Q8

Zhen Zhang

4/25/2020

```
setwd("../data/")
auto = read.csv("Auto.csv", na.strings="?")
auto$origin = factor(auto$origin, 1:3, c("US", "Europe", "Japan"))
names(auto)
```

```
## [1] "mpg"          "cylinders"    "displacement" "horsepower"   "weight"
## [6] "acceleration" "year"         "origin"       "name"
```

```
head(auto)
```

```
##   mpg cylinders displacement horsepower weight acceleration year origin
## 1  18         8          307        130   3504         12.0   70     US
## 2  15         8          350        165   3693         11.5   70     US
## 3  18         8          318        150   3436         11.0   70     US
## 4  16         8          304        150   3433         12.0   70     US
## 5  17         8          302        140   3449         10.5   70     US
## 6  15         8          429        198   4341         10.0   70     US
##                       name
## 1 chevrolet chevelle malibu
## 2         buick skylark 320
## 3        plymouth satellite
## 4             amc rebel sst
## 5               ford torino
## 6          ford galaxie 500
```

(a) You should see that there are two dummy variables for the origin variable. If origin were dropped from the model (i.e., the two dummies were set equal to 0), by how much would RSS increase?

```
fit = lm(log(mpg) ~ log(weight) + year + I(year^2) + as.factor(origin), auto)
drop1(fit, test="F")
```

```
## Single term deletions
##
## Model:
## log(mpg) ~ log(weight) + year + I(year^2) + as.factor(origin)
```

```
##                   Df Sum of Sq     RSS     AIC   F value    Pr(>F)
## <none>                          5.0449 -1721.1
## log(weight)        1   13.5126 18.5574 -1206.0 1047.2853 < 2.2e-16 ***
## year               1    0.1667  5.2116 -1710.2   12.9211 0.0003664 ***
## I(year^2)          1    0.2132  5.2580 -1706.7   16.5224 5.814e-05 ***
## as.factor(origin)  2    0.1857  5.2306 -1710.8    7.1968 0.0008523 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

0.1857

(b) Can you reject the null hypothesis that both origin dummies are 0, so that none of the origin levels have different effects?

$0.0008523 < .05$, no.