**Computational Statistics II**

**Subash Kharel**

20 Feb 2020

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.**  
   *Answer: Table 3.4 shows the null hypothesis which says advertising budget on TV, radio or newspaper do not have any contribution on sales. Studying the p value at the last column of table we can reject the null hypothesis assumed for TV and radio. But for newspaper, null hypothesis cannot be rejected since the p-value is significant and we can say budget on advertising through newspaper can be discarded.*
2. **Question number 3: Suppose we have a data set with five predictors, X1 = GPA, X2 = IQ, X3 = Gender (1 for Female and 0 for Male), X4 = Interaction between GPA and IQ, and X5 = Interaction between GPA and Gender. 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.**
   1. **Which answer is correct, and why?**
      1. **For a fixed value of IQ and GPA, males earn more on average than females.**  
         *Answer: No, we cannot conclude that. It depends upon the value of GPA. If the value of GPA is < 3.5 female earns more than male. If its > 3.5 male earns more than females. And if its = 3.5, male and female will earn same amount.*
      2. **For a fixed value of IQ and GPA, females earn more on average than males.**  
         *Answer: No, we cannot say that. Females only earn more on average than males if the value of GPA is < 3.5.*
      3. **For a fixed value of IQ and GPA, males earn more on average than females provided that the GPA is high enough.**  
         *Answer: Yes, male will only earn more on average than females only when GPA is higher I.e. > 3.5.*
      4. **For a fixed value of IQ and GPA, females earn more on average than males provided that the GPA is high enough.**  
         *Answer: No, females will earn more on average than males when GPA is lower enough. The threshold value is 3.5, i.e. the value of GPA must be less than 3.5.*
   2. **Predict the salary of a female with IQ of 110 and a GPA of 4.0.**  
      *Answer: The salary of a female is calculated below:*  
      *We have the equation*,  
       **Salary = 50 + 20 \* GPA + 0.07 \* IQ + 35 \* Gender + 0.01 \* GPA \* IQ – 10 \* GPA \* Gender**

**= 50 + 20 \* 4 + 0.07 \* 110 + 35 \* 1 + 0.01 \* 4 \* 110 – 10 \* 4 \* 1**

**= 137.1 Answer**

* 1. **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.**  
     Answer: No, we cannot say that the interaction effect is low looking at the coefficients only. If we compare it to the coefficient of IQ, it also has significant significance to the salary. This can be confirmed by the null hypothesis test only.

1. **Question number 4: 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.**
   1. **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 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.**  
      *Answer: As the true relationship between X and Y is linear, we can guess the training RSS for linear regression to be lower than that of cubic regression. The reason for saying this is that, linear regression can perfectly fit the training data (RSS will be 0) but cubic regression may not (producing some value for RSS). Still we cannot confirm this without knowing the nature of training data. Since sometime same linear data may be represented by non-linear equation which can be shown by the figure below:*
   2. **Answer (a) using test rather than training RSS.**  
      *Answer: Cubic (Non-linear) regression are supposed to have high chance of overfitting the training data and may not predict test data well resulting in high test RSS. Since the relation is linear, we expect linear regression to produce less test RSS. But none of this can be confirmed unless the nature of test data is determined.*
   3. **Suppose that the true relationship between X and Y is not linear, but we don’t know how far it is from linear. Consider the training RSS for the linear regression, and 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.**  
      *Answer: The higher tendency of polynomial regression to fit the training data well will result in less RSS compared to that of linