2 Var(y_i)}{(\sum x_i^2)^2}$$

$$= \frac{\sum x_i^2 \sigma^2}{(\sum x_i^2)^2} = \frac{\sigma^2}{\sum x_i^2}$$

$$\widehat{RSS} = \sum (y_i - \widehat{\beta}_1 x_i)^2$$

$$= \sum y_i^2 + \widehat{\beta}_1^2 \sum x_i^2 - 2\widehat{\beta}_1 \sum x_i y_i$$

$$= \sum y_i^2 - \frac{(\sum x_i y_i)^2}{\sum x_i^2}$$

As there is only one parameter estimate in the regression, the degree of freedom of the estimated variance is (n-1).

$$\text{So} \quad \widehat{\sigma^2} = \frac{\sum y_i^2 - \frac{(\sum x_i y_i)^2}{\sum x_i^2}}{n-1}$$

2.7.2

For 2.16, $\widehat{RSS_1} = SYY - \frac{SXY^2}{SXX}$. For 2.30, $\widehat{RSS_2} = \sum y_i^2 - \frac{(\sum x_i y_i)^2}{\sum x_i^2}$.

The ANOVA Table is as follows:

Source	Df	SS	MS	F
Regression	**1**	$\widehat{RSS_2}$ - $\widehat{RSS_1}$	$\widehat{RSS_2}$ - $\widehat{RSS_1}$	$\left(\widehat{RSS_2} - \widehat{RSS_1}\right) \Big/ \frac{\widehat{RSS_1}}{n-2}$

Residuals	**n-2**	$\widehat{RSS}_1$	$\dfrac{\widehat{RSS}_1}{n-2}$	
Total	**n-1**	$\widehat{RSS}_2$		

R codes

```
library(data)
data(snake)
m2=lm(Y~X-1,data=snake)
m1= update(m2,~.+1)
m1
anova(m2,m1)
```

You can finally find that $t^2 = F$.

2.7.3

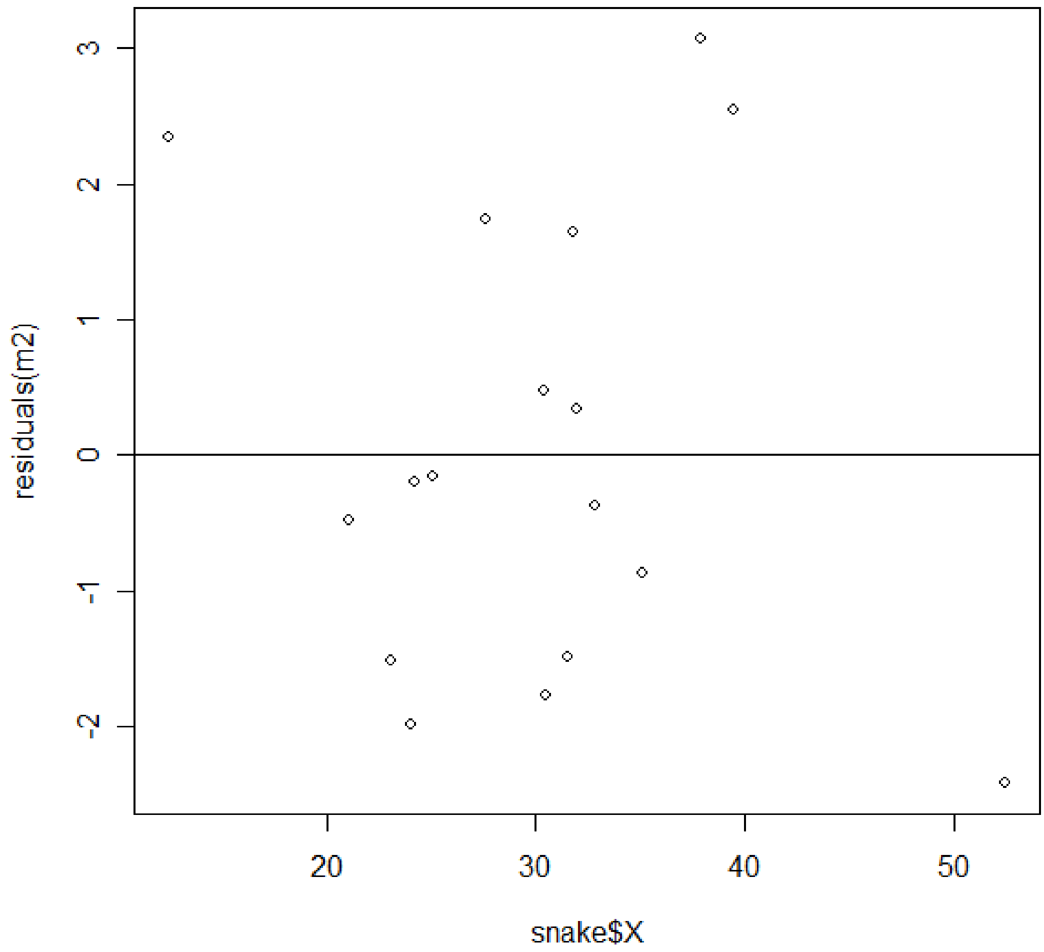
R codes

```
library(data)
data(snake)
m2=lm(Y~X-1,data=snake)
m1= update(m2,~.+1)
summary(m1)
anova(m2,m1)
```

