Quiz #2 (Ch. 3) – Solutions

1. Consider the problem of predicting how well a student does in her second year of college/ university, given how well she did in her first year. Specifically, let X_1 be equal to the number of "A" grades (including A-, A, and A+ grades) that the student receives in her first year of college (freshmen year). We would like to predict the value of Y, which we define as the number of "A" grades she gets in her second year (sophomore year). Recall that in linear regression, our linear model is $f(X) = b_0+b_1X_1$.

Here is our training dataset: X_1 = c(3, 2, 4, 0); Y = c(4, 1, 3, 1).

Based on this training set, what is the residual sum of squares (RSS) for $b_0=0$, $b_1=1$? (Your answer should be a whole number, so do not write any decimal places.)

Solution: 4

When $b_0=0$ and $b_1=1$, the fitted values are Y-hat = 0 + 1 * X_1 = X_1. As a result, RSS = sum((Y – Y-hat)^2) = sum((Y - X_1)^2) = (4 – 3)^2 + (1 – 2)^2 + (3 – 4)^2 + (1 – 0)^2 = 4

2. The ordinary least squares solution for the situation described in Question 1 is $b_0=0.7714$, $b_1=0.6571$, corresponding to RSS=2.97. What is f(4)? (Round your answer to one decimal place.)

Solution: 3.4

When $b_0=0.7714$ and $b_1=0.6571$, f(4) = 0.7714 + 0.6571(4) = 3.4 (to one decimal place).

3. Suppose that for the situation described in Questions 1 and 2 we have a second predictor variable, X_2, available and we fit the model $f(X)=b_0+b_1X_1+b_2X_2$ using ordinary least squares. What will the value of RSS be for this model? (Select the single best answer.)

a) Less than (or equal to) 2.97
b) Equal to 2.97
c) Greater than (or equal to) 2.97
d) It depends on whether X_2 is correlated with Y
e) It depends on whether X_2 is correlated with X_1

Solution: A

The answer for this question was (a): Less than (or equal to) 2.97. When adding a predictor to a multiple linear regression model, RSS always either decreases or stays the same. In this case, for example, think of RSS for the model with just X_1 as measuring the sum of squared vertical distances between the data points (Y) and their fitted values (Y-hat) on the regression line. When we add X_2 to the model, the regression line becomes a regression plane (like in Figure 3.5 on page 81). Now, RSS for the model with X_1 and X_2 is the sum of squared vertical distances

between the data points (Y) and their fitted values (Y-hat) on the regression plane. This will be less than (or equal to) RSS for the model with just X_1 since we have a second dimension (given by the X_2 axis) available to tilt the regression plane to get closer to the data points. It doesn't matter whether X_2 is correlated with Y or X_1, this result always holds. The only correlation that is relevant here is if the residuals from the simple linear regression model with just X_1 were uncorrelated with X_2. Then RSS for both models would be equal.

4. Suppose that for some linear regression problem (say, predicting housing prices), we have a training set, and for our training set we managed to find values of b_0 and b_1 such that RSS=0. Which of the statements below must then be true? (Check all that apply.)

a) We can perfectly predict the value of Y even for new examples that we have not yet seen. (e.g., we can perfectly predict prices of new houses that we have not yet seen.)
b) This is not possible. By the definition of RSS, it is not possible for there to exist b_0 and b_1 so that RSS=0.
c) For these values of b_0 and b_1 that satisfy RSS=0, we have that $f(X_i)=Y_i$ for every training example (X_i, Y_i).
d) For this to be true, we must have $b_0=0$ and $b_1=0$ so that $f(X)=0$.

Solution: C

- (a) is false because even if RSS=0 in the training set, the observations in the test set are unlikely to fit this model exactly (this is what happens in cases of extreme over-fitting where the training set is fit perfectly but the model is terrible at predicting test data).
- (b) is false because it is possible (albeit unlikely) that we could have some data that lies exactly along a straight line, so that there exists b_0 and b_1 for which RSS=0.
- (c) is true because we are summing non-negative quantities so the only way to have this equal zero is if all the individual residuals are zero.
- (d) is false because if b_0 and b_1 are both zero so that f(X)=0, then RSS = sum(Y^2) and this would only be zero if all the response values were zero.

5. Consider using multiple linear regression modeling to predict the amount that a customer will spend in response to receiving a catalog. You will then select customers with high predicted amounts to mail the catalog to. Suppose you have enough data to partition the data into training, validation, and test samples. Which of the following approaches would never be appropriate during the analysis? (Select all that apply.)

a) Use the training sample to fit different models under consideration.
b) Compare RSS values for different models in the training sample to select a subset of the predictors to use in the final model.
c) Use mean squared error in the validation sample to compare models.
d) Consider transforming the response or predictor variables to improve models.

Solution: B

- (a) is appropriate because this is exactly what the training sample is used for.
- (b) is not appropriate because using RSS as a criteria in this way would always tell you to just use all the predictors.
- (c) is appropriate because this is exactly what the validation sample is used for.
- (d) is appropriate because we may be able to find better models by transforming some variables.

6. Suppose we have a data set with five predictors, X_1 = GPA, X_2 = IQ, X_3 = Gender (1 for Female and 0 for Male), X_4 = Interaction between GPA and IQ, and X_5 = Interaction between GPA and Gender. The response is starting salary after graduation (in thousands of dollars). Suppose we use ordinary least squares to fit the model, and get $\beta_0 = 50$, $\beta_1 = 20$, $\beta_2 = 0.07$, $\beta_3 = 35$, $\beta_4 = 0.01$, $\beta_5 = -10$ (assume these β's have "hats"). Which of the following is a true statement? (Select the single best answer.)

a) For a fixed value of IQ and GPA, males earn more on average than females.
b) For a fixed value of IQ and GPA, females earn more on average than males.
c) For a fixed value of IQ and GPA, males earn more on average than females provided that the GPA is higher than 3.5.
d) For a fixed value of IQ and GPA, females earn more on average than males provided that GPA is higher than 3.5

Solution: C

Note that the female intercept is 35 units more than the male intercept (since $\beta_3 = 35$) but the female GPA slope is 10 units less than the male GPA slope (since $\beta_5 = -10$). The net difference between females and males is thus 35 – 10*GPA, which is positive for GPAs less than 3.5, equal to zero for GPAs equal to 3.5, and negative for GPAs more than 3.5. In other words, females earn more on average than males for GPAs less than 3.5 but less for GPAs more than 3.5. (Again, note that these β's have "hats".)

7. Predict the salary (in thousands of dollars) of a female with IQ of 110 and a GPA of 4.0 for the situation described in Question 6. Round your answer to the nearest whole number.

Solution: 137

Predicted salary = 50 + 20(4.0) + 0.07(110) + 35(1) + 0.01(4.0)(110) – 10(4.0)(1) = 137

8. For the situation described in Questions 6 and 7, is the following statement true or false? "Since the coefficient for the GPA/IQ interaction term is very small, there is very little evidence of an interaction effect."

Solution: False

The magnitude of the t-statistic (which depends on the coefficient estimate's standard error) determines the significance of a regression coefficient. The coefficient estimate tells us nothing by itself since it's scale depends on the measurement scale of the predictor.