# Homework 4

## Problem 1

Let $\mu_1 = 10, \mu_2 = 14, \sigma^2 = 4, p_1 = p_2 = 1/2$, then

$$P(B_1|A_2) = P(Y \le c|A_2) = P\left(\frac{Y-\mu_2}{\sigma} \le \frac{c-\mu_2}{\sigma}|A_2\right) = \Phi\left(\frac{c-\mu_2}{\sigma}\right),$$

$$P(A_2B_1) = P(B_1|A_2)P(A_2) = p_2\Phi\left(\frac{c-\mu_2}{\sigma}\right),$$

$$P(B_2|A_1) = P(Y > c|A_1) = P\left(\frac{Y-\mu_1}{\sigma} > \frac{c-\mu_1}{\sigma}|A_1\right) = 1 - \Phi\left(\frac{c-\mu_1}{\sigma}\right),$$

$$P(A_1B_2) = P(B_2|A_1)P(A_1) = p_1 - p_1\Phi\left(\frac{c-\mu_1}{\sigma}\right),$$

$$TPM = P(A_1B_2) + P(A_2B_1), \quad ECM = c(2|1)P(A_1B_2) + c(1|2)P(A_2B_1).$$

Direct calculation yields

```r
rm(list = ls())
cc <- seq(9, 14)
p1o2 <- pnorm(cc, 14, 2)
p2o1 <- 1 - pnorm(cc, 10 ,2)
p1a2 <- p2o1 * .5
p2a1 <- p1o2 * .5
TPM <- p1a2 + p2a1
ECMa <- TPM * 10
ECMb <- p1a2 * 5 + p2a1 * 15
res <- data.frame(cc, p1o2, p2o1, p1a2, p2a1, TPM, ECMa, ECMb)
res <- round(res, 4)
colnames(res) <- c("c", "p(1|2)", "p(2|1)", "p(1,2)", "p(2,1)",
                   "TPM", "ECM(a)", "ECM(b)")
res
```

```
##    c p(1|2) p(2|1) p(1,2) p(2,1)    TPM ECM(a) ECM(b)
## 1  9 0.0062 0.6915 0.3457 0.0031 0.3488 3.4884 1.7752
## 2 10 0.0228 0.5000 0.2500 0.0114 0.2614 2.6138 1.4206
## 3 11 0.0668 0.3085 0.1543 0.0334 0.1877 1.8767 1.2724
## 4 12 0.1587 0.1587 0.0793 0.0793 0.1587 1.5866 1.5866
## 5 13 0.3085 0.0668 0.0334 0.1543 0.1877 1.8767 2.4810
## 6 14 0.5000 0.0228 0.0114 0.2500 0.2614 2.6138 3.8069
```

# Problem 2

## (a)

We have

$$\hat{\boldsymbol{\beta}} = \left(X'X\right)^{-1} X'\boldsymbol{y}, \quad \hat{\boldsymbol{y}} = X\left(X'X\right)^{-1} X'\boldsymbol{y}, \quad \hat{\boldsymbol{\epsilon}} = \left(I - X\left(X'X\right)^{-1} X'\right)\boldsymbol{y}.$$

Thus,

$$X'\hat{\boldsymbol{\epsilon}} = X'\left(I - X\left(X'X\right)^{-1} X'\right)\boldsymbol{y} = \mathbf{0}.$$

## (b)

We have

$$\hat{\boldsymbol{y}}'\hat{\boldsymbol{\epsilon}} = \boldsymbol{y}X\left(X'X\right)^{-1} X'\left(I - X\left(X'X\right)^{-1} X'\right)\boldsymbol{y} = 0.$$

## (c)

We have

$$\begin{aligned} Cov\left(\hat{\boldsymbol{\epsilon}}, \hat{\boldsymbol{\beta}}\right) &= \left(I - X\left(X'X\right)^{-1} X'\right) Cov\left(\boldsymbol{y}, \boldsymbol{y}\right) X\left(X'X\right)^{-1} \\ &= \left(I - X\left(X'X\right)^{-1} X'\right)\sigma^2 I X\left(X'X\right)^{-1} = \mathbf{0}, \end{aligned}$$

which means that $\hat{\boldsymbol{\epsilon}}$ and $\hat{\boldsymbol{\beta}}$ are uncorrelated.

## (d)

We have

$$E\left(\hat{\boldsymbol{\beta}}\right) = \left(X'X\right)^{-1} X'E\left(\boldsymbol{y}\right) = \left(X'X\right)^{-1} X'X\boldsymbol{\beta} = \boldsymbol{\beta}$$

and

$$Var\left(\hat{\boldsymbol{\beta}}\right) = \left(X'X\right)^{-1} X'Var\left(\boldsymbol{y}\right) X\left(X'X\right)^{-1} = \left(X'X\right)^{-1} X'\sigma^2 I X\left(X'X\right)^{-1} = \sigma^2\left(X'X\right)^{-1}.$$

Since $\hat{\boldsymbol{\beta}}$ is a linear transformation of $\boldsymbol{y}$, we can conclude that $\hat{\boldsymbol{\beta}} \sim N_{q+1}\left(\boldsymbol{\beta}, \sigma^2\left(X'X\right)^{-1}\right)$.

## (e)

Denote $P = I - X\left(X'X\right)^{-1} X'$, then $P^2 = P$ and $P$ is symmetric. Now we have $\hat{\boldsymbol{\epsilon}} = P\boldsymbol{y} = P\left(X\boldsymbol{\beta} + \boldsymbol{\epsilon}\right) = P\boldsymbol{\epsilon}$. Suppose $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ are eigenvalues of $P$, we have $\lambda_1^2, \lambda_2^2, \cdots, \lambda_n^2$ are eigenvalues of $P^2$, which leads to $\lambda_i^2 = \lambda_i \Longrightarrow \lambda_i = 0$ or 1. Note that

$$\lambda_1 + \lambda_2 + \cdots + \lambda_n = \operatorname{tr}\left(P\right) = n - \operatorname{tr}\left(\left(X'X\right)^{-1} X'X\right) = n - q - 1,$$

thus we have $\lambda_i = 1$ $(1 \leq i \leq n-q-1)$ and $\lambda_i = 0$ $(n-q \leq i \leq n)$. So, we obtain that the spectral decomposition of $P$ has the form

$$P = Q'\begin{pmatrix} I_{n-q-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} Q,$$

where $Q$ is orthogonal matrix. Let $\boldsymbol{e} = Q\boldsymbol{\epsilon}/\sigma \sim N_n\left(0, Q\sigma^2 I Q'/\sigma^2\right) \sim N_n\left(0, I\right)$, we have

$$\frac{\text{ESS}}{\sigma^2} = \frac{\hat{\boldsymbol{\epsilon}}'\hat{\boldsymbol{\epsilon}}}{\sigma^2} = \frac{\boldsymbol{\epsilon}'P'P\boldsymbol{\epsilon}}{\sigma^2} = \frac{\boldsymbol{\epsilon}'P\boldsymbol{\epsilon}}{\sigma^2} = \boldsymbol{e}'\begin{pmatrix} I_{n-q-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}\boldsymbol{e} = \sum_{i=1}^{n-q-1} e_i^2 \sim \chi^2_{n-q-1}.$$

