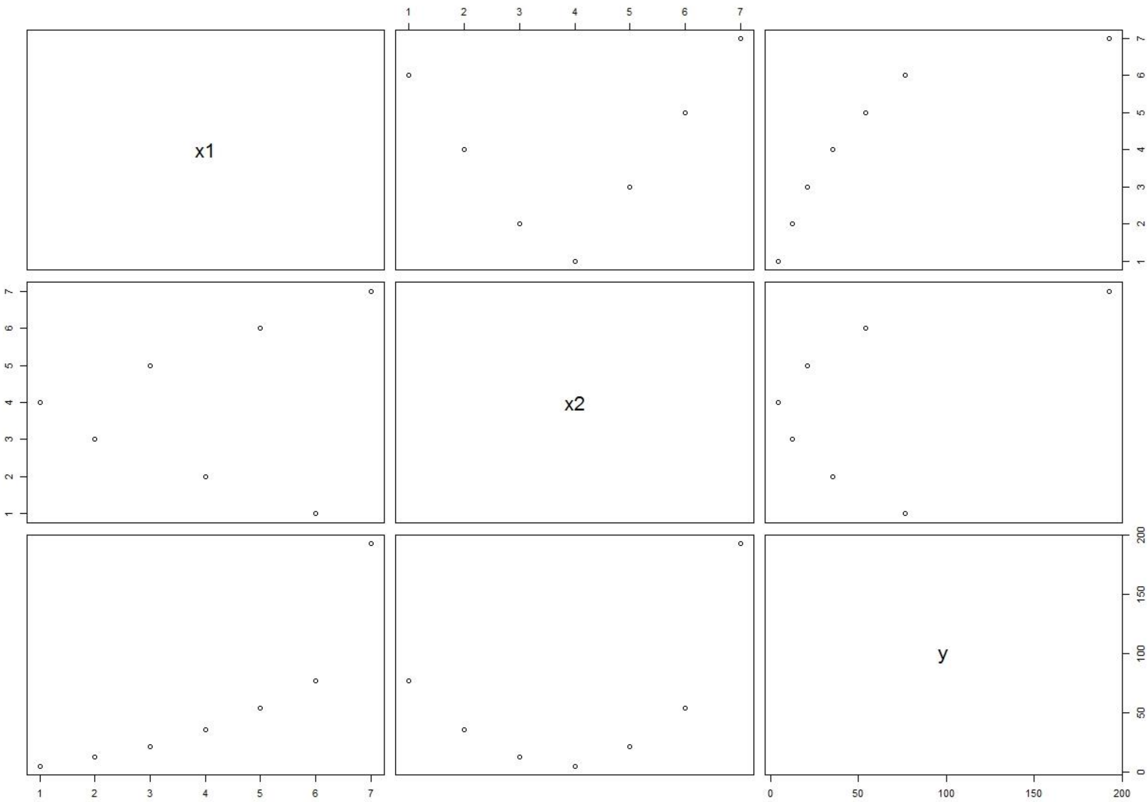
STAT3008 Assignment 4 Solutions

1.

i)R Codes:

```
x1=c(1,2,3,4,5,6,7)
x2=c(4,3,5,2,6,1,7)
y=c(4.562,12.280,21.018,35.643,54.084,76.642,192.780)
Data<-data.frame(x1,x2,y)
plot(Data)
```

Scatterplot Matrix



ii)R Codes:

```
lambda=c(-1,-1/2,0,1/3,1/2,1,2)
ytrans=NULL
for(i in 1:7){
        if(lambda[i]==0)temp=prod(y)^(1/7)*log(y)
        if(lambda[i]!=0)temp=(prod(y)^(1/7))^(1-lambda[i])*(y^lambda[i]-1)/lambda[i]
        ytrans=cbind(ytrans,temp)
```

```
}
sum(lm(ytrans[,1]~x1+x2)$residuals^2)
sum(lm(ytrans[,2]~x1+x2)$residuals^2)
sum(lm(ytrans[,3]~x1+x2)$residuals^2)
sum(lm(ytrans[,4]~x1+x2)$residuals^2)
sum(lm(ytrans[,5]~x1+x2)$residuals^2)
sum(lm(ytrans[,6]~x1+x2)$residuals^2)
sum(lm(ytrans[,7]~x1+x2)$residuals^2)
```

λ	-1	-1/2	0	1/3	1/2	1	2
RSS	10594.23	1905.136	186.0754	294.197	670.252	4356.311	86324.43

From the table, the RSS is minimized when $\lambda = 0$. Hence we choose $\lambda = 0$.

