STAT 393
Test 2

Irshad Ul Ala
Student ID 300397080

October 22, 2020

1 Question 1 — Part 1

1.1 a

The distribution of ϵ is as follows

$$\epsilon \sim N(0, \sigma^2\mathbf{I})$$

1.2 b

1.2.1 $E(\hat{\beta})$

$$E(\hat{\beta}) = E((\mathbf{X^TX})^{-1}\mathbf{X^Ty}) = (\mathbf{X^TX})^{-1}\mathbf{X^T} \cdot E(\mathbf{y}) = (\mathbf{X^TX})^{-1}\mathbf{X^T} \cdot \mathbf{X}\beta = (\mathbf{X^TX})^{-1}\mathbf{X^TX} \cdot \beta = \beta$$

1.2.2 $Var(\hat{\beta})$

Letting $Var(\mathbf{y}) = \sigma^2\mathbf{I_n}$

$$Var(\hat{\beta}) = Var((\mathbf{X^TX})^{-1}\mathbf{X^Ty}) = (\mathbf{X^TX})^{-1}\mathbf{X^T}Var(\mathbf{y})((\mathbf{X^TX})^{-1}\mathbf{X^T})^T = (\mathbf{X^TX})^{-1}\mathbf{X^T}\sigma^2\mathbf{I_n}(\mathbf{X}(\mathbf{X^TX})^{-1}) =$$

$$\sigma^2(\mathbf{X^TX})^{-1}\mathbf{X^TX}(\mathbf{X^TX})^{-1} = \sigma^2(\mathbf{X^TX})^{-1}$$

1.2.3 Distribution of $\hat{\beta}$

Therefore the distribution of $\hat{\beta}$ is

$$\hat{\beta} \sim N(\beta, \sigma^2(\mathbf{X^TX})^{-1})$$

1.3 c

Considering $\hat{\mathbf{y}} = \mathbf{X}\hat{\beta}$

$$E(\hat{\mathbf{y}}) = E(\mathbf{X}\hat{\beta}) = \mathbf{X} \cdot E(\hat{\beta}) = \mathbf{X}\beta$$

$$Var(\hat{\mathbf{y}}) = Var(\mathbf{X}\hat{\beta}) = \mathbf{X}Var(\hat{\beta})\mathbf{X^T} = \mathbf{X}\sigma^2(\mathbf{X^TX})^{-1}\mathbf{X^T} = \sigma^2\mathbf{X}(\mathbf{X^TX})^{-1}\mathbf{X^T}$$

$$\therefore \hat{\mathbf{y}} \sim N(\mathbf{X}\beta, \sigma^2\mathbf{X}(\mathbf{X^TX})^{-1}\mathbf{X^T})$$

1.4 d — Number of Degrees of freedom for SSE

The number of degrees of freedom for $SSE = (\mathbf{y} - \hat{\mathbf{y}})^{\mathbf{T}}(\mathbf{y} - \hat{\mathbf{y}})$ is n-p. This is because there are p explanatory factors that remove the amount of variability (by p-1) in the prediction of the data, deducting from the initial n-1 degrees of freedom one would associate with just the results in $\hat{\mathbf{y}}$ from SST.

2 Question 1 — Part 2

2.1 Design matrix

```
X<-model.matrix(mpg~weight+year, data=Auto)
head(X,10)

[Output]
> head(X,10)
   (Intercept) weight year
1            1   3504   70
2            1   3693   70
3            1   3436   70
4            1   3433   70
5            1   3449   70
6            1   4341   70
7            1   4354   70
8            1   4312   70
9            1   4425   70
10           1   3850   70
```

2.2 b—LSE of β

```
y<-Auto$mpg
beta_hat<-solve(t(X)%*%X)%*%t(X)%*%y

[Output]
> beta_hat
                     [,1]
(Intercept) -14.347253018
weight        -0.006632075
year           0.757318281
```

2.3 c—Predicted values, $\hat{y}$

```
y_hat<-X%*%beta_hat
head(y_hat,10)

[Output]
> head(y_hat,10)
        [,1]
1  15.426235
2  14.172773
3  15.877216
4  15.897112
5  15.790999
6   9.875188
7   9.788971
8  10.067518
9   9.318093
10 13.131537
```

2.4 c—SSE and RSE

SSE follows the following formula

$$SSE = (\mathbf{y} - \hat{\mathbf{y}})^{\mathbf{T}}(\mathbf{y} - \hat{\mathbf{y}}) \tag{1}$$

Additionally,

$$RSE = \frac{1}{n-p} SSE = \frac{1}{n-3} SSE \tag{2}$$

In our example, p=3 because there are 2 explanatory variables and one intercept.

