2021

Please note for Question 5 c), the MLE *does* concide. I forgot to add the r_i component to the exponent.

Question 1

```
Call:
lm(formula = log.mpg ~ log.disp)

Residuals:
    Min       1Q   Median       3Q      Max
    -0.56908 -0.12283 -0.00699  0.14049  0.36851

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  4.38377    0.26310  16.662  < 2e-16 ***
log.disp    -0.27539    0.05004  -5.504 5.62e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.2176 on 30 degrees of freedom
Multiple R-squared:  0.5024, Adjusted R-squared:  0.4858
F-statistic: 30.29 on 1 and 30 DF,  p-value: 5.617e-06
```

Part 1

a)

```
plot_residuals <- function(model) {
    par(mfrow=c(2,2))
    plot(model)
    par(mfrow=c(1, 1))
    }

normal_model <- lm(mpg~disp)
log_model <- lm(log.mpg~log.disp)
plot_residuals(normal_model)
plot_residuals(log_model)
```

Part 2

b)

50.24%

c)

```
# H_0: log.disp = 0
# H_1: log.disp != 0

# F-statistic: 30.29
# Null-distribution: F(1, 30)
# p-value = 5.617 * 10^{-6}

With a 1% level of significance, the p-value shows we can
reject the null hypothesis - implying that the
model with log.disp as a predictor is better than the
intercept only model.
```

d)

```
# This is just the slope
# -0.27539
```

e)

```
confint(log_model, level=0.97)
# (-0.3893795, -0.1613927)
```

f)

```
predict(log_model, data.frame(log.disp=log(290)), interval="confidence", level=0.99)
# (2.698533, 2.946202)
```

Question 2

```
Call:
lm(formula = log.mpg ~ log.disp + dummy, data = cars)

Residuals:
    Min       1Q   Median       3Q      Max
    -0.58657 -0.12023 -0.00934  0.12338  0.37026

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  4.42299    0.26624  16.613 2.34e-16 ***
log.disp    -0.28425    0.05087  -5.588 4.95e-06 ***
dummy        0.22116    0.22473   0.984    0.333
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.2177 on 29 degrees of freedom
Multiple R-squared:  0.5185, Adjusted R-squared:  0.4853
F-statistic: 15.61 on 2 and 29 DF,  p-value: 2.498e-05
```

b)

```
t value = 0.984
null distribution = t(29)
p-value = 0.333

p-value > 0.01, therefore, we cannot reject the null hypothesis
that the dummy has statistical significance to the model.
```

c)

```
# No. The test is testing whether the 25-th observation has
# statistical significance to the model. It also does not
# externally calculate the variance.

rstudent(log_model)[25]
# r_i = 0.9841012, under t(29), as we externally studentize
# the observation

1 - pt(0.9841012, 29)
# 0.1666027. Still not statistically significant.
```

Question 3

```
Call:
lm(formula = mpg ~ wt + cyl + disp + hp, data = mpg)

Residuals:
    Min      1Q  Median      3Q     Max
    -4.9930 -2.1404  0.3625  1.1596  6.5199

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 43.67842    3.18573  13.711 1.11e-13 ***
wt          -4.06476    1.22240  -3.325  0.00255 **
cyl         -2.39820    0.70630  -3.395  0.00214 **
disp         0.02960    0.01275   2.321  0.02806 *
hp          -0.01834    0.01480  -1.239  0.22588
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.029 on 27 degrees of freedom
Multiple R-squared:  0.7888, Adjusted R-squared:  0.7575
F-statistic: 25.21 on 4 and 27 DF,  p-value: 8.912e-09
```

a)

```
q3_model <- lm(mpg~wt+cyl+disp+hp, data=mpg)
# a)
# F-test
# H_0: wt = cyl = disp = hp = 0
# H_1: not all betas are 0
# F-statistic = 25.21
# Distribution: F(4, 27)
# p-value: 8.912 * 10^{-9}

With a 5% level of significance, the F-test shows enough evidence
to reject the null hypothesis; concluding that the model
with the predictors is better than the model only containing
the intercept term.
```

b)

```
# F-test
# H_0: cycl = disp = hp = 0
# H_1: not all are 0
anova(lm(mpg~wt, data=mpg), q3_model)
```

```
# F-statistic = 8.0194
# Distribution: F(3, 27)
# p-value = 0.0005578
```

```
The F-test concludes, given a significance level of 5%, that
there is enough evidence to reject the null hypothesis.
This indicates that not all of the betas of cycl, disp
and hp are 0, and thus contribute to the model.
```

c)

```
# F-test
# H_0: displacement = 0
# H_1: displacement != 0
anova(lm(mpg~wt, data=mpg), lm(mpg~wt+disp, data=mpg))
# F-statistic = 0.7911
# Distribution: F(1, 29)
# p-value = 0.3811
```

```
With a 5% level of significance, and a p-value of 0.3811,
there is not enough evidence to reject the null hypothesis.
This means that the addition of displacement to the model
with weight, does not benefit the model with statistical
significance.
```

d)

```
predict(q3_model, data.frame(wt=3.8,cyl=6,disp=220,hp=160), interval="prediction", level=0.92)
# (11.60024, 23.24071)
```

