THE UNIVERSITY OF MELBOURNE

Student number 1067799

Semester 1 Assessment, 2021

School of Mathematics and Statistics

MAST30025 Linear Statistical Models Assignment 3

Submission deadline: **Friday May 28, 5pm**

This assignment consists of 3 pages (including this page)

Instructions to Students

Writing

- There are 5 questions with marks as shown. The total number of marks available is 48.
- 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.

a)

According to defination 6.1, $A = AA^cA$ when A^c is the conditional inverse of A.

$$\Rightarrow r(A) = r(AA^cA)$$

Also according to rank property

$r(A^cA) \leq r(A^c)$ and $r(A)$ ①

$r(AA^cA) \leq r(A)$ and $r(A^cA)$ ②

We can have $r(A^cA) \leq rA$ from ①

and $r(A) \leq r(A^cA)$ from ②

Therefore $r(A^cA) = r(A)$

b) $A^cA \cdot A^cA = A^c \cdot (AA^cA) = A^cA$ (Also according to defination 6.1)

It is idempotent.

c) $A(A^TA)^cA^T$ is unique.

As $A = A(A^TA)^c(A^TA)$, from conditional inverse properties.

$$A(A^TA)^cA^T = A(A^TA^c)(A^TA)\cdot A^{-1} = A\cdot A^{-1} = I$$

The result is always I, therefore it's unique

lsm3_Q2

Yuxin Ma

25/05/2021

#(a)

```
y <- c(22, 23, 24, 22, 26, 16, 18, 19, 28, 27, 29, 29)
X <- matrix(c(rep(1, 12), rep(0, 36)), 12,4)
X[1:5, 2] <- 1
X[6:8, 3] <- 1
X[9:12, 4] <- 1
XtX <- t(X)%*%X
XtXc <- matrix(0,4,4)
XtXc[2:4, 2:4] <- t(solve(XtX[2:4, 2:4]))
XtXc <- t(XtXc)
XtXc
```

```
##      [,1] [,2]      [,3] [,4]
## [1,]    0  0.0 0.0000000 0.00
## [2,]    0  0.2 0.0000000 0.00
## [3,]    0  0.0 0.3333333 0.00
## [4,]    0  0.0 0.0000000 0.25
```

#(b)

```
library(MASS)
b <- ginv(XtX)%*%t(X)%*%y
e <- y - X%*%b
SSRes <- sum(e^2)
n <- length(y)
(s2 <- SSRes/(n-3))
```

```
## [1] 2.068519
```

(b). s2 = 2.068519

#(c)

```
C <- matrix(c(1, 2, 1, 0), 1, 4)
C%*%XtXc%*%XtX
```

```
##      [,1] [,2] [,3] [,4]
## [1,]    3    2    1    0
```

(c). It is not estimable since C%%*XtXc*%%XtX is not equal to C

#(d)

```
r <- 3
tt <- c(1, 0, 1, 0)
ta <- qt(0.95,df=n-r)
hw <- ta*sqrt(s2*t(tt)%*%XtXc%*%tt)
tt%*%b - hw #lower tail
```

```
##          [,1]
## [1,] 16.14451
```

```
tt%*%b + hw #upper tail
```

```
##          [,1]
## [1,] 19.18882
```

(d). 90% CI for the lifetime of the 2nd typee of bulb is (16.14451, 19.18882)

#(e)

```
#The hypothesis can be written as H0: Cß = 0, where C = [0 1 0 -1]
C <- matrix(c(0, 1, 0, -1), 1, 4)
Fstat <- t(C%*%b) %*% solve(C%*%XtXc%*%t(C)) %*% C%*%b / s2
pf(Fstat,1,n-r,lower.tail=FALSE)
```

```
##              [,1]
## [1,] 0.0007123037
```

(e). As p-value = 0.0007123037 < 0.025, we can reject H0, that there is difference in life time between 1st and 3rd types of bulb

Q3

useful.

rank property:

① $\text{rank}(XY) \le \text{rank}(X), \text{rank}(Y)$

② $r(X) = r(X^T) = r(X^TX)$

To prove $t^T\beta$ is estimable, we need to solve for z in $X^TXz = t$.

$$\left[\begin{array}{c} X_1 \\ X_2 \end{array}\right]\left[X_1 \mid X_2\right] \cdot z = \left[\begin{array}{c} t_1 \\ t_2 \end{array}\right]$$

$$\left[\begin{array}{c} X_1^TX_1z_1 \\ X_2^TX_2z_2 \end{array}\right] = \left[\begin{array}{c} t_1 \\ t_2 \end{array}\right]$$

