THE UNIVERSITY OF MELBOURNE

Student number 998008

Semester 1 Assessment, 2021

School of Mathematics and Statistics

MAST30025 Linear Statistical Models Assignment 2

Submission deadline: **Friday April 30, 5pm**

This assignment consists of 4 pages (including this page)

Instructions to Students

Writing

- There are 5 questions with marks as shown. The total number of marks available is 40.
- This assignment is worth 7% 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, including the `lm` function unless specified. 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 2

Kim Seang CHY

Question 1: Maximum likelihood of σ^2

Recall that in the general full model $\boldsymbol{\varepsilon} \sim MVN(0, \sigma^2 I)$, hence the likelihood function is given by:

$$
\begin{aligned}
L(\beta, \sigma) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(\frac{-\varepsilon_i^2}{2\sigma^2}\right) \\
&= (2\pi)^{-\frac{n}{2}} \sigma^{-n} \exp\left(\sum_{i=1}^{n} \frac{-\varepsilon_i^2}{2\sigma^2}\right)
\end{aligned}
$$

Let $\ell(\beta, \sigma)$ be the log likelihood function of $L(\beta, \sigma)$.

$$
\ell(\beta, \sigma) = -\frac{n}{2}\log(2\pi) - n\log(\sigma) - \frac{1}{2\sigma^2}\left(\sum_{i=1}^{n} \varepsilon_i^2\right)
$$

Since $\sum_{i=1}^{n} \varepsilon_i^2 = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = SS_{Res}$:

$$
\begin{aligned}
&\ell(\beta, \sigma) = -\frac{n}{2}\log(2\pi) - n\log(\sigma) - \frac{SS_{Res}}{2\sigma^2} \\
&\frac{\partial \ell}{\partial \sigma} = -\frac{1}{\sigma} + \frac{SS_{Res}}{\sigma^3}
\end{aligned}
$$

Setting $\frac{\partial \ell}{\partial \sigma} = 0$ and solve for σ we get:

$$
\begin{aligned}
-\frac{1}{\sigma} + \frac{SS_{Res}}{\sigma^3} &= 0 \\
\implies \sigma &= \frac{SS_{Res}}{n}
\end{aligned}
$$

Question 2:

```
#Loading Data
q2data <- data.matrix(read.csv(file = "Q2_Data.csv"))
q2frame <- read.csv(file="Q2_Data.csv")
y <- matrix(q2data[,1],7,1)
y
```

```
##      [,1]
## [1,] 37.9
## [2,] 42.2
## [3,] 47.3
## [4,] 43.1
## [5,] 54.8
## [6,] 47.1
## [7,] 40.3
```

```
x <- matrix(c(rep(1,7),q2data[,-1]),7,4)
x
```

```
##      [,1] [,2]   [,3] [,4]
## [1,]    1 32.0   84.9   19
## [2,]    1 19.5  306.6    9
## [3,]    1 13.3  562.0    5
## [4,]    1 13.3  562.0    5
## [5,]    1  5.0  390.6    5
## [6,]    1  7.1 2175.0    3
## [7,]    1 34.5  623.5    7
```

```
df <- 7-4
```

a: Fit a linear model to the data and estimate the parameters and variance.

```
#Finding Beta using BLUE
b <- solve(t(x)%*%x,t(x)%*%y)
b
```

```
##              [,1]
## [1,] 54.776606226
## [2,] -0.389598784
## [3,] -0.001973937
## [4,] -0.242767764
```

```
#Finding variance
#sum-Square
e <- (y-x%*%b)
SSres <- sum(e^2)
s2 <- SSres/(df)
s <- sqrt(s2)

#Beta Variance
C2x <- solve(t(x)%*%x)*s2
diag(C2x)
```

```
## [1] 1.964791e+01 3.378471e-02 7.330554e-06 1.870117e-01
```

```
#Beta standard Error
sqrt(diag(C2x))
```

```
## [1] 4.4325965 0.1838062 0.0027075 0.4324485
```

