應數碩一 M072040019 梅瀚中

4. I collect a set of data ($n = 100$ observations) containing a single predictor and a quantitative response. I then fit a linear regression model to the data, as well as a separate cubic regression, i.e. $Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \beta_3 X^3 + \epsilon$.

   (a) Suppose that the true relationship between X and Y is linear, i.e. $Y = \beta_0 + \beta_1 X + \epsilon$. Consider the training residual sum of squares (RSS) for the linear regression, and also the training RSS for the cubic regression. Would we expect one to be lower than the other, would we expect them to be the same, or is there not enough information to tell? Justify your answer.

   (b) Answer (a) using test rather than training RSS.

   (c) Suppose that the true relationship between X and Y is not linear, but we don't know how far it is from linear. Consider the training RSS for the linear regression, and also the training RSS for the cubic regression. Would we expect one to be lower than the other, would we expect them to be the same, or is there not enough information to tell? Justify your answer.

   (d) Answer (c) using test rather than training RSS.

(a) 如果真實 X 和 Y 的關係是線性的,那使用 linear regression 在 training set 的 RSS 會低於使用 cubic regression 的 RSS。

(b) 高次方的 regression 通常在 training set 上的準確率通常會很好,但是在 testing set 上的表現不會太好,稱為 overfitting。我認為高次方的 testing RSS 會比 training RSS 更高。

(c) 真實 X 和 Y 的關係是非線性,那使用 cubic regression 在 training set 的 RSS 會低於使用 linear regression 的 RSS

(d) 如果真實 X 和 Y 的關係是非線性的,那使用 linear regression 通常會在 training set 和 testing set 上 underfitting 。但是我們也不知道它指的”非線性”的程度到底是多大, 如果它更接近線性而不是立方，則 linear 的 testing RSS 可能低於 cubic 的 testing RSS。

5. Consider the fitted values that result from performing linear regression without an intercept. In this setting, the ith fitted value takes the form

$$\hat{y}_i = x_i\hat{\beta},$$

不太懂為何這邊要有i',何不直接用i=1到n就好??

where

$$\hat{\beta} = \left(\sum_{i=1}^{n} x_i y_i\right) \Big/ \left(\sum_{i'=1}^{n} x_{i'}^2\right). \tag{3.38}$$

Show that we can write

$$\hat{y}_i = \sum_{i'=1}^{n} a_{i'} y_{i'}.$$

What is $a_{i'}$?

Note: We interpret this result by saying that the fitted values from linear regression are linear combinations *of the response values.*

$$\hat{y}_i = x_i \frac{\left(\sum_{i=1}^{n} x_i y_i\right)}{\left(\sum_{i'=1}^{n} x_{i'}^2\right)}$$

$$= x_i \frac{\sum_{k=1}^{n} x_k y_k}{\sum_{j=1}^{n} x_j^2}$$

$$= \sum_{k=1}^{n} \frac{x_k x_i}{\sum_{j=1}^{n} x_j^2} y_k$$

$$\because \sum_{k=1}^{n} \frac{x_k x_i}{\sum_{j=1}^{n} x_j^2} y_k = \sum_{i'=1}^{n} a_{i'} y_{i'}$$

$$\therefore \frac{x_{i'} x_i}{\sum_{j=1}^{n} x_j^2} = a_{i'}$$

6. Using (3.4), argue that in the case of simple linear regression, the least squares line always passes through the point $(\bar{x}, \bar{y})$.

$$\min_{\hat\beta_0, \hat\beta_1} SSE = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 = \sum_{i=1}^{n} (Y_i - \hat\beta_0 - \hat\beta_1 X_i)^2$$

f.o.c. $$\frac{\partial SSE}{\partial \hat\beta_0} = \sum_{i=1}^{n} 2(Y_i - \hat\beta_0 - \hat\beta_1 X_i)(-1) = 0$$

$$\Rightarrow n\hat\beta_0 + \sum_{i=1}^{n} X_i \hat\beta_1 = \sum_{i=1}^{n} Y_i$$