The only way $X^TXz = t$ is consistent is that

$\text{rank}(X^TX \mid t) = \text{rank}(X^TX)$. according to theorem 6.3

proof:

$$\text{rank}\left(\left[X^TX \mid t\right]\right) = \text{rank}\left(\left[X^TX \mid \begin{array}{c} X_1^TX_1z_1 \\ X_2^TX_2z_2 \end{array}\right]\right)$$

$$= \text{rank}\left(\left[\begin{array}{cc} X_1^T & 0 \\ 0 & X_2^T \end{array}\right]\left[\begin{array}{cc|c} X_1 & X_2 & X_1z_1 \\ X_1 & X_2 & X_2z_2 \end{array}\right]\right)$$

$$\le \text{rank}\left(\left[\begin{array}{cc} X_1^T & 0 \\ 0 & X_2^T \end{array}\right]\right) \quad ①$$

$$= \text{rank}(X_1^T) + \text{rank}(X_2^T)$$

$$= \text{rank}(X_1) + \text{rank}(X_2) \quad ②$$

$$= \text{rank}(X)$$

$$= \text{rank}(X^TX) \quad ②$$

From above that: $\text{rank}(X^TX \mid t) \le \text{rank}(X^TX)$

and it is obvious that

$$\text{rank}(X^TX \mid t) \ge \text{rank}(X^TX)$$

Therefore $\text{rank}(X^TX \mid t) = \text{rank}(X^TX)$

$X^TXz = t$ is consistent.

$t^T\beta$ is therefore estimable.

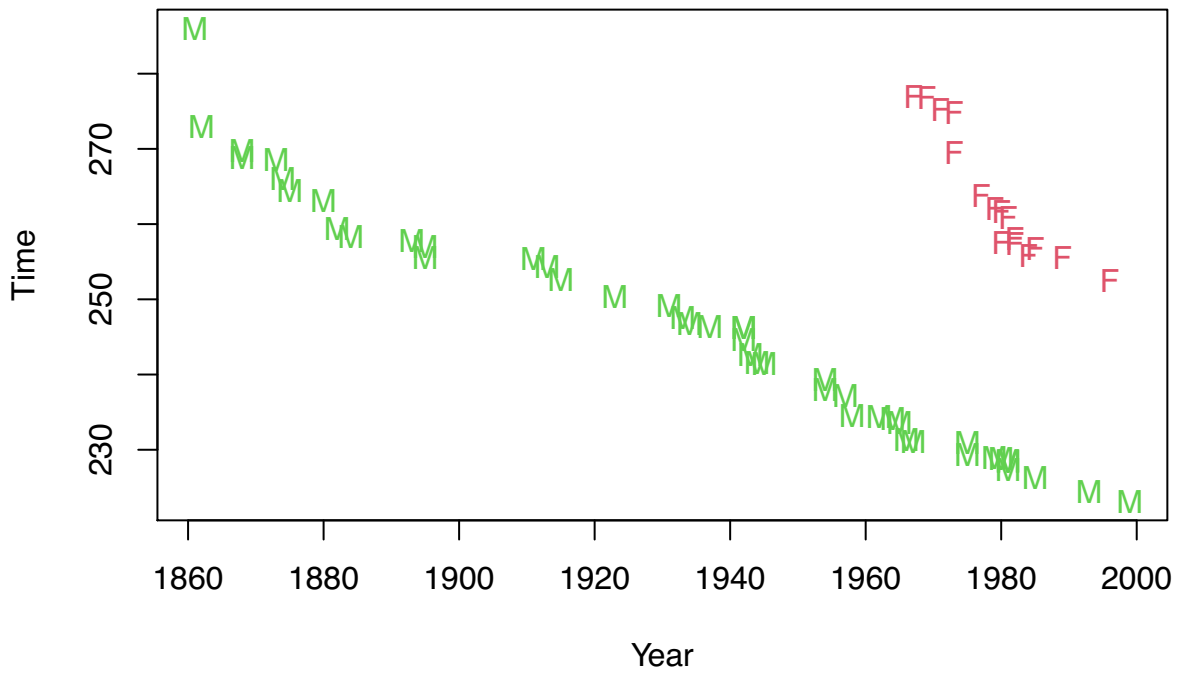
lsm3_Q4

Yuxin Ma

26/05/2021

#(a)

```
mile <- read.csv('mile.csv', header = TRUE)
mile$Gender <- factor(mile$Gender)
plot(Time ~ Year, data = mile, col = as.numeric(Gender)+1, pch = as.character(Gender))
```



(a) This does not satisfy linear model assumptions. Both time of male and female decrease with the increase of year, thus, there's a violation of independence assumption.

