THE UNIVERSITY OF MELBOURNE

Student number 1067799

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.

©University of Melbourne 2021

Can be placed in Baillieu Library

Question 1 (4 marks)

Prove Theorem 4.8: show that the maximum likelihood estimator of the error variance σ^2 is

$$\hat{\sigma}^2 = \frac{SS_{Res}}{n}.$$

Question 2 (11 marks)

We wish to predict the price of apartments in Melbourne using some of their features. Let y be the apartment price per square metre, x_1 be the apartment age (in years), x_2 be the distance (in metres) to the nearest train station, and x_3 be the number of convenience stores nearby. The following data is collected:

age x_1	dist x_2	#stores x_3	y $(\times 10^3)$
32	84.9	10	37.9
19.5	306.6	9	42.2
13.3	562.0	5	47.3
13.3	562.0	5	43.1
5	390.6	5	54.8
7.1	2175.0	3	47.1
34.5	623.5	7	40.3

For this question, you may NOT use the `lm` function in R.

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

(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.

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

(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.

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

Question 3 (5 marks)

Consider two full rank linear models $\mathbf{y} = X_1\boldsymbol{\gamma}_1 + \boldsymbol{\varepsilon}_1$ and $\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}_2$, where all predictors in the first model ($\boldsymbol{\gamma}_1$) are also contained in the second model ($\boldsymbol{\beta}$). Show that the SS_{Res} for the first model is at least the SS_{Res} for the second model.

Q1

$$y = X\beta + \varepsilon \sim MVN(X\beta, \sigma^2 I)$$

$$SS_{res} = \sum_{i=1}^{N} \varepsilon_i^2$$

the log likelihood is

$$l(\beta, \sigma^2 | y) = -\frac{N}{2}\log(2\pi) - \frac{N}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}|y - X\beta|^T |y - X\beta|$$

$$\frac{\partial l(\beta, \sigma^2 | y)}{\partial \sigma^2} = \frac{1}{2}|y - X\beta|^T |y - X\beta| \cdot \sigma^{-4} - \frac{N}{2}\sigma^{-2} = 0$$

$$\Rightarrow \frac{1}{2}\sigma^{-2}\left(|y - X\beta|^T |y - X\beta| \sigma^{-2} - N\right) = 0$$

$$\hat{\sigma}^2 = \frac{1}{N}|y - X\beta|^T |y - X\beta|$$

As $y - X\beta$ is the residual vector ε.

$$|y - X\beta|^T |y - X\beta| = \varepsilon^T \varepsilon = SS_{res}$$

Thus, $\hat{\sigma}^2 = \frac{1}{N} SS_{res}$.

Q2

assign2_lsm

Yuxin Ma

27/04/2021

#Q2 #(a)

```
y<- c(37.9, 42.2, 47.3, 43.1, 54.8, 47.1, 40.3)
x <- matrix(c(rep(1,7), 32, 19.5, 13.3, 13.3, 5, 7.1, 34.5, 84.9, 306.6, 562.0,
              562.0, 390.6, 2175.0, 623.5, 10, 9, 5, 5, 5, 3, 7), 7, 4)
(b <- solve(t(x)%*%x, t(x)%*%y))
```

```
##              [,1]
## [1,] 58.369312708
## [2,] -0.346291960
## [3,] -0.002900359
## [4,] -0.887671692
```

```
#estimate variance
e <- y - x%*%b
SSRes <- sum(e^2)
(s2 <- SSRes/(7-4))
```

```
## [1] 13.06871
```

#(b)

```
s <- sqrt(s2)
xst <- c(1, 10, 100, 6)
ta <- qt(0.95, df = 7-4)
(xst%*%b - ta*s*sqrt(t(xst)%*%solve(t(x)%*%x)%*%xst))
```

```
##          [,1]
## [1,] 43.27252
```

```
(xst%*%b + ta*s*sqrt(t(xst)%*%solve(t(x)%*%x)%*%xst))
```

```
##          [,1]
## [1,] 55.30814
```

```
# the 90% CI is (43.27252, 55.30814)
```