### (f)

Note that

$$\begin{pmatrix} \hat{\boldsymbol\beta} \\ \hat{\boldsymbol\epsilon} \end{pmatrix} = \begin{pmatrix} (X'X)^{-1}X' \\ I - X(X'X)^{-1}X' \end{pmatrix} (X\boldsymbol\beta + \boldsymbol\epsilon) = \begin{pmatrix} \boldsymbol\beta \\ I - X\boldsymbol\beta \end{pmatrix} + \begin{pmatrix} (X'X)^{-1}X' \\ I - X(X'X)^{-1}X' \end{pmatrix} \boldsymbol\epsilon,$$

so $\hat{\boldsymbol\beta}$ and $\hat{\boldsymbol\epsilon}$ are jointly normally distributed. And from (c), we have $Cov\left(\hat{\boldsymbol\epsilon}, \hat{\boldsymbol\beta}\right) = \mathbf{0}$, thus $\hat{\boldsymbol\beta}$ and $\hat{\boldsymbol\epsilon}$ are independent, which means that ESS $= \hat{\boldsymbol\epsilon}'\hat{\boldsymbol\epsilon}$ and $\hat{\boldsymbol\beta}$ are independent.

## Problem 3

### (a)

To begin with, we have $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_p)$ and $\mathbf{B} = \left(\boldsymbol\beta_1, \boldsymbol\beta_2, \ldots, \boldsymbol\beta_p\right)$, thus

$$\mathbf{Y} - \mathbf{XB} = \left(Y_1 - \mathbf{X}\boldsymbol\beta_1, Y_2 - \mathbf{X}\boldsymbol\beta_2, \ldots, Y_p - \mathbf{X}\boldsymbol\beta_p\right).$$

Now we have

$$\begin{aligned} \operatorname{tr}\left\{(\mathbf{Y} - \mathbf{XB})'(\mathbf{Y} - \mathbf{XB})\right\} = \operatorname{tr}\left\{(\mathbf{Y} - \mathbf{XB})(\mathbf{Y} - \mathbf{XB})'\right\} &= \operatorname{tr}\left\{\sum_{i=1}^{p} (Y_i - \mathbf{X}\boldsymbol\beta_i)(Y_i - \mathbf{X}\boldsymbol\beta_i)'\right\} \\ &= \sum_{i=1}^{p} \operatorname{tr}\left\{(Y_i - \mathbf{X}\boldsymbol\beta_i)(Y_i - \mathbf{X}\boldsymbol\beta_i)'\right\} \\ &= \sum_{i=1}^{p} \operatorname{tr}\left\{(Y_i - \mathbf{X}\boldsymbol\beta_i)'(Y_i - \mathbf{X}\boldsymbol\beta_i)\right\} \\ &= \sum_{i=1}^{p} (Y_i - \mathbf{X}\boldsymbol\beta_i)'(Y_i - \mathbf{X}\boldsymbol\beta_i) \\ &\geq \sum_{i=1}^{p} \left(Y_i - \mathbf{X}\hat{\boldsymbol\beta}_i\right)'\left(Y_i - \mathbf{X}\hat{\boldsymbol\beta}_i\right), \end{aligned}$$

where $\hat{\boldsymbol\beta}_i = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'Y_i$ and the equality holds if and only if $\boldsymbol\beta_i = \hat{\boldsymbol\beta}_i$ for all $1 \leq i \leq p$, which is equivalent to that $\mathbf{B} = \left(\hat{\boldsymbol\beta}_1, \hat{\boldsymbol\beta}_2, \ldots, \hat{\boldsymbol\beta}_p\right) = \hat{\mathbf{B}}$.

### (b)

The proof is based on the fact that for two positive semi-definite matrices $A$ and $B$, we have $|A + B| \geq |A| + |B|$.

Note that we have

$$\begin{aligned} (\mathbf{Y} - \mathbf{XB})'(\mathbf{Y} - \mathbf{XB}) &= \left(\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}} + \mathbf{X}\hat{\mathbf{B}} - \mathbf{XB}\right)'\left(\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}} + \mathbf{X}\hat{\mathbf{B}} - \mathbf{XB}\right) \\ &= \left(\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}}\right)'\left(\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}}\right) + \left(\mathbf{X}\hat{\mathbf{B}} - \mathbf{XB}\right)'\left(\mathbf{X}\hat{\mathbf{B}} - \mathbf{XB}\right) + 2\left(\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}}\right)'\left(\mathbf{X}\hat{\mathbf{B}} - \mathbf{XB}\right), \end{aligned}$$

and

$$\begin{aligned} \left(\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}}\right)'\left(\mathbf{X}\hat{\mathbf{B}} - \mathbf{XB}\right) &= \left(\mathbf{Y} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}\right)'\left(\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} - \mathbf{XB}\right) \\ &= \mathbf{Y}'\left(\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right)'\mathbf{X}\left((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} - \mathbf{B}\right) \\ &= \mathbf{0}. \end{aligned}$$

Thus,

$$(\mathbf{Y}-\mathbf{XB})'(\mathbf{Y}-\mathbf{XB}) = \left(\mathbf{Y}-\mathbf{X}\hat{\mathbf{B}}\right)'\left(\mathbf{Y}-\mathbf{X}\hat{\mathbf{B}}\right) + \left(\mathbf{X}\hat{\mathbf{B}}-\mathbf{XB}\right)'\left(\mathbf{X}\hat{\mathbf{B}}-\mathbf{XB}\right).$$