Thus the model is given by

$$\begin{array}{lllll} y = & 54.776606226 & -0.389598784X_1 & -0.001973937X_2 & -0.242767764X_3 \\ & (4.4325965) & (0.1838062) & (0.0027075) & (0.4324485) \end{array}$$

b. Find a 90% confidence interval for the expected price per square metre of a 10 year old apartment that is 100 meters away from the train station and has 6 convenience stores nearby.

```r
#Part B Computing CI

alpha <- 0.1
x.star <- c(1,10,100,6)
y.star <- x.star%*%b
ta <- qt(1-alpha/2, df)

#Computing 90CI for x1=10, x2= 100 ,x3 =6
CI = c(y.star - s*sqrt(t(x.star)%*%solve(t(x)%*%x)%*%x.star),
       y.star + s*sqrt(t(x.star)%*%solve(t(x)%*%x)%*%x.star))
CI
```

```
## [1] 46.59336 51.85988
```

The 90% confidence invterval of 10 years old appeartment that is 100 meters away from train station and has 6 convenience stores nearby is $(46.59336, 51.85988)$.

c. Find the standard error of $\beta_1 - \beta_3$

```r
#General Linear Hypothesis for B1-B3
#Setting C and delta star
C <- c(0,1,0,-1)
cdelta.star <- matrix(0)

#Computing the variance and standard error for B1-B3
Cb.var <- t(C)%*%solve(t(x)%*%x)%*%C*s2
Cb.var
```

```
##          [,1]
## [1,] 0.316463
```

```r
Cb.ste <- sqrt(Cb.var)
Cb.ste
```

```
##           [,1]
## [1,] 0.5625504
```

The standard error of $\beta_1 - \beta_3$ is 0.5625504.

d. Test the hypothesis that the price per square metre falls by \$1000 for every year that the apartment ages, at the 5% significance level.

Testing $H_0 = \beta_1 = -1$ vs $H_1 = \beta_1 \neq -1$

```r
#General Linear Hypothesis
#General Linear Hypothesis
C <- matrix(c(0,1,0,0),1,4)
dst <- matrix(-1)
```

```r
#Conducting an F-test for y=-1 given B1=1
num <- (t(C%*%b-dst)*solve((C%*%solve(t(x)%*%x)%*%t(C)))%*%(C%*%b-dst))
Fstat <- num/(SSres/df)
pf(Fstat,1,3, lower.tail = FALSE)
```

```
##             [,1]
## [1,] 0.04502395
```

Since the P-value is 0.04502395, which is less than 0.05, which should reject the null that the price will fall by \$1000 for each year the apartment age at 5% statistical significant.2

e. Test for model relevance using a corrected sum of squares.

Testing for $H_0 = \beta_1 = \beta_2 = \beta_3 = 0$ vs $H_1 = \beta_1$ or β_2 or β_3 is non-zero using corrected sum of squares.

```r
#Computing model 2
x2 <- x[,-1]
b2 <- solve(t(x2)%*%x2,t(x2)%*%y)

#Breaking Rg1g2 and Rg2 for correct sum squared
SSres2 <- sum((y-x2%*%b2)^2)
Rg2 <- t(y)%*%x2%*%b2
SSreg <- t(y)%*%y
Rg1g2 <- SSreg - Rg2
Rg1g2
```

```
##          [,1]
## [1,] 2158.632
```

```r
#F Test
r <- 1
Fstat <- (Rg1g2/r)/(SSres/(df))
Fstat
```

```
##          [,1]
## [1,] 155.7122
```

```r
pf(Fstat,r,df, lower.tail=FALSE)
```

```
##              [,1]
## [1,] 0.001109267
```

Since the p-value for the test is $0.001109267 < 0.05$, we can say the model is statically significant using the corrected sum of squared method. Hence, we should reject the null

Question 3: Show that the SS_{Res} for the first model is at least the SS_{Res} for the second model.