#(c)

```r
# se(ß1 - ß3) = sqrt(var(ß1 - ß3)) = sqrt(var(b1 - b3))
# matrix of variance b
varb <- solve(t(x)%*%x)*s2
varb1b3 <- varb[2, 2] + varb[4, 4] - 2*varb[2, 4]
(seß1ß3 <- sqrt(varb1b3))
```

```
## [1] 1.388968
```

```r
#the standard error for ß1 - ß3 is 1.388968
```

#(d)

```r
#H0 : ß1 = -1
c <- matrix(c(0, 1, 0, 0), 1, 4)
dst <- -1
r <- 1
num <- t(c%*%b-dst)%*%solve(c%*%solve(t(x)%*%x)%*%t(c))%*%(c%*%b-dst)/r
Fstat<- num/(SSRes/(7-4))
(pf(Fstat, r, 7-4, lower = F))
```

```
##             [,1]
## [1,] 0.04945829
```

```r
#Since p-value = 0.0494 < 0.05, we can reject H0, that there's no evidence to show that price per squar
```

#(e)

```r
# H0: model irrelevance with alpha = 5%
SSReg <- t(y)%*%x%*%b - 1/7*sum(y)^2
MSReg <- SSReg/(4-1)
SSRes <- t(y)%*%y - t(y)%*%x%*%b
MSRes <- SSRes/(7-3-1)
Fstat <- MSReg/MSRes
(pf(Fstat,3,7-4,lower.tail=FALSE))
```

```
##           [,1]
## [1,] 0.1500833
```

```r
# We can NOT reject H0 as p-value = 0.15 > 0.05
```

```r
knitr::opts_chunk$set(echo = TRUE)
```

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

Question 3 (5 marks)

Consider two full rank linear models $\mathbf{y} = X_1\gamma_1 + \varepsilon_1$ and $\mathbf{y} = X\beta + \varepsilon_2$, where all predictors in the first model (γ_1) are also contained in the second model (β). Show that the SS_{Res} for the first model is at least the SS_{Res} for the second model.

As γ_1 is a reduced model of β, and β can be interpreted as $\begin{bmatrix} \gamma_1 \\ \gamma_2 \end{bmatrix}$, the order is not important.

γ_1 is the reduced model of β that the 1st model can be written as $y = X\begin{bmatrix} \gamma_1 \\ 0 \end{bmatrix} + \varepsilon_1$

As β is the best linear unbiased estimator. it minimise the error term, which also minimise the residual sum of squares, which is:

$$SSres_2 = (y - X\beta)^T (y - X\beta) = \varepsilon_2^2$$

$$SSres_1 = (y - X_1\gamma_1)^T (y - X_1\gamma_1) = \varepsilon_1^2$$

Thus, ε_2^2 has to be smaller or equal to ε_1^2

Since γ_1 is an estimator that assumed to be not as good as β.