Since $\left(\mathbf{Y}-\mathbf{X}\hat{\mathbf{B}}\right)'\left(\mathbf{Y}-\mathbf{X}\hat{\mathbf{B}}\right)$ and $\left(\mathbf{X}\hat{\mathbf{B}}-\mathbf{XB}\right)'\left(\mathbf{X}\hat{\mathbf{B}}-\mathbf{XB}\right)$ are both positive semi-definite, we have

$$\begin{aligned}\left|(\mathbf{Y}-\mathbf{XB})'(\mathbf{Y}-\mathbf{XB})\right| &\geq \left|\left(\mathbf{Y}-\mathbf{X}\hat{\mathbf{B}}\right)'\left(\mathbf{Y}-\mathbf{X}\hat{\mathbf{B}}\right)\right| + \left|\left(\mathbf{X}\hat{\mathbf{B}}-\mathbf{XB}\right)'\left(\mathbf{X}\hat{\mathbf{B}}-\mathbf{XB}\right)\right| \\ &\geq \left|\left(\mathbf{Y}-\mathbf{X}\hat{\mathbf{B}}\right)'\left(\mathbf{Y}-\mathbf{X}\hat{\mathbf{B}}\right)\right|.\end{aligned}$$

Thus, $\hat{\mathbf{B}}$ minimize $(\mathbf{Y}-\mathbf{XB})'(\mathbf{Y}-\mathbf{XB})$ in the sense of its determinant.

### (c)

From (b) we have $(\mathbf{Y}-\mathbf{XB})'(\mathbf{Y}-\mathbf{XB}) = \left(\mathbf{Y}-\mathbf{X}\hat{\mathbf{B}}\right)'\left(\mathbf{Y}-\mathbf{X}\hat{\mathbf{B}}\right) + \left(\mathbf{X}\hat{\mathbf{B}}-\mathbf{XB}\right)'\left(\mathbf{X}\hat{\mathbf{B}}-\mathbf{XB}\right)$ and $\left(\mathbf{X}\hat{\mathbf{B}}-\mathbf{XB}\right)'\left(\mathbf{X}\hat{\mathbf{B}}-\mathbf{XB}\right)$ is positive semi-definite, thus

$$(\mathbf{Y}-\mathbf{XB})'(\mathbf{Y}-\mathbf{XB}) \succeq \left(\mathbf{Y}-\mathbf{X}\hat{\mathbf{B}}\right)'\left(\mathbf{Y}-\mathbf{X}\hat{\mathbf{B}}\right).$$

## Problem 4

### (a)

We have

$$\Lambda = \frac{|E|}{|E+H|} = \frac{1}{|E^{-1}|\,|E+H|} = \frac{1}{|I+E^{-1}H|} = \prod_{i=1}^{s}\frac{1}{1+\lambda_i}.$$

### (b)

Note that

$$\begin{aligned}(H+E)^{-1}H &= (H+E)^{-1}(H+E-E) = I-(H+E)^{-1}E = I-(H+E)^{-1}\left(E^{-1}\right)^{-1} \\ &= I-\left(E^{-1}(H+E)\right)^{-1} = I-\left(E^{-1}H+I\right)^{-1},\end{aligned}$$

and the eigenvalues of $(E^{-1}H+I)$ are $(\lambda_1+1)^{-1},(\lambda_2+1)^{-1},\ldots,(\lambda_s+1)^{-1}$, thus we have

$$\operatorname{tr}\left((H+E)^{-1}H\right) = s-\sum_{i=1}^{s}\frac{1}{1+\lambda_i} = \sum_{i=1}^{s}\frac{\lambda_i}{1+\lambda_i}.$$

### (c)

We know that the trace of a matrix is the sum of its eigenvalues, so the conclusion holds.

## Problem 5

```r
rm(list = ls())
y <- read.table('./battery.DAT')
y
```

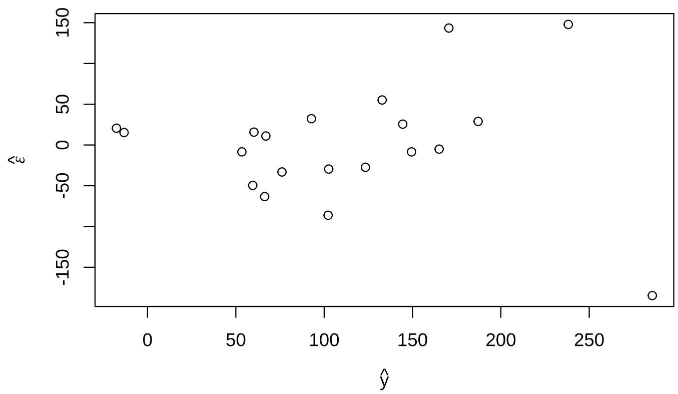
```
##       V1   V2    V3 V4   V5  V6
## 1  0.375 3.13  60.0 40 2.00 101
## 2  1.000 3.13  76.8 30 1.99 141
## 3  1.000 3.13  60.0 20 2.00  96
## 4  1.000 3.13  60.0 20 1.98 125
## 5  1.625 3.13  43.2 10 2.01  43
## 6  1.625 3.13  60.0 20 2.00  16
## 7  1.625 3.13  60.0 20 2.02 188
## 8  0.375 5.00  76.8 10 2.01  10
## 9  1.000 5.00  43.2 10 1.99   3
## 10 1.000 5.00  43.2 30 2.01 386
## 11 1.000 5.00 100.0 20 2.00  45
## 12 1.625 5.00  76.8 10 1.99   2
## 13 0.375 1.25  76.8 10 2.01  76
## 14 1.000 1.25  43.2 10 1.99  78
## 15 1.000 1.25  76.8 30 2.00 160
## 16 1.000 1.25  60.0  0 2.00   3
## 17 1.625 1.25  43.2 30 1.99 216
## 18 1.625 1.25  60.0 20 2.00  73
## 19 0.375 3.13  76.8 30 1.99 314
## 20 0.375 3.13  60.0 20 2.00 170
```