同除n $$\Rightarrow \bar{Y} = \hat\beta_0 + \hat\beta_1 \bar{X}.$$

11. In this problem we will investigate the t-statistic for the null hypothesis $H_0 : \beta = 0$ in simple linear regression without an intercept. To begin, we generate a predictor `x` and a response `y` as follows.

```
> set.seed(1)
> x=rnorm(100)
> y=2*x+rnorm(100)
```

(a) Perform a simple linear regression of `y` onto `x`, *without* an intercept. Report the coefficient estimate $\hat\beta$, the standard error of this coefficient estimate, and the t-statistic and p-value associated with the null hypothesis $H_0 : \beta = 0$. Comment on these results. (You can perform regression without an intercept using the command `lm(y~x+0)`.)

(b) Now perform a simple linear regression of `x` onto `y` without an intercept, and report the coefficient estimate, its standard error, and the corresponding t-statistic and p-values associated with the null hypothesis $H_0 : \beta = 0$. Comment on these results.

(c) What is the relationship between the results obtained in (a) and (b)?

(d) For the regression of Y onto X without an intercept, the t-statistic for $H_0 : \beta = 0$ takes the form $\hat\beta/\mathrm{SE}(\hat\beta)$, where $\hat\beta$ is given by (3.38), and where

$$\mathrm{SE}(\hat\beta) = \sqrt{\frac{\sum_{i=1}^{n}(y_i - x_i\hat\beta)^2}{(n-1)\sum_{i'=1}^{n} x_{i'}^2}}.$$

(These formulas are slightly different from those given in Sections 3.1.1 and 3.1.2, since here we are performing regression without an intercept.) Show algebraically, and confirm numerically in R, that the t-statistic can be written as

$$\frac{(\sqrt{n-1})\sum_{i=1}^{n} x_i y_i}{\sqrt{(\sum_{i=1}^{n} x_i^2)(\sum_{i'=1}^{n} y_{i'}^2) - (\sum_{i'=1}^{n} x_{i'} y_{i'})^2}}.$$

(e) Using the results from (d), argue that the t-statistic for the regression of y onto x is the same as the t-statistic for the regression of x onto y.

(f) In R, show that when regression is performed *with* an intercept, the t-statistic for $H_0 : \beta_1 = 0$ is the same for the regression of y onto x as it is for the regression of x onto y.

(a)

```
Coefficients:
  Estimate Std. Error t value Pr(>|t|)
x   1.9939     0.1065   18.73   <2e-16 ***
```

根據 p-value 值非常的小,我們有顯著證據拒絕 H0。

(B)

```
Coefficients:
  Estimate Std. Error t value Pr(>|t|)
y  0.39111    0.02089   18.73   <2e-16 ***
```

根據 p-value 值非常的小,我們有顯著證據拒絕 H0。

(C)

我們從上圖可以看出把自變量和應變量互換去做線性回歸,得出來的 P-VALUE 和 t test 的值是一樣的。

(D)

$$T = \frac{\hat{\beta}}{SE(\hat{\beta})} = \frac{\frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2}}{\sqrt{\frac{\sum_{i=1}^{n}(y_i - x_i\hat{\beta})^2}{(n-1)\sum_{i=1}^{n} x_i^2}}} = \frac{\sqrt{n-1}\sum_{i=1}^{n} x_i y_i}{\sqrt{\sum_{i=1}^{n}(y_i - x_i\hat{\beta})^2 (\sum_{i=1}^{n} x_i^2)}} = \frac{\sqrt{n-1}\sum_{i=1}^{n} x_i y_i}{(\sum_{i=1}^{n} x_i^2)(\sum_{i=1}^{n} y_i^2) - (\sum_{i=1}^{n} x_i y_i)^2}$$