Therefore, we can do the following in R to attain it.

```
(SSE<-t(y-y_hat)%*%(y-y_hat))
n<-length(y)
(RSE<-SSE/(n-3))

[Output]
> (SSE<-t(y-y_hat)%*%(y-y_hat))
[1,] 4568.952
> (RSE<-SSE/(n-3))
         [,1]
[1,] 11.74538
```

Therefore, the calculated SSE is 4568.952 while the calculated RSE is 11.74538.

3 Question 1 — Part 3

3.1 a—summary output

```
> summary(lm(mpg~weight+year, data=Auto))

Call:
lm(formula = mpg ~ weight + year, data = Auto)

Residuals:
    Min      1Q  Median      3Q     Max
-8.8505 -2.3014 -0.1167  2.0367 14.3555

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.435e+01  4.007e+00  -3.581 0.000386 ***
weight      -6.632e-03  2.146e-04 -30.911  < 2e-16 ***
year         7.573e-01  4.947e-02  15.308  < 2e-16 ***
---
Signif. codes:  0    ***    0.001    **    0.01    *    0.05    .    0.1         1

Residual standard error: 3.427 on 389 degrees of freedom
Multiple R-squared:  0.8082,  Adjusted R-squared:  0.8072
F-statistic: 819.5 on 2 and 389 DF,  p-value: < 2.2e-16
```

Above is the printed summary function result for fitting mpg to year and weight as explanatory variables.

3.2 b— 99% confidence intervals

```
cons<-qnorm((1-0.99)/2, lower.tail = F)
modelq1<-(lm(mpg~weight+year, data=Auto))

std<-summary(modelq1)$coef[,2]
b_smm<-summary(modelq1)$coef[,1]
(cbind(b_smm-cons*std, b_smm+cons*std))

[Output]
                    [,1]         [,2]
(Intercept) -24.667360938 -4.027145098
weight       -0.007184735 -0.006079416
year          0.629885083  0.884751479
```

Above is the 99% confidence interval for the coefficients

3.3 c— Interpretation

From the summary table, it is evident that for each of the coefficients, the following hypothesis is being tested

$$H_0 : \beta_i = 0$$
$$H_1 : \beta_i \neq 0$$

Firstly, for the weight variable, the corresponding t-value with respect to it is -30.911 at 1 degree of freedom and it returned a p-value lower than the machine epsilon 2×10^{-16}, which provided substantial evidence to reject the null hypothesis and conclude that the weight variable is a significant explanatory variable that is non-zero to explain miles per gallon.

Secondly, for the year variable, again, we get a t-value of 15.308 with a p-value lower than the machine epsilon 2×10^{-16}, which provided substantial evidence to reject the null hypothesis and conclude that the year variable is a significant explanatory variable that is non-zero to explain miles per gallon.

The 99% confidence intervals produced earlier did not include 0 in them, which strongly supports the notion that these variables are significant and non-zero when considering mpg as a response variable.

4 Question 1 — Part 4

The predicted values for the new data, and their corresponding 99% confidence and prediction intervals are calculated as follows

```
predict(modelq1, newdata=new_cars)
predict(modelq1, newdata=new_cars, interval="confidence", level=0.99)
predict(modelq1, newdata=new_cars, interval="prediction", level=0.99)

[Output]
> predict(modelq1, newdata=new_cars)
       1        2
19.99667 13.83515
> predict(modelq1, newdata=new_cars, interval="confidence", level=0.99)
       fit      lwr      upr
1 19.99667 19.46242 20.53092
2 13.83515 12.31128 15.35902
> predict(modelq1, newdata=new_cars, interval="prediction", level=0.99)
       fit       lwr      upr
1 19.99667 11.109323 28.88402
2 13.83515  4.833943 22.83636
```

5 Question 1 — Part 5 — Null model vs Full model

We fit the null model to a variable 'null', and then run an anova sequential sum of squares test against out previously fitted full model, with year and weight as the explanatory variables.

```
null<-(lm(mpg~1, data=Auto))
anova(null, modelq1)

[Output]
> anova(null, modelq1)
Analysis of Variance Table

Model 1: mpg ~ 1
Model 2: mpg ~ weight + year
  Res.Df   RSS Df Sum of Sq      F    Pr(>F)
1    391 23819
2    389  4569  2     19250 819.47 < 2.2e-16 ***
---
Signif. codes:  0    ***    0.001    **    0.01    *    0.05    .    0.1         1
```

The above test is testing between the following hypotheses

$$H_0: \text{ Null model is true — Extra parameters are insignificant}$$

$$H_0: \text{ Full model is true — Extra parameters are significant}$$

Since the returned p-value was below 2.2×10^{-16}, we conclude that there is very significant evidence to reject the null hypothesis and conclude that the full model is true compared to the null model and is a better explanation for the variation in the data. In other words, the explanatory variables weight an year are significant.

