

Q1

Zhen Zhang

4/20/2020

```
library(car)
```

```
## Loading required package: carData
```

```
auto = read.csv("../data/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) Regress mpg on cylinders, displacement, weight, and year. Comment on the signs of the estimated coefficients and note which are significantly different from 0. What is value of R²?

```
fit = lm(mpg ~ cylinders + displacement + weight + year, auto)
summary(fit)
```

```
##
## Call:
## lm(formula = mpg ~ cylinders + displacement + weight + year,
##     data = auto)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -8.995 -2.270 -0.165  2.053 14.368
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -14.076941   4.055159  -3.471 0.000575 ***
## cylinders      -0.289589   0.329225  -0.880 0.379611
## displacement   0.004973   0.006701   0.742 0.458425
## weight        -0.006702   0.000572 -11.717 < 2e-16 ***
## year           0.764751   0.050684  15.089 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.436 on 392 degrees of freedom
## Multiple R-squared:  0.8091, Adjusted R-squared:  0.8072
## F-statistic: 415.5 on 4 and 392 DF,  p-value: < 2.2e-16
```

$P(\text{weight})$ and $P(\text{year}) < .05$, so weight and year are significantly different from 0. And Multiple R-squared: 0.8091.

(b) Compute the variance inflation factors. What do they tell you?

```
vif(fit)

##      cylinders displacement      weight      year
##      10.524432      16.406259      7.888061      1.173000
```

Only year shows no multicollinearity.

(c) Drop weight from the model. What happens to the parameter estimates and R²?

```
fit2 = lm(mpg ~ cylinders + displacement + year, auto)
summary(fit2)

##
## Call:
## lm(formula = mpg ~ cylinders + displacement + year, data = auto)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
```

```
## -10.0801 -2.6445 -0.2925 2.1004 14.9103
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -18.199719  4.688296  -3.882 0.000122 ***
## cylinders    -0.620910  0.380657  -1.631 0.103658
## displacement -0.041545  0.006265  -6.632 1.1e-10 ***
## year          0.699324  0.058461  11.962 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.988 on 393 degrees of freedom
## Multiple R-squared:  0.7423, Adjusted R-squared:  0.7403
## F-statistic: 377.3 on 3 and 393 DF, p-value: < 2.2e-16
```

Displacement becomes significant. And R square is decreased.

(d) Drop weight and displacement from the model. What happens to the parameter estimates and R²?

```
fit3 = lm(mpg ~ cylinders + year, auto)
summary(fit3)
```

```
##
## Call:
## lm(formula = mpg ~ cylinders + year, data = auto)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -10.6462  -2.8847  -0.1399   2.5095  15.6875
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -17.30285    4.93534  -3.506 0.000507 ***
## cylinders    -3.00405    0.13223 -22.718 < 2e-16 ***
## year          0.75289    0.06098  12.347 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.2 on 394 degrees of freedom
## Multiple R-squared:  0.7135, Adjusted R-squared:  0.712
## F-statistic: 490.5 on 2 and 394 DF, p-value: < 2.2e-16
```

Cylinders becomes significant. And R square is decreased.