其中 $\sum_{i=1}^{n}(y_i - x_i\hat{\beta})^2 = \sum_{i=1}^{n} y_i^2 - 2\sum_{i=1}^{n} y_i x_i \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2} + \sum_{i=1}^{n} x_i^2 \left(\frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2}\right)^2$

$$= \sum_{i=1}^{n} y_i^2 - 2\frac{(\Sigma x_i y_i)^2}{\Sigma x_i^2} + \frac{(\Sigma x_i y_i)^2}{\Sigma x_i^2}$$

$$= \sum_{i=1}^{n} y_i^2 - \frac{(\Sigma (x_i y_i))^2}{\Sigma x_i^2}$$

(e)

由(d)答案可知 f(x,y)=f(y,x)

(f)

由(a) , (b)可知有相同的 t-test 的檢定結果。

13. In this exercise you will create some simulated data and will fit simple linear regression models to it. Make sure to use `set.seed(1)` prior to starting part (a) to ensure consistent results.

    (a) Using the `rnorm()` function, create a vector, `x`, containing 100 observations drawn from a $N(0, 1)$ distribution. This represents a feature, X.

    (b) Using the `rnorm()` function, create a vector, `eps`, containing 100 observations drawn from a $N(0, 0.25)$ distribution i.e. a normal distribution with mean zero and variance 0.25.

    (c) Using `x` and `eps`, generate a vector `y` according to the model

    $$Y = -1 + 0.5X + \epsilon. \tag{3.39}$$

    What is the length of the vector `y`? What are the values of β_0 and β_1 in this linear model?

(d) Create a scatterplot displaying the relationship between `x` and `y`. Comment on what you observe.

(e) Fit a least squares linear model to predict `y` using `x`. Comment on the model obtained. How do $\hat{\beta}_0$ and $\hat{\beta}_1$ compare to β_0 and β_1?

(f) Display the least squares line on the scatterplot obtained in (d). Draw the population regression line on the plot, in a different color. Use the `legend()` command to create an appropriate legend.

(g) Now fit a polynomial regression model that predicts `y` using `x` and $\mathtt{x}^2$. Is there evidence that the quadratic term improves the model fit? Explain your answer.

(h) Repeat (a)–(f) after modifying the data generation process in such a way that there is *less* noise in the data. The model (3.39) should remain the same. You can do this by decreasing the variance of the normal distribution used to generate the error term ϵ in (b). Describe your results.

(i) Repeat (a)–(f) after modifying the data generation process in such a way that there is *more* noise in the data. The model (3.39) should remain the same. You can do this by increasing the variance of the normal distribution used to generate the error term ϵ in (b). Describe your results.

(j) What are the confidence intervals for β_0 and β_1 based on the original data set, the noisier data set, and the less noisy data set? Comment on your results.

(a)

```
set.seed(1)
x=rnorm(100)
```

(b)

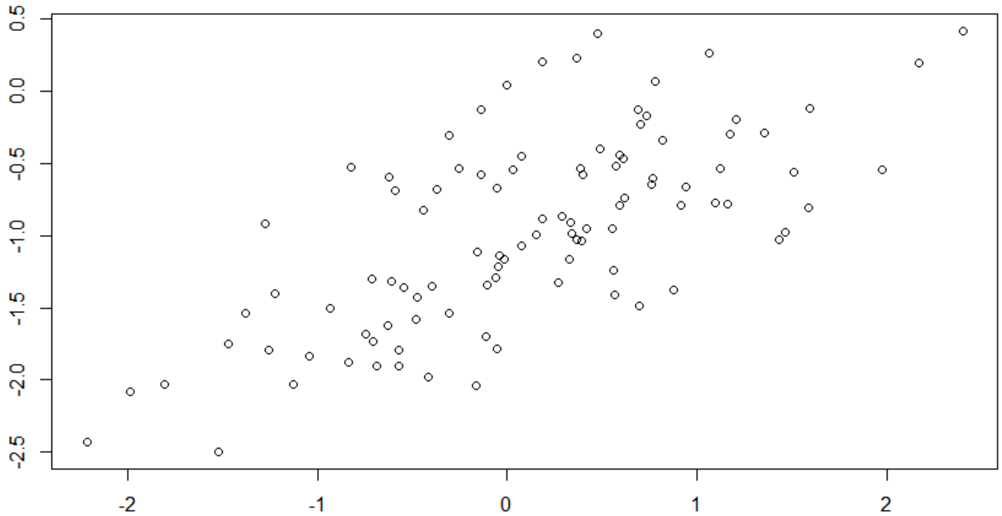
```
set.seed(1)
eps=rnorm(100, mean = 0, sd = 0.5)
```