For the second model our sum of residual is given by $SS_{Res} = \mathbf{y}^T\mathbf{y} - \mathbf{y}^T\mathbf{H}\mathbf{y} = \mathbf{y}^T\mathbf{y} - \mathbf{y}^T\mathbf{X}\boldsymbol{\beta}$. Since $\mathbf{H}\mathbf{y} = \mathbf{X}\boldsymbol{\beta}$, then $SS_{Res} = \mathbf{y}^T\mathbf{y} - \mathbf{y}^T\mathbf{X}\boldsymbol{\beta}$

Let $SS_{Res_{\gamma_1}} = \mathbf{y}^T\mathbf{y} - \mathbf{y}^T\mathbf{H}_1\mathbf{y}$ be the sum of residual of the first model.
We can partition $\mathbf{X}$ and $\boldsymbol{\beta}$ as follow:

$\mathbf{X} = \left[\begin{array}{c|c} \mathbf{X}_1 & \mathbf{X}_2 \end{array}\right]$ and $\boldsymbol{\beta} = \left[\begin{array}{c} \gamma_1 \\ \hline \gamma_2 \end{array}\right]$

Hence, $\mathbf{X}\boldsymbol{\beta} = \mathbf{X}_1\boldsymbol{\gamma}_1 + \mathbf{X}_1\boldsymbol{\gamma}_2 = \mathbf{H}_1\mathbf{y} + \mathbf{H}_2\mathbf{y}$, where $\mathbf{H}_1$ is the hat matrix the first model and $\mathbf{H}_2$ is hat matrix for the rest of the predictor that was in the second model but not in first model. Thus SS_{Res} of the full model is given by:

$$\begin{aligned} SS_{Res} &= \mathbf{y}^T\mathbf{y} - \mathbf{y}^T(\mathbf{H}_1 + \mathbf{H}_2)\mathbf{y} \\ &= \mathbf{y}^T\mathbf{y} - \mathbf{y}^T\mathbf{H}_1\mathbf{y} - \mathbf{y}^T\mathbf{H}_2\mathbf{y} \\ &= SS_{Res_{\gamma_1}} - \mathbf{y}^T\mathbf{H}_2\mathbf{y} \end{aligned}$$

