MVA_HW4

陈子睿 15220212202842

5/11/24

1 Misclassification Costs

(a) From the question we can see that $\mu_1 = 10$, $\mu_2 = 14$, $\sigma^2 = 4$, $p_1 = p_2 = 0.5$, then we have

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

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

$$P(A_1B_2) = P(B_1|A_2) \times P(A_2) = p_2\Phi(\frac{c-\mu_2}{\sigma}),$$

$$P(A_2B_1) = P(B_2|A_1) \times P(A_1) = p_1 - p_1\Phi(\frac{c-\mu_1}{\sigma}).$$

R code is given by:

```
rm(list = ls())
critical_values <- seq(9, 14)
p1c2 <- pnorm(critical_values, mean = 14, sd = 2)
p2c1 <- 1 - pnorm(critical_values, mean = 10, sd = 2)
p1a2 <- p1c2 * 0.5
p2a1 <- p2c1 * 0.5
TPM <- p1a2 + p2a1
ECM.a <- TPM * 10
ECM.b <- p1a2 * 5 + p2a1 * 15
results <- data.frame(
  "Critical_Value" = critical_values,
  "P(1|2)" = p1c2,
  "P(2|1)" = p2c1,
  "P(1,2)" = p1a2,
  "P(2,1)" = p2a1,
  "TPM" = TPM,
  "ECM(a)" = ECM.a,
```

```r
  "ECM(b)" = ECM.b
)
results <- round(results, 4)
library(dplyr)
library(kableExtra)
latex_table <- kable(results, format = "latex", booktabs = TRUE, caption = "Table of
↪  Classification Analysis") %>%
  kable_styling(latex_options = c("striped", "scale_down"))
print(latex_table)
```

Critical_Value	P.1.2.	P.2.1.	P.1.2..1	P.2.1..1	TPM	ECM.a.	ECM.b.
9	0.0062	0.6915	0.0031	0.3457	0.3488	3.4884	5.2015
10	0.0228	0.5000	0.0114	0.2500	0.2614	2.6138	3.8069
11	0.0668	0.3085	0.0334	0.1543	0.1877	1.8767	2.4810
12	0.1587	0.1587	0.0793	0.0793	0.1587	1.5866	1.5866
13	0.3085	0.0668	0.1543	0.0334	0.1877	1.8767	1.2724
14	0.5000	0.0228	0.2500	0.0114	0.2614	2.6138	1.4206

2 Simple Linear Regression Model

Date

Problem 2

(a) To show that $X'e=0$, we have

$$e = y - \hat{y} = (X\beta + \varepsilon) - X\hat{\beta}$$
$$= X(\beta - \hat{\beta})\ \varepsilon$$
$$= -X(X'X)^{-1}X'\varepsilon + \varepsilon$$
$$= \left(I - X(X'X)^{-1}X'\right)\varepsilon$$

Then

$$X'e = X'\left(I - X(X'X)^{-1}X'\right)\varepsilon$$
$$= 0 \cdot \varepsilon$$
$$= 0$$

(b) To show $\hat{y}'e = 0$, since $\hat{y} = X\hat{\beta} = X(X'X)^{-1}X'y$, then

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

(c) Given that $\hat{\beta} - \beta = (X'X)^{-1}X'\varepsilon$, $e = Y - X\hat{\beta} = MY = M\varepsilon$, where

$$M = I - X(X'X)^{-1}X'.$$

then we have

$$Cov(\hat{\beta}, e \mid X) := E\left[(\hat{\beta} - E(\hat{\beta}))(e - E(e))' \mid X\right]$$
$$= E\left[(\hat{\beta} - \beta)e' \mid X\right]$$
$$= E\left[(X'X)^{-1}X'\varepsilon\varepsilon'M \mid X\right]$$
$$= (X'X)^{-1}X'E\left[\varepsilon\varepsilon' \mid X\right]M$$
$$= \sigma^2 (X'X)^{-1}X'M$$
$$= 0$$

by $MX = 0$.

