**Section 3.7: Exercise 6**

**Question:** Using (3.4), argue that in the case of simple linear regression, the least squares line always passes through the point (¯x, y¯).

**Solution:**

Yes, in case of simple linear regression the line of best fit always has to pass through from the points of x and y bars, if there is a strong relation between the variables. From the summary of the data it is clear that the variables no of units sold and advertising medium are correlated due to significance of the regression slope or coefficient.

**Section 3.7: Exercises 1,3,4-a**

**Question 1:** Describe the null hypotheses to which the p-values given in Table 3.4 correspond. Explain what conclusions you can draw based on these p-values. Your explanation should be phrased in terms of sales, TV, radio, and newspaper, rather than in terms of the coefficients of the linear model.

**Solution:**

The null hypothesis ***H0*** : ***Bi* = 0.**

Intercept: **p < 0.0001** means that the null hypothesis is rejected, i.e., sales is not zero when TV, radio and newspaper are zero.

TV: **p < 0.0001** means that the null hypothesis is rejected. The change in TV will affect sales.

Radio: **p < 0.0001**means that the null hypothesis is rejected. The change in Radio will affect the sales value.

Newspaper: **p = 0.8599**means the null hypothesis is accepted, there is no relationship between sales and newspaper.

**Question 3:** Suppose we have a data set with five predictors, X1 = GPA, X2 = IQ, X3 = Level (1 for College and 0 for High School), X4 = Interaction between GPA and IQ, and X5 = Interaction between GPA and Level. The response is starting salary after graduation (in thousands of dollars). Suppose we use least squares to fit the model, and get βˆ0 = 50, βˆ1 = 20, βˆ2 = 0.07, βˆ3 = 35, βˆ4 = 0.01, βˆ5 = −10.

(a) Which answer is correct, and why?

i. For a fixed value of IQ and GPA, high school graduates earn

more, on average, than college graduates.

ii. For a fixed value of IQ and GPA, college graduates earn more, on average, than high school graduates.

iii. For a fixed value of IQ and GPA, high school graduates earn more, on average, than college graduates provided that the GPA is high enough.

iv. For a fixed value of IQ and GPA, college graduates earn more, on average, than high school graduates provided that the GPA is high enough.

**Solution:** ii is the solution to this problem since B3 = 35 for college graduates and B3 = 0 for high school graduates. Thus college graduates earn more.

(b) Predict the salary of a college graduate with IQ of 110 and a GPA of 4.0.

**Solution:** College Graduate Salary = 50+20\*4+0.07\*110+35\*1+0.01\*4\*110-10\*4\*1 = $137.1 in the thousands.

(c) True or false: Since the coefficient for the GPA/IQ interaction term is very small, there is very little evidence of an interaction effect. Justify your answer.

**Solution:** False, because the significance of the interaction is dependent on the standard error and the test statistic. Looking at the coefficient alone, we cannot tell whether there is no evidence of an interaction.

**Question 4-a: .** I collect a set of data (n = 100 observations) containing a single predictor and a quantitative response. I then fit a linear regression model to the data, as well as a separate cubic regression, i.e. Y = β0 + β1X + β2X2 + β3X3 + ϵ.

(a) Suppose that the true relationship between X and Y is linear, i.e. Y = β0 + β1X + ϵ. Consider the training residual sum of squares (RSS) for the linear regression, and also the training RSS for the cubic regression. Would we expect one to be lower than the other, would we expect them to be the same, or is there not enough information to tell? Justify your answer.

**Solution:** Given that the relationship between X and Y is linear , one can easily infer that the least square fitted line would be close to the true regression line when you compare the same with the cubic regression line. Now , since RSS is the residuals of squared sum , it would be safe to assume that the RSS for the liner would be lower when you compare the same with the cubic regression line