(c)

```
y=-1+0.5x+eps
length(y)
```

y 長度 100,β0=-1 , β1=0.5

(d)



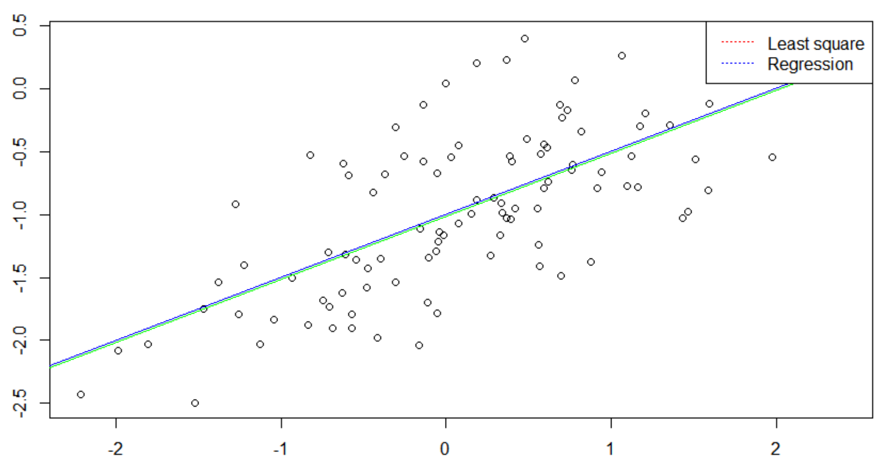
X 和 y 有線性的趨勢。

(e)

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) -1.01885    0.04849 -21.010  < 2e-16 ***
x            0.49947    0.05386   9.273 4.58e-15 ***
```

$\hat{\beta}_0$=-1.01885 , $\hat{\beta}_1$=049947 ,這跟 β0=-1 , β1=0.5 非常接近,且 p-value 非常小,我們有顯著證據拒絕 H0。

(f)



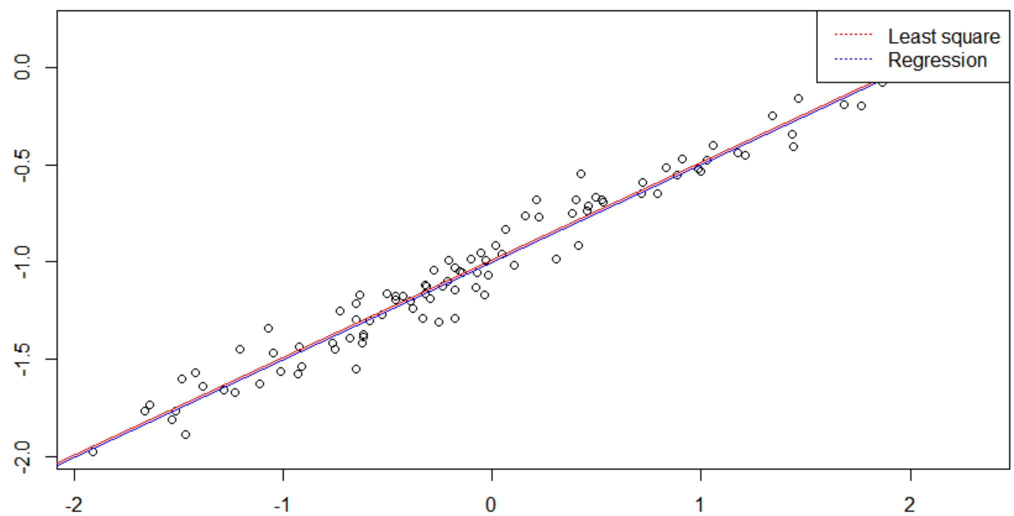
(g)

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) -0.97164    0.05883 -16.517  < 2e-16 ***
x            0.50858    0.05399   9.420  2.4e-15 ***
I(x^2)      -0.05946    0.04238  -1.403    0.164    
---
```