(d) Since $E(\hat\beta|X)=\beta$, we have

$$\begin{aligned}
var(\hat\beta|X) &= E\left[(\hat\beta-E(\hat\beta))(\hat\beta-E(\hat\beta))'|X\right] \\
&= E\left[(\hat\beta-\beta)(\hat\beta-\beta)'|X\right] \\
&= E\left[(X'X)^{-1}X'\varepsilon\varepsilon'X(X'X)^{-1}|X\right] \\
&= (X'X)^{-1}X'E\left[\varepsilon\varepsilon'|X\right]X(X'X)^{-1} \\
&= \sigma^2(X'X)^{-1}
\end{aligned}$$

Since $\hat\beta=(X'X)^{-1}X'y$ is a linear combinatton of y. Under normality assumption of ε, we have

$$\hat\beta \sim N_{q+1}\left(\beta, \sigma^2(X'X)^{-1}\right).$$

(e) Let $P=X(X'X)^{-1}X'$, $M=I-P$. It's easy to verify that M is symmetric and $M^2=M$. And $e=M\varepsilon$. Since the eigenvalue of M is either 1 or 0, and $tr(M)=n-tr\left(X(X'X)^{-1}X'\right)=n-tr\left((X'X)^{-1}X'X\right)=n-q-1$. Thus we can decompose M as

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

Thus

$$\frac{SSE}{\sigma^2}=\frac{e'e}{\sigma^2}=\frac{\varepsilon'M'M\varepsilon}{\sigma^2}=\frac{\varepsilon'M\varepsilon}{\sigma^2}=\frac{(Q\varepsilon)'\begin{pmatrix} I_{n-q-1} & 0 \\ 0 & 0\end{pmatrix}(Q\varepsilon)}{\sigma^2}\sim\chi^2_{n-q-1}.$$

(f) Since

$$\begin{pmatrix}\hat\beta \\ e\end{pmatrix}=\begin{pmatrix}(X'X)^{-1}X'y \\ y-X\hat\beta\end{pmatrix}=\begin{pmatrix}(X'X)^{-1}X' \\ I-X(X'X)^{-1}X'\end{pmatrix}(X\beta+\varepsilon)=\begin{pmatrix}\beta \\ I-X\beta\end{pmatrix}+\begin{pmatrix}(X'X)^{-1}X' \\ I-X(X'X)^{-1}X'\end{pmatrix}\varepsilon$$

And by (c), $cov(\hat\beta,e)=0$, thus $\hat\beta$ and e are independent.

3 Multivariate Linear Regression Model

Date

Problem 3

(a) Let $Y=(y_1, y_2, \cdots, y_p)$, $B=(\beta_1, \beta_2, \cdots, \beta_p)$, then

$$Y-XB=(y_1-X\beta_1, \cdots, y_p-X\beta_p).$$

Thus

$$\begin{aligned} tr\left((Y-XB)'(Y-XB)\right) &= tr\left((Y-XB)(Y-XB)'\right) \\ &= tr\left(\sum_{i=1}^{p}(y_i-X\beta_i)(y_i-X\beta_i)'\right) \\ &= \sum_{i=1}^{p} tr\left((y_i-X\beta_i)'(y_i-X\beta_i)\right) \\ &= \sum_{i=1}^{p}(y_i-X\beta_i)'(y_i-X\beta_i) \end{aligned}$$

Thus by Guass-Markov Theorem, $\hat{\beta}_i=(X'X)^{-1}X'y_i$ minimizes the trace.

Hence $\hat{B}=(\hat{\beta}_1, \cdots, \hat{\beta}_p)$.

(b) Naturally, we decompose

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

While the cross term is

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

Thus

$$(Y-XB)'(Y-XB) = (Y-X\hat{B})'(Y-X\hat{B})+(X\hat{B}-XB)'(X\hat{B}-XB)$$

Since both $(Y-X\hat{B})'(Y-X\hat{B})$ and $(X\hat{B}-XB)'(X\hat{B}-XB)$ are both p.s.d, then

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

Thus $\hat{B}$ minimizes the determinant of $(Y-XB)'(Y-XB)$.