Q4

lsm_ass2Q4

Yuxin Ma

28/04/2021

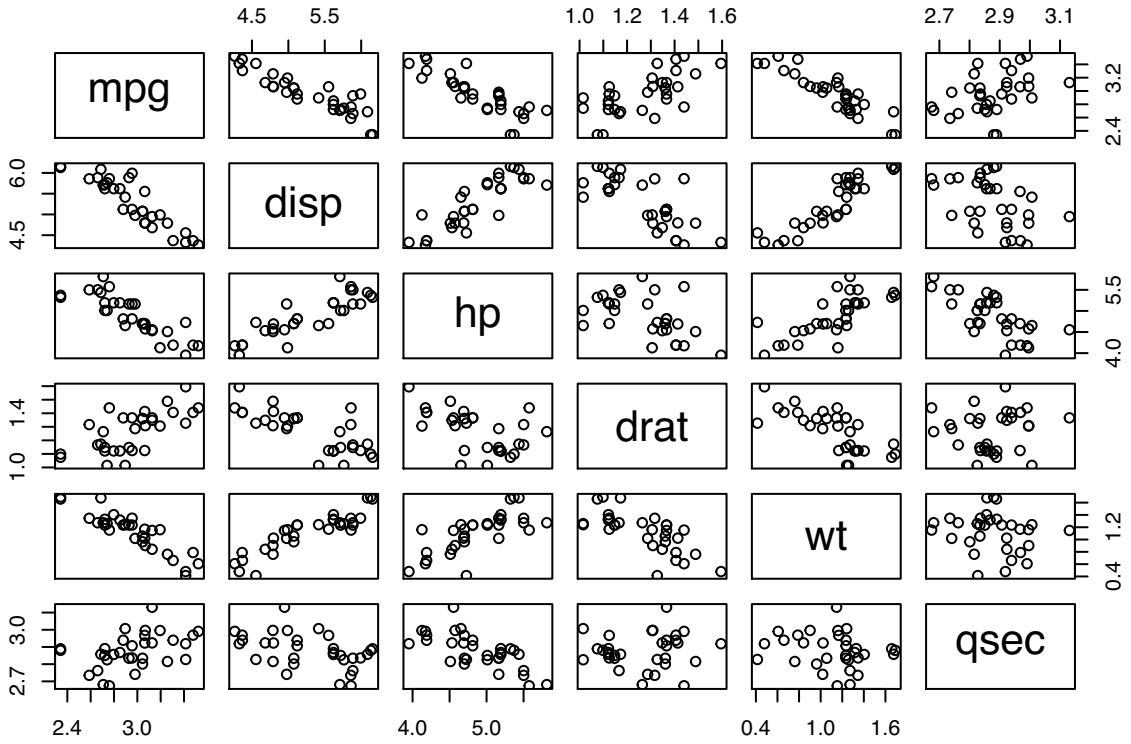
R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

#(a)

```
data(mtcars)
mtcars <- log(mtcars[, c(1, 3:7)])
y <- mtcars$mpg
disp <- mtcars$disp
hp <- mtcars$hp
drat <- mtcars$drat
wt <- mtcars$wt
qsec <- mtcars$qsec
pairs(mtcars)
```



#comment: the plots shows relationship between log-x-variables and response variable mpg. From the graphs, disp, #hp and wt present linear relationship with mpg whereas drat and qsec has no obvious linear relationship with mpg.

#(b)

```
basemodel <- lm(y~1, data = mtcars)
add1(basemodel, scope = ~.+disp+hp+drat+wt+qsec, test = "F")
```

```
## Single term additions
##
## Model:
## y ~ 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
```

```
#disp is added to model2 as it has the largest F-value and has p-value < 0.05
model2 <- lm(y~disp, data = mtcars)
add1(model2, scope = ~.+hp+drat+wt+qsec, test = "F")
```

```
## Single term additions
##
## Model:
## y ~ 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
```

```
# wt is added as it has the largest F-value among the remaining variables and has p-value < 0.05
model3 <- lm(y~disp+wt, data = mtcars)
add1(model3, scope = ~.+hp+drat+qsec, test = "F")
```

```
## Single term additions
##
## Model:
## y ~ 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
```

```
#hp is added for the same reason as above
model4 <- lm(y~disp+wt+hp, data = mtcars)
add1(model4, scope = ~.+drat+qsec, test = "F")
```

```
## Single term additions
##
## Model:
## y ~ disp + wt + hp
##        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
```

#no more variable will be added as they both have p-value>0.05 #the final model is y~disp+wt+hp

