



Semester 1 Assessment, 2021

School of Mathematics and Statistics

MAST30025 Linear Statistical Models Assignment 1

Submission deadline: **Friday March 26, 5pm**

This assignment consists of 3 pages (including this page)

Instructions to Students

- If you have a printer, print the assignment template.

Writing

- There are 5 questions with marks as shown. The total number of marks available is 40.
- This assignment is worth 6% of your total mark.
- You may choose to either typeset your assignment in \LaTeX or handwrite and scan it to produce an electronic version.
- You may use R for this assignment, but for matrix calculations only (you may not use the `lm` function). If you do, include your R commands and output.
- Write your answers on A4 paper. Page 1 should only have your student number, the subject code and the subject name. Write on one side of each sheet only. Each question should be on a new page. The question number must be written at the top of the page.

Scanning

- Put the pages in question order and all the same way up. Use a scanning app to scan all pages to PDF. Scan directly from above. Crop pages to A4. Check PDF is readable.

Submitting

- Go to the Gradescope window. Choose the Canvas assignment for this assignment. Submit your file as a single PDF document only. Get Gradescope confirmation on email.
- It is your responsibility to ensure that your assignments are submitted correctly and on time, and problems with online submissions are not a valid excuse for submitting a late or incorrect version of an assignment.

Linear Statistical Models Assignment 1

Kim Seang CHY

Question 1

a. For a symmetric matrix \mathbf{A} , if $\mathbf{A}^2 = \mathbf{A}^3$, then \mathbf{A} is idempotent.

There are two cases the first is the first is matrix \mathbf{A} over \mathbb{R} and the second case is matrix \mathbf{A} over \mathbb{C} .

Case 1: Let \mathbf{A} be a symmetric matrix over \mathbb{R} such that $\mathbf{A}^2 = \mathbf{A}^3$.

By eigenvalue properties, \mathbf{A} is diagonalisable with real eigenvalues and orthogonal eigenvectors, thus there exist a diagonalise matrix \mathbf{P} such that:

$$\mathbf{P}^T \mathbf{A} \mathbf{P} = \mathbf{D}$$

where $\mathbf{D} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_i, \dots)$ and λ_i is the eigenvalue of \mathbf{A} for all $i \in \mathbb{N} \setminus \{0\}$

Since $\mathbf{A}^2 = \mathbf{A}^3$, this implied:

$$\begin{aligned} \mathbf{D}^2 &= \mathbf{D}^3 && \text{(Diagonalise relationship)} \\ \implies (\lambda_i)^2 &= (\lambda_i)^3, \forall i \in \mathbb{N} \setminus \{0\} \\ \implies \lambda_i^2(\lambda_i - 1) &= 0 \\ \implies \lambda_i &= 0 \text{ or } \lambda_i = 1 \end{aligned}$$

Since the eigenvalue can only take value of 0 or 1. We can inferred:

$$\begin{aligned} \mathbf{D} &= \mathbf{D}^n, n = 1, 2, 3 \dots && \text{(Diagonal matrix properties)} \\ \implies \mathbf{A} &= \mathbf{A}^n && \text{(By diagonalisation properties)} \end{aligned}$$

Given $\mathbf{A} = \mathbf{A}^n$. Hence, \mathbf{A} is an idempotent matrix by the definition of idempotent matrix and its diagonalise properties. The statement is true if \mathbf{A} is a real matrix and symmetric.

Case 2: Proof by counter example.

Let

$$\mathbf{A} = \begin{bmatrix} -\mathbf{i} & 1 \\ 1 & \mathbf{i} \end{bmatrix}$$

Notice that $\mathbf{A} = \mathbf{A}^T$, this implied \mathbf{A} is a symmetric matrix. Now notice that:

$$\mathbf{A}^2 = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} = \mathbf{0}$$

This implied $\mathbf{A}^3 = \mathbf{0A} = \mathbf{0} = \mathbf{A}^2$. However, $\mathbf{A} \neq \mathbf{A}^2$, thus \mathbf{A} is not idempotent. Hence, the statement is false for a complex symmetric matrix.

Therefore, a symmetric matrix \mathbf{A} such that $\mathbf{A}^2 = \mathbf{A}^3$, implied \mathbf{A} is idempotent if and only if \mathbf{A} has all real element.

b. For a symmetric matrix \mathbf{A} , if $\mathbf{A} = \mathbf{A}^3$, then \mathbf{A} is idempotent.