2.7.4

R Codes:

```
plot(snake$X,residuals(m2))
abline(h=0)
```



The residual plot is okay. Only few points are very far away from zero.

Q2.

Problem 3.4

3.4.1

$$\hat{\beta}_1 = \frac{SX_1Y}{SX_1X_1}$$

$$\hat{\beta}_2 = \frac{SX_2Y}{SX_2X_2}$$

$$\hat{\beta}_3 = 0$$

3.4.2

$$\hat{e}_{1i} = y_i - \hat{\alpha} - \hat{\beta}_1 x_{i1} = y_i - \overline{y} - \hat{\beta}_1 (x_{i1} - \overline{x}_1)$$

$$\hat{e}_{3i} = x_{i2} - \hat{\gamma} - \hat{\beta}_1 x_{i1} = x_{i2} - \overline{x}_2$$

where α and γ are the intercepts of model 1 and model 3 respectively.

3.4.3

Since $\sum \hat{e}_{3i} = \sum (x_{i2} - \bar{x}_2) = 0$ and $\sum \hat{e}_{1i} = 0$,

$$Slope = \frac{\sum (\hat{e}_{1i} - 0)(\hat{e}_{3i} - 0)}{\sum \hat{e}_{3i}^{\ 2}} = \frac{\sum (y_i - \bar{y} - \widehat{\beta}_1 (x_{i1} - \bar{x}_1))(x_{i2} - \bar{x}_2)}{\sum (x_{i2} - \bar{x}_2)^2}$$

$$= \frac{\sum (y_i - \bar{y})(x_{i2} - \bar{x}_2) - \widehat{\beta}_1 \sum (x_{i1} - \bar{x}_1)(x_{i2} - \bar{x}_2)}{SX_2X_2} = \frac{SX_2Y - 0}{SX_2X_2} = \frac{SX_2Y}{SX_2X_2}$$

Q3.

The 95% confidence region of β is:

$$\frac{(\beta - \widehat{\beta})^T (X^T X)(\beta - \widehat{\beta})}{2\widehat{\sigma^2}} \le F(0.05, 2, 100-2)$$

