# Multivariate Analysis

Homework 5

M052040003 鍾冠毅

7.21.

a.

i.

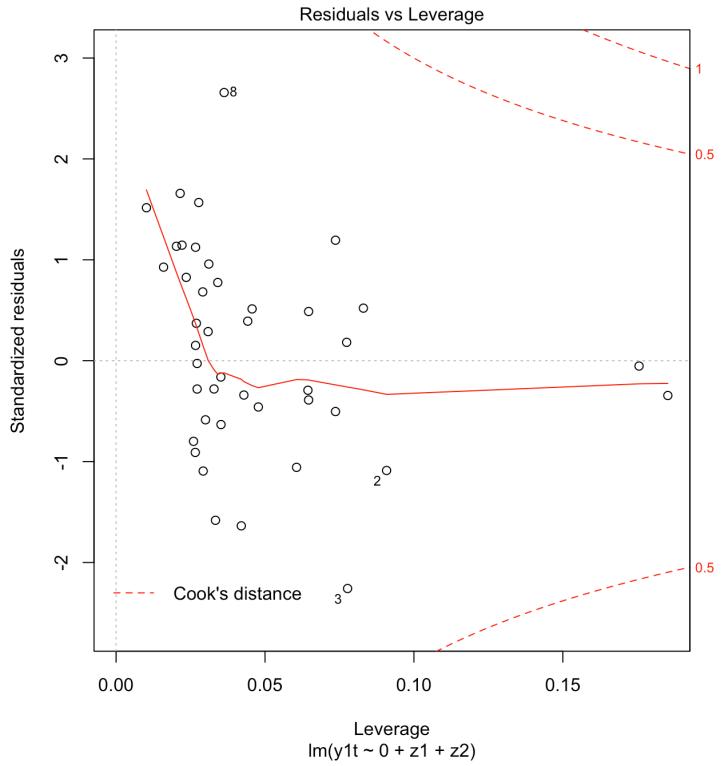
y1原本不符合常態之假設，經由Box-Cox轉換後，lambda = -0.0955692 ，可得新的y1t符合常態之假設。並以此作為反應變數配適模型，因原模型之R-square太低，故嘗試刪除截距項，可得到z1、z2皆顯著，且R-square有高達0.9663的驚人表現。

```
> shapiro.test(y1)

        Shapiro-Wilk normality test

data:  y1
W = 0.93004, p-value = 0.013

> lambda <- powerTransform(y1)$lambda
> y1t <- ((y1)^lambda-1)/lambda
> shapiro.test(y1t)

        Shapiro-Wilk normality test

data:  y1t
W = 0.98194, p-value = 0.7364
> summary(lm(y1t ~ z1 + z2))

Call:
lm(formula = y1t ~ z1 + z2)

Residuals:
     Min       1Q   Median       3Q      Max
-0.61049 -0.15592 -0.00617  0.14928  0.60172

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.897785   0.279749   6.784 4.25e-08 ***
z1          -0.007769   0.026167  -0.297    0.768
z2           0.002513   0.002387   1.053    0.299
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.2636 on 39 degrees of freedom
Multiple R-squared:  0.03162,   Adjusted R-squared:  -0.01804
F-statistic: 0.6368 on 2 and 39 DF,  p-value: 0.5344
```

```
> summary(lm(y1t ~  0 + z1 + z2))

Call:
lm(formula = y1t ~ 0 + z1 + z2)

Residuals:
     Min       1Q   Median       3Q      Max
-0.83334 -0.18280  0.02347  0.30836  1.00260

Coefficients:
   Estimate Std. Error t value Pr(>|t|)
z1 0.128109   0.024549   5.219 5.89e-06 ***
z2 0.013867   0.002481   5.590 1.78e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.3842 on 40 degrees of freedom
Multiple R-squared:  0.9663,    Adjusted R-squared:  0.9646
F-statistic: 572.8 on 2 and 40 DF,  p-value: < 2.2e-16
```

ii.

