Mark Gameng

CS 422 – HW 2

Part 1

1. The null hypotheses are that TV, radio, or newspaper advertising budgets do not influence the sales of the product. The p value for TV and radio are less than 0.0001 while newspaper has a p value of 0.8599. Thus, the p values of TV and radio are statistically significant while newspaper is not. This means that for newspaper, we accept the null hypothesis that newspaper advertising budget do not affect the sales. However, TV and radio advertising does affect the sales of a product.

3. Given values: X1, X2, X3, X4, X5 = GPA, IQ, Gender, GPA ~ IQ, GPA ~ Gender. Response is starting salary after graduation. B0, B1, B2, B3, B4, B5 = 50, 20, 0.07, 35, 0.01, -10.
   This results in the equation:
   Salary = 50 + 20*GPA + 0.07*IQ + 35*Gender + 0.01*GPA*IQ – 10*GPA*Gender
   Thus, for males (Gender = 0)
   SalaryM = 50 + 20*GPA + 0.07*IQ + 0.01*GPA*IQ
   for females (Gender = 1)
   SalaryF = 50 + 20*GPA + 0.07*IQ + 35 + 0.01*GPA*IQ – 10*GPA
   = 85 + 10*GPA + 0.07*IQ + 0.01*GPA*IQ
   a. Setting the two equations together, we get that 50 + 20*GPA = 85 + 10*GPA which simplifies to GPA = 3.5. Thus at 3.5 GPA, salary is equal, if IQ is also equal. Above 3.5 GPA, the starting salary for males is higher. **Thus, iii is correct, for a fixed value of IQ and GPA, males earn more on average than females provided that the GPA is high enough.**
   b. With IQ of 110 and GPA of 4.0, substituting to the equation for females, we get that the predicted salary is 137.1, which is $137,100.
   c. False, the coefficient doesn't matter. The p value is what matters. If the p value is small enough for the null hypothesis B4 = 0 (Interaction between GPA and IQ), then there is statistically significant evidence that there is an effect.

4.
   a. Since the true relationship between X and Y is linear, the RSS for the linear regression is expected to be lower than the RSS for the cubic regression. A smaller residual sum of squares means a better fit of the model to the data. Thus, when the true relationship is linear, a linear regression would most likely result in a lower RSS than for cubic regression.