### (a) Find the estimated linear regression of...

```
m1 <- lm(V6 ~ V1 + V2 + V3 + V4 + V5, y)
summary(m1)
```

```
##
## Call:
## lm(formula = V6 ~ V1 + V2 + V3 + V4 + V5, data = y)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -184.715  -30.446    2.968   26.375  147.850
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2937.7571  4040.6401  -0.727  0.47918
## V1            -33.7934    43.3653  -0.779  0.44879
## V2             -0.1798    13.9073  -0.013  0.98987
## V3             -1.7397     1.3414  -1.297  0.21564
## V4              7.0627     1.9728   3.580  0.00302 **
## V5           1529.2897  2020.2396   0.757  0.46161
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 84.49 on 14 degrees of freedom
## Multiple R-squared:  0.5201, Adjusted R-squared:  0.3487
## F-statistic: 3.034 on 5 and 14 DF,  p-value: 0.04627
```

```r
plot(m1$fitted.values, m1$residuals,
     xlab = expression(hat(y)), ylab = expression(hat(epsilon)))
```




From the residual plot above, we can see that when $\hat{y}$ getting larger, the variation of $\hat{\varepsilon}$ gets stronger. This may due to the heteroskedasticity and suggests a violation of our assumptions.

### (b) Find the estimated linear regression of...

```r
m2 <- lm(log(V6) ~ V1 + V2 + V3 + V4 + V5, y)
m2.step <- step(m2, direction = "both")
```

```
## Start:  AIC=7.58
## log(V6) ~ V1 + V2 + V3 + V4 + V5
##
##         Df Sum of Sq    RSS     AIC
## - V3     1    0.4917 16.523  6.1810
## - V1     1    0.8006 16.832  6.5514
## <none>               16.032  7.5768
## - V5     1    1.9995 18.031  7.9275
## - V2     1    3.9387 19.971  9.9705
## - V4     1   24.5815 40.613 24.1673
##
## Step:  AIC=6.18
## log(V6) ~ V1 + V2 + V4 + V5
##
##         Df Sum of Sq    RSS     AIC
## - V1     1    0.5160 17.039  4.7960
## <none>               16.523  6.1810
## - V5     1    1.9690 18.492  6.4327
## + V3     1    0.4917 16.032  7.5768
## - V2     1    4.5731 21.097  9.0675
## - V4     1   24.3922 40.916 22.3156
##
## Step:  AIC=4.8
## log(V6) ~ V2 + V4 + V5
##
##         Df Sum of Sq    RSS     AIC
## <none>               17.039  4.7960
## - V5     1    1.9747 19.014  4.9890
```

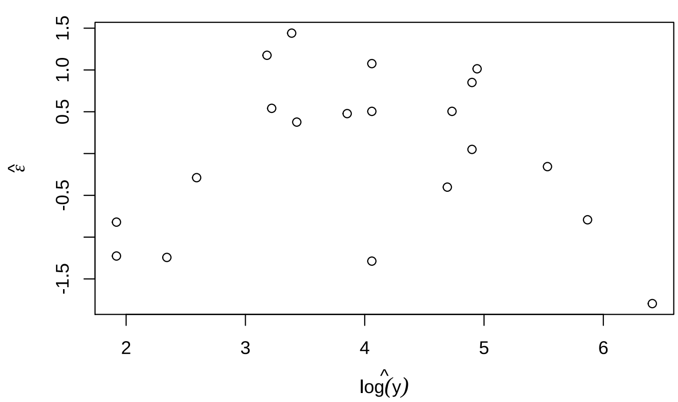
```
## + V1    1    0.5160 16.523  6.1810
## + V3    1    0.2071 16.832  6.5514
## - V2    1    4.3410 21.380  7.3349
## - V4    1   25.8384 42.878 21.2524
```

```
summary(m2.step)
```

```
##
## Call:
## lm(formula = log(V6) ~ V2 + V4 + V5, data = y)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.7954 -0.7995  0.2129  0.6183  1.4406
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -64.43215   49.34720  -1.306 0.210121
## V2           -0.33647    0.16666  -2.019 0.060573 .
## V4            0.11754    0.02386   4.926 0.000152 ***
## V5           33.59708   24.67298   1.362 0.192161
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.032 on 16 degrees of freedom
## Multiple R-squared:  0.6421, Adjusted R-squared:  0.575
## F-statistic: 9.568 on 3 and 16 DF,  p-value: 0.0007419
```

**(c) Plot the residuals from the fitted model chosen in (b) to check the model assumption, including the normality assumption.**

```
plot(m2.step$fitted.values, m2.step$residuals,
     xlab = expression(hat(log(y))), ylab = expression(hat(epsilon)))
```




From the residual plot above, we can see that when $\widehat{\log y}$ getting larger, the variation of $\hat{\varepsilon}$ does not change to much. This coincides with our assumption of homoskedasticity.

```
shapiro.test(m2.step$residuals)
```

```
##
```

```
##  Shapiro-Wilk normality test
##
## data:  m2.step$residuals
## W = 0.95007, p-value = 0.3682
```

Based on the Shapiro-Wilks test above, the normality assumption is also valid.

**(d) Conduct statistical inference for the model in (b).**

```
summary(m2.step)
```

```
##
## Call:
## lm(formula = log(V6) ~ V2 + V4 + V5, data = y)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.7954 -0.7995  0.2129  0.6183  1.4406
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -64.43215   49.34720  -1.306 0.210121
## V2           -0.33647    0.16666  -2.019 0.060573 .
## V4            0.11754    0.02386   4.926 0.000152 ***
## V5           33.59708   24.67298   1.362 0.192161
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.032 on 16 degrees of freedom
## Multiple R-squared:  0.6421, Adjusted R-squared:  0.575
## F-statistic: 9.568 on 3 and 16 DF,  p-value: 0.0007419
```

The model in (b) can be written as

$$\log(Y) = \beta_0 + \beta_2 Z_2 + \beta_4 Z_4 + \beta_5 Z_5 + \varepsilon.$$

Based on the results above, we can see that the null hypothesis of overall testing

$$H_0 : \beta_2 = \beta_4 = \beta_5 = 0 \leftrightarrow H_1 : \exists j \in 2, 4, 5, \text{ s.t. } \beta_j \neq 0$$