MATH2831 2021 FINAL

Q4.

a) $b = (X^T X)^{-1} X^T y$

$= (X^T X)^{-1} X^T (X\beta + \varepsilon)$

$= \beta + (X^T X)^{-1} X^T \varepsilon$

$= \beta + S\varepsilon$

b) $Var(z) = E[(\beta + S\varepsilon - \beta)(\beta + S\varepsilon - \beta)^T]$

$= E[(S\varepsilon)(S\varepsilon)^T]$

$= S E[\varepsilon\varepsilon^T] S^T$

$= \sigma^2 S S^T$

c) $(y - X(\beta + S\varepsilon))^T (y - X(\beta + S\varepsilon))$

$= (y - X\beta - XS\varepsilon)^T (y - X\beta - XS\varepsilon)$

$= (X\beta + \varepsilon - X\beta - XS\varepsilon)^T (X\beta + \varepsilon - X\beta - XS\varepsilon)$

$= (\varepsilon - XS\varepsilon)^T (\varepsilon - XS\varepsilon)$

$= (\varepsilon^T (I - XS)^T)(I - XS)\varepsilon$

$= \varepsilon^T (I - XS)^T (I - XS)\varepsilon$.

Consider $(I - XS)^T (I - XS)$:

$(I^T - S^T X^T)(I - XS) = (I - IXS - SXI + S^T X^T XS)$

$= (I - XS - SX + S^T X^T X S)$

$= (I - XS - (X^T X)^{-1} X^T X + S^T X^T X S)$

$= (-XS + S^T X^T X S)$

$X(X^T X)^{-1} X^T X S$

$= I$

$= (I - XS)$

$= \varepsilon^T (I - XS)\varepsilon$.

d)

$$\hat{\sigma}^2 = \frac{(y - X\beta)^T (y - X\beta)}{n - p}$$

$$E(\hat{\sigma}^2) = \frac{E(\varepsilon^T (I - XS)\varepsilon)}{n-p}$$ [Now, applying E of quad. form]

$$= \frac{tr((I - XS)\sigma^2 I) + 0.}{n-p}$$

$$= \frac{\sigma^2 tr(I) - \sigma^2 tr(XS)}{n-p}$$

$$= \frac{n\sigma^2 - \sigma^2 tr(X(X^TX)^{-1}X^T)}{n-p}$$

$p \times \cdots \times p.$

$$= \frac{n\sigma^2 - \sigma^2 tr((X^TX)^{-1}X^TX)}{n-p}$$

$$= \frac{n\sigma^2 - p\sigma^2}{n-p}$$

$$= \sigma^2$$

QUESTION 5.

a) $$S(\beta) = \sum_{i=1}^{n} (y_i - \beta x_i)^2$$

$$\frac{\partial S(\beta)}{\partial \beta} = -2 \sum_{i=1}^{n} x_i (y_i - \beta x_i)$$

$$\frac{\partial^2 S(\beta)}{\partial \beta^2} = 2 \sum_{i=1}^{n} x_i^2 > 0 \quad \therefore \text{ MINIMUM}$$

$$-2 \sum_{i=1}^{n} x_i (y_i - \beta x_i) = 0$$

$$= \sum_{i=1}^{n} x_i y_i - \beta \sum_{i=1}^{n} x_i^2 = 0$$

$$\beta = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2}$$

b)

$$S_w(\beta) = \sum_{i=1}^{n} k_i (y_i - \beta x_i)^2$$

$$\frac{\partial S_w(\beta)}{\partial \beta} = -2 \sum_{i=1}^{n} k_i x_i (y - \beta x_i)$$

$$\frac{\partial^2 S_w(\beta)}{\partial \beta^2} = 2 \sum_{i=1}^{n} k_i x_i^2 > 0, \text{ as } k_i > 0. \therefore \text{ MIN.}$$

$$-2 \sum_{i=1}^{n} k_i x_i (y_i - \beta x_i) = 0$$

$$\sum_{i=1}^{n} k_i x_i (y_i - \beta x_i) = 0$$

$$= \sum_{i=1}^{n} k_i x_i y_i - \beta \sum_{i=1}^{n} k_i x_i^2 = 0$$

$$= \beta \sum_{i=1}^{n} k_i x_i^2 = \sum_{i=1}^{n} k_i x_i y_i$$

$$= \beta = \frac{\sum_{i=1}^{n} k_i x_i y_i}{\sum_{i=1}^{n} k_i x_i^2}$$

c) $\varepsilon_i \sim N\left(0, \frac{\sigma^2}{k_i}\right)$

$\therefore\ y_i \sim N\left(\beta x_i, \frac{\sigma^2}{k_i}\right)$

$$f_y(y_i) = \frac{\sqrt{k_i}}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(y_i - \beta x_i)^2\right)$$