#(b)

```
#H0: no interaction between two predictors, significance level 0.05
imodel <- lm(Time ~ Year * Gender, data = mile)
amodel <- lm(Time ~ Year + Gender, data = mile)
anova(amodel, imodel)
```

```
## Analysis of Variance Table
##
## Model 1: Time ~ Year + Gender
## Model 2: Time ~ Year * Gender
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1     59 895.62
## 2     58 518.03  1    377.59 42.276 2.001e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

(b) p-value = 2.001e-08 « 0.025, we can reject H0 that there exists interaction between two predictors.

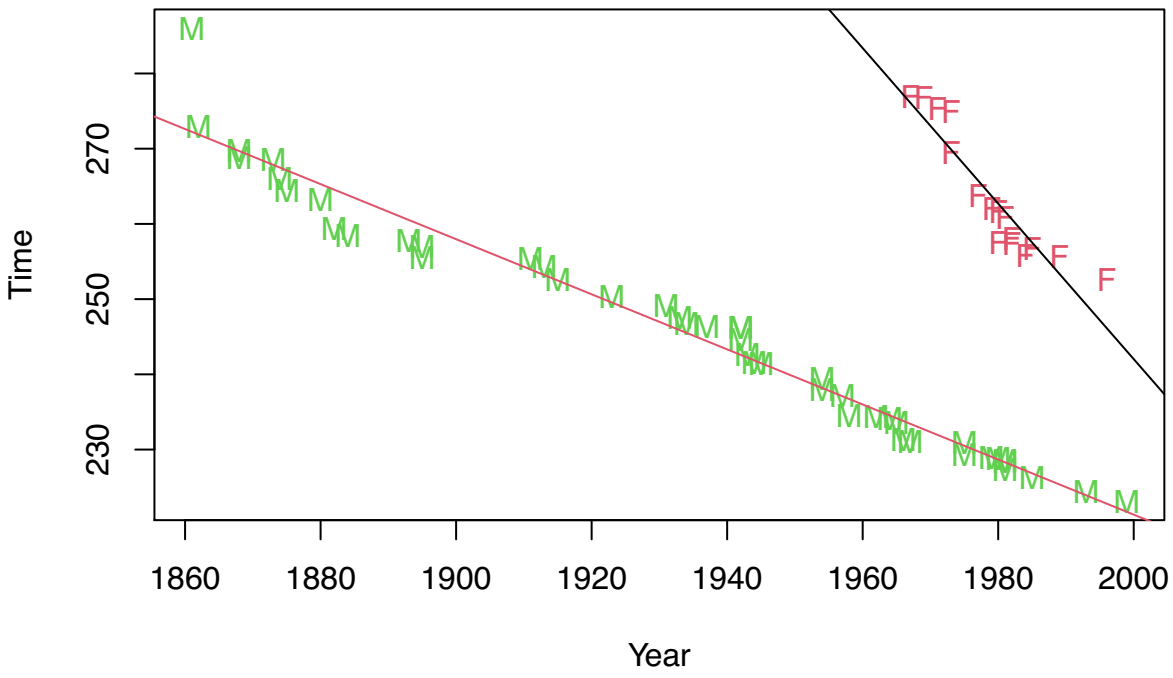
#(c)

```
summary(imodel)
```

```
##
## Call:
## lm(formula = Time ~ Year * Gender, data = mile)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.4512 -1.6160 -0.1137  1.1784 13.7265
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)      2309.4247   202.0583  11.429  < 2e-16 ***
## Year               -1.0337     0.1021 -10.126 1.95e-14 ***
## GenderMale      -1355.6778   203.1441  -6.673 1.03e-08 ***
## Year:GenderMale     0.6675     0.1027   6.502 2.00e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.989 on 58 degrees of freedom
## Multiple R-squared:  0.9663, Adjusted R-squared:  0.9645
## F-statistic: 553.8 on 3 and 58 DF,  p-value: < 2.2e-16
```

```
# from the given data we can conclude that
# female: time = 2309.4247 - 1.0337*Year
# male: time = 2309.4247 - 1355.6778 + (-1.0337 + 0.6675)*year = 953.7469 - 0.3662*Year
```

```
plot(Time ~ Year, data = mile, col = as.numeric(Gender)+1, pch = as.character(Gender)) +
abline(a = 2309.4247, b = -1.0337, col = 1) +
abline(a = 953.7469, b = -0.3662, col = 2)
```




