MATH533: Assignment 3

Kabilan Sriranjan

November 20, 2017

Question 1

```
data = read.csv("http://www.math.mcgill.ca/yyang/regression/data/cigs.csv" , header=TRUE)
y = data$CO
x1 = data$TAR
x2 = data$NICOTINE
x3 = data$WEIGHT
```

We are going to compare models that use different combinations of our predictors, namely tar, nicotine, and weight.

a)

```
full_model = lm(y~x1+x2+x3)
SS_res_full = anova(full_model)[4,2]
SS_res_full
```

```
## [1] 43.89259
```

Here we used the anova function to find $SS_{Res}(\beta_0, \beta_1, \beta_2, \beta_3)$

b)

```
reduced_model = lm(y~x1+x2)
SS_res_reduced = anova(reduced_model)[3,2]
SS_res_reduced
```

```
## [1] 43.89494
```

Now using the anova function to find $SS_{Res}(\beta_0, \beta_1, \beta_2)$

c)

```
n = length(x1)
p = 4
r=1
F_stat  = ((SS_res_reduced - SS_res_full)/r)/(SS_res_full/(n-p))
F_stat
```

```
## [1] 0.001127825
```

Computing the F-statistic to compare the full model with the model without x_3

d)

```
table_1 = anova(lm(y~x3+x2+x1))
SSR3_0 = table_1[1,2]
SSR2_03 = table_1[2,2]
SSR1_032 = table_1[3,2]
decomp_1 = c(SSR3_0, SSR2_03, SSR1_032, SSR3_0+SSR2_03+SSR1_032)
decomp_1
```

```
## [1] 116.05651 346.19988  33.00142 495.25781
```

We add the variables to the model in the order $x_3 >> x_2 >> x_1$ so that we can use the anova function to get the decomposition

$$\overline{SS}_R(\beta_1, \beta_2, \beta_3|\beta_0) = \overline{SS}_R(\beta_3|\beta_0) + \overline{SS}_R(\beta_2|\beta_0, \beta_3) + \overline{SS}_R(\beta_1|\beta_0, \beta_3, \beta_2)$$

e)

```
table_2 = anova(reduced_model)
SSR1_0 = table_2[1,2]
SSR2_01 = table_2[2,2]
decomp_2 = c(SSR1_0, SSR2_01, SSR1_0+SSR2_01)
decomp_2
```

```
## [1] 494.2813099   0.9741472 495.2554571
```

We add the variables to the model in the order $x_1 >> x_2$ so that we can get the decomposition

$$\overline{SS}_R(\beta_1, \beta_2|\beta_0) = \overline{SS}_R(\beta_1|\beta_0) + \overline{SS}_R(\beta_2|\beta_0, \beta_1)$$

f)

```
reduced_model_2 = lm(y~x1)
SS_res_red2 = anova(reduced_model_2)[2,2]
p=3
r=1
F_stat2 = ((SS_res_red2 - SS_res_reduced)/r)/(SS_res_reduced/(n-p))
F_stat2
```

```
## [1] 0.4882394
```

If we consdier our full model to now only include x_1 and x_2 then the above code computes the F-statistic comparing the full model with the model that only uses x_1

g)

```
F_stat3 = summary(reduced_model)$fstatistic[1]
F_stat3
```

```
##    value
## 124.1102
```

Here we computed the F-statistic for comparing our new full model to the mean only model. Note that this is equivalent to just finding the F-statistic of the full model normally.

MATH533 Extra Question 1

a)

We want to show that if Z is an orthogonal $n \times M$ matrix then $\hat{y} = \bar{y}\mathbf{1} + \sum_{m=1}^{M} \hat{\theta}_m z_m$ where $\hat{\theta}_m = z_m^T y / z_m^T z_m$