由下圖可發現，沒有離群值或是高槓桿點。



iii.

```
> pre.NO2.CI
                       fit       lwr       upr
original y1     15.08768 5.645278 44.654918
transformed y1   2.39049 1.595246  3.185734
```

b.

i.

檢測結果並非為雙變量常態分佈，故使用Box-cox轉換，並以轉換後的 y1t、y2t作為新的反應變數，並配適無截距項之模型，則有很高的模型解釋力。

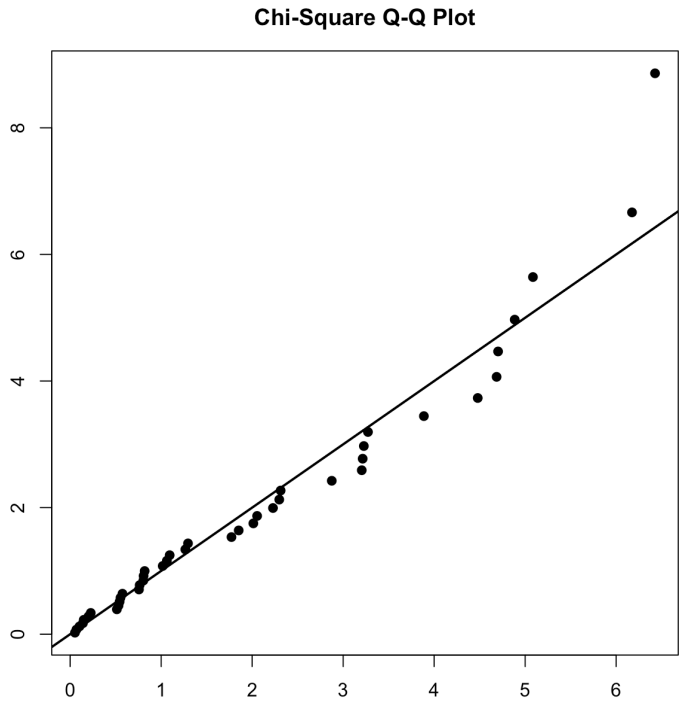
```
> mardiaTest(cbind(y1, y2))
   Mardia's Multivariate Normality Test
---------------------------------------
   data : cbind(y1, y2)

   g1p            : 2.707408
   chi.skew       : 18.95186
   p.value.skew   : 0.0008032467

   g2p            : 9.97511
   z.kurtosis     : 1.600022
   p.value.kurt   : 0.1095937

   chi.small.skew : 21.29608
   p.value.small  : 0.0002766104

   Result          : Data are not multivariate normal.
---------------------------------------
```

```
Response y1t :

Call:
lm(formula = y1t ~ 0 + z1 + z2)

Residuals:
     Min       1Q   Median       3Q      Max
-0.73923 -0.16314  0.02438  0.28579  0.87792

Coefficients:
   Estimate Std. Error t value Pr(>|t|)
z1 0.120067   0.022011   5.455 2.75e-06 ***
z2 0.012749   0.002224   5.732 1.12e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.3445 on 40 degrees of freedom
Multiple R-squared:  0.9684,    Adjusted R-squared:  0.9669
F-statistic: 613.6 on 2 and 40 DF,  p-value: < 2.2e-16
```

```
Response y2t :

Call:
lm(formula = y2t ~ 0 + z1 + z2)

Residuals:
    Min      1Q  Median      3Q     Max
-1.8611 -0.6982  0.1491  0.4897  2.0474

Coefficients:
   Estimate Std. Error t value Pr(>|t|)
z1 0.024860   0.060846   0.409    0.685
z2 0.034102   0.006148   5.547 2.05e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.9524 on 40 degrees of freedom
Multiple R-squared:  0.8986,    Adjusted R-squared:  0.8935
F-statistic: 177.3 on 2 and 40 DF,  p-value: < 2.2e-16
```