$$L(\beta, \sigma^2; y_i) = \frac{\sqrt{\prod k_i}}{\sqrt{(2\pi\sigma^2)^n}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \beta x_i)^2\right)$$

$$l(\beta, \sigma^2; y_i) = \frac{1}{2}\sum_{i=1}^{n}\ln(k_i) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \beta x_i)^2 - \frac{n}{2}\ln(2\pi\sigma^2).$$

$$\frac{\partial l}{\partial \beta} = \frac{1}{\sigma^2}\sum_{i=1}^{n} x_i(y_i - \beta x_i)$$

NO. NOT THE SAME.

d) $$b_w = \frac{\sum_{i=1}^{n} k_i x_i y_i}{\sum_{i=1}^{n} k_i x_i^2}$$

$$E(b_w) = \frac{\sum_{i=1}^{n} k_i x_i E(y_i)}{\sum_{i=1}^{n} k_i x_i^2}$$

$$= \frac{\sum_{i=1}^{n} k_i x_i^2 \beta}{\sum_{i=1}^{n} k_i x_i^2}$$

$$= \underline{\beta}$$

$$Var(b_w) = \frac{\sum_{i=1}^{n} Var(k_i x_i y_i)}{\sum_{i=1}^{n} k_i x_i^2}$$

$$= \frac{\sum_{i=1}^{n} k_i^2 x_i^2 Var(y_i)}{\sum_{i=1}^{n} k_i x_i^2}$$

$$= \frac{\sum_{i=1}^{n} k_i^2 x_i^2 \times \frac{\sigma^2}{k_i}}{\sum_{i=1}^{n} k_i x_i^2}$$

$$= \frac{\sigma^2 \cancel{\sum_{i=1}^{n} k_i x_i^2}}{\cancel{\sum_{i=1}^{n} k_i x_i^2}}$$

$$= \sigma^2.$$

c) Consider the first-order Taylor series expansion about $E(Y_i)$.

$$\sqrt{k_i}\, Y_i = \sqrt{k_i}\, \beta x_i + \sqrt{k_i}\, \varepsilon_i$$

$$\therefore Var(\sqrt{k_i}\, Y_i) = k_i Var(Y_i)$$

$$= k_i \times \frac{\sigma^2}{k_i}$$

$$= \sigma^2.$$

$$Var(\sqrt{k_i}\, \varepsilon_i) = k_i Var(\varepsilon_i)$$

$$= \sigma^2. \quad \therefore \text{OLS}.$$

$$S(\beta) = \sum_{i=1}^{n} (\sqrt{k_i}\, Y_i - \sqrt{k_i}\, \beta x_i)^2$$

$$= \sum_{i=1}^{n} k_i (Y_i - \beta x_i)^2$$

$\therefore$ SAME.

a) OBSERVATION 14 has a Cook's Distance > 0.9.
This indicates that the removal of the observation significantly changes the estimates.
This can be observed in the right figure.

b) CEEBS.

c)

$$-4.5839057 \times \sqrt{\frac{0.23572744}{1-0.23572744}}$$

$$= -2.546 < -2, \therefore \text{INFLUENTIAL}$$

d) The i-th predicted response is given by the i-th row of predictors (x_i) linearly multiplied by b.

$$\hat{y}_i = x_i^T b.$$

For $\hat{y}_{i,-i}$, b_{-i} ensures the ith observation is not considered.

e)

$$\hat{Y}_i - \hat{Y}_{i,-i} = x_i^T b - x_i^T b + \frac{x_i^T (X^TX)^{-1} x_i e_i}{1-h_{ii}}$$

$$= \frac{x_i^T (X^TX)^{-1} x_i e_i}{1-h_{ii}}$$

Considering $x_i = (x_{i1} \ x_{i2} \ \cdots \ x_{in})^T$

$$H = X(X^TX)^{-1}X^T$$

$$h_{ii} = (x_i)^T (X^TX)^{-1} x_i$$

$$= \frac{h_{ii} e_i}{1-h_{ii}}$$

f)

$$t_i = \frac{e_i}{\hat{\sigma}_{-i}\sqrt{1-h_{ii}}}$$

$$\frac{\frac{h_{ii} e_i}{1-h_{ii}}}{\hat{\sigma}_{-i}\sqrt{h_{ii}}} = \frac{e_i \sqrt{h_{ii}}}{\hat{\sigma}_{-i}(1-h_{ii})} = t_i \times \sqrt{\frac{h_{ii}}{1-h_{ii}}}$$