Proof: Proof by counter example.

$$\text{Let } \mathbf{A} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

Notice that $\mathbf{A} = \mathbf{A}^T$, this implied \mathbf{A} is a symmetric by definition of symmetric.

$$\mathbf{A}^2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$\mathbf{A}^3 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

$\mathbf{A} \neq \mathbf{A}^2$, hence \mathbf{A} is not idempotent. Since $\mathbf{A} = \mathbf{A}^3$ but \mathbf{A} is not idempotent. Therefore, the statement is false.

Question 2:

Prove for A_1, A_2, \dots, A_m set of symmetric $k \times k$ matrices and there exist an orthogonal matrix P such that $P^T A_i P$ is diagonal for all i then $A_i A_j = A_j A_i$ for every pair of $i, j = 1, 2, \dots, m$.

Proof:

Supposed A_1 and A_2 are symmetric $k \times k$ matrices, simultaneously diagonalisable by orthogonal matrix P . This implied there exist:

$$A_1 = P D_1 P^T$$

and

$$A_2 = P D_2 P^T$$

Where D_1 is the diagonal matrix contain its respective eigenvalue for A_1 and D_2 is the diagonal matrix contain its respective eigenvalue for A_2

Hence:

$$\begin{aligned} A_1 A_2 &= P D_1 P^T P D_2 P^T \\ &= P D_1 D_2 P^T \quad (P^T P = I \text{ by orthogonal matrix properties}) \end{aligned}$$

Since diagonal matrix are commutable this implied $D_1 D_2 = D_2 D_1$. Therefore:

$$\begin{aligned} A_1 A_2 &= P D_2 D_1 P^T \\ &= P D_2 P^T P D_1 P = A_2 A_1 \end{aligned}$$

The above can expand to be show for any set of symmetric matrix that is simultaneously diagonalisable by the same orthogonal matrix. Therefore, if a set of symmetric matrices A_1, A_2, \dots, A_m is simultaneously diagonalisable by P then $A_i A_j = A_j A_i$ for every pair of $i, j = 1, 2, \dots, m$.

Question 3:

Let \mathbf{A} be an $n \times p$ matrix with rank p and all element belong to \mathbb{R} . This implied column vector of \mathbf{A} is linearly independent. Let \mathbf{y} be a $p \times 1$ matrix such that $\mathbf{y} = [y_1, y_2, \dots, y_p]^T$.

Note that:

$$\mathbf{y}^T \mathbf{A}^T \mathbf{A} \mathbf{y} = (\mathbf{A} \mathbf{y})^T \mathbf{A} \mathbf{y} \quad (\text{By the symmetric properties})$$

For $\mathbf{A} \mathbf{y}$ matrix, it can be represented in a form $\mathbf{A} \mathbf{y} = \mathbf{a}_1 y_1 + \mathbf{a}_2 y_2 + \dots + \mathbf{a}_p y_p$ where \mathbf{a}_i represent the column vectors for matrix \mathbf{A} for column i where $i = 1, 2, \dots, p$.

Since the column are linearly independent, this implied there a non-zero solution to $\mathbf{A} \mathbf{y} = 0$ expect for $\mathbf{y} = \mathbf{0}$, thus implied there are also non-zero solution to $\mathbf{y}^T \mathbf{A}^T \mathbf{A} \mathbf{y} = 0$. Since $\mathbf{y}^T \mathbf{A}^T \mathbf{A} \mathbf{y} = (\mathbf{a}_1 y_1)^2 + (\mathbf{a}_2 y_2)^2 + \dots + (\mathbf{a}_p y_p)^2$, this is just a sum of squares, which implied $\mathbf{y}^T \mathbf{A}^T \mathbf{A} \mathbf{y} > 0$ for all $y \in \mathbb{R} \setminus \{0\}$. Therefore, $\mathbf{y}^T \mathbf{A}^T \mathbf{A} \mathbf{y}$ is positively defined matrix.

Question 4:

a. Let $\mathbf{y} = \mathbf{A} \mathbf{x}$. Find \mathbf{A} .

Want to find a vector \mathbf{A} such that: $\mathbf{y} = \mathbf{A} \mathbf{x} = \mathbf{x} - \bar{x} \mathbf{1}$ where $\mathbf{1}$ is a column vector of 1 with three rows.

