STAT 350

Assignment 2

Yuhan Zhang
301345627

1. $Var(\hat{\beta}_1) = \dfrac{\sigma^2}{S_{xx}}$.

$\therefore$ To minimize $Var(\hat{\beta}_1)$, we will maximize S_{xx} or minimize ϵ

$\because \epsilon \sim N(0, \sigma^2)$.

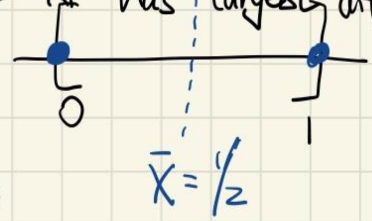
As σ^2 is determined by ϵ.

$\therefore$ To minimize $Var(\hat{\beta}_1)$, we can only maximize S_{xx}.

$\because S_{xx} = \sum (X_i - \bar{X})^2 = \sum \left(X_i - \frac{\sum X_i}{10}\right)^2$

$\therefore$ When X_i spreads to the extreme value evenly.

$S_{xx} = \sum (X_i - \bar{X})^2$ is maximize. $\Rightarrow$ X_i has largest difference



$\therefore$ There should be 5 0s and 5 1s e.g. $X_1 = X_2 = X_3 = X_4 = X_5 = 0$

$X_6 = X_7 = X_8 = X_9 = X_{10} = 1$.

2. $L = f(\vec{\epsilon})$ ($\because \epsilon_i$ independent)

$= (2\pi\sigma^2)^{-n/2} e^{[-(y-X\beta)'(y-X\beta)/2\sigma^2]}$

$l = \ln L = -\frac{n}{2}\ln(2\pi) - n\ln(\sigma^2) - [\frac{1}{2\sigma^2}(y-X\beta)'(y-X\beta)]$

$0 = \frac{dl}{d\beta} = -\frac{1}{2\sigma^2}\frac{d(y-X\beta)'(y-X\beta)}{d\beta} = \frac{d(y'y - y'X\beta - \beta'X'y - \beta'X'X\beta)}{d\beta}$

$\Rightarrow$ equivalent to LSE.

$\therefore 0 = \frac{dl}{d\beta} = -2X'y + 2\beta X'X$

$X'y = X'X\beta$

$\therefore \hat{\beta} = (X'X)^{-1}X'y$

$0 = \frac{dl}{d\sigma^2} = -\frac{n}{2}\left(\frac{2\pi}{2\pi\sigma^2}\right) + \frac{(y-\beta x)'(y-\beta x)}{2\sigma^4}$

$\therefore \frac{n}{2}\cdot\frac{1}{\sigma^2} = \frac{(y-\beta x)'(y-\beta x)}{2\sigma^4}$

$\sigma n^2/2 = (y-\beta x)'(y-\beta x)/2$

$\therefore \hat{\sigma}^2 = (y-\beta x)'(y-\beta x)/n$ ~ biased

3. $y = X\beta + \epsilon$

a)

$$Var(\hat{\beta}) = Var[(X'X)^{-1}X'y]$$
$$= (X'X)^{-1}X' \, Var(y) \, [(X'X)^{-1}X']'$$
$$= \sigma^2 (X'X)^{-1}X'X(X'X)^{-1}$$
$$= \sigma^2 (X'X)^{-1}$$

$$\therefore Var(\hat{y}) = X \, Var(\hat{\beta}) \, X' = X\sigma^2 (X'X)^{-1} X'$$
$$= \sigma^2 X(X'X)^{-1}X' = \sigma^2 H. \quad (H = X(X'X)^{-1}X')$$

b) $\underline{\underline{\hat{y}_0 = x_0'\hat{\beta}}}$

$$E(y|x_0) = E(\beta_0 + \beta_1 x_{10} + \beta_2 x_{20} + \cdots + \beta_k x_{k0} + \epsilon | x_0).$$
$$= \beta_0 + \beta_1 x_{10} + \beta_2 x_{20} + \cdots + \beta_k x_{k0} + E(\epsilon) \Rightarrow 0$$
$$= x_0'\hat{\beta}$$

$$E(\hat{y}_0) = E(x_0'\hat{\beta}) = x_0' E(\hat{\beta}) = x_0'\beta.$$

$$[\because E(\hat{\beta}) = E[(X'X)^{-1}X'Y] = (X'X)^{-1}X'E(Y)$$
$$= E[(X'X)^{-1}X'(X\beta + \epsilon)]$$
$$= E[(X'X)^{-1}X'X\beta + (X'X)^{-1}X'\epsilon] = E[\beta + (X'X)^{-1}X'\epsilon]$$
$$= \beta + (X'X)^{-1}X'\underbrace{E(\epsilon)}_{0} = \beta \quad \therefore \hat{\beta} \text{ is unbiased for } E(\hat{\beta})]$$

Q4a

```
###a.
set.seed(789)

x_1 <- rnorm(n=200,mean=0,sd=2)
x_2 <- rnorm(200,0,2)
e <- rnorm(n=200,mean=0,sd=1)
y <- 1+2*x1+5*x2+ e

fit <- lm(y~x_1+x_2)
fit
```

```
Coefficients:
(Intercept)          x_1          x_2
      1.062        2.015        5.008
```

The estimated line is y=1.062+2.015x1+5.008x2+error.