```
## integer(0)
```

(c) female: Time = 2309.4247 - 1.0337 *Year* *male:* *Time* = *953.7469 - 0.3662* Year

#(d)

```
(x <- (2309.4247 - 953.7469)/(1.0337 - 0.3662))
```

```
## [1] 2030.978
```

```
(y <- 953.7469 - 0.3662*x)
```

```
## [1] 210.0028
```

(d) around year 2031, we can exepect a coordinate and the time will be 210.0028. However, this might not be realistic since 2031 is far away from our observation Year range, that there may exist deviation.

(e) Not estimable as it's a ratio of (t1 - t2) and (beta1 - beta2). But as each (t/beta) is estimable, it is still cosistant with (d)

#(f)

```
confint(imodel)[4,]
```

```
##     2.5 %    97.5 %
## 0.4620087 0.8730100
```

(f) 95% CI for the amount by which the gap between the male and female world records narrow every year is (0.4620087, 0.8730100)

#(g)

```
#H0: C%*%b = -0.4
C <- c(0, 1, 0, 1)
library(car)
```

```
## Loading required package: carData
```

```
linearHypothesis(imodel, C, -0.4)
```

```
## Linear hypothesis test
##
## Hypothesis:
## Year  + Year:GenderMale = - 0.4
##
## Model 1: restricted model
## Model 2: Time ~ Year * Gender
##
##   Res.Df    RSS Df Sum of Sq      F   Pr(>F)
## 1     59 604.84
## 2     58 518.03  1    86.806 9.7191 0.002837 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

(g) p-value = 0.002387 < 0.025, we can reject H0 that the male record does not decrease by 0.4 secs each year.

Q5 (a)

As we have 3 treatment and we want to compare the first two treatments' performance over the placebo 3, we have

$$\text{var}\,\hat{\tau_1 - \tau_3} = \sigma^2\left(\frac{1}{n_3} + \frac{1}{n_1}\right)$$

$$\text{var}\,\hat{\tau_2 - \tau_3} = \sigma^2\left(\frac{1}{n_3} + \frac{1}{n_2}\right)$$

Constraint: $5000n_1 + 2000n_2 + 1000n_3 = 100{,}000$

$$5n_1 + 2n_2 + n_3 = 100.$$

So we minimize,

$$f(n_1, n_2, n_3, \lambda) = \frac{1}{\sigma^2}\left(\frac{1}{n_1} + \frac{1}{n_2} + \frac{2}{n_3}\right) + \lambda(5n_1 + 2n_2 + n_3 - 100)$$

$$\frac{\partial f}{\partial n_3} = -2\frac{\sigma^2}{n_3^2} + \lambda = 0$$

$$n_3^2 = \frac{2\sigma^2}{\lambda}$$

$$\frac{\partial f}{\partial n_1} = -\frac{\sigma^2}{n_1^2} + 5\lambda = 0$$

$$n_1^2 = \frac{\sigma^2}{5\lambda} = \frac{1}{10}n_3^2 \quad \Rightarrow \quad n_1 = \frac{1}{\sqrt{10}}n_3$$

$$\frac{\partial f}{\partial n_2} = -\frac{\sigma^2}{n_2^2} + 2\lambda = 0$$

$$n_2^2 = \frac{\sigma^2}{2\lambda} = \frac{1}{4}n_3 \quad \Rightarrow \quad n_2 = \frac{1}{2}n_3$$

Substitude into constraint. $5\left(\frac{1}{\sqrt{10}}n_3\right) + 2\left(\frac{1}{2}n_3\right) + n_3 = 100$

$$n_3 = 27.92$$

$$n_1 = 8.83$$

$$n_2 = 13.96.$$

Therefore, random allocation for each treatment is: treatment 1: 9

2: 14

3: 27

price $= 9 \times 5000 + 14 \times 2000 + 27 \times 1000 = \$100{,}000.$

lsm3_Q5(b)

Yuxin Ma

26/05/2021

```
x <- sample(9+14+27)
(x[1:9]) # treatment 1
```

```
## [1] 45 20 29 34 44 32  5 27  7
```

```
(x[10:23]) #treatment 2
```

```
##  [1] 10 25 13 24 43 26 38 39  2  4 33 49  1 47
```

```
(x[24:50]) #treatment 3
```

```
##  [1] 16 11  8  3 46 50 41 15 18 40  9 17 31 42 21  6 48 28 35 37 30 22 12 14 36
## [26] 19 23
```