Notice that:

$$\begin{aligned} \bar{x} &= \frac{1}{3} \mathbf{1}^T \mathbf{x} \\ \implies \mathbf{x} - \bar{x} \mathbf{1} &= \mathbf{x} - \frac{1}{3} \mathbf{1} \mathbf{1}^T \mathbf{x} \\ &= (\mathbf{I} - \frac{1}{3} \mathbf{1} \mathbf{1}^T) \mathbf{x} \\ \implies \mathbf{A} &= \mathbf{I} - \frac{1}{3} \mathbf{1} \mathbf{1}^T \\ \mathbf{I} - \frac{1}{3} \mathbf{1} \mathbf{1}^T &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} - \frac{1}{3} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \end{bmatrix}^T = \begin{bmatrix} \frac{2}{3} & \frac{-1}{3} & \frac{-1}{3} \\ \frac{-1}{3} & \frac{2}{3} & \frac{-1}{3} \\ \frac{-1}{3} & \frac{-1}{3} & \frac{2}{3} \end{bmatrix} \end{aligned}$$

Hence

$$A = \begin{bmatrix} \frac{2}{3} & \frac{-1}{3} & \frac{-1}{3} \\ \frac{-1}{3} & \frac{2}{3} & \frac{-1}{3} \\ \frac{-1}{3} & \frac{-1}{3} & \frac{2}{3} \end{bmatrix}$$

b. Find the rank of A.

The rank(A) = 2, since the sum of column 1 and column 2 is equal to negative of column 3 and column 1 is not a multiple of column 2.

c. Find $E(\mathbf{y}^T \mathbf{y})$.

$$\begin{aligned} E(\mathbf{y}^T \mathbf{y}) &= E \left(\begin{bmatrix} x_1 - \bar{x} \\ x_2 - \bar{x} \\ x_3 - \bar{x} \end{bmatrix}^T \begin{bmatrix} x_1 - \bar{x} \\ x_2 - \bar{x} \\ x_3 - \bar{x} \end{bmatrix} \right) \\ &= E \left(\sum_{i=1}^3 (x_i - \bar{x})^2 \right) \\ &= E \left[\sum_{i=1}^3 x_i^2 \right] - E[3\bar{x}^2] \\ &= \sum_{i=1}^3 E[x_i^2] - 3E[\bar{x}^2] \end{aligned}$$

Notice that $\text{Var}(x) = E(x^2) - (E(x))^2$, this implied $E(x^2) = \text{Var}(x) + (E(x))^2 = \sigma^2 + \mu^2$. Similarly, $\text{Var}(\bar{x}) = E(\bar{x}^2) - (E(\bar{x}))^2$, this implied $E(\bar{x}^2) = \text{Var}(\bar{x}) + E(\bar{x})^2 = \frac{\sigma^2}{3} + \mu^2$, since x_1, x_2, x_3 is IID Therefore,

$$\begin{aligned} E(\mathbf{y}^T \mathbf{y}) &= \left[\sum_{i=1}^3 (\sigma^2 + \mu^2) - 3 \left(\frac{\sigma^2}{3} + \mu^2 \right) \right] \\ &= \left[(\sigma^2 + \mu^2) - 3 \left(\frac{\sigma^2}{3} + \mu^2 \right) \right] = [2\sigma^2] \end{aligned}$$

d. Using theorem 3.5, find the distribution of $\frac{\mathbf{y}^T \mathbf{y}}{\sigma^2}$.

From part 1 we know $\mathbf{y} = A\mathbf{x}$ and $\mathbf{x} \sim MVN(\mu, \sigma^2)$ is 3×1 random vectors and

$$\begin{aligned} A &= \begin{bmatrix} \frac{2}{3} & \frac{-1}{3} & \frac{-1}{3} \\ \frac{-1}{3} & \frac{2}{3} & \frac{-1}{3} \\ \frac{-1}{3} & \frac{-1}{3} & \frac{2}{3} \end{bmatrix} \\ \implies A^2 &= \begin{bmatrix} \frac{2}{3} & \frac{-1}{3} & \frac{-1}{3} \\ \frac{-1}{3} & \frac{2}{3} & \frac{-1}{3} \\ \frac{-1}{3} & \frac{-1}{3} & \frac{2}{3} \end{bmatrix} \end{aligned}$$