#(c)

```
fullmodel <- lm(y~disp+hp+drat+wt+qsec, data = mtcars)
model2 <- step(fullmodel, scope = ~., steps = 1)
```

```
## Start:  AIC=-136.21
## y ~ disp + hp + drat + wt + qsec
##
```

```
##        Df Sum of Sq     RSS     AIC
## - drat  1  0.000402 0.31207 -138.17
## - disp  1  0.002104 0.31377 -138.00
## - qsec  1  0.003699 0.31536 -137.83
## <none>              0.31166 -136.21
## - hp    1  0.023697 0.33536 -135.87
## - wt    1  0.103076 0.41474 -129.07
##
## Step:  AIC=-138.17
## y ~ disp + hp + wt + qsec
```

```
step(model2, scope=~.+drat)
```

```
## Start:  AIC=-138.17
## y ~ disp + hp + wt + qsec
##
##        Df Sum of Sq     RSS     AIC
## - qsec  1  0.003307 0.31537 -139.83
## - disp  1  0.004372 0.31644 -139.72
## <none>              0.31207 -138.17
## - hp    1  0.024123 0.33619 -137.79
## + drat  1  0.000402 0.31166 -136.21
## - wt    1  0.103779 0.41584 -130.98
##
## Step:  AIC=-139.83
## y ~ disp + hp + wt
##
##        Df Sum of Sq     RSS     AIC
## - disp  1  0.006635 0.32201 -141.16
## <none>              0.31537 -139.83
## + qsec  1  0.003307 0.31207 -138.17
## + drat  1  0.000010 0.31536 -137.83
## - hp    1  0.078605 0.39398 -134.71
## - wt    1  0.131870 0.44724 -130.65
##
## Step:  AIC=-141.17
## y ~ hp + wt
##
##        Df Sum of Sq     RSS     AIC
## <none>              0.32201 -141.16
## + disp  1   0.00664 0.31537 -139.83
## + qsec  1   0.00557 0.31644 -139.72
## + drat  1   0.00112 0.32089 -139.28
## - hp    1   0.21221 0.53422 -126.97
## - wt    1   0.45939 0.78140 -114.80
```

```
##
## Call:
## lm(formula = y ~ hp + wt, data = mtcars)
##
## Coefficients:
## (Intercept)           hp           wt
##      4.8347      -0.2553      -0.5623
```

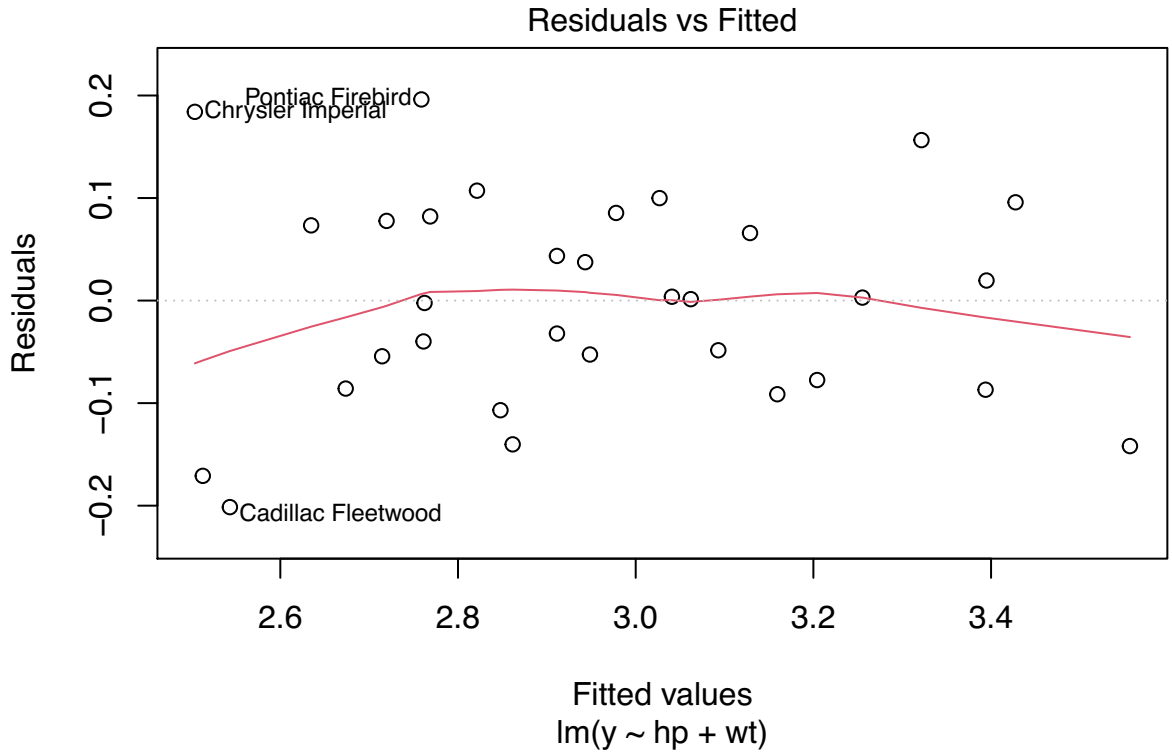
At last we have hp and wt included in the final linear model with stepwise function, starting from full model.