(iii)Choose to transform x_1.

R Codes:

```
lambda=c(2,3,4,5,6,7,8)
x1trans=NULL
for(i in 1:7){
    if(lambda[i]==0)temp=prod(x1)^(1/7)*log(x1)
    if(lambda[i]!=0)temp=(prod(x1)^(1/7))^(1-lambda[i])*(x1^lambda[i]-1)/lambda[i]
    x1trans=cbind(x1trans,temp)
}
sum(lm(y~x1trans[,1]+x2)$residuals^2)
sum(lm(y~x1trans[,2]+x2)$residuals^2)
sum(lm(y~x1trans[,3]+x2)$residuals^2)
sum(lm(y~x1trans[,4]+x2)$residuals^2)
sum(lm(y~x1trans[,5]+x2)$residuals^2)
sum(lm(y~x1trans[,6]+x2)$residuals^2)
sum(lm(y~x1trans[,7]+x2)$residuals^2)
```

λ	2	3	4	5	6	7	8
RSS	2019.282	918.8331	498.8717	428.9743	537.3498	732.9166	965.7738

From the table, the RSS is minimized when $\lambda = 5$. Hence we choose $\lambda = 5$.

iv)From ii), RSS(best fit)=186.0754; from iii), RSS(best fit)=428.9743. The method in ii) gives a smaller RSS.

2. The model we are considering is with the form of:

$$Y = X\beta + e, \qquad e \sim N(0, \sigma^2 W^{-1})$$

(i)

Solution: $\hat{\beta} = (X'WX)^{-1}X'WY$,
$\hat{Y} = X\hat{\beta} = X(X'WX)^{-1}X'WY = HY$
$\therefore H = X(X'WX)^{-1}X'W$.

(ii)

Solution: $H' = [X(X'WX)^{-1}X'W]' = W'X(X'WX)^{-1}X' \neq H$.
$\therefore$ H is not symmetric.

(iii)

Solution: $HH = [X(X'WX)^{-1}X'W][X(X'WX)^{-1}X'W] = X(X'WX)^{-1}X' = H$.

(iv)

Solution: $HX = [X(X'WX)^{-1}X'W]X = X$.

(v)

Solution: $X'H = X'[X(X'WX)^{-1}X'W] \neq X'$.

(vi)

Solution: $tr(H) = tr'[X(X'WX)^{-1}X'W] = tr[(X'WX)^{-1}X'WX] = tr(I_{p+1}) = p+1$.

3. The model we are considering is with the form of:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + e, \qquad e \sim N(0, \sigma^2 I)$$

(i)

Solution: $E[\sum(\hat{Y}_i - \bar{Y})^2] = E[Y'(H - J/n)Y]$
$= tr\{E[Y'(H - J/n)Y]\}$
$= E\{tr[Y'(H - J/n)Y]\}$
$= E\{tr[(H - J/n)YY']\}$
$= tr[(H - J/n)E(YY')]$
$= tr[(H - J/n)(\sigma^2 I + X\beta\beta' X')]$
$= \sigma^2 tr(H - J/n) + tr\{(H - J/n)X\beta\beta' X'\}$
$= \sigma^2(p + 1 - 1) + tr\{\beta' X'(H - J/n)X\beta\}$
$= 2\sigma^2 + tr\{\beta' X'(I - J/n)X\beta\} \qquad \because X'HX = X'X$
$= 2\sigma^2 + \beta' X'(I - J/n)X\beta.$

(ii)