Notice that $A = A^T$ and $A = A^2$, this implies A is a symmetric and idempotent matrix. Thus:

$$\begin{aligned}
\frac{\mathbf{y}^T \mathbf{y}}{\sigma^2} &= \frac{1}{\sigma^2} (\mathbf{A}\mathbf{x})^T (\mathbf{A}\mathbf{x}) \\
&= \frac{1}{\sigma^2} \mathbf{x}^T \mathbf{A}^T \mathbf{A} \mathbf{x} \\
&= \frac{1}{\sigma^2} \mathbf{x}^T \mathbf{A}^2 \mathbf{x} \\
&= \frac{\mathbf{x}^T}{\sigma} \mathbf{A} \frac{\mathbf{x}}{\sigma}
\end{aligned}$$

Since σ is a constant this implied $\frac{1}{\sigma}\mathbf{x} \sim MVN(\frac{1}{\sigma}\mu, \sigma^2)$. Therefore, by Theorem 3.5, implied $\frac{\mathbf{y}^T \mathbf{y}}{\sigma^2}$ has a noncentral χ^2 with $\text{rank}(\mathbf{A}) = 2$ degree of freedoms, and a noncentral parameter $\lambda = \frac{1}{2} \left(\frac{\mu}{\sigma}\right)^T \mathbf{A} \left(\frac{\mu}{\sigma}\right) = \frac{1}{2\sigma^2} \mu^T \mathbf{A} \mu$. In this case:

$$\lambda = \frac{1}{2\sigma^2} \begin{bmatrix} \mu \\ \mu \\ \mu \end{bmatrix}^T \begin{bmatrix} \frac{2}{3} & \frac{-1}{3} & \frac{-1}{3} \\ \frac{-1}{3} & \frac{2}{3} & \frac{-1}{3} \\ \frac{-1}{3} & \frac{-1}{3} & \frac{2}{3} \end{bmatrix} \begin{bmatrix} \mu \\ \mu \\ \mu \end{bmatrix} = [0]$$

Since $\lambda = 0$, this implied $\frac{\mathbf{y}^T \mathbf{y}}{\sigma^2}$ is regular χ^2 distribution with 2 degree of freedom.

Question 5:

All the matrix calculation and other calculation for this question will be in appendix.

a. Write down the matrices and vector involved in the linear model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$.

Since, we are trying to model the price of fish in 1980, based on the 1970 price then the predictor variables(y_i) is the price of fish in 1980 and the explanation variable x_{i2} is the price of fish in 1970. The model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ is given by:

$$\begin{bmatrix} 27.3 \\ 42.4 \\ 38.7 \\ 4.5 \\ 23 \\ 166.3 \\ 109.7 \\ 80.1 \\ 150.7 \\ 20.3 \\ 189.7 \\ 131.3 \\ 404.2 \\ 149 \end{bmatrix} = \begin{bmatrix} 1 & 13.1 \\ 1 & 15.3 \\ 1 & 25.8 \\ 1 & 1.8 \\ 1 & 4.9 \\ 1 & 55.4 \\ 1 & 39.3 \\ 1 & 26.7 \\ 1 & 47.5 \\ 1 & 6.6 \\ 1 & 94.7 \\ 1 & 61.1 \\ 1 & 135.6 \\ 1 & 47.6 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_{14} \end{bmatrix}$$

Where β_0, β_1 are the parameter and ϵ_i are the error term in row i .

b. Find the least squares estimators of the parameters.

$$\text{Let } \mathbf{b} = \begin{bmatrix} b_0 \\ b_1 \end{bmatrix}$$

By theorem 4.4, the best least estimator is give by: $\mathbf{b} = (X^T X)^{-1} X^T \mathbf{y}$.

$$X^T X = \begin{bmatrix} 14.0 & 575.4 \\ 575.4 & 42079.36 \end{bmatrix} \text{ and } (X^T X)^{-1} = \begin{bmatrix} 0.16308 & -0.0022 \\ -0.0022 & 5.43\text{e-}5 \end{bmatrix}$$

$$\begin{aligned} \Rightarrow \mathbf{b} &= (X^T X)^{-1} X^T \mathbf{y} \\ &= \begin{bmatrix} -1.2338 \\ 2.7016 \end{bmatrix} \end{aligned}$$

The least square estimator $b_0 = -1.2338$ with $\text{var}(b_0) = 0.163\sigma^2$ and $b_1 = 2.7016$ with $\text{var}(b_1) = (5.43\text{e-}5)\sigma^2$.

c. Find the sample variances s^2 .