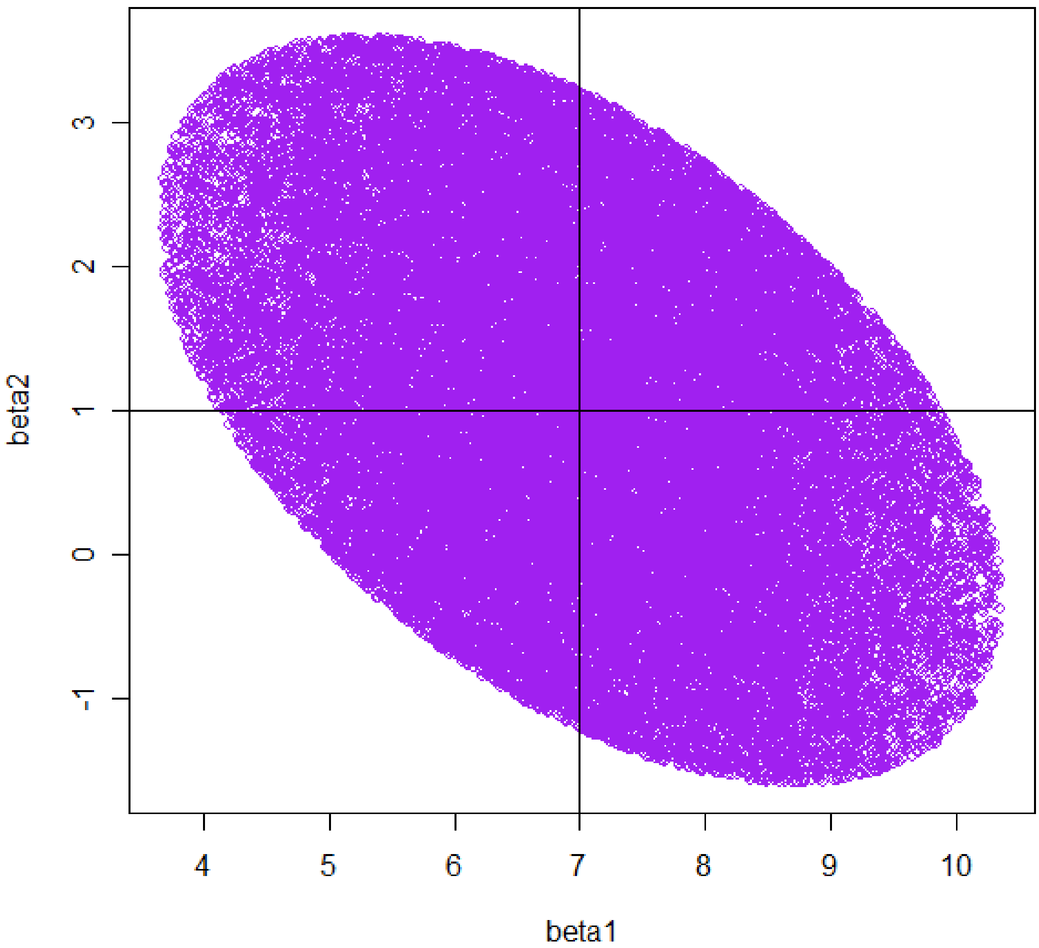
$$\frac{\left(\begin{pmatrix}\beta_1 \\ \beta_2\end{pmatrix} - \begin{pmatrix}7 \\ 1\end{pmatrix}\right)^T \begin{pmatrix}3 & 2 \\ 2 & 5\end{pmatrix}\left(\begin{pmatrix}\beta_1 \\ \beta_2\end{pmatrix} - \begin{pmatrix}7 \\ 1\end{pmatrix}\right)}{2 \times 2^2} \le 3.0892$$

$$\begin{pmatrix}\beta_1 - 7 & \beta_2 - 1\end{pmatrix}\begin{pmatrix}3 & 2 \\ 2 & 5\end{pmatrix}\begin{pmatrix}\beta_1 - 7 \\ \beta_2 - 1\end{pmatrix} \le 3.0892 \times 8$$

$$\begin{pmatrix}3\beta_1 + 2\beta_2 - 23 & 2\beta_1 + 5\beta_2 - 19\end{pmatrix}\begin{pmatrix}\beta_1 - 7 \\ \beta_2 - 1\end{pmatrix} \le 3.0892 \times 8$$

$$3\beta_1^{\ 2} + 5\beta_2^{\ 2} + 4\beta_1\beta_2 - 46\beta_1 - 38\beta_2 + 180 \le 3.0892 \times 8$$

$$3\beta_1^{\ 2} + 5\beta_2^{\ 2} + 4\beta_1\beta_2 - 46\beta_1 - 38\beta_2 + 155.2864 \le 0$$



beta2
3
2
1
0
-1
4
5
6
7
8
9
10
beta1