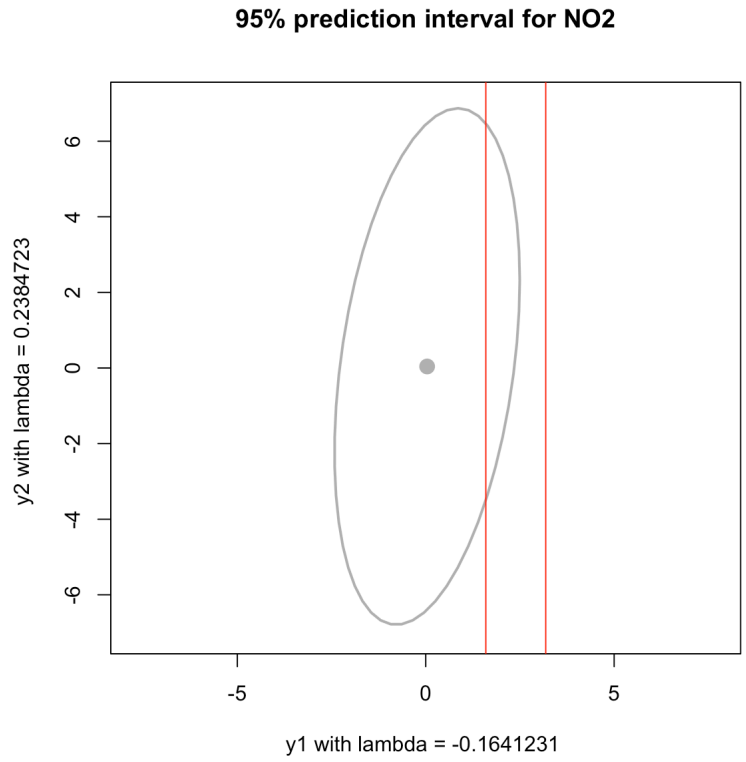
ii.

由下圖可看出殘差呈現一直線，屬於雙變數常態分佈 。




iii.

下圖為預測橢圓與a小題之信賴區間之比較，可發現信賴橢圓之中心點未必落在單一反應變數之信賴區間內。



7.25.

a.

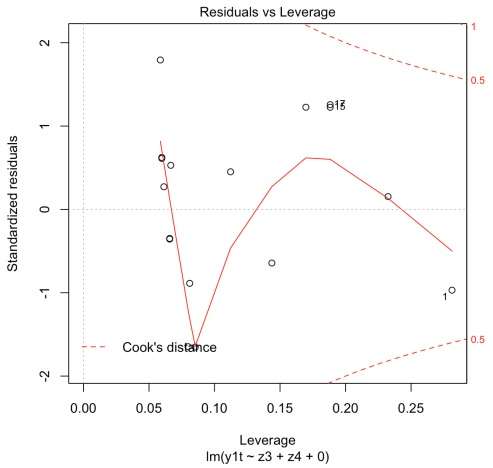
i.

步驟同7.21.a.i，lambda = -0.7666967，惟刪除截距項之模型有部分預測變數不顯著，故刪除該變數後重新配適，可得到相當好的模型解釋力。

```
> summary(lm(y1t ~ z1 + z2 + z3 + z4 + z5))

Call:
lm(formula = y1t ~ z1 + z2 + z3 + z4 + z5)

Residuals:
       Min         1Q     Median         3Q        Max
-0.0023321 -0.0006218  0.0001097  0.0009064  0.0020569

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 1.281e+00  4.790e-03 267.514   <2e-16 ***
z1          1.917e-03  8.691e-04   2.206   0.0496 *
z2          5.681e-07  3.266e-07   1.739   0.1099
z3          5.588e-05  2.282e-05   2.449   0.0323 *
z4          2.593e-05  1.730e-05   1.499   0.1620
z5          2.205e-05  2.064e-05   1.068   0.3082
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.001508 on 11 degrees of freedom
Multiple R-squared:  0.695,     Adjusted R-squared:  0.5564
F-statistic: 5.013 on 5 and 11 DF,  p-value: 0.01228
```

```
> summary(lm(y1t ~ z1 + z2 + z3 + z4 + z5 + 0))

Call:
lm(formula = y1t ~ z1 + z2 + z3 + z4 + z5 + 0)

Residuals:
     Min       1Q   Median       3Q      Max
-0.11643 -0.05997 -0.02089  0.07633  0.22265

Coefficients:
    Estimate Std. Error t value Pr(>|t|)
z1  9.997e-02  6.086e-02   1.643 0.126352
z2 -1.748e-05  2.468e-05  -0.708 0.492392
z3  5.127e-03  9.806e-04   5.229 0.000212 ***
z4  2.636e-03  1.103e-03   2.390 0.034110 *
z5  2.294e-03  1.453e-03   1.579 0.140241
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1165 on 12 degrees of freedom
Multiple R-squared:  0.9943,    Adjusted R-squared:  0.9919
F-statistic: 419.4 on 5 and 12 DF,  p-value: 4.941e-13
```

```
> summary(lm(y1t ~ z3 + z4 + 0))

Call:
lm(formula = y1t ~ z3 + z4 + 0)

Residuals:
     Min       1Q   Median       3Q      Max
-0.20569 -0.07752  0.03424  0.07869  0.22653

Coefficients:
    Estimate Std. Error t value Pr(>|t|)
z3 0.0064033  0.0003139  20.401 2.37e-12 ***
z4 0.0031894  0.0009187   3.472  0.00342 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1302 on 15 degrees of freedom
Multiple R-squared:  0.9911,    Adjusted R-squared:  0.9899
F-statistic: 836.4 on 2 and 15 DF,  p-value: 4.129e-16
```




