R Lab #3 – Solutions

1. Section 5.3.1: The Validation Set Approach

Use "set.seed(3)" then immediately select a random training set (without replacement) of 196 observations out of the original 392 observations in the Auto data set. Fit linear, quadratic, and cubic regressions to the 196 observations in the training set and then calculate the MSE of the other 196 observations in the validation set. Enter the values of estimated test MSE to the linear, quadratic, and cubic regression models (round to two decimal places).

a) Linear: _____
b) Quadratic: _____
c) Cubic: _____

Solution: Linear (26.29), Quadratic (21.50), Cubic (21.51)

```
set.seed(3)
train <- sample(392, 196)
lm.fit <- lm(mpg ~ horsepower, data = Auto, subset = train)
mean((mpg - predict(lm.fit, Auto))[-train]^2) # MSE (Test Error) = 26.29
lm.fit2 <- lm(mpg ~ poly(horsepower, 2), data = Auto, subset = train)
mean((mpg - predict(lm.fit2, Auto))[-train]^2) # MSE (Test Error) = 21.50
lm.fit3 <- lm(mpg ~ poly(horsepower, 3), data = Auto, subset = train)
mean((mpg - predict(lm.fit3, Auto))[-train]^2) # MSE (Test Error) = 21.51
```

2. Section 5.3.2: Leave-One-Out Cross-Validation

What is the leave-one-out cross-validation estimate of the test MSE for the sixth-order polynomial regression model for the Auto data set (using "mpg" as the response variable and "horsepower" as the predictor variable)? Round your answer to 2 decimal places.

Solution: 18.98

```
glm.fit <- glm(mpg ~ poly(horsepower, 6), data = Auto)
cv.err <- cv.glm(Auto, glm.fit)
cv.err$delta # 18.98 18.98
```

3. Section 5.3.2: Leave-One-Out Cross-Validation

True or False? If we want to reproduce leave-one-out cross-validation results precisely at a later time, then we need to set a seed for R's random number generator using the "set.seed" function.

Solution: False, since there is no random component to leave-one-out cross-validation.