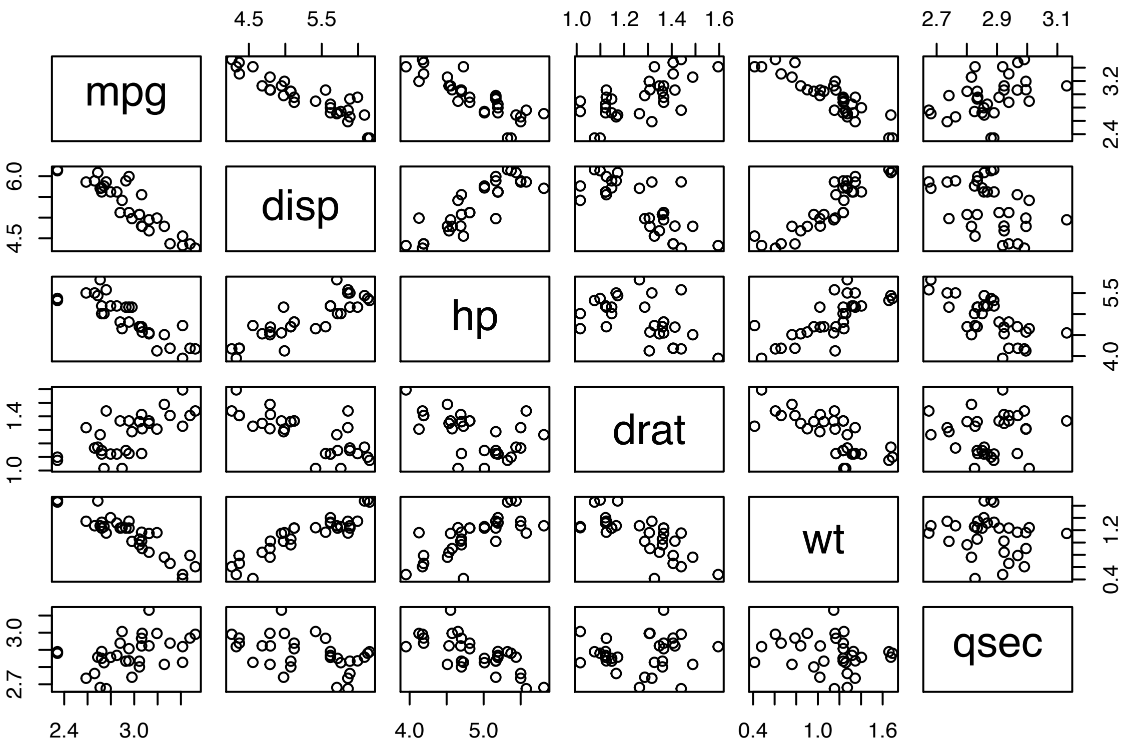
Since $\mathbf{H}_2$ is an idempotent and symmetric matrix, this implied its eigenvalue are either 0 or 1, hence it is positive semi-define, implies $\mathbf{y}^T\mathbf{H}_2\mathbf{y} \geq 0$. Hence $SS_{Res_{\gamma_1}} \geq SS_{Res}$.

Question 4:

```
#Loading and Scaling Data
data(mtcars)
mtcars.new = log(mtcars[, c(1,3:7)])
```

a. Plot the data and Comment

```
#Plotting Pair Graph
pairs(mtcars.new)
```



Form the pairs plots we can see there is a negative linear relationship between **mpg** and **disp**; **mpg** and **hp**; **mpg** and **wt** with all of thems having a small additive error

Between **mpg** and **drat** there is a positive linear relationship; however there seem to be a big error additive error. Similarly, the positive linear relationship also exist between **mpg** and **qsec** but with a multiplicative error instead of an additive one like the other explanatory variables.

For **disp** it is positively correlated with **hp** and **wt** but is negatively correlated with wt. There is a linear relationship between weight and gross horse power.