By theorem 4.6, the unbiased for σ^2 is given by:

$$s^2 = \frac{(\mathbf{y} - X\mathbf{b})^T (\mathbf{y} - X\mathbf{b})}{n - (k + 1)}$$

Where n is row of \mathbf{x} and \mathbf{y} and $k + 1$ is the column of \mathbf{x} . Hence for our data sets

$$\begin{aligned} s^2 &= \frac{(\mathbf{y} - X\mathbf{b})^T (\mathbf{y} - X\mathbf{b})}{14 - 2} \\ &= \frac{9325.833}{12} = 777.1528 \end{aligned}$$

The sample variance for our model is 777.1528.

Since the variance of the parameter is given by $\text{var}(\mathbf{b}) = (X^T X)^{-1} \sigma^2$ or $(X^T X)^{-1} s^2$, this implied $\text{var}(b_0) = 126.74$ and $\text{var}(b_1) = 0.04217$

d. Predict the price for ocean trout in 1980 given a fisher sold ocean trout for 28c/pound in 1970.

$$\text{Let } t = \begin{bmatrix} 1 \\ 28 \end{bmatrix}$$

By theorem 4.5, we can uses $\mathbf{t}^T \mathbf{b}$ we can use to predict the price as it is best linear unbiased estimator for $\mathbf{t}^T \boldsymbol{\beta}$.

Therefore, to predict the of ocean trout given fisher sold ocean trout for 28c/pound and our model:

$$\begin{aligned} \text{price}_{1980} &= -1.2338 + 2.7016 \text{ price}_{1970} \\ &= \begin{bmatrix} 1 \\ 28 \end{bmatrix}^T \begin{bmatrix} -1.2338 \\ 2.7016 \end{bmatrix} = [74.40965] \end{aligned}$$

The price for ocean trout the fisherman expected get in 1980 is \$74.41c /pound.

e. Calculate the standardised residual for sea scallops.

Let $H = X(X^T X)^{-1} X^T$ and $\mathbf{e} = \mathbf{y} - X\mathbf{b}$. Hence $H_{13\ 13} = 0.5560$ and $\mathbf{e}_{13\ 1} = 39.10$ since the sea scallop is in the 13th index of our data

The standardised residual is given by:

$$z_i = \frac{e_{1i}}{\sqrt{s^2(1 - H_{ii})}}$$

$$\Rightarrow z_{13} = \frac{39.10}{\sqrt{777.1528(1 - 0.5560)}} = 2.104999$$

The standardised residual for sea scallops is 2.1050.

f. Calculate the Cook's distance for sea scallops.

The Cook's distance for sea scallops is:

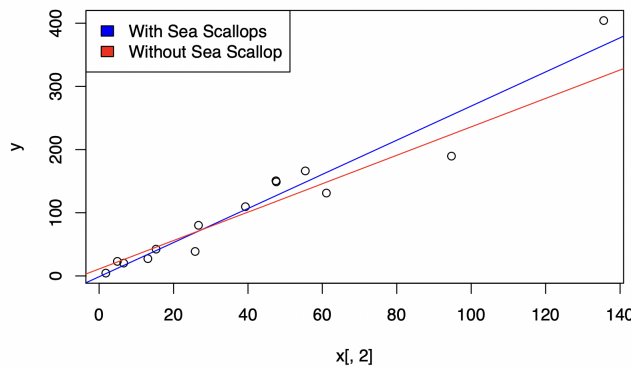
$$D_{13} = \frac{z_{13}^2}{1 + 1} \left(\frac{H_{13\ 13}}{1 - H_{13\ 13}} \right)$$

$$= \frac{2.1050^2}{2} \left(\frac{0.5560}{1 - 0.5560} \right) = 2.7740$$

g. Does sea scallops fit the linear model? Justify your argument.

The sea scallops Cook's distance is 2.774 which is significantly greater than 1 thus implied the sea scallops observation is a point of high leverage and high standardised residual. Such point is considered to be an influential outlier that significant impact the fit a model hence the sea scallops does not fit the linear model.

This was also reflected in best fit line of the graph below where the best fit line without sea scallops observation is steeper than the best fit line with sea scallops observation.



