Ch3

June 5, 2022

1 Chapter 3

1.1 q1

The null hypotheses associated with table 3.4 are that advertising budgets of "TV", "radio" or "newspaper" do not have an effect on sales. More precisely $H_0^{(1)} : \beta_1 = 0, H_0^{(2)} : \beta_2 = 0$ and $H_0^{(3)} : \beta_3 = 0$. The corresponding p-values are highly significant for "TV" and "radio" and not significant for "newspaper"; so we reject $H_0^{(1)}$ and $H_0^{(2)}$ and we do not reject $H_0^{(3)}$. We may conclude that newspaper advertising budget do not affect sales.

1.2 q3

a) The least square line is given by

$$\hat{y} = 50 + 20GPA + 0.07IQ + 35Gender + 0.01GPA \times IQ - 10GPA \times \text{ Gender}$$

which becomes for the males

$$\hat{y} = 50 + 20GPA + 0.07IQ + 0.01GPA \times IQ$$

and for the females

$$\hat{y} = 85 + 10GPA + 0.07IQ + 0.01GPA \times IQ$$

So the starting salary for males is higher than for females on average iff $50+20GPA \geq 85+10GPA$ which is equivalent to $GPA \geq 3.5$. Therefore iii. is the right answer.

b) It suffices to plug in the given values in the least square line for females given above and we obtain

$$\hat{y} = 85 + 40 + 7.7 + 4.4 = 137.1$$

which gives us a starting salary of 137100$.

c) False. To verify if the GPA/IQ has an impact on the quality of the model we need to test the hypothesis $H_0 : \hat{\beta}_4 = 0$ and look at the p-value associated with the t or the F statistic to draw a conclusion.

1.3 q4

a) Without knowing more details about the training data, it is difficult to know which training RSS is lower between linear or cubic. However, as the true relationship between X and Y is linear, we may expect the least squares line to be close to the true regression line, and consequently the RSS for the linear regression may be lower than for the cubic regression.

b) In this case the test RSS depends upon the test data, so we have not enough information to conclude. However, we may assume that polynomial regression will have a higher test RSS as the overfit from training would have more error than the linear regression.

c) Polynomial regression has lower train RSS than the linear fit because of higher flexibility: no matter what the underlying true relationshop is the more flexible model will closer follow points and reduce train RSS.

d) There is not enough information to tell which test RSS would be lower for either regression given the problem statement is defined as not knowing "how far it is from linear". If it is closer to linear than cubic, the linear regression test RSS could be lower than the cubic regression test RSS. Or, if it is closer to cubic than linear, the cubic regression test RSS could be lower than the linear regression test RSS. It is dues to bias-variance tradeoff: it is not clear what level of flexibility will fit data better.