#(d)

```
# with the coefficient listed above, we have:
#mpg = 4.8347 - 0.2553*hp - 0.5623*wt
```

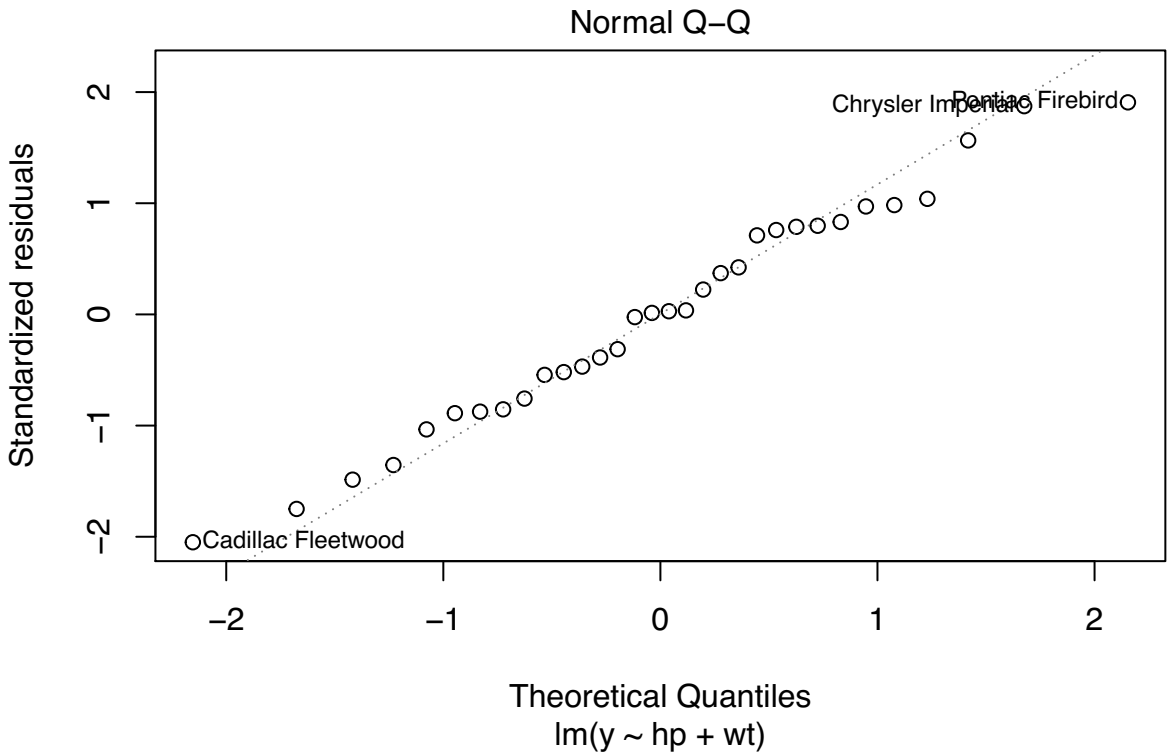
#(e)

```
finalmodel <- lm(y ~ hp + wt, data = mtcars)
plot(finalmodel, which = 1)
```