Appendix.R

kschy

2021-03-24

```
#Question 5
```

```
#Part A
```

```
#Setting Data Point
```

```
y <- matrix(c(27.3, 42.4 ,38.7, 4.5, 23, 166.3, 109.7, 80.1, 150.7, 20.3, 189.7,  
              131.3, 404.2, 149),14,1,byrow = TRUE)
```

```
y
```

```
##      [,1]  
## [1,] 27.3  
## [2,] 42.4  
## [3,] 38.7  
## [4,]  4.5  
## [5,] 23.0  
## [6,] 166.3  
## [7,] 109.7  
## [8,]  80.1  
## [9,] 150.7  
## [10,]  20.3  
## [11,] 189.7  
## [12,] 131.3  
## [13,] 404.2  
## [14,] 149.0
```

```
x <- matrix(c(rep(1,14),13.1, 15.3, 25.8, 1.8, 4.9, 55.4,  
              39.3, 26.7, 47.5, 6.6, 94.7, 61.1, 135.6, 47.6), 14, 2)
```

```
x
```

```
##      [,1] [,2]  
## [1,]    1 13.1  
## [2,]    1 15.3  
## [3,]    1 25.8  
## [4,]    1  1.8  
## [5,]    1  4.9  
## [6,]    1 55.4  
## [7,]    1 39.3  
## [8,]    1 26.7  
## [9,]    1 47.5  
## [10,]    1  6.6  
## [11,]    1 94.7  
## [12,]    1 61.1  
## [13,]    1 135.6  
## [14,]    1  47.6
```

```

#Part B
#Finding Least Square Estimator
b <- solve(t(x)%*%x,t(x)%*%y)
b

##           [,1]
## [1,] -1.233836
## [2,]  2.701553

#Inverse of diagonal  $X^T X$ 
xtx.inverse <- solve(t(x)%*%x)
xtx.inverse

##           [,1]           [,2]
## [1,]  0.163081936 -2.230009e-03
## [2,] -0.002230009  5.425812e-05

```

```

#Part C
#Calculating Sample Variance
e <- y-x%*%b
e

##           [,1]
## [1,] -6.8565106
## [2,]  2.3000724
## [3,] -29.7662361
## [4,]  0.8710405
## [5,] 10.9962256
## [6,] 17.8677893
## [7,]  4.7627957
## [8,]  9.2023660
## [9,] 23.6100596
## [10,]  3.7035852
## [11,] -64.9032511
## [12,] -32.5310639
## [13,] 39.1032233
## [14,] 21.6399042

```

```

#Calculating Sum-squared
SSRes <- sum(e^2)
SSRes

```

```
## [1] 9325.833
```

```

#Calculating sample variance
s2 <- SSRes/(14-2)
s2

```

```
## [1] 777.1528
```

```

#Calculating the variance of the least square estimator
b.var <- solve(t(x)%*%x)*s2
diag.b.var <- diag(solve(t(x)%*%x))*s2
diag.b.var

```

```
## [1] 126.73957949  0.04216685
```

```

#Part D
#Calculating ocean trout expected price in 1980

```

```
t <- matrix(c(1,28),2,1)
t
```

```
##      [,1]
## [1,]    1
## [2,]   28
```

```
y.bar <- t(t)%*%b
y.bar
```

```
##      [,1]
## [1,] 74.40965
```

