Homework Assignment 2

Trevor Klar and Blaine Quackenbush

October 26, 2021

Linear regression (12 pts)

In this problem, we will make use of the *Auto* data set, which is part of the ISLR package and can be directly accessed by the name `Auto` once the `ISLR` package is loaded. The dataset contains 9 variables of 392 observations of automobiles. The qualitative variable **origin** takes three values: 1, 2, and 3, where 1 stands for American car, 2 stands for European car, and 3 stands for Japanese car.

```
head(Auto)
```

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

Here we just remind ourselves how `origin` is coded:

Origin	1	2	3
	American	European	Japanese

1. (2 pts) Fit a linear model to the data, in order to predict mpg using all of the other predictors except for name. Present the estimated coefficients. (2 pts) With a 0.01 threshold, comment on whether you can reject the null hypothesis that there is no linear association between mpg with any of the predictors.

Here we fit a linear model to the data, using all variables except `name` as predictors for `mpg`. We will also consider, with a 0.01 threshold, whether there is a statistically significant linear association between mpg and any of the predictors.

```
auto.lmfit <- lm(mpg ~ cylinders + displacement + horsepower + weight + acceleration + year + origin, Au
summary(auto.lmfit)
```

```
##
## Call:
## lm(formula = mpg ~ cylinders + displacement + horsepower + weight +
##     acceleration + year + origin, data = Auto)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -9.590 -2.157 -0.117  1.869 13.060
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -1.72e+01   4.64e+00   -3.71  0.00024 ***
## cylinders    -4.93e-01   3.23e-01   -1.53  0.12780
## displacement  1.99e-02   7.51e-03    2.65  0.00844 **
## horsepower   -1.70e-02   1.38e-02   -1.23  0.21963
## weight       -6.47e-03   6.52e-04   -9.93  < 2e-16 ***
## acceleration  8.06e-02   9.88e-02    0.82  0.41548
## year          7.51e-01   5.10e-02   14.73  < 2e-16 ***
## origin        1.43e+00   2.78e-01    5.13  4.7e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.33 on 384 degrees of freedom
## Multiple R-squared:  0.821,  Adjusted R-squared:  0.818
## F-statistic:  252 on 7 and 384 DF,  p-value: <2e-16
```

Note that the F-statistic is quite large (284), and indeed the p-value associated with this F-statistic is is less than 2×10^{-16}. This is much smaller than 0.01, so we conclude (with 99% certainty) that there is a linear relationship between `mpg` and at least one of these variables.

2. (2 pts) Take the whole dataset as training set. What is the training mean squared error of this model?

```
MSE <- function(model) {
  mean(residuals(model)^2)
}

MSE(auto.lmfit)
```

```
## [1] 10.85
```

3. (2 pts) What gas mileage do you predict for an European car with 4 cylinders, displacement 122, horsepower of 105, weight of 3100, acceleration of 32, built in the year 1991? (Be sure to check how year is coded in the dataset).

```
predict(auto.lmfit, data.frame(cylinders = 4, displacement = 122, horsepower = 105, weight = 3100, accel
```

```
##     1
## 35.14
```

4. (1 pts) On average, holding all other covariates fixed, what is the difference between the mpg of a Japanese car and the mpg of an American car? (1 pts) What is the difference between the mpg of a European car and the mpg of an American car? # simple table creation here | Origin | 1 | 2 | 3 | |———|:——-:|:——:|:——:| | |American|European|Japanese|

5. (2 pts) On average, holding all other predictor variables fixed, what is the change in mpg associated with a 10-unit increase in displacement?