# Problem 6

## (a)

```
data = read.table("amitriptyline.dat",col.names = c("Y1","Y2","X1","X2","X3","X4","X5"))
```

### (a.1)

We first try to fit all the independent variables:

```
m1 = lm(Y1~X1+X2+X3+X4+X5,data=data)
summary(m1)
```

```
##
## Call:
## lm(formula = Y1 ~ X1 + X2 + X3 + X4 + X5, data = data)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -399.2 -180.1    4.5  164.1  366.8
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.879e+03  8.933e+02  -3.224 0.008108 **
## X1           6.757e+02  1.621e+02   4.169 0.001565 **
## X2           2.848e-01  6.091e-02   4.677 0.000675 ***
## X3           1.027e+01  4.255e+00   2.414 0.034358 *
## X4           7.251e+00  3.225e+00   2.248 0.046026 *
## X5           7.598e+00  3.849e+00   1.974 0.074006 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 281.2 on 11 degrees of freedom
## Multiple R-squared:  0.8871, Adjusted R-squared:  0.8358
## F-statistic: 17.29 on 5 and 11 DF,  p-value: 6.983e-05
```

We can see that all the coefficients are pretty significant, so we can treat this model as our appropriate model. We can, however, still conduct variable selection to valid our conclusion.

```
m1.step = step(m1,direction="both")
```

```
## Start:  AIC=196.33
## Y1 ~ X1 + X2 + X3 + X4 + X5
##
##        Df Sum of Sq     RSS    AIC
## <none>               870008 196.33
## - X5    1    308241 1178249 199.49
## - X4    1    399803 1269811 200.76
## - X3    1    460973 1330981 201.56
## - X1    1   1374824 2244832 210.45
## - X2    1   1729764 2599772 212.94
```

From the output, we suggest that the appropriate model is

$$Y_1 = \beta_{01} + X_1\beta_{11} + X_2\beta_{21} + X_3\beta_{31} + X_4\beta_{41} + X_5\beta_{51} + \epsilon_1,$$

and the fitted model is

$$Y_1 = -2879 + 675.7X_1 + 0.2848X_2 + 10.27X_3 + 7.251X_4 + 7.598X_5.$$

### (a.2)

We plot the residuals:

```
plot(m1)
```

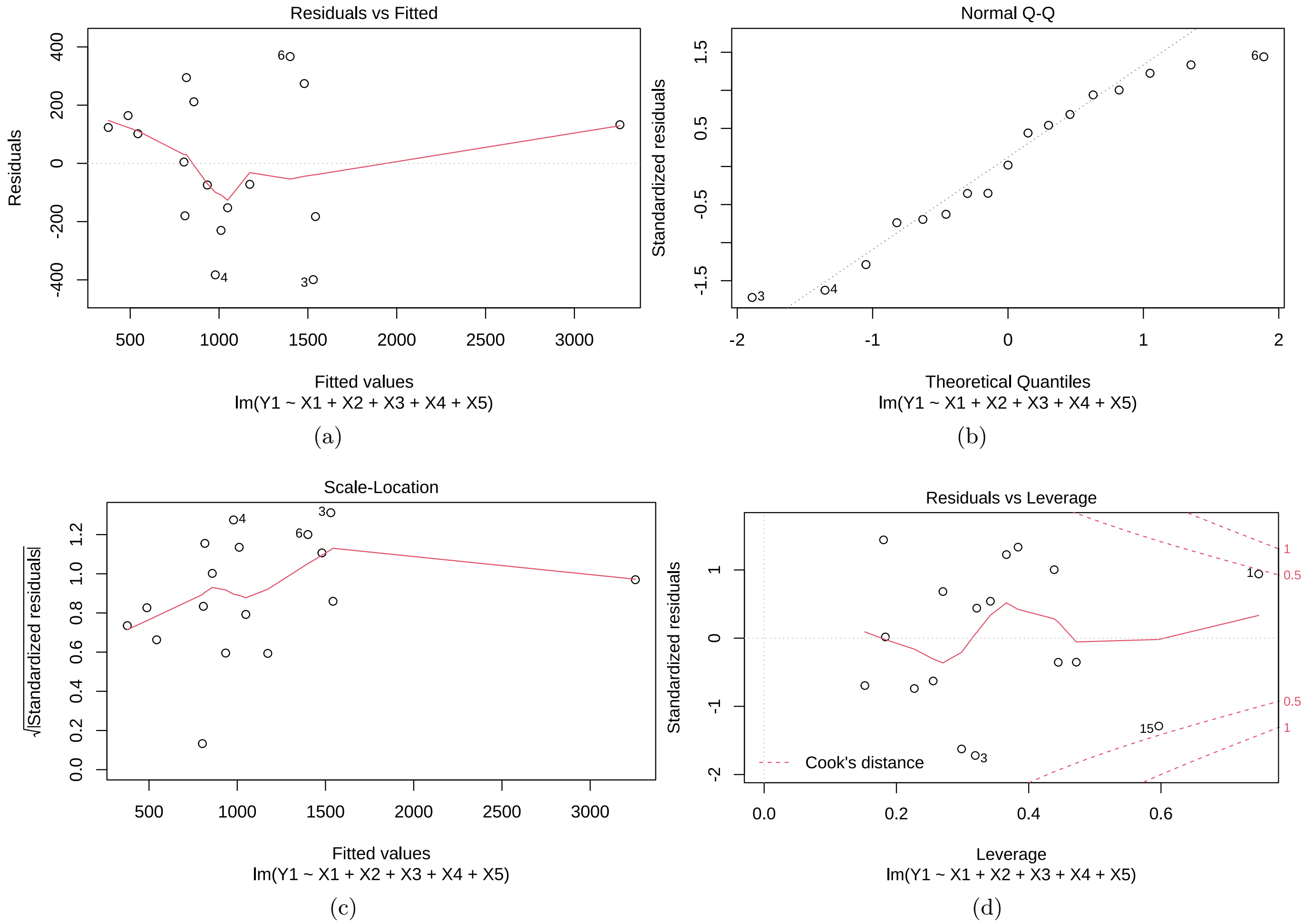



Figure 1: residual plot 1

See Figure 3. From figure (a), we cannot see clear patterns, but we can see that the variance of residuals increases as the increasing of the fitted values. So, we may suspect that there may be heteroscedasticity in the residuals. Besides, we can identify an obvious outlier in figure (a) and from figure (d) we can conclude that point 1 is an influential point. From figure (b), we can claim that the residuals are normally distributed, but we can also conduct Shapiro–Wilk test to verify our conclusion:

```
shapiro.test(m1$residuals)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  m1$residuals
## W = 0.95892, p-value = 0.6114
```

From the output, we cannot reject that the residuals are normally distributed. Thus, we can conclude that the normality assumption holds.