6 Question 1 — Part 5 — Full model vs Saturated model

```
satm2<-lm(mpg~cylinders+displacement+horsepower+weight+acceleration+year+origin+name, data=Auto)
anova(modelq1, satm2)

[Output]
> anova(modelq1, satm2)
Analysis of Variance Table

Model 1: mpg ~ weight + year
Model 2: mpg ~ cylinders + displacement + horsepower + weight + acceleration +
    year + origin + name
  Res.Df    RSS  Df Sum of Sq      F    Pr(>F)
1    389 4569.0
2     85  438.7 304    4130.2 2.6322 2.553e-07 ***
---
Signif. codes:  0    ***    0.001    **    0.01    *    0.05    .    0.1         1
```

Similar to before, we are testing between the following hypotheses

$$H_0: \text{ Full model is true — Extra parameters are insignificant}$$
$$H_0: \text{ Saturated model is true — Extra parameters are significant}$$

The returned p-value of 2.553×10^{-7} means that we have significant evidence to reject the null hypothesis and conclude that the saturated model with all the explanatory variables in the dataset is a better fit for the data than our constructed full model and therefore, we conclude that the extra variables are significant, and reject the null hypothesis that the full model we had is a better way to explain the data.

7 Question 2 — Part 1

7.1 a

```
library(nlme)
model2<-lme(score~treat+time+treat*time,random =~1|id, data=dflong)
X<-model.matrix(model2, dflong)

[Output]
> X
   (Intercept) treattreat timepre treattreat:timepre
1            1          1       1                  1
2            1          1       0                  0
3            1          1       1                  1
4            1          1       0                  0
5            1          1       1                  1
6            1          1       0                  0
7            1          1       1                  1
8            1          1       0                  0
9            1          1       1                  1
10           1          1       0                  0
11           1          0       1                  0
12           1          0       0                  0
13           1          0       1                  0
14           1          0       0                  0
15           1          0       1                  0
16           1          0       0                  0
17           1          0       1                  0
18           1          0       0                  0
19           1          0       1                  0
20           1          0       0                  0
attr(,"assign")
[1] 0 1 2 3
attr(,"contrasts")
attr(,"contrasts")$treat
[1] "contr.treatment"

attr(,"contrasts")$time
[1] "contr.treatment"
```

```
Z<-model.matrix(score~id-1,dflong)
Z=Z[,order(colnames(Z))]
Z

[Output]
> Z
   id1 id10 id2 id3 id4 id5 id6 id7 id8 id9
1    1    0   0   0   0   0   0   0   0   0
2    1    0   0   0   0   0   0   0   0   0
3    0    0   1   0   0   0   0   0   0   0
4    0    0   1   0   0   0   0   0   0   0
5    0    0   0   1   0   0   0   0   0   0
6    0    0   0   1   0   0   0   0   0   0
7    0    0   0   0   1   0   0   0   0   0
8    0    0   0   0   1   0   0   0   0   0
9    0    0   0   0   0   1   0   0   0   0
10   0    0   0   0   0   1   0   0   0   0
11   0    0   0   0   0   0   1   0   0   0
12   0    0   0   0   0   0   1   0   0   0
13   0    0   0   0   0   0   0   1   0   0
14   0    0   0   0   0   0   0   1   0   0
15   0    0   0   0   0   0   0   0   1   0
16   0    0   0   0   0   0   0   0   1   0
17   0    0   0   0   0   0   0   0   0   1
18   0    0   0   0   0   0   0   0   0   1
19   0    1   0   0   0   0   0   0   0   0
20   0    1   0   0   0   0   0   0   0   0
```