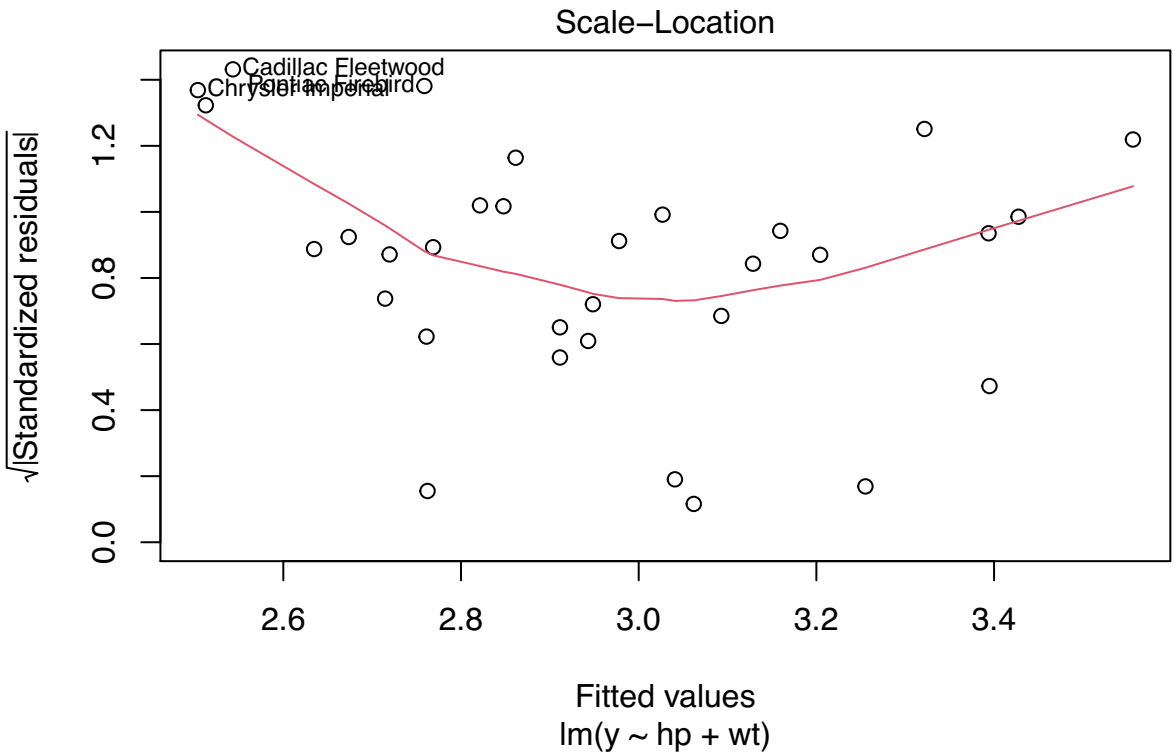
We can observe equally spread residuals around a horizontal line without distinct patterns and even outliers # have small residuals. The model is fitted well.

```
plot(finalmodel, which=2)
```



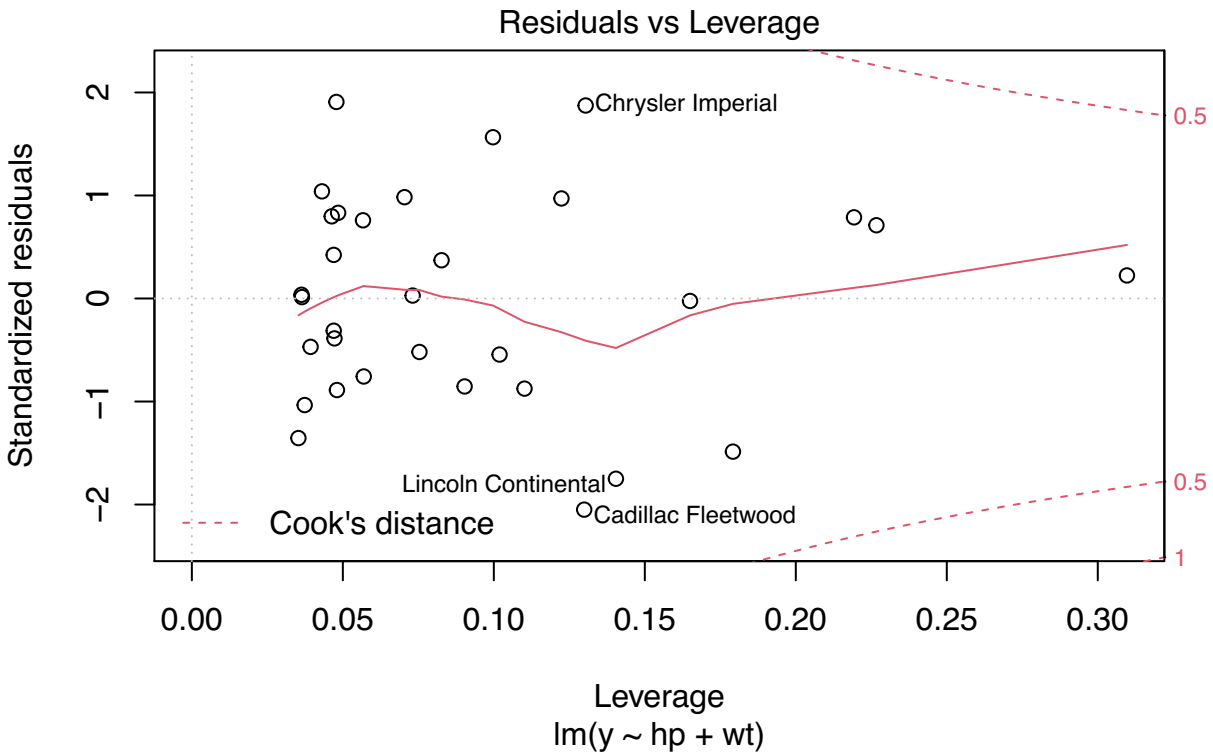
The points follow the dashed straight line with a small number of outliers. Thus it's a good fit.

```
plot(finalmodel, which=3)
```