Solution: $E[\sum(Y_i - \hat{Y}_i)^2] = E[Y'(I - H)Y]$
$= tr\{E[Y'(I - H)Y]\}$
$= E\{tr[Y'(I - H)Y]\}$
$= E\{tr[(I - H)YY']\}$
$= tr[(I - H)E(YY')]$
$= tr[(I - H)(\sigma^2 I + X\beta\beta' X')]$
$= \sigma^2 tr(I - H) + tr\{(I - H)X\beta\beta' X'\}$
$= \sigma^2[n - (p + 1)] + tr\{X\beta\beta' X' - HX\beta\beta' X'\}$
$= \sigma^2[n - (p + 1)] + tr\{X\beta\beta' X' - X\beta\beta' X'\} \qquad \because HX = X$
$= (n - p - 1)\sigma^2 + 0 = (n - 3)\sigma^2.$

(iii)

Solution: $E[\sum(Y_i - \bar{Y})^2] = E[Y'(I - J/n)Y]$
$= tr\{E[Y'(I - J/n)Y]\}$
$= E\{tr[Y'(I - J/n)Y]\}$
$= E\{tr[(I - J/n)YY']\}$
$= tr[(I - J/n)E(YY')]$
$= tr[(I - J/n)(\sigma^2 I + X\beta\beta' X')]$
$= \sigma^2 tr(I - J/n) + tr\{(I - J/n)X\beta\beta' X'\}$
$= \sigma^2(n - 1) + tr\{\beta' X'(I - J/n)X\beta\}$
$= (n - 1)\sigma^2 + \beta' X'(I - J/n)X\beta$

4. (i)

Solution:

j	R_j^2	VIF
1	0.7508	4.0136
2	0.7084	3.4294
3	0.7358	3.7844

(ii)

Solution: Forward selection algorithm with AIC:
$V_0 = 12.3447$ (with intercept only,$E(Y|X) = \beta_0$),
Step 1:
$E(Y|X) = \beta_0 + \beta_1 X_1$, $AIC = 9.5679 < V_0$,
$E(Y|X) = \beta_0 + \beta_2 X_2$, $AIC = 14.5510 > V_0$,
$E(Y|X) = \beta_0 + \beta_3 X_3$, $AIC = 9.2560 < V_0$,
$\therefore V_1 = 9.2560$, we add X_3 into the previous model, that means $E(Y|X) = \beta_0 + \beta_3 X_3$.
Step 2:
$E(Y|X) = \beta_0 + \beta_3 X_3 + \beta_1 X_1$, $AIC = 8.0612 < V_1$,
$E(Y|X) = \beta_0 + \beta_3 X_3 + \beta_2 X_2$, $AIC = 9.3443 > V_1$,
$\therefore V_2 = 8.0612$, we add X_1 into the previous model, that means $E(Y|X) = \beta_0 + \beta_3 X_3 + \beta_1 X_1$.
Step 3:
$E(Y|X) = \beta_0 + \beta_3 X_3 + \beta_1 X_1 + \beta_2 X_2$, $AIC = 10.0612 > V_2$,
We stop adding terms into the model.
Therefore, the best model is $E(Y|X) = \beta_0 + \beta_3 X_3 + \beta_1 X_1$.

(iii)

Solution: Backward selection algorithm with BIC:
$V_0 = 9.2282$ (with all prredictors,$E(Y|X) = \beta_0 + \beta_3 X_3 + \beta_1 X_1 + \beta_2 X_2$),
Step 1:
$E(Y|X) = \beta_0 + \beta_2 X_2 + \beta_3 X_3$, $BIC = 8.7196 < V_0$,
$E(Y|X) = \beta_0 + \beta_1 X_1 + \beta_3 X_3$, $BIC = 7.4365 < V_0$,
$E(Y|X) = \beta_0 + \beta_1 X_1 + \beta_2 X_2$, $BIC = 8.9617 < V_0$,
$\therefore V_1 = 7.4365$, we remove X_2 from the previous model, that means $E(Y|X) = \beta_0 + \beta_1 X_1 + \beta_3 X_3$.
Step 2:
$E(Y|X) = \beta_0 + \beta_3 X_3$, $BIC = 8.8395 > V_1$,
$E(Y|X) = \beta_0 + \beta_1 X_1$, $BIC = 9.1515 > V_1$,
We stop removing terms from the model.
Therefore, the best model is $E(Y|X) = \beta_0 + \beta_1 X_1 + \beta_3 X_3$.