7.2 b

ϵ has a distribution as follows

$$\epsilon \sim N(0, \sigma^2 \mathbf{I})$$

7.3 c

for β

$$\begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix}$$

For **b**

$$\begin{bmatrix} b_1 \\ b_2 \\ \dots \\ b_{10} \end{bmatrix}$$

8 Question 2 — Part 2

8.1 a

Firstly, we know that

$$b \sim N(0, \sigma_b^2 I)$$

$$\therefore Zb \sim N(0, Z\sigma_b^2 I Z^T) \cong N(0, \sigma_b^2 ZZ^T)$$

The overall covariance of the model however includes the error term ϵ so

$$\Sigma \sim N(0, Var(Zb + \epsilon) \cong N(0, \sigma_b^2 ZZ^T + \sigma^2 I)$$

8.2 b

Using the given values for σ_b and σ, we get the following

```
sig<-14.14214^2*Z%*%t(Z) + 11.40175^2*diag(1, length(dflong$score))

[Output]
> sig
         1        2        3        4        5        6        7        8
1  330.0000 200.0001   0.0000   0.0000   0.0000   0.0000   0.0000   0.0000
2  200.0001 330.0000   0.0000   0.0000   0.0000   0.0000   0.0000   0.0000
3    0.0000   0.0000 330.0000 200.0001   0.0000   0.0000   0.0000   0.0000
4    0.0000   0.0000 200.0001 330.0000   0.0000   0.0000   0.0000   0.0000
5    0.0000   0.0000   0.0000   0.0000 330.0000 200.0001   0.0000   0.0000
6    0.0000   0.0000   0.0000   0.0000 200.0001 330.0000   0.0000   0.0000
7    0.0000   0.0000   0.0000   0.0000   0.0000   0.0000 330.0000 200.0001
8    0.0000   0.0000   0.0000   0.0000   0.0000   0.0000 200.0001 330.0000
9    0.0000   0.0000   0.0000   0.0000   0.0000   0.0000   0.0000   0.0000
10   0.0000   0.0000   0.0000   0.0000   0.0000   0.0000   0.0000   0.0000
11   0.0000   0.0000   0.0000   0.0000   0.0000   0.0000   0.0000   0.0000
12   0.0000   0.0000   0.0000   0.0000   0.0000   0.0000   0.0000   0.0000
13   0.0000   0.0000   0.0000   0.0000   0.0000   0.0000   0.0000   0.0000
14   0.0000   0.0000   0.0000   0.0000   0.0000   0.0000   0.0000   0.0000
15   0.0000   0.0000   0.0000   0.0000   0.0000   0.0000   0.0000   0.0000
16   0.0000   0.0000   0.0000   0.0000   0.0000   0.0000   0.0000   0.0000
17   0.0000   0.0000   0.0000   0.0000   0.0000   0.0000   0.0000   0.0000
18   0.0000   0.0000   0.0000   0.0000   0.0000   0.0000   0.0000   0.0000
19   0.0000   0.0000   0.0000   0.0000   0.0000   0.0000   0.0000   0.0000
20   0.0000   0.0000   0.0000   0.0000   0.0000   0.0000   0.0000   0.0000
         9       10       11       12       13       14       15       16
1    0.0000   0.0000   0.0000   0.0000   0.0000   0.0000   0.0000   0.0000
2    0.0000   0.0000   0.0000   0.0000   0.0000   0.0000   0.0000   0.0000
```

```
3     0.0000   0.0000   0.0000   0.0000   0.0000   0.0000   0.0000   0.0000
4     0.0000   0.0000   0.0000   0.0000   0.0000   0.0000   0.0000   0.0000
5     0.0000   0.0000   0.0000   0.0000   0.0000   0.0000   0.0000   0.0000
6     0.0000   0.0000   0.0000   0.0000   0.0000   0.0000   0.0000   0.0000
7     0.0000   0.0000   0.0000   0.0000   0.0000   0.0000   0.0000   0.0000
8     0.0000   0.0000   0.0000   0.0000   0.0000   0.0000   0.0000   0.0000
9   330.0000 200.0001   0.0000   0.0000   0.0000   0.0000   0.0000   0.0000
10  200.0001 330.0000   0.0000   0.0000   0.0000   0.0000   0.0000   0.0000
11    0.0000   0.0000 330.0000 200.0001   0.0000   0.0000   0.0000   0.0000
12    0.0000   0.0000 200.0001 330.0000   0.0000   0.0000   0.0000   0.0000
13    0.0000   0.0000   0.0000   0.0000 330.0000 200.0001   0.0000   0.0000
14    0.0000   0.0000   0.0000   0.0000 200.0001 330.0000   0.0000   0.0000
15    0.0000   0.0000   0.0000   0.0000   0.0000   0.0000 330.0000 200.0001
16    0.0000   0.0000   0.0000   0.0000   0.0000   0.0000 200.0001 330.0000
17    0.0000   0.0000   0.0000   0.0000   0.0000   0.0000   0.0000   0.0000
18    0.0000   0.0000   0.0000   0.0000   0.0000   0.0000   0.0000   0.0000
19    0.0000   0.0000   0.0000   0.0000   0.0000   0.0000   0.0000   0.0000
20    0.0000   0.0000   0.0000   0.0000   0.0000   0.0000   0.0000   0.0000
           17       18       19       20
1     0.0000   0.0000   0.0000   0.0000
2     0.0000   0.0000   0.0000   0.0000
3     0.0000   0.0000   0.0000   0.0000
4     0.0000   0.0000   0.0000   0.0000
5     0.0000   0.0000   0.0000   0.0000
6     0.0000   0.0000   0.0000   0.0000
7     0.0000   0.0000   0.0000   0.0000
8     0.0000   0.0000   0.0000   0.0000
9     0.0000   0.0000   0.0000   0.0000
10    0.0000   0.0000   0.0000   0.0000
11    0.0000   0.0000   0.0000   0.0000
12    0.0000   0.0000   0.0000   0.0000
13    0.0000   0.0000   0.0000   0.0000
14    0.0000   0.0000   0.0000   0.0000
15    0.0000   0.0000   0.0000   0.0000
16    0.0000   0.0000   0.0000   0.0000
17  330.0000 200.0001   0.0000   0.0000
18  200.0001 330.0000   0.0000   0.0000
19    0.0000   0.0000 330.0000 200.0001
20    0.0000   0.0000 200.0001 330.0000
```