2 次項的 p-value=0.164,所以我們沒有顯著證據 reject H0。有二次項的回歸方程式雖然 R^2 會比簡單線性回歸高,但其實回歸的解釋資料能力沒有比較好。

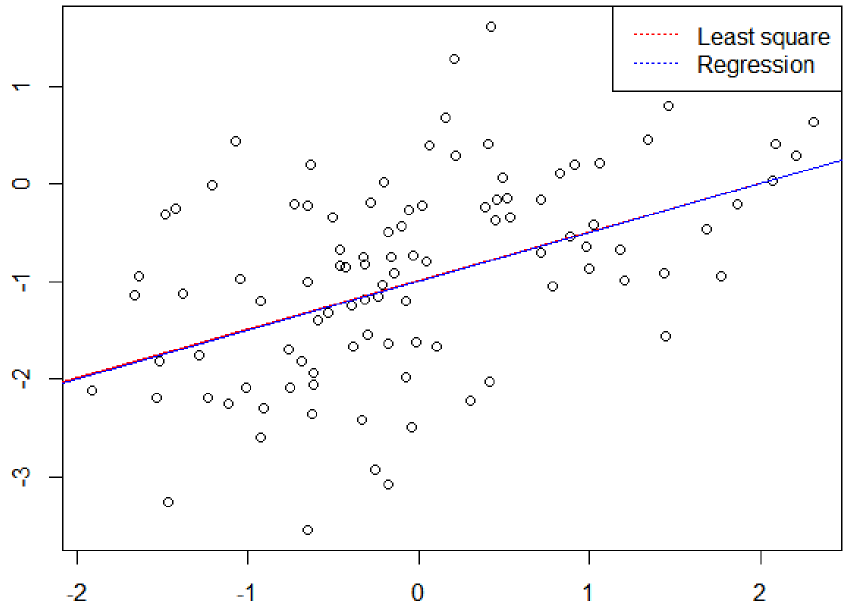
(h)

```
set.seed(1)
eps <- rnorm(100,mean=0,sd = 0.1)
x <- rnorm(100)
y <- -1 + 0.5 * x + eps
plot(x, y)
haha <- lm(y ~ x)
summary(haha)
abline(haha, col = "red")
abline(-1, 0.5, col = "blue")
legend("topright", c("Least square", "Regression"), col = c("red", "blue"), lty = c(3, 3)
```



(i)

```
set.seed(1)
eps <- rnorm(100,mean=0,sd = 1)
x <- rnorm(100)
y <- -1 + 0.5 * x + eps
plot(x, y)
haha1 <- lm(y ~ x)
summary(haha1)
abline(haha, col = "red")
abline(-1, 0.5, col = "blue")
legend("topright", c("Least square", "Regression"), col = c("red", "blue"), lty = c(3, 3)
```



(j)

```
> confint(a)
                 2.5 %     97.5 %
(Intercept) -1.1150804 -0.9226122
x            0.3925794  0.6063602
> confint(haha)
                 2.5 %     97.5 %
(Intercept) -1.0070441 -0.9711855
x            0.4811096  0.5187039
> confint(haha1)
                 2.5 %     97.5 %
(Intercept) -1.0704405 -0.7118552
x            0.3110958  0.6870395
```

隨著 noise 越多,信賴區間越寬,反之;noise 越少,信賴區間越窄,代表有比較好的預測。