This plot shows if residuals are spread equally along the ranges of predictors. # We can observe a generally horizontal line with equally randomly spread points with low residuals. # Thus, it is generally a good fit.

```
plot(finalmodel, which = 5)
```



the residuals, leverage and cook's distance are generally small. # Accorrding to the diagnostic plots, the final model computed from stepwise selection is acceptable.

Question 5 (10 marks)

For ridge regression, we choose parameter estimators $\mathbf{b}$ which minimise

$$\sum_{i=1}^{n} e_i^2 + \lambda \sum_{j=0}^{k} b_j^2,$$

where λ is a constant penalty parameter.

(a) Show that these estimators are given by

$$\mathbf{b} = (X^T X + \lambda I)^{-1} X^T \mathbf{y}.$$

(b) Show that $\mathbf{b}$ is biased if $\lambda \neq 0$.

(c) One way to calculate the optimal value for the penalty parameter is to minimise the AIC. Since the number of parameters p does not change, we use a slightly modified version:

$$AIC = n \ln \frac{SS_{Res}}{n} + 2\, df,$$

where df is the "effective degrees of freedom" defined by

$$df = tr(H) = tr(X(X^T X + \lambda I)^{-1} X^T).$$

We will use the data from Q2. In order to avoid penalising some parameters unfairly, we must first standardise the variables; this also means an intercept parameter is not used. You can do this with `scale`:

```
> X <- scale(X[,-1],center=T,scale=T)
> y <- scale(y,center=T,scale=T)
> p <- 3
```

Construct a plot of λ against AIC. Thereby find the optimal value for λ.

End of Assignment — Total Available Marks = 40

a) let $w = \sum_{i=1}^{N} e_i^2 + \lambda \sum_{j=0}^{k} b_j^2$

$= (y - Xb)^T (y - Xb) + \lambda b^T b$

$= (y^T - b^T X^T)(y - Xb) + \lambda b^T b$

$= \underline{y^T y - 2y^T Xb + b^T X^T X b + \lambda I b^T b}$

$\frac{\partial w}{\partial b} = 0 - 2y^T X + 2X^T Xb + 2\lambda Ib$

$= -2X^T y + 2X^T Xb + 2\lambda Ib$

$= 0$

According to handout 4 P 90.

$(2X^T X + 2\lambda I) b = 2X^T y$

$b = (X^T X + \lambda I)^{-1} X^T y$

b) $b = (X^TX + \lambda I)^{-1} X^Ty$

$E(b) = E[(X^TX + \lambda I)^{-1} X^Ty]$

$= E[(X^TX + \lambda I)^{-1} X^TX\beta]$

if b is unbiased, $E(b) = \beta$.

It is shown that $(X^TX + 0)^{-1}(X^TX\beta) = \beta$;

which means if and only if $\lambda I = 0$, the estimator is unbiased.