8.3 c — Covariance structure

This might be block compound structure, since there is an off diagonal element.

9 Question 2 — Part 3

9.1 a—GLS estimate of β

```
> (BBeta<-solve(t(X)%*%solve(sig)%*%X)%*%t(X)%*%solve(sig)%*%y)
                    [,1]
(Intercept)           24
treattreat            40
timepre                2
treattreat:timepre   -42
```

9.2 b—GLS estimate of b

```
> (b<-14.14214^2*diag(1,10)%*%t(Z)%*%solve(sig)%*%(y-X%*%BBeta))
             [,1]
 [1,]   0.7547172
 [2,] -11.3207585
 [3,] -10.5660413
 [4,]  23.3962342
 [5,]  -3.0188689
 [6,] -10.5660413
 [7,]   7.5471723
 [8,] -11.3207585
 [9,]   7.5471723
[10,]   7.5471723
```

9.3 c—Predicted value of y

```
(Y<-X%*%BBeta + Z%*%b)

[Output]
```

```
> (Y<-X%*%BBeta + Z%*%b)
       [,1]
1  24.75472
2  64.75472
3  13.43396
4  53.43396
5  47.39623
6  87.39623
7  20.98113
8  60.98113
9  13.43396
10 53.43396
11 33.54717
12 31.54717
13 14.67924
14 12.67924
15 33.54717
16 31.54717
17 33.54717
18 31.54717
19 14.67924
20 12.67924
```

10 Question 2 — Part 4

```
model2<-lme(score~treat+time+treat*time,random =~1|id, data=dflong)

[Output]
> summary(model2)
Linear mixed-effects model fit by REML
 Data: dflong
       AIC      BIC    logLik
  152.9671 157.6026 -70.48354

Random effects:
 Formula: ~1 | id
        (Intercept) Residual
StdDev:    14.14214 11.40175

Fixed effects: score ~ treat + time + treat * time
                   Value Std.Error DF   t-value p-value
(Intercept)           24  8.124038  8  2.954196  0.0183
treattreat            40 11.489125  8  3.481553  0.0083
timepre                2  7.211103  8  0.277350  0.7885
treattreat:timepre   -42 10.198039  8 -4.118439  0.0034
 Correlation:
                   (Intr) trttrt timepr
treattreat         -0.707
timepre            -0.444  0.314
treattreat:timepre  0.314 -0.444 -0.707

Standardized Within-Group Residuals:
       Min         Q1        Med         Q3        Max
-1.1881654 -0.3284831 -0.2349854  0.2862850  1.6184203

Number of Observations: 20
Number of Groups: 10
```

From the p-value of 0.0083, it would appear that there is sufficient evidence to reject the null hypothesis and conclude that treatment has a significant effect on the health scores.