$$
\begin{aligned}
\hat{y} &= \mathbf{Z}(\mathbf{Z^T Z})^{\mathbf{-1}}\mathbf{Z^T} y \\
&= \mathbf{Z}\left(\begin{bmatrix} z_0 \\ \vdots \\ z_M \end{bmatrix} \begin{bmatrix} z_0 & \dots & z_M \end{bmatrix} \right)^{-1} \begin{bmatrix} z_0 \\ \vdots \\ z_M \end{bmatrix} \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \\
&= \mathbf{Z} \begin{bmatrix} z_0^T z_0 & \dots & z_0^T z_j & \dots & z_0^T z_M \\ \vdots & & \vdots & & \vdots \\ z_i^T z_0 & \dots & z_i^T z_j & \dots & z_i^T z_M \\ \vdots & & \vdots & & \vdots \\ z_M^T z_0 & \dots & z_M^T z_j & \dots & z_M^T z_M \end{bmatrix}^{-1} \begin{bmatrix} z_0^T y \\ \vdots \\ z_M^T y \end{bmatrix} \\
&= \mathbf{Z} \begin{bmatrix} z_0^T z_0 & & \mathbf{0} \\ & \dots & \\ \mathbf{0} & & z_M^T z_M \end{bmatrix}^{-1} \begin{bmatrix} z_0^T y \\ \vdots \\ z_M^T y \end{bmatrix} (orthogonality) \\
&= \mathbf{Z} \begin{bmatrix} \mathbf{1}^T \mathbf{1} & & \mathbf{0} \\ & \dots & \\ \mathbf{0} & & z_M^T z_M \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{1}^T y \\ \vdots \\ z_M^T y \end{bmatrix} \\
&= \mathbf{Z} \begin{bmatrix} n & & \mathbf{0} \\ & \dots & \\ \mathbf{0} & & z_M^T z_M \end{bmatrix}^{-1} \begin{bmatrix} \sum_{i=1}^{n} y_i \\ \vdots \\ z_m^T y \end{bmatrix} \\
&= \mathbf{Z} \begin{bmatrix} \frac{1}{n} & & \mathbf{0} \\ & \dots & \\ \mathbf{0} & & \frac{1}{z_M^T z_M} \end{bmatrix} \begin{bmatrix} \sum_{i=1}^{n} y_i \\ \vdots \\ z_M^T y \end{bmatrix} \\
&= \mathbf{Z} \begin{bmatrix} \bar{y} \\ z_1^T y / z_1^T z_1 \\ \vdots \\ z_M^T y / z_M^T z_M \end{bmatrix} \\
&= \mathbf{Z} \begin{bmatrix} \bar{y} \\ \hat{\theta}_1 \\ \vdots \\ \hat{\theta}_M \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{1} & z_1 & \dots & z_M \end{bmatrix} \begin{bmatrix} \bar{y} \\ \hat{\theta}_1 \\ \vdots \\ \hat{\theta}_M \end{bmatrix} \\
&= \bar{y}\mathbf{1} + \sum_{m=1}^{M} \hat{\theta}_m z_m
\end{aligned}
$$

b)

If V is a matrix whose columns are the v_m and Θ is given by $(Z^TZ)^{-1}Z^Ty$ then we have

$$
\begin{aligned}
\mathbf{X}\hat{\beta}^{pcr}(M) &= \mathbf{Z}\Theta \\
\mathbf{X}\hat{\beta}^{pcr}(M) &= \mathbf{X}V\Theta \\
\hat{\beta}^{pcr}(M) &= V\Theta \\
\hat{\beta}^{pcr}(M) &= \begin{bmatrix} v_1 & ... & v_M \end{bmatrix} \begin{bmatrix} \theta_1 \\ \vdots \\ \theta_M \end{bmatrix} \\
\hat{\beta}^{pcr}(M) &= \begin{bmatrix} \sum_{i=1}^M V_{i1}\theta_i \\ \vdots \\ \sum_{i=1}^M V_{iM}\theta_M \end{bmatrix} \\
\hat{\beta}^{pcr}(M) &= \sum_{i=1}^M \begin{bmatrix} V_{i1}\theta_i \\ \vdots \\ V_{iM}\theta_i \end{bmatrix} \\
\hat{\beta}^{pcr}(M) &= \sum_{m=1}^M \hat{\theta}_m v_m
\end{aligned}
$$

V is an orthogonal matrix and if we let $M = p$, it is also invertible. We will see what the PCR estimators are in this special case

$$
\begin{aligned}
\hat{\beta}^{pcr}(p) &= V\Theta \\
&= V(\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{Z}^Ty \\
&= V((\mathbf{X}V)^T\mathbf{X}V)^{-1}(\mathbf{X}V)^Ty \\
&= V(V^T\mathbf{X}^T\mathbf{X}V)^{-1}V^T\mathbf{X}^Ty \\
&= VV^{-1}(\mathbf{X}^T\mathbf{X})^{-1}(V^T)^{-1}V^T\mathbf{X}^Ty \\
&= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^Ty \\
&= \hat{\beta}^{ls}
\end{aligned}
$$