b. Preform using forward Selection

```
#Preforming forward selection of mtcars model
basemodel <- lm(mpg~1, data=mtcars.new)
add1(basemodel, scope = ~.+disp+hp+drat+wt+qsec, test="F")
```

```
## Single term additions
##
## Model:
## mpg ~ 1
##        Df Sum of Sq     RSS      AIC  F value    Pr(>F)
## <none>              2.74874  -76.547
## disp    1   2.25596 0.49277 -129.550 137.3427 1.006e-12 ***
```

```
## hp      1   1.96733 0.78140 -114.797  75.5310 1.080e-09 ***
## drat    1   1.23131 1.51742  -93.559  24.3435 2.807e-05 ***
## wt      1   2.21452 0.53422 -126.966 124.3596 3.406e-12 ***
## qsec    1   0.47755 2.27119  -80.654   6.3079   0.01763 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
q4model2 <- lm(mpg ~ disp, data=mtcars.new)
add1(q4model2, scope = ~.+hp+drat+wt+qsec, test="F")
```

```
## Single term additions
##
## Model:
## mpg ~ disp
##        Df Sum of Sq     RSS     AIC F value  Pr(>F)
## <none>              0.49277 -129.55
## hp      1  0.045531 0.44724 -130.65  2.9523 0.09641 .
## drat    1  0.001383 0.49139 -127.64  0.0816 0.77711
## wt      1  0.098796 0.39398 -134.71  7.2722 0.01154 *
## qsec    1  0.000308 0.49247 -127.57  0.0181 0.89382
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
q4model3 <- lm(mpg~disp+wt, data=mtcars.new)
add1(q4model3, scope = ~.+hp+drat+qsec, test="F")
```

```
## Single term additions
##
## Model:
## mpg ~ disp + wt
##        Df Sum of Sq     RSS     AIC F value  Pr(>F)
## <none>              0.39398 -134.71
## hp      1  0.078605 0.31537 -139.83  6.9789 0.01334 *
## drat    1  0.007358 0.38662 -133.31  0.5329 0.47146
## qsec    1  0.057788 0.33619 -137.79  4.8130 0.03671 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
q4model4 <- lm(mpg~disp+hp+wt, data=mtcars.new)
add1(q4model4, scope = ~.+drat+qsec, test="F")
```

```
## Single term additions
##
## Model:
## mpg ~ disp + hp + wt
##        Df Sum of Sq     RSS     AIC F value Pr(>F)
## <none>              0.31537 -139.83
## drat    1 0.0000095 0.31536 -137.83  0.0008 0.9774
## qsec    1 0.0033067 0.31206 -138.17  0.2861 0.5971
```

```
summary(q4model4)
```

```
##
## Call:
## lm(formula = mpg ~ disp + hp + wt, data = mtcars.new)
##
## Residuals:
```

```
##       Min        1Q    Median        3Q       Max 
## -0.196932 -0.086109  0.005329  0.073336  0.220450 
## 
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept)  4.94620    0.26867  18.410  < 2e-16 ***
## disp        -0.07792    0.10152  -0.768  0.44919    
## hp          -0.21299    0.08063  -2.642  0.01334 *  
## wt          -0.47880    0.13993  -3.422  0.00193 ** 
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 0.1061 on 28 degrees of freedom
## Multiple R-squared:  0.8853, Adjusted R-squared:  0.873 
## F-statistic: 72.01 on 3 and 28 DF,  p-value: 2.805e-13
```

Model 4 is the optimal model using forward selection as **drat** and **qsec** no longer have any significant after adding **disp**, **hp** and **wt**.

c. Starting from the full model, perform model selection using stepwise selection with AIC}

```
AICbasemodel <- lm(mpg ~ disp+hp+drat+wt+qsec ,data=mtcars)
q4modelAIC <- step(AICbasemodel, scope = ~., steps=4)
```

```
## Start:  AIC=65.47
## mpg ~ disp + hp + drat + wt + qsec
## 
##        Df Sum of Sq    RSS    AIC
## - disp  1     3.974 174.10 64.205
## <none>              170.13 65.466
## - hp    1    11.886 182.01 65.627
## - qsec  1    12.708 182.84 65.772
## - drat  1    15.506 185.63 66.258
## - wt    1    81.394 251.52 75.978
## 
## Step:  AIC=64.21
## mpg ~ hp + drat + wt + qsec
## 
##        Df Sum of Sq    RSS    AIC
## - hp    1     9.418 183.52 63.891
## - qsec  1     9.578 183.68 63.919
## <none>              174.10 64.205
## - drat  1    11.956 186.06 64.331
## + disp  1     3.974 170.13 65.466
## - wt    1   113.882 287.99 78.310
## 
## Step:  AIC=63.89
## mpg ~ drat + wt + qsec
## 
##        Df Sum of Sq    RSS    AIC
## <none>              183.52 63.891
## - drat  1    11.942 195.46 63.908
## + hp    1     9.418 174.10 64.205
## + disp  1     1.506 182.02 65.627
```

```
## - qsec  1    85.720 269.24 74.156
## - wt    1   275.686 459.21 91.241
```

The best model for AIC was achieved after 4 steps; doing nothing allow us to have the lowest possible value for $AIC = -141.16$

d. Write down the final fitted model from stepwise selection.