**(a.3)**

```r
predict.lm(m1,data.frame("X1" = 1, "X2" = 1200, "X3" = 140, "X4" = 70,
                "X5" = 85),interval = "prediction")
```

```
##        fit      lwr      upr
## 1 729.5248 41.34785 1417.702
```

According to the output, the 95% prediction interval is $[41.34785, 1417.702]$.

## (b)

### (b.1)

We first try to fit all the independent variables:

```r
m2 = lm(Y2~X1+X2+X3+X4+X5,data=data)
summary(m2)
```

```
##
## Call:
## lm(formula = Y2 ~ X1 + X2 + X3 + X4 + X5, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -373.85 -247.29  -83.74  217.13  462.72
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.729e+03  9.288e+02  -2.938 0.013502 *
## X1           7.630e+02  1.685e+02   4.528 0.000861 ***
## X2           3.064e-01  6.334e-02   4.837 0.000521 ***
## X3           8.896e+00  4.424e+00   2.011 0.069515 .
## X4           7.206e+00  3.354e+00   2.149 0.054782 .
## X5           4.987e+00  4.002e+00   1.246 0.238622
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 292.4 on 11 degrees of freedom
## Multiple R-squared:  0.8764, Adjusted R-squared:  0.8202
## F-statistic:  15.6 on 5 and 11 DF,  p-value: 0.0001132
```

We can see that not all the coefficients are significant, so We can conduct variable selection to find a better model:

```r
m2.step = step(m2,direction="both")
```

```
## Start:  AIC=197.66
## Y2 ~ X1 + X2 + X3 + X4 + X5
##
##        Df Sum of Sq     RSS    AIC
## <none>               940709 197.66
## - X5    1    132786 1073495 197.91
## - X3    1    345750 1286459 200.98
## - X4    1    394789 1335498 201.62
## - X1    1   1753418 2694127 213.55
```

```
## - X2    1   2001028 2941737 215.04
```

From the output, we still suggest that the appropriate model is

$$Y_2 = \beta_{02} + X_1\beta_{12} + X_2\beta_{22} + X_3\beta_{32} + X_4\beta_{42} + X_5\beta_{52} + \epsilon_1,$$

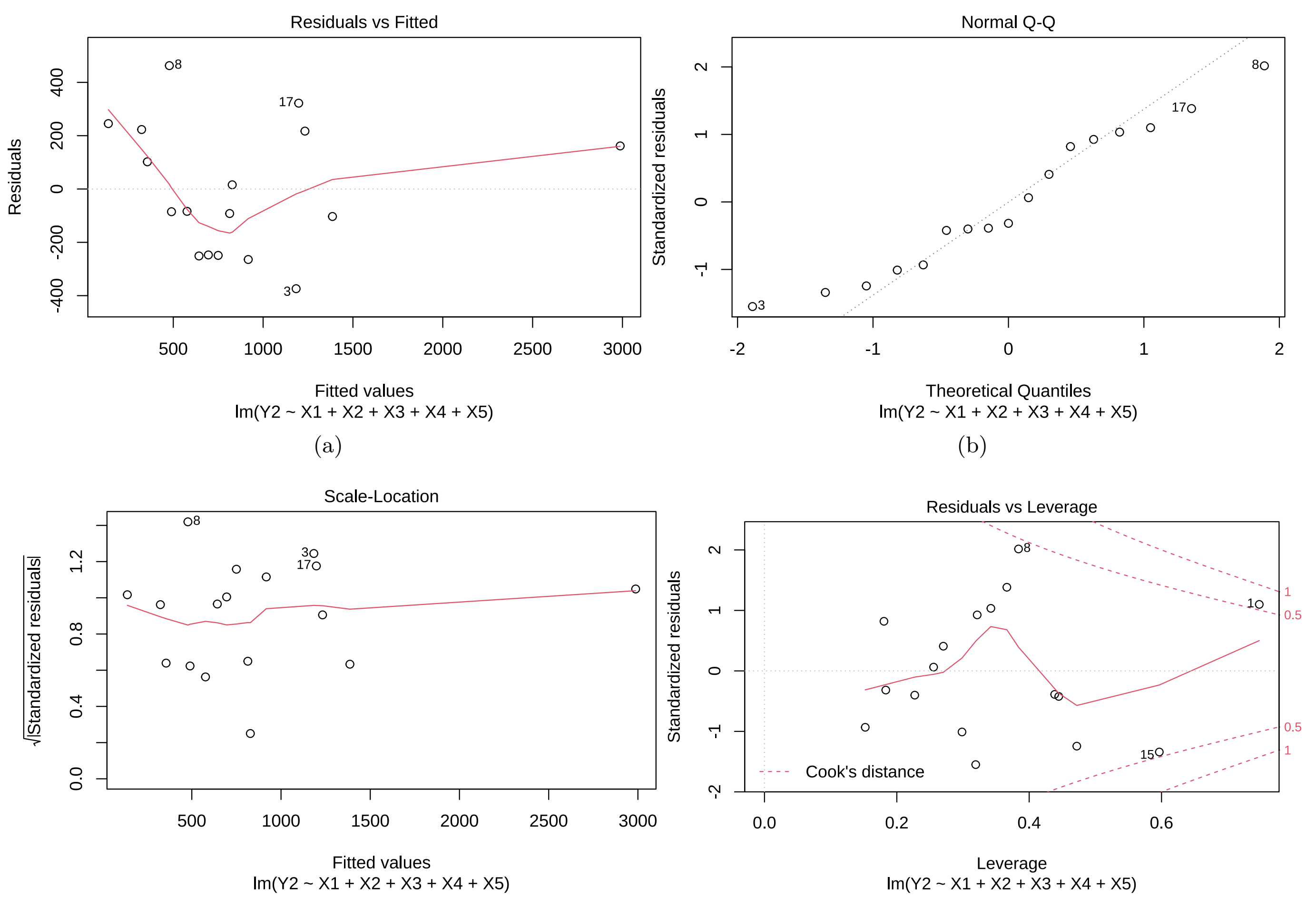
and the fitted model is

$$Y_2 = -2729 + 763X_1 + 0.3064X_2 + 8.896X_3 + 7.206X_4 + 4.987X_5.$$

### (b.2)

We plot the residuals:

```
plot(m2)
```




(a) (b) (c) (d)

Figure 2: residual plot 2

See figure 4. From figure (a), we can see a downward linear trend, and we can identify an obvious outlier in figure (a) and from figure (d) we can conclude that point 1 is an influential point. From figure (b), we can claim that the residuals are normally distributed, but we can also conduct Shapiro–Wilk test to verify our conclusion:

```
shapiro.test(m2$residuals)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  m2$residuals
## W = 0.94966, p-value = 0.4512
```

From the output, we cannot reject that the residuals are normally distributed. Thus, we can conclude that the normality assumption holds.