Q4b

```
> summary(fit)

Call:
lm(formula = y ~ x_1 + x_2)

Residuals:
    Min      1Q  Median      3Q     Max 
-3.1839 -0.5840  0.0062  0.5170  2.6338 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  1.06244    0.06532   16.26   <2e-16 ***
x_1          2.01535    0.03248   62.06   <2e-16 ***
x_2          5.00825    0.03219  155.59   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.9223 on 197 degrees of freedom
Multiple R-squared:  0.9931,	Adjusted R-squared:  0.9931 
F-statistic: 1.422e+04 on 2 and 197 DF,  p-value: < 2.2e-16

> round(vcov(fit),5)
            (Intercept)      x_1      x_2
(Intercept)     0.00427  0.00012 -0.00004
x_1             0.00012  0.00105 -0.00002
x_2            -0.00004 -0.00002  0.00104
> X <- cbind(rep(1,200),x1,x2)
> X
                 x1          x2
  [1,] 1  1.04819342 -6.32306492
  [2,] 1 -4.52153577 -3.31152460
  [3,] 1 -0.03935944 -2.33647279
  [4,] 1  0.36627979  4.09359632
  [5,] 1 -0.72270296 -0.85719799
  [6,] 1 -0.96896798 -1.62992707
  [7,] 1 -1.33262589 -0.21218891
  [8,] 1 -0.34893856  1.89985560
  [9,] 1 -2.02191933 -3.31900830
 [10,] 1  1.47939211  0.46607089
 [11,] 1 -0.80459261 -0.42947869
 [12,] 1 -2.00556065  0.03154355
 [13,] 1 -0.35543842  1.72507086
 [14,] 1 -0.97579913 -0.30737995
 [15,] 1  1.85581478 -0.43156995
 [16,] 1 -1.54885045  1.64194669
 [17,] 1  0.84574225  1.99833563
 [18,] 1 -1.21392678 -3.22299162
 [19,] 1  0.41874995 -0.53306720
 [20,] 1  1.55467438  0.11502108
```

```
[16,] 1 -1.54885045  1.64194669
[17,] 1  0.84574225  1.99833563
[18,] 1 -1.21392678 -3.22299162
[19,] 1  0.41874995 -0.53306720
[20,] 1 -1.55467438  0.11502198
[21,] 1 -1.40410024 -1.09057491
[22,] 1  1.36693834  2.06982454
[23,] 1 -1.71537515 -1.54738717
[24,] 1  0.73552167 -3.21604250
[25,] 1 -2.85938406 -0.78979727
[26,] 1 -1.02464965  0.98331873
[27,] 1 -0.53622075 -3.31996800
[28,] 1 -0.39846424  2.61508910
[29,] 1  1.71314852 -1.04127501
[30,] 1 -0.33357872  0.24354470
[31,] 1 -0.75134313  0.04812526
[32,] 1 -2.26995375  2.67482562
[33,] 1  1.21862951 -1.34346968
[34,] 1  1.13171063  2.24317986
[35,] 1 -1.47209718 -1.12493500
[36,] 1 -2.11710379  0.55574625
[37,] 1 -2.63572855  1.64649232
[38,] 1 -0.40877957  1.25530940
[39,] 1 -1.20277788 -1.28833076
[40,] 1  1.32120523 -2.41788233
[41,] 1 -0.32937966 -3.16081390
[42,] 1  3.51906206  1.02209020
[43,] 1  3.22409349 -0.69480981
[44,] 1  1.60639721 -1.54138442
[45,] 1  2.53047046  0.38010928
[46,] 1  1.02194958  2.76954295
[47,] 1 -6.17524020  1.65611259
[48,] 1  0.94845824  0.53733907
[49,] 1  1.25178799 -0.34962850
[50,] 1  1.96235328  1.78743692
[51,] 1 -0.09541972 -1.23513582
[52,] 1 -3.03922966  0.29921109
[53,] 1  1.58987288 -0.04045490
[54,] 1 -0.28846502 -2.46045771
[55,] 1 -1.41299582 -2.98734011
[56,] 1  1.22142494  1.57665194
[57,] 1  2.17012328 -2.04625009
[58,] 1 -1.42269849 -1.56932617
[59,] 1  2.31257609  0.55571763
[60,] 1  2.47112559 -2.97681228
[61,] 1 -0.64491062 -4.31777809
[62,] 1  1.46555080  1.08101661
[63,] 1 -0.57505766  0.18443716
[64,] 1  4.69702656 -0.94481978
[65,] 1  0.69532804 -1.47379564
```

```
[152,] 1 -0.95141421  3.46268408
[153,] 1  0.37383410  0.67181832
[154,] 1  2.78191672  0.50641396
[155,] 1 -1.53146119  0.70195499
[156,] 1  0.29581871  2.24609597
[157,] 1 -0.80426813 -3.55927515
[158,] 1 -1.33781504  3.49091233
[159,] 1 -0.15723165 -1.45348501
[160,] 1  1.42029253  1.95604810
[161,] 1  2.16958781 -0.06998524
[162,] 1 -0.78612851 -1.63233110
[163,] 1 -1.60311244 -0.87807951
[164,] 1 -0.03830980 -0.56059845
[165,] 1  2.27618998 -0.29334188
[166,] 1 -2.79898038  1.56480581
[167,] 1 -3.31647434  0.70162119
[168,] 1 -1.35623063 -1.85690884
[169,] 1  3.06060188  0.24421505
[170,] 1  0.07192984 -0.58726138
[171,] 1 -2.83817972  2.68252471
[172,] 1 -0.98323456  1.30242232
[173,] 1  0.22180919 -0.84865623
[174,] 1  3.96678729 -0.19489112
[175,] 1  2.48112161  2.48217984
[176,] 1 -1.66217233  2.76482679
[177,] 1  0.47810974 -0.95108037
[178,] 1 -4.41727089  2.63431885
[179,] 1 -3.79091511 -1.49114438
[180,] 1  1.03721384 -5.99139669
[181,] 1  3.38691902 -1.59017166
[182,] 1 -0.13587408  0.88097706
[183,] 1  1.37220773 -3.15918993
[184,] 1 -2.98877413 -2.75748653
[185,] 1  0.92629998  0.19296107
[186,] 1  3.05365440  1.12463149
[187,] 1 -0.03321762  1.47586510
[188,] 1 -1.22013495  1.57705760
[189,] 1 -0.03011107  3.36680426
[190,] 1 -0.56800676  0.77958091
[191,] 1 -1.95920994 -0.10351078
[192,] 1 -1.81218219  1.91627170
[193,] 1 -2.90850317 -0.49717367
[194,] 1  3.31017348  0.99621934
[195,] 1  2.06309434 -0.39074794
[196,] 1  2.04969184  1.83513879
[197,] 1 -1.92657031  0.78802576
[198,] 1 -0.11938121 -0.83211183
[199,] 1 -4.30359918  3.70370809
[200,] 1 -0.68199598 -3.25926559
> sigma2 1
```

```
> sigma2=1
> round(vcov(fit),5)
            (Intercept)      x_1      x_2
(Intercept)     0.00427  0.00012 -0.00004
x_1             0.00012  0.00105 -0.00002
x_2            -0.00004 -0.00002  0.00104
> round(sigma2*solve(t(X) %*% X),5)
                   x1       x2
    0.00502  0.00014 -0.00004
x1  0.00014  0.00124 -0.00002
x2 -0.00004 -0.00002  0.00122
```