ii.

由上圖可發現沒有離群值與高槓桿點。

iii.

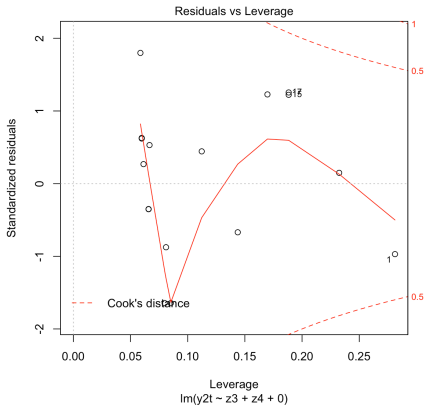
```
> pre.25a.CI
                        fit       lwr      upr
original y1      12.811812 3.7492043      NaN
transformed y1    1.119721 0.8307758 1.408666
```

b.

i.

重複a小題之步驟，得lambda = -0.8616772，變數選取後得到的模型亦有很高的模型解釋力。

```
> summary(lm(y2t ~ z3 + z4 + 0))

Call:
lm(formula = y2t ~ z3 + z4 + 0)

Residuals:
     Min       1Q   Median       3Q      Max
-0.1834 mardiaTest(data, cov = TRUE, qqplot = FALSE) 6

Coefficients:
    Estimate Std. Error t value Pr(>|t|)
z3 0.0057051  0.0002804  20.343 2.47e-12 ***
z4 0.0028466  0.0008209   3.468  0.00344 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1163 on 15 degrees of freedom
Multiple R-squared:  0.9911,    Adjusted R-squared:  0.9899
F-statistic:   832 on 2 and 15 DF,  p-value: 4.295e-16
```




ii.

由上圖可發現沒有潛在的離群值與高槓桿點。

iii.

```
> pre.25b.CI
                       fit       lwr      upr
original y1     9.7878880 3.2462768      NaN
transformed y1 0.9979721 0.7397948 1.256149
```

c.

i.

做法同7.21.b.i，lambda1 = -0.1772955 、lambda1 = -0.1852451，選模時將不顯著的變數刪除，重新配適後亦有很好的模型解釋力。