(c) It's obvious from (b).

4 Properties for Different Test Statistics

Problem 4

(a) For Λ.

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

(b) Notice that

$$(H+E)^{-1}H = (H+E)^{-1}(H+E-E) = I-(H+E)^{-1}E$$
$$= I-(H+E)^{-1}(E^{-1})^{-1} = I-(E^{-1}(H+E))^{-1}$$
$$= I-(E^{-1}H-I)^{-1}$$

Thus

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

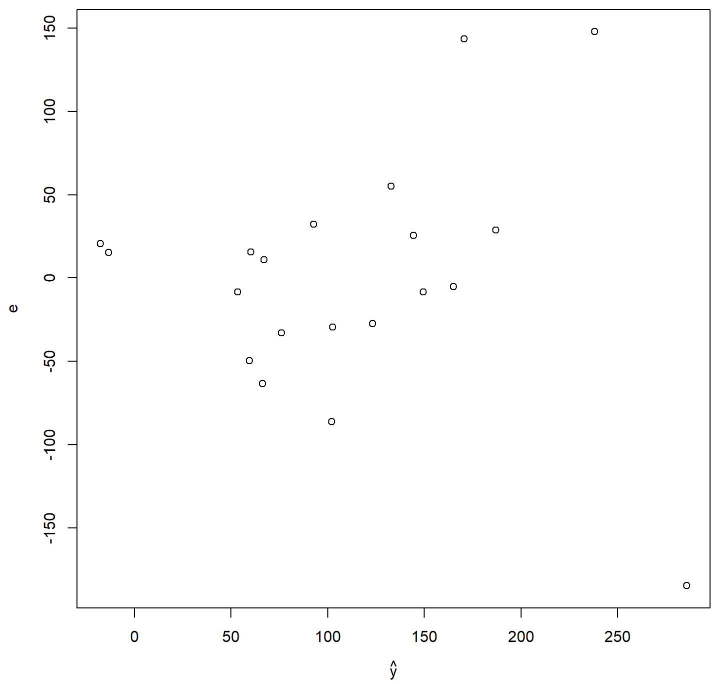
(c) Trivial

5 Satellite Applications

(a) To find the estimated regression,

```
> rm(list = ls())
> y <- read.table('C:/Users/Ray Chen/Desktop/MVA/battery.DAT')
```

```
> View(y)
> model1 <- lm(V6 ~ V1 + V2 + V3 + V4 + V5, data = y)
> View(model1)
> summary(model1)

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,^^IAdjusted R-squared:  0.3487
F-statistic: 3.034 on 5 and 14 DF,  p-value: 0.04627
```



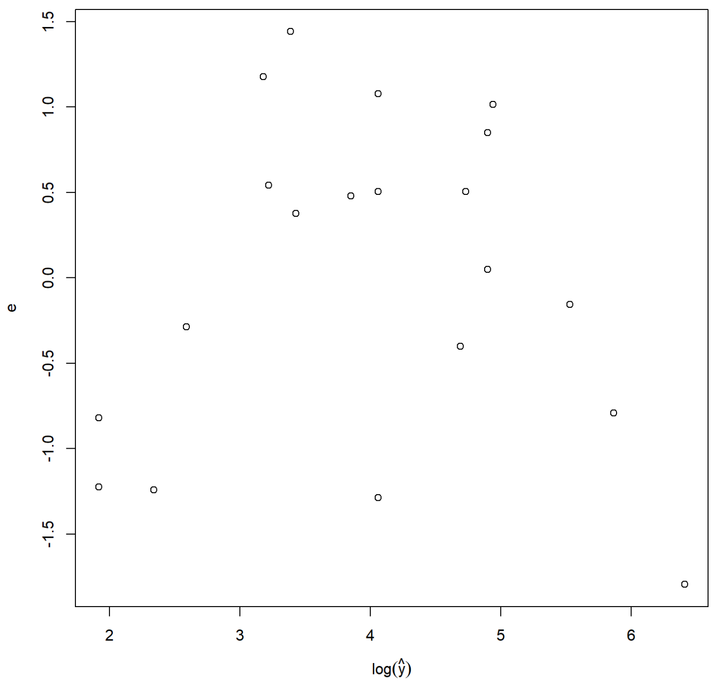
We can check that as $\hat{y}$ gets larger, the residual e gets larger. This may due to heteroskedasticity which violates our assumption.