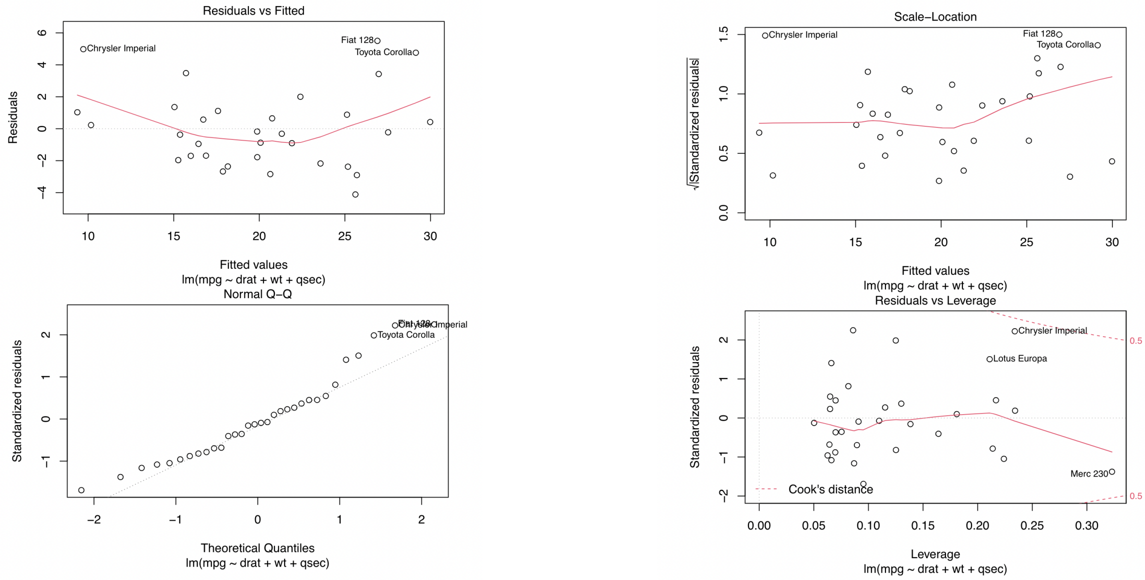
```
summary(q4modelAIC)
```

```
##
## Call:
## lm(formula = mpg ~ drat + wt + qsec, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.1152 -1.8273 -0.2696  1.0502  5.5010
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  11.3945     8.0689   1.412  0.16892
## drat          1.6561     1.2269   1.350  0.18789
## wt           -4.3978     0.6781  -6.485 5.01e-07 ***
## qsec          0.9462     0.2616   3.616  0.00116 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.56 on 28 degrees of freedom
## Multiple R-squared:  0.837,  Adjusted R-squared:  0.8196
## F-statistic: 47.93 on 3 and 28 DF,  p-value: 3.723e-11
```

The model final fitted model from stepwise using AIC as a goodness of fit is given by:

$$\begin{array}{llll} mpg = & 4.83469 & -0.25532hp & -0.56228wt \\ & (0.22440) & (0.05840) & (0.08742) \end{array}$$

e. Produce diagnostic plots for your final model from stepwise selection and comment.



The residuals vs fitted graph trend line is around 0. However, the two end which trend downward(upward) at the start points(end points). The start points does not seem to be a problem as indicated by the scale-location graph. However for the end point there seem to be a increasing variance which indicate we may not have a constant variances.

From the QQplots, also reflect this as while the starting point not be perfectly normal they are close. On the other hand, the end points is not normally distributed .
The residuals vs leverage graph, nothing really stand out, as point that may considered to be troublesome like Chrysler Imperial and Merc 230 are still with 0.5 crooks distance.