Thus, if $\lambda \neq 0$, b is a biased estimator.

lsm_ass2_Q5

Yuxin Ma

28/04/2021

R Markdown

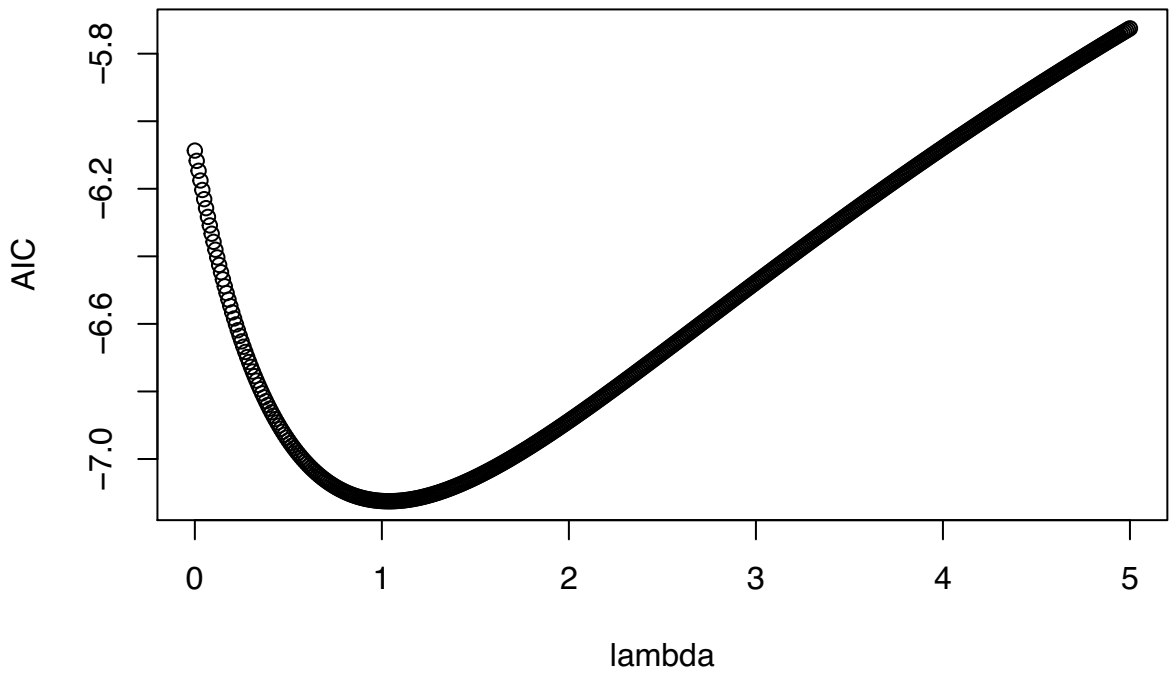
This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

(c)

```
y<- c(37.9, 42.2, 47.3, 43.1, 54.8, 47.1, 40.3)
X <- matrix(c(rep(1,7), 32, 19.5, 13.3, 13.3, 5, 7.1, 34.5, 84.9, 306.6, 562.0,
              562.0, 390.6, 2175.0, 623.5, 10, 9, 5, 5, 5, 3, 7), 7, 4)
X <- scale(X[,-1],center=T,scale=T)
y <- scale(y,center=T,scale=T)
p <- 3

n <- length(y)
```

```
# assume lambda are all positive.
lambda <- seq(0, 5, 0.01)
AIC <- replicate(length(lambda), 0)

for (i in 1:length(AIC)){
  b <- solve(t(X)%*%X+diag(lambda[i],p), t(X)%*%y)
  e <- y - X%*%b
  SSRes <- sum(e^2)
  H <- X%*%solve(t(X)%*%X+diag(lambda[i],p))%*%t(X)
  df <- sum(diag(H))
  AIC[i] <- n*log(SSRes/n) + 2*df

}

plot(lambda, AIC)
```




```
#lambda with the smallest AIC
lambda[which(AIC == min(AIC))]
```

```
## [1] 1.04
```