(b)(c)

```
> model2 <- lm(log(V6) ~ V1 + V2 + V3 + V4 + V5, data = y)
> model2.subset <- step(model2, 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(model2.subset)

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,^^IAdjusted R-squared:  0.575
F-statistic: 9.568 on 3 and 16 DF,  p-value: 0.0007419
> shapiro.test(model2.subset$residuals)
```

```
^^IShapiro-Wilk normality test

data:  model2.subset$residuals
W = 0.95007, p-value = 0.3682
```



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

(d) To conduct statistical inference on the model from (b), we can write the model in (b) as:

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

Then the null hypothesis and the alternative hypothesis is

$$H_0 : \beta_2 = \beta_4 = \beta_5 = 0 \leftrightarrow H_1 : \exists \beta_j \neq 0.$$

6 Amitriptyline

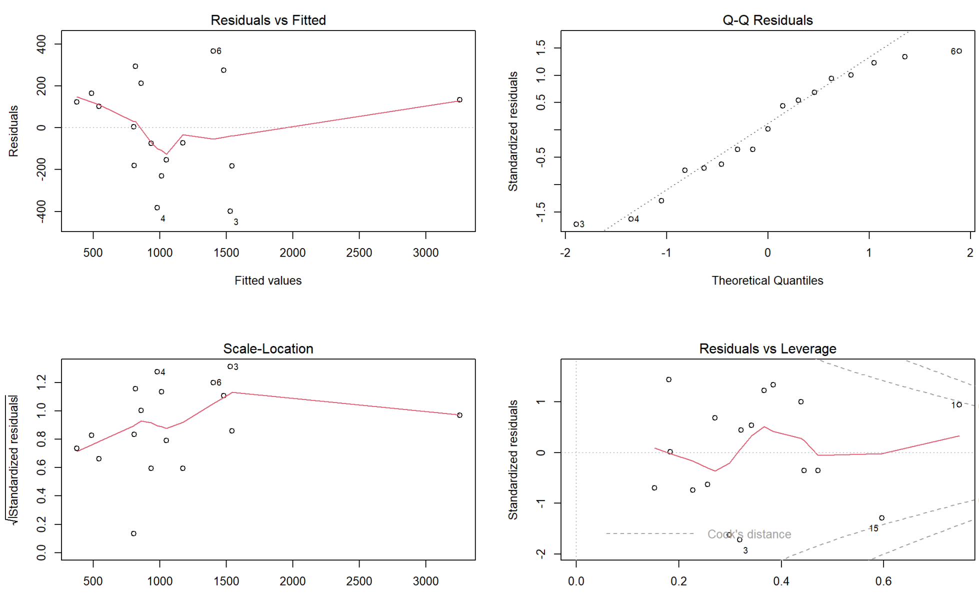
(a.1)

```
> y <- read.table('C:/Users/Ray Chen/Desktop/MVA/amitriptyline.DAT')
> colnames(y) <-  c("Y1","Y2","X1","X2","X3","X4","X5")
> View(y)
```

```
> model1 <-lm(Y1 ~ X1+X2+X3+X4+X5,data=y)
> summary(model1)

Call:
lm(formula = Y1 ~ X1 + X2 + X3 + X4 + X5, data = y)

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,^^IAdjusted R-squared:  0.8358
F-statistic: 17.29 on 5 and 11 DF,  p-value: 6.983e-05

> model1.subset <- step(model1, 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
> summary(model1.subset)

Call:
```

```
lm(formula = Y1 ~ X1 + X2 + X3 + X4 + X5, data = y)

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,^^IAdjusted R-squared:  0.8358
F-statistic: 17.29 on 5 and 11 DF,  p-value: 6.983e-05
```