Question 5: Ridge Regression

a. Show the estimator:

Since $\sum_{i=1}^{n} e_i^2 = (y - X\beta)^T(y - X\beta)$ and $\sum_{i=1}^{n} b_i^2 = \beta^T\beta$. The parameter above is give by:

$$\begin{aligned} L(\beta) &= (y - X\beta)^T(y - X\beta) + \lambda(\beta^T\beta) \\ &= y^Ty - 2y^TX\beta + \beta^T(X^TX)\beta + \lambda(\beta^T\beta) \end{aligned}$$

To find the maximum likelihood estimator we find $\frac{\partial L}{\partial \beta}$ setting it to 0 then solve for β.

$$\begin{aligned} &\frac{\partial L}{\partial \beta} = -2(X^Ty) + 2(X^TX)\beta + 2\lambda\beta = 0 \\ &(X^TX + \lambda I)\beta = X^Ty \\ &\beta = (X^TX + \lambda I)^{-1}X^Ty \end{aligned}$$

Hence the estimator is give by $b = (X^TX + \lambda I)^{-1}X^Ty$.

b. Show that b is unbiased if $\lambda \neq 0$.

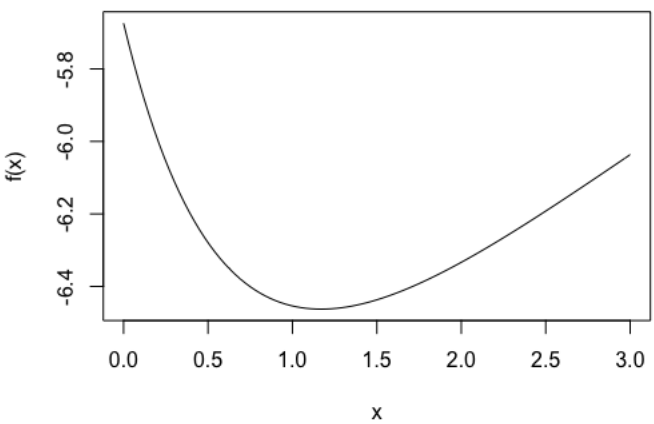
$$\begin{aligned} E(b) &= E(X^TX + \lambda I)^{-1}X^Ty) \\ &= (X^TX)^{-1}X^TE(y) + (\lambda I)^{-1}X^TE(y) \\ &= (X^TX)^{-1}X^TXb + (\lambda I)^{-1}X^TXb && \text{Since E(y)=Xb} \\ &= b + (\lambda I)^{-1}X^TXb \end{aligned}$$

Since $(\lambda I)^{-1}X^TXb = 0$ if and only if $\lambda = 0$ then $E(b) \neq b$ if $\lambda \neq 0$ hence it is biased when $\lambda \neq 0$.

c. Optimal λ value.

```
f1 <-function(lambda){
  #Calling Data Frame
  q2data <- data.matrix(read.csv(file = "Q2_Data.csv"))
  q2frame <- read.csv(file="Q2_Data.csv")
  #Converting Data frame to matrix
  y <- matrix(q2data[,1],7,1)
  x <- matrix(c(rep(1,7),q2data[,-1]),7,4)
  #Scaling data
  x <- scale(x[,-1],center=T,scale=T)
  y <- scale(y,center=T,scale=T)
  p <- 3
  #Definting Degree of freedom
  lambda = matrix(c(lambda,0,0,0,lambda,0,0,0,lambda),3,3)
  H <- x%*%solve(t(x)%*%x+lambda)%*%t(x)
  df <- sum(diag(H))
  #Computing Beta
  b <- solve((t(x)%*%x)+lambda)%*%t(x)%*%y
  #Computing Sum-squared
  e <- (y-x%*%b)
  SSres <- sum(e^2)
  n <- dim(y)[1]
  gof <- n*log(SSres/n)+2*df
  return(gof)
}
```

```
> f <- Vectorize(f1); curve(f, 0 ,3)
> f(1.17)<f(1.171)
[1] TRUE
> f(1.17)<f(1.169)
[1] TRUE
```



Hence, the optimal value for λ is approximately 1.17.