### (b.3)

```
predict.lm(m2,data.frame("X1" = 1, "X2" = 1200, "X3" = 140, "X4" = 70,
                "X5" = 85),interval = "prediction")
```

```
##        fit       lwr      upr
## 1 575.7255 -139.8674 1291.318
```

According to the output, the 95% prediction interval is $[41.34785, 1417.702]$.

## (c)

### (c.1)

From (a) and (b), we suggest the appropriate model is

$$(Y_1, Y_2) = (\mathbf{1}, X_1, X_2, X_3, X_4, X_5) \begin{pmatrix} \beta_{01} & \beta_{02} \\ \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \\ \beta_{31} & \beta_{32} \\ \beta_{41} & \beta_{42} \\ \beta_{51} & \beta_{52} \end{pmatrix} + (\epsilon_1, \epsilon_2) \, .$$

Now, we fit it:

```
m3 = lm(cbind(Y1,Y2)~cbind(X1,X2,X3,X4,X5), data=data)
m3$coefficients
```

```
##                                     Y1            Y2
## (Intercept)                 -2879.4782461 -2728.7085444
## cbind(X1, X2, X3, X4, X5)X1   675.6507805   763.0297617
## cbind(X1, X2, X3, X4, X5)X2     0.2848511     0.3063734
## cbind(X1, X2, X3, X4, X5)X3    10.2721328     8.8961977
## cbind(X1, X2, X3, X4, X5)X4     7.2511714     7.2055597
## cbind(X1, X2, X3, X4, X5)X5     7.5982397     4.9870508
```

### (c.2)

We check the multi-normality of residuals:

```
library(mvnormtest)
mshapiro.test(t(m3$residuals))
```

```
##
##  Shapiro-Wilk normality test
```

```
##
## data:  Z
## W = 0.94353, p-value = 0.3625
```

From the output, we can conclude that the residuals are multinormally distributed.

### (C.3)

We discuss under the normality assumption. Since $\hat{\mathbf{B}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ and $\Sigma = (\sigma_{ij})_{n\times n}$, we have

$$Cov\left(\hat{\boldsymbol{\beta}}_{(j)}, \hat{\boldsymbol{\beta}}_{(k)}\right) = Cov\left((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{y}_{(j)}, (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{y}_{(k)}\right) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\sigma_{jk}\mathbf{I}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} = \sigma_{jk}(\mathbf{X}'\mathbf{X})^{-1}.$$

Now, we obtain

$$Cov\left(\hat{\boldsymbol{\beta}}'_{(j)}\boldsymbol{x}_0, \hat{\boldsymbol{\beta}}'_{(k)}\boldsymbol{x}_0\right) = Cov\left(\boldsymbol{x}'_0\hat{\boldsymbol{\beta}}_{(j)}, \boldsymbol{x}'_0\hat{\boldsymbol{\beta}}_{(k)}\right) = \sigma_{jk}\boldsymbol{x}'_0(\mathbf{X}'\mathbf{X})^{-1}\boldsymbol{x}_0,$$

which leads to

$$Var(\bar{\boldsymbol{y}}_0) = Var\left(\hat{\mathbf{B}}'\boldsymbol{x}_0\right) = \boldsymbol{x}'_0(\mathbf{X}'\mathbf{X})^{-1}\boldsymbol{x}_0\Sigma.$$

Thus,

$$Var(\hat{\boldsymbol{y}}_0 - \bar{\boldsymbol{y}}_0) = Var(\hat{\boldsymbol{y}}_0) + Var(\bar{\boldsymbol{y}}_0) = \left(\boldsymbol{x}'_0(\mathbf{X}'\mathbf{X})^{-1}\boldsymbol{x}_0 + 1\right)\Sigma; \quad \hat{\boldsymbol{y}}_0 \sim N\left(\mathbf{B}'\boldsymbol{x}_0, \left(\boldsymbol{x}'_0(\mathbf{X}'\mathbf{X})^{-1}\boldsymbol{x}_0 + 1\right)\Sigma\right).$$

Note that

$$\Sigma^{-1/2}E\Sigma^{-1/2} = \Sigma^{-1/2}\left(\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}}\right)'\left(\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}}\right)\Sigma^{-1/2} = \Sigma^{-1/2}\boldsymbol{\Xi}'(I - H)\boldsymbol{\Xi}\Sigma^{-1/2},$$

where $H = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ and $\mathrm{rank}(H) = n - q - 1$. By spectral decomposition, we have

$$H = Q\begin{pmatrix} I_{n-q-1} & 0 \\ 0 & 0 \end{pmatrix}Q',$$

where $Q = (q_{ij})_{n\times n}$ is orthogonal. So,

$$\Sigma^{-1/2}E\Sigma^{-1/2} = \left(\Sigma^{-1/2}\boldsymbol{\epsilon}_1, \Sigma^{-1/2}\boldsymbol{\epsilon}_2, \cdots, \Sigma^{-1/2}\boldsymbol{\epsilon}_n\right)Q\begin{pmatrix} I_{n-q-1} & 0 \\ 0 & 0 \end{pmatrix}Q'\begin{pmatrix} \boldsymbol{\epsilon}'_1\Sigma^{-1/2} \\ \boldsymbol{\epsilon}'_2\Sigma^{-1/2} \\ \vdots \\ \boldsymbol{\epsilon}'_n\Sigma^{-1/2} \end{pmatrix},$$

where $\boldsymbol{\epsilon}_i$ is the $i$th column of $\Xi$ and denoted by column vector. Now, $\Sigma^{-1/2}\boldsymbol{\epsilon}_i \sim N(0, I_p)$ and $Cov\left(\Sigma^{-1/2}\boldsymbol{\epsilon}_i, \Sigma^{-1/2}\boldsymbol{\epsilon}_j\right) = \mathbf{0}$ for $i \neq j$. Now, denote