(a.2)



In figure (1,1), although clear patterns are not evident, there appears to be an increasing variance of residuals with higher fitted values, suggesting the presence of heteroscedasticity. Moreover, an obvious outlier can be identified in figure (1,1), and based on figure (2,2), point 1 is deemed to be an influential point. In figure (1,2), it seems that the residuals follow a normal distribution; however, conducting a Shapiro-Wilk test would be prudent to confirm this observation:

```
> shapiro.test(model1$residuals)

^^IShapiro-Wilk normality test

data:  model1$residuals
W = 0.95892, p-value = 0.6114
```

(a.3)

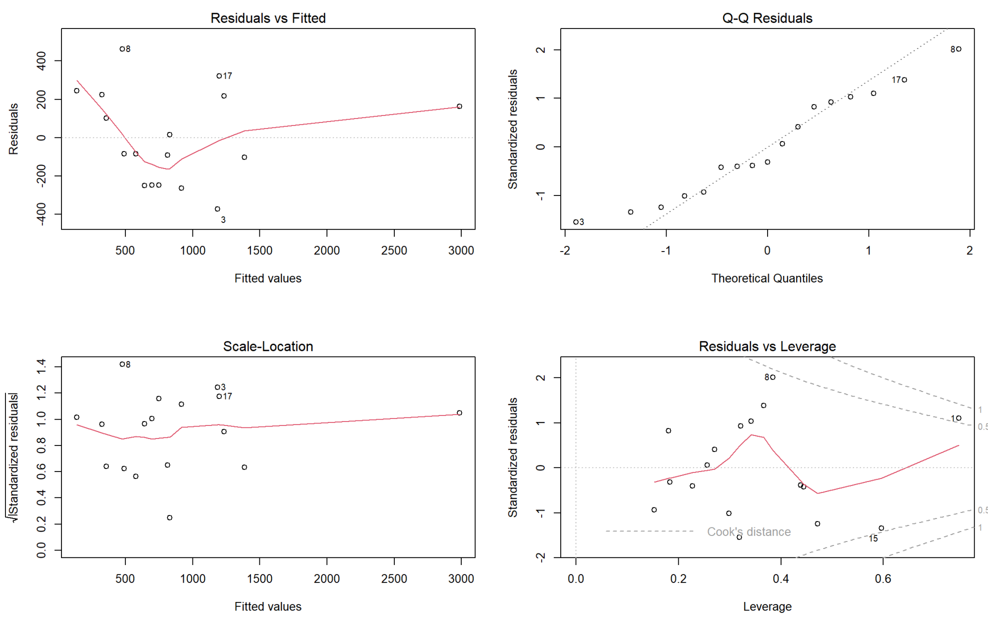
```
> predict.lm(model1,data.frame("X1" = 1, "X2" = 1200, "X3" = 140, "X4" = 70, "X5" =
↪  85),interval = "prediction")
       fit      lwr      upr
1 729.5248 41.34785 1417.702
```

(b)

```
> model2 <-lm(Y2 ~ X1+X2+X3+X4+X5,data=y)
> summary(model2)

Call:
lm(formula = Y2 ~ X1 + X2 + X3 + X4 + X5, data = y)

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,^^IAdjusted R-squared:  0.8202
F-statistic:  15.6 on 5 and 11 DF,  p-value: 0.0001132

> model2.subset <- step(model2, 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
> summary(model2.subset)

Call:
lm(formula = Y2 ~ X1 + X2 + X3 + X4 + X5, data = y)

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,^^IAdjusted R-squared:  0.8202
F-statistic:  15.6 on 5 and 11 DF,  p-value: 0.0001132
```

```
> par(mfrow=c(2,2))
> plot(model2)
> shapiro.test(model2$residuals)

^^IShapiro-Wilk normality test

data:  model2$residuals
W = 0.94966, p-value = 0.4512

> predict.lm(model2,data.frame("X1" = 1, "X2" = 1200, "X3" = 140, "X4" = 70, "X5" =
↪  85),interval = "prediction")
       fit       lwr      upr
1 575.7255 -139.8674 1291.318
```