#Part E

#Calculating the H(hat) matrix

```
hat <- x%*%solve(t(x)%*%x)%*%t(x)
hat
```

```
##      [,1]      [,2]      [,3]      [,4]      [,5]      [,6]
## [1,] 0.11396694 0.11062464 0.094672751 0.13113421 0.12642461 0.04970362
## [2,] 0.11062464 0.107544949 0.092846423 0.12644305 0.12210349 0.05141058
## [3,] 0.09467275 0.092846423 0.084129856 0.10405344 0.10147998 0.05955744
## [4,] 0.13113421 0.126443053 0.104053438 0.15522970 0.14861943 0.04093605
## [5,] 0.12642461 0.122103488 0.101479976 0.14861943 0.14253059 0.04334131
## [6,] 0.04970362 0.051410579 0.059557437 0.04093605 0.04334131 0.08252382
## [7,] 0.07416318 0.073948319 0.072922840 0.07526679 0.07496403 0.07003197
## [8,] 0.09330545 0.091586549 0.083382721 0.10213433 0.09971225 0.06025574
## [9,] 0.06170552 0.062469470 0.066115616 0.05778157 0.05885805 0.07639427
## [10,] 0.12384192 0.119723727 0.100068722 0.14499445 0.13919154 0.04466033
## [11,] -0.01000202 -0.003603902 0.026932570 -0.04286508 -0.03384955 0.11301634
## [12,] 0.04104402 0.043431380 0.054825586 0.02878169 0.03214569 0.08694639
## [13,] -0.07213842 -0.060858159 -0.007020536 -0.13007796 -0.11418304 0.14475029
## [14,] 0.06155359 0.062329484 0.066032601 0.05756833 0.05866164 0.07647186
##      [,7]      [,8]      [,9]      [,10]      [,11]      [,12]
## [1,] 0.07416318 0.093305447 0.06170552 0.12384192 -0.010002020 0.04104402
## [2,] 0.07394832 0.091586549 0.06246947 0.11972373 -0.003603902 0.04343138
## [3,] 0.07292284 0.083382721 0.06611562 0.10006872 0.026932570 0.05482559
## [4,] 0.07526679 0.102134329 0.05778157 0.14499445 -0.042865080 0.02878169
## [5,] 0.07496403 0.099712246 0.05885805 0.13919154 -0.033849550 0.03214569
## [6,] 0.07003197 0.060255739 0.07639427 0.04466033 0.113016338 0.08694639
## [7,] 0.07160437 0.072834942 0.07080352 0.07479800 0.066193748 0.06947528
## [8,] 0.07283494 0.082679536 0.06642814 0.09838401 0.029549982 0.05580223
## [9,] 0.07080352 0.066428143 0.07365098 0.05944838 0.090041278 0.07837361
## [10,] 0.07479800 0.098384007 0.05944838 0.13600930 -0.028905550 0.03399047
## [11,] 0.06619375 0.029549982 0.09004128 -0.02890555 0.227309989 0.12959328
## [12,] 0.06947528 0.055802232 0.07837361 0.03399047 0.129593280 0.09313182
## [13,] 0.06219926 -0.002405883 0.10424388 -0.10546647 0.346256817 0.17397642
## [14,] 0.07079375 0.066350011 0.07368571 0.05926119 0.090332102 0.07848213
##      [,13]      [,14]
## [1,] -0.072138423 0.06155359
## [2,] -0.060858159 0.06232948
## [3,] -0.007020536 0.06603260
## [4,] -0.130077960 0.05756833
## [5,] -0.114183042 0.05866164
## [6,] 0.144750286 0.07647186
## [7,] 0.062199265 0.07079375
```

```
## [8,] -0.002405883 0.06635001
## [9,] 0.104243884 0.07368571
## [10,] -0.105466475 0.05926119
## [11,] 0.346256817 0.09033210
## [12,] 0.173976424 0.07848213
## [13,] 0.555967177 0.10475662
## [14,] 0.104756624 0.07372098
```

```
#Sea scallops leverage
```

```
hat[13,13]
```

```
## [1] 0.5559672
```

```
#Calculating Residual variance
```

```
e.var <- s2*(diag(14)-hat)
```

```
e.var
```

```
##           [,1]      [,2]      [,3]      [,4]      [,5]      [,6]
## [1,] 688.583054 -85.972246 -73.575192 -101.91132 -98.25123 -38.62731
## [2,] -85.972246 693.573923 -72.155856 -98.26557 -94.89307 -39.95387
## [3,] -73.575192 -72.155856 711.771027 -80.86542 -78.86544 -46.28523
## [4,] -101.911316 -98.265570 -80.865419 656.51559 -115.50001 -31.81356
## [5,] -98.251233 -94.893065 -78.865445 -115.50001 666.38474 -33.68282
## [6,] -38.627305 -39.953875 -46.285227 -31.81356 -33.68282 713.01917
## [7,] -57.636122 -57.469141 -56.672188 -58.49380 -58.25850 -54.42554
## [8,] -72.512587 -71.176741 -64.801113 -79.37398 -77.49165 -46.82791
## [9,] -47.954613 -48.548322 -51.381935 -44.90511 -45.74170 -59.37002
## [10,] -96.244091 -93.043627 -77.768685 -112.68284 -108.17309 -34.70790
## [11,] 7.773098 2.800782 -20.930721 33.31272 26.30627 -87.83096
## [12,] -31.897476 -33.752817 -42.607856 -22.36777 -24.98211 -67.57063
## [13,] 56.062576 47.296087 5.456029 101.09045 88.73767 -112.49309
## [14,] -47.836546 -48.439532 -51.317419 -44.73939 -45.58905 -59.43032
##           [,7]      [,8]      [,9]      [,10]      [,11]      [,12]
## [1,] -57.63612 -72.512587 -47.95461 -96.24409 7.773098 -31.89748
## [2,] -57.46914 -71.176741 -48.54832 -93.04363 2.800782 -33.75282
## [3,] -56.67219 -64.801113 -51.38193 -77.76869 -20.930721 -42.60786
## [4,] -58.49380 -79.373977 -44.90511 -112.68284 33.312716 -22.36777
## [5,] -58.25850 -77.491649 -45.74170 -108.17309 26.306272 -24.98211
## [6,] -54.42554 -46.827915 -59.37002 -34.70790 -87.830961 -67.57063
## [7,] 721.50524 -56.603878 -55.02515 -58.12947 -51.442655 -53.99291
## [8,] -56.60388 712.898147 -51.62482 -76.45940 -22.964850 -43.36686
## [9,] -55.02515 -51.624816 719.91471 -46.20047 -69.975829 -60.90827
## [10,] -58.12947 -76.459404 -46.20047 671.45277 22.464029 -26.41579
## [11,] -51.44265 -22.964850 -69.97583 22.46403 600.498189 -100.71378
## [12,] -53.99291 -43.366860 -60.90827 -26.41579 -100.713777 704.77512
## [13,] -48.33833 1.869738 -81.01342 81.96356 -269.094448 -135.20626
## [14,] -55.01756 -51.564095 -57.26505 -46.05500 -70.201844 -60.99260
##           [,13]      [,14]
## [1,] 56.062576 -47.83655
## [2,] 47.296087 -48.43953
## [3,] 5.456029 -51.31742
## [4,] 101.090448 -44.73939
## [5,] 88.737669 -45.58905
## [6,] -112.493087 -59.43032
## [7,] -48.338331 -55.01756
## [8,] 1.869738 -51.56410
```

```
## [9,] -81.013424 -57.26505
## [10,] 81.963564 -46.05500
## [11,] -269.094448 -70.20184
## [12,] -135.206262 -60.99260
## [13,] 345.081342 -81.41190
## [14,] -81.411901 719.86032

#Standardised Residual calculation function
z <- function(i){
  e[i,1]/sqrt(s2*(1-hat[i,i]))
}

#Calculating Standardised Residual for sea scallops
z(13)
```

```
## [1] 2.104999
```

```
#Part F
#Cook's distance function
d <-function(i){
  (((z(i))^2)/2)*((hat[i,i])/(1-hat[i,i]))
}

#Cook's distance for sea scallops
d(13)
```

```
## [1] 2.774008
```

```
#Part G
#Cook's Distance matrix for all observation
cook.d <- matrix(c(0,14), 14, 1)
i <- 0
while (i < 15) {
  cook.d[i,1]=d(i)
  i = i+1
}
cook.d
```

```
##           [,1]
## [1,] 0.0043908534
## [2,] 0.0004595830
## [3,] 0.0571733691
## [4,] 0.0001061788
## [5,] 0.0150807062
## [6,] 0.0201370081
## [7,] 0.0012124417
## [8,] 0.0053532616
## [9,] 0.0307813052
## [10,] 0.0016078990
## [11,] 1.0318211614
## [12,] 0.0771027556
## [13,] 2.7740080577
## [14,] 0.0258869983
```

```
#Omitting Sea Scallop observation fitting
x.omit <- x[-13,]
x.omit
```

```
##      [,1] [,2]
## [1,]    1 13.1
## [2,]    1 15.3
## [3,]    1 25.8
## [4,]    1  1.8
## [5,]    1  4.9
## [6,]    1 55.4
## [7,]    1 39.3
## [8,]    1 26.7
## [9,]    1 47.5
## [10,]   1  6.6
## [11,]   1 94.7
## [12,]   1 61.1
## [13,]   1 47.6

y.omit <- y[-13,1]
y.omit

## [1] 27.3 42.4 38.7  4.5 23.0 166.3 109.7  80.1 150.7  20.3 189.7 131.3
## [13] 149.0

b.omit <- solve(t(x.omit)%*%x.omit,t(x.omit)%*%y.omit)
b.omit

##      [,1]
## [1,] 11.034093
## [2,]  2.250015

#Plotting Regression Graph with and without sea scallop
plot(x[,2],y)
abline(b[1,1],b[2,1], col = "blue")
abline(b.omit[1,1],b.omit[2,1], col = "red")
legend("topleft",
      c("With Sea Scallops","Without Sea Scallop"),
      fill=c("blue","red"))
```