$$(\boldsymbol{e}_1, \boldsymbol{e}_2, \cdots, \boldsymbol{e}_n) = \left(\Sigma^{-1/2}\boldsymbol{\epsilon}_1, \Sigma^{-1/2}\boldsymbol{\epsilon}_2, \cdots, \Sigma^{-1/2}\boldsymbol{\epsilon}_n\right)Q,$$

we have

$$Cov(\boldsymbol{e}_i, \boldsymbol{e}_j) = Cov\left(\sum_{s=1}^n q_{si}\Sigma^{-1/2}\boldsymbol{\epsilon}_s, \sum_{t=1}^n q_{tj}\Sigma^{-1/2}\boldsymbol{\epsilon}_t\right) = \sum_{s=1}^n q_{si}q_{sj}Cov(\boldsymbol{\epsilon}_s, \boldsymbol{\epsilon}_s) = \delta_{ij}I_p,$$

where $\delta_{ij} = 1$ if $i = j$; $\delta_{ij} = 0$ otherwise. Now, we can conclude that $\boldsymbol{e}_i$'s are i.i.d. and $\boldsymbol{e}_i \sim N(\mathbf{0}, I_p)$. Thus,

$$\Sigma^{-1/2}E\Sigma^{-1/2} = (\boldsymbol{e}_1, \boldsymbol{e}_2, \cdots, \boldsymbol{e}_n)\begin{pmatrix} I_{n-q-1} & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} \boldsymbol{e}'_1 \\ \boldsymbol{e}'_2 \\ \vdots \\ \boldsymbol{e}'_n \end{pmatrix} = \sum_{i=1}^{n-q-1}\boldsymbol{e}_i\boldsymbol{e}'_i \sim W_p(n - q - 1, I_p).$$

Note that

$$\left(\boldsymbol{x}'_0(\mathbf{X}'\mathbf{X})^{-1}\boldsymbol{x}_0 + 1\right)^{-1/2}\Sigma^{-1/2}(\hat{\boldsymbol{y}}_0 - \mathbf{B}'\boldsymbol{x}_0) \sim N(0, I_p),$$

hence according to the definition of Hotelling t distribution, we have

$$\left(\hat{\boldsymbol{y}}_0 - \mathbf{B}'\boldsymbol{x}_0\right)' \Sigma^{-1/2} \left(\boldsymbol{x}_0' \left(\mathbf{X}'\mathbf{X}\right)^{-1} \boldsymbol{x}_0 + 1\right)^{-1/2} \left(\frac{\Sigma^{-1/2} E \Sigma^{-1/2}}{n-q-1}\right)^{-1} \left(\boldsymbol{x}_0' \left(\mathbf{X}'\mathbf{X}\right)^{-1} \boldsymbol{x}_0 + 1\right)^{-1/2} \Sigma^{-1/2} \left(\hat{\boldsymbol{y}}_0 - \mathbf{B}'\boldsymbol{x}_0\right) \sim T^2\left(p, n-q-1\right),$$

i.e.,

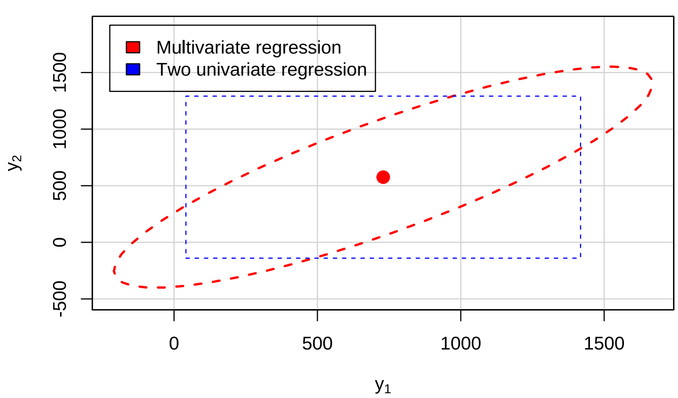
$$\left(\hat{\boldsymbol{y}}_0 - \mathbf{B}'\boldsymbol{x}_0\right)' \left(\frac{\left(\boldsymbol{x}_0' \left(\mathbf{X}'\mathbf{X}\right)^{-1} \boldsymbol{x}_0 + 1\right) E}{n-q-1}\right)^{-1} \left(\hat{\boldsymbol{y}}_0 - \mathbf{B}'\boldsymbol{x}_0\right) \sim T^2\left(n-q-1\right).$$

Therefore, the 95% prediction ellipse is

$$\left(\hat{\boldsymbol{y}}_0 - \mathbf{B}'\boldsymbol{x}_0\right)' \left(\frac{E}{n-q-1}\right)^{-1} \left(\hat{\boldsymbol{y}}_0 - \mathbf{B}'\boldsymbol{x}_0\right) \leq T^2_{0.05}\left(n-q-1\right) \left(\boldsymbol{x}_0' \left(\mathbf{X}'\mathbf{X}\right)^{-1} \boldsymbol{x}_0 + 1\right).$$

Now, we depict the ellipse:

```r
n = dim(data)[1]
p = 2
q = 5
x0 = c(1, 1, 1200,140, 70, 85)
library(car)
E = summary(Manova(m3))$SSPE
critical = qf(0.95,p,n-q-p) * (p) * (n-q-1) /(n-p-q)
X = cbind(1,as.matrix(data[,3:7]))
XX = solve(t(X)%*%X)
fa =  ((t(as.matrix(x0))%*%XX%*%as.matrix(x0) +1) * critical)[1]
cm = t(t(as.matrix(m3$coefficients))%*%matrix(x0))
ellipse(c(cm), shape=E*fa/(n-q-1), radius=1,col="red",lty=2,add=FALSE,ylim=c(-500,1900),
        xlab=expression(paste(y[1])),ylab=expression(paste(y[2])))
rect(41.34785, -139.8674, 1417.702, 1291.318, density = 0, col="blue",lty = 2, lwd = par("lwd"))
legend("topleft", inset=0.03,c("Multivariate regression","Two univariate regression"),
       fill=c("red","blue"))
```




We can see two regions overlap, in the bottom right region, multivariate regression is more powerful than two univariate regression. And we can see that multivariate indeed have considered the correlation between $Y_1$ and $Y_2$.