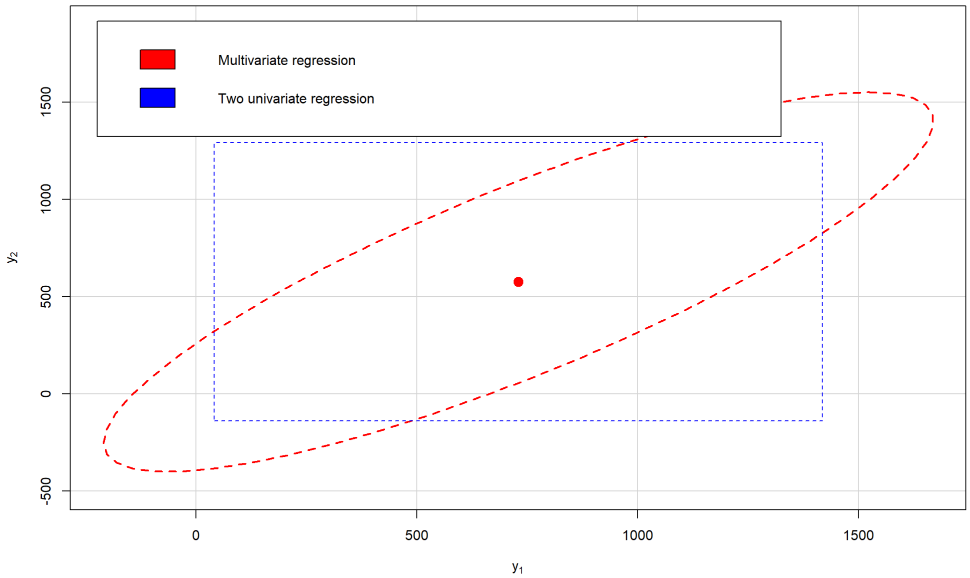


In figure (1,1), a clear downward linear trend is visible, and an obvious outlier can be identified. Additionally, based on figure (2,2), it can be concluded that point 1 is an influential point. Figure (1,2) suggests that the residuals follow a normal distribution.

(C)

```
> model3 <-lm(cbind(Y1,Y2) ~ X1+X2+X3+X4+X5,data=y)
> model3$coefficients
                       Y1            Y2
(Intercept) -2879.4782461 -2728.7085444
X1            675.6507805   763.0297617
```

```
X2              0.2848511     0.3063734
X3             10.2721328     8.8961977
X4              7.2511714     7.2055597
X5              7.5982397     4.9870508
> library(mvnormtest)
> mshapiro.test(t(model3$residuals))

^^IShapiro-Wilk normality test

data:  Z
W = 0.94353, p-value = 0.3625
> n <- dim(y)[1]
> p <- 2
> q <- 5
> x0 <- c(1, 1, 1200,140, 70, 85)
> library(car)
> E <- summary(Manova(model3))$SSPE
> critical <- qf(0.95,p,n-q-p) * (p) * (n-q-1) /(n-p-q)
> X <- cbind(1,as.matrix(y[,3:7]))
> XX <- solve(t(X)%*%X)
> fa <- ((t(as.matrix(x0))%*%XX%*%as.matrix(x0) +1) * critical)[1]
> cm <- t(t(as.matrix(model3$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"))
```


Multivariate regression
Two univariate regression
-500
0
500
1000
1500
y_2
0
500
1000
1500
y_1