Math 423
Linear Regression

Homework III

Frédéric Boileau

Prof. David A. Stephens

16th November 2016

a

```
setwd(dirname(rstudioapi::getActiveDocumentContext()$path))

## Error:  RStudio not running

smoke<-read.csv("data.csv",header=TRUE)
regressors = smoke[,1:ncol(smoke)-1]

CO = smoke$CO
tar = smoke$TAR
nico = smoke$NICOTINE
weight = smoke$WEIGHT
k = dim(regressors)[2]
n = dim(smoke)[1]
X = cbind(rep(1,n),regressors)
y = smoke[ncol(smoke)]
Xcs = scale(regressors,center = TRUE, scale = TRUE)
Xc = scale(regressors,center = TRUE, scale = FALSE)
```

```
fit.full = lm(CO ~ tar + nico + weight)
anova.full = anova(fit.full)
m.full = summary(fit.full)
SSres.full = (m.full$sigma)^2 * (m.full$df[2])
SSres.full_2 = sum(residuals(fit.full)^2)
MSres = sqrt(SSres.full/(n - k))
beta.0 = coefficients(fit.full)[1]
beta.1 = coefficients(fit.full)[2]
beta.2 = coefficients(fit.full)[3]
beta.3 = coefficients(fit.full)[4]
BigBeta = coefficients(fit.full)
```

$$SS_{Res}(\beta_0, \beta_1, \beta_2, \beta_3) = 33.5983067 \tag{1}$$

b

```
fit.noWeight = lm(CO ~tar + nico)
m.NoWeight = summary(fit.noWeight)
SSres.noWeight = sum(residuals(fit.noWeight)^2)
```

$$SS_{Res}(\beta_0, \beta_1, \beta_2) = 33.8740523 \tag{2}$$

c

The corresponding f-statistic is

$$F_3 = \frac{SS_R(\beta_3|\beta_2, \beta_1, \beta_0)}{MS_{res}} \tag{3}$$

```
f3 = anova.full[3,4] #directly from anova table
f3_2 = (SSres.noWeight - SSres.full)/(SSres.full / (n-k-1))
f3_3 = anova.full[3,2]/(SSres.full / (n-k))

f3

## [1] 0.1723496

f3_2

## [1] 0.1723496

f3_3

## [1] 0.1805568
```

d

```
fit.full_backwards = lm(CO ~ weight + nico + tar)
anova.full_backwards = anova(fit.full_backwards)
anova.full_backwards

## Analysis of Variance Table
##
## Response: CO
##           Df Sum Sq Mean Sq F value    Pr(>F)
## weight     1 178.75  178.75 111.722 7.277e-10 ***
## nico       1 465.74  465.74 291.104 8.762e-14 ***
## tar        1  28.14   28.14  17.588 0.0004088 ***
## Residuals 21  33.60    1.60
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We have

$$\overline{SS}_R(\beta_3|\beta_0) = 178.7470223 \tag{4}$$

$$\overline{SS}_R(\beta_2|\beta_0, \beta_3) = 465.7428416 \tag{5}$$

$$\overline{SS}_R(\beta_1|\beta_0, \beta_3, \beta_2) = 28.1394854 \tag{6}$$

e

```
anova.reduced = anova(fit.noWeight)
anova.reduced
```

```
## Analysis of Variance Table
##
## Response: CO
##           Df Sum Sq Mean Sq  F value    Pr(>F)
## tar        1 672.29  672.29 436.6261 5.323e-16 ***
## nico       1   0.07    0.07   0.0438    0.8361
## Residuals 22  33.87    1.54
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We have

$$\overline{SS}_R(\beta_1|\beta_0) = 672.2860929 \tag{7}$$

$$\overline{SS}_R(\beta_2|\beta_0, \beta_1) = 0.0675108 \tag{8}$$

f

```
anova.noWeight = anova(fit.noWeight)
anova.noWeight
```

```
## Analysis of Variance Table
##
## Response: CO
##           Df Sum Sq Mean Sq  F value    Pr(>F)
## tar        1 672.29  672.29 436.6261 5.323e-16 ***
## nico       1   0.07    0.07   0.0438    0.8361
## Residuals 22  33.87    1.54
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

$$F_2 = 0.0438459 \tag{9}$$

g

We want to know if regressing the CO value against tar and nicotine is better than no regressing at all.

```
fit.nothing = lm(CO ~ 1)
anova.nothing = anova(fit.nothing)
SSres.nothing = anova.nothing[1,2]
r = 2
f_last = ((SSres.nothing - SSres.noWeight) / r) /
          (SSres.noWeight/(n-k-r))
f_last

## [1] 198.4863
```