The variance for three regression coefficients is similar to the theoretical values of the variance of the predictors.

Q4c

```r
###c

summary(fit)

summary(fit)$coef[2,4] < 0.05

anova(fit)
```

```
> summary(fit)$coef[2,4] < 0.05
[1] TRUE
> anova(fit)
Analysis of Variance Table

Response: y
           Df  Sum Sq Mean Sq F value    Pr(>F)
x_1         1  3596.0  3596.0  4227.4 < 2.2e-16 ***
x_2         1 20593.8 20593.8 24209.7 < 2.2e-16 ***
Residuals 197   167.6     0.9
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
>
```

As the F values of both x1 and x2 are smaller than the critical value, we can conclude that we can reject H0 at significance level of 0.05.

Q4d

```
> rejectH0
   [1] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
  [26] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
  [51] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
  [76] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
 [101] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
 [126] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
 [151] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
 [176] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
 [201] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
 [226] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
 [251] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
 [276] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
 [301] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
 [326] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
 [351] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
 [376] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
 [401] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
 [426] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
 [451] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
 [476] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
 [501] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
 [526] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
 [551] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
 [576] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
 [601] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
 [626] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
 [651] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
 [676] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
 [701] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
 [726] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
 [751] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
 [776] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
 [801] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
 [826] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
 [851] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
 [876] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
 [901] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
 [926] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
 [951] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
 [976] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
> mean(rejectH0)
[1] 1
> sd(rejectH0)
[1] 0
```

```r
###d

rejectH0 <- NULL

for (i in 1:1000){
  x_1 <- rnorm(n=200,mean=0,sd=2)
  x_2 <- rnorm(n=200,mean=0,sd=2)
  e <- rnorm(n=200,mean=0,sd=1)
  y <- 1+2*x1+5*x2+ e
  fit <- lm(y~x1+x2)
  rejectH0[i] = summary(fit)$coef[2,4]<0.05
}


rejectH0
mean(rejectH0)
sd(rejectH0)
```

As all the outputs are TRUE, we can 100% reject H0.

Besides, the mean of times reject H0 is o and the variance is 0, which also means all the results return TRUE.