姓名:葉冠宏 學號:108753208

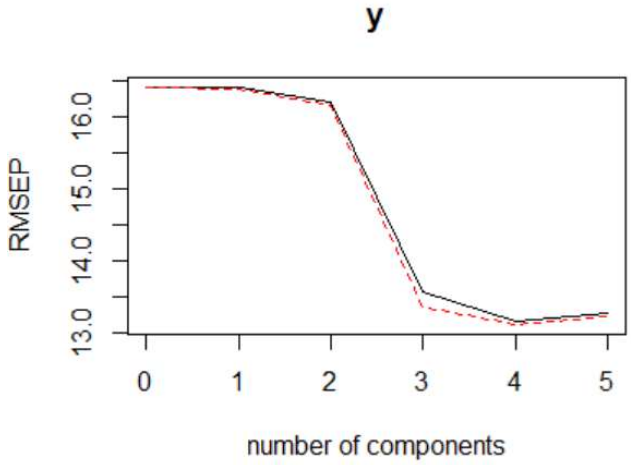
以下是我針對 Least Squares (LS) regression model, Principal Component Regression (PCR), Partial Least Squares (PLS) regression 所做的統計

1. Least Squares (LS) regression model

```
Mean absolute error        :  11.02971
Sample standard deviation  :  3.32377

Mean squared error         :  210.3135
Sample standard deviation  :  140.0214

Root mean squared error    :  13.87959
Sample standard deviation  :  4.431012
Residuals:
    Min      1Q  Median      3Q     Max
-55.932  -6.988   1.016   8.438  39.685

Coefficients:
   Estimate Std. Error t value Pr(>|t|)
x1  0.71002    0.06408  11.080  < 2e-16 ***
x2  0.12843    0.04651   2.761  0.00691 **
x3  0.23094    0.05616   4.112 8.33e-05 ***
x4 -0.03496    0.06727  -0.520  0.60449
x5  0.04547    0.06908   0.658  0.51198
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 14.22 on 95 degrees of freedom
Multiple R-squared:  0.9726,    Adjusted R-squared:  0.9712
F-statistic: 674.9 on 5 and 95 DF,  p-value: < 2.2e-16
```

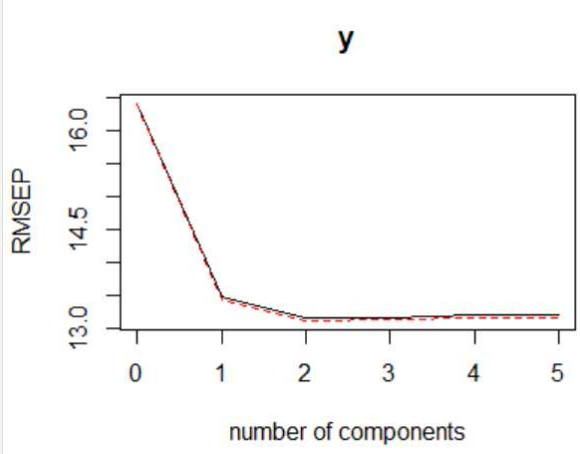
2. Principal Component Regression (PCR)

```
VALIDATION: RMSEP
Cross-validated using 10 random segments.
       (Intercept)  1 comps  2 comps  3 comps  4 comps  5 comps
CV           16.41    16.56    16.44    13.57    12.99    13.00
adjCV        16.41    16.53    16.41    13.49    12.96    12.97

TRAINING: % variance explained
   1 comps  2 comps  3 comps  4 comps  5 comps
X   37.131   61.290    79.22    94.29   100.00
y    2.487    9.336    38.77    41.62    41.67
```



3. Partial Least Squares (PLS) regression

```
VALIDATION: RMSEP
Cross-validated using 10 random segments.
       (Intercept)  1 comps  2 comps  3 comps  4 comps  5 comps
CV           16.41    13.46    13.14    13.16    13.19    13.19
adjCV        16.41    13.41    13.11    13.12    13.15    13.15

TRAINING: % variance explained
   1 comps  2 comps  3 comps  4 comps  5 comps
X    23.00    54.33    77.90    91.42   100.00
y    39.17    41.55    41.66    41.67    41.67
```



Question1:

三種模型是 different 的。

Least Squares (LS) regression model 是 find Beta^ so that Corr(Y,XBeta^) is maximized. 你的模型有可能會隨著你的 sample data 而使 beta^ 的 variance 比較大。

Principal Component Regression (PCR)是先 find w so that Var(Xw)=Var(Z) is maximized.Then find Beta^ so that Corr(Y,ZBeta^) is maximized. 我們的首要任務是去盡量找出一個投影向量使投影後 data 的 variance 不會和原來差太多。

Partial Least Squares (PLS) regression 是先 find w so that Cov(Y,XW)=Cov(Y,Z) is maximized. Then find beta^ so that Corr(Y,Zbeta^) is maximized. 當我們先去找 max (Cov(Y,Z))^2=Var(Xbeta)*(Corr(Y,Xbeta))^2 時，我們就同時兼顧到讓 Var(Xbeta) 和 Corr(Y,Xbeta) 都蠻大的。

Question2:

我們可以看到 Least Squares (LS) regression model, 做出的結果是 RMSE 為 13.87。

Principal Component Regression (PCR) 如果取 4 個 component ，可以使其 variance 佔原來的 94.29%，做出來的 RMSE 為最好的 12.99。所以我會選 4 個 component。

Partial Least Squares (PLS) regression 中，我們可以觀察到取 2 個 component，其可以達到最好的 rmse 為 13.14，然而此時的 variance explained 僅有 54.33%。如果取到 4 個 component, rmse 也才上升一點點至 13.19，但 variance explained 可以提高到 91.42%。所以我會選擇 4 個 component。

綜觀上述的結果，我們可以觀察到 Principal Component Regression (PCR)是預測“Academic Reputation”最好的模型，因為其可以達到三者模型中最低的 rmse，以及高的 variance explained。