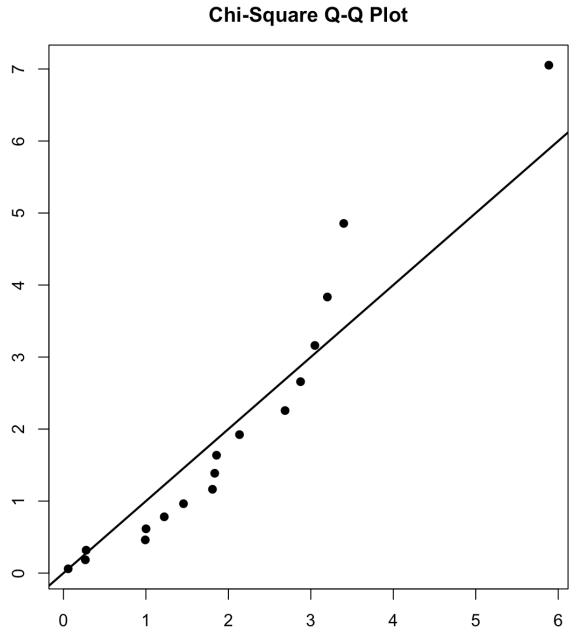
```
> mardiaTest(cbind(y1, y2))
   Mardia's Multivariate Normality Test
---------------------------------------
   data : cbind(y1, y2)

   g1p            : 7.283451
   chi.skew       : 20.63644
   p.value.skew   : 0.0003737932

   g2p            : 13.2787
   z.kurtosis     : 2.720579
   p.value.kurt   : 0.006516762

   chi.small.skew : 27.31294
   p.value.small  : 1.718303e-05

   Result         : Data are not multivariate normal.
---------------------------------------
```

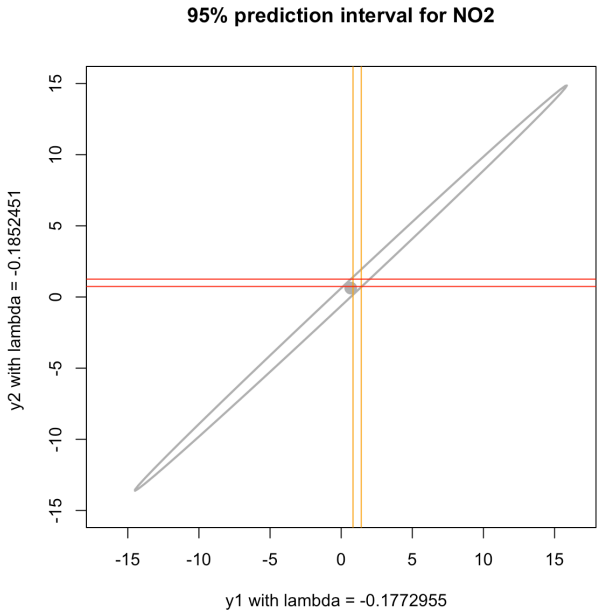
```
Response y1t :

Call:
lm(formula = y1t ~ z1 + z3 + z4 + 0)

Residuals:
     Min       1Q   Median       3Q      Max
-0.43639 -0.14771 -0.01829  0.30388  0.51171

Coefficients:
    Estimate Std. Error t value Pr(>|t|)
z1 0.3788044  0.1528008   2.479  0.02652 *
z3 0.0188722  0.0009088  20.766 6.46e-12 ***
z4 0.0073800  0.0021527   3.428  0.00408 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.3037 on 14 degrees of freedom
Multiple R-squared:  0.9952,    Adjusted R-squared:  0.9942
F-statistic: 966.2 on 3 and 14 DF,  p-value: < 2.2e-16
```

```
Response y2t :

Call:
lm(formula = y2t ~ z1 + z3 + z4 + 0)

Residuals:
     Min       1Q   Median       3Q      Max
-0.35062 -0.17976 -0.04664  0.33248  0.45736

Coefficients:
    Estimate Std. Error t value Pr(>|t|)
z1 0.4108960  0.1436616   2.860   0.0126 *
z3 0.0179685  0.0008544  21.030 5.44e-12 ***
z4 0.0066344  0.0020240   3.278   0.0055 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.2856 on 14 degrees of freedom
Multiple R-squared:  0.9954,    Adjusted R-squared:  0.9944
F-statistic:   999 on 3 and 14 DF,  p-value: < 2.2e-16
```

ii.

由下圖左可知，殘差呈現一直線，故為雙變數常態分佈。




iii.

由上圖右可看出，分別對y1t與y2t作出CI並不一定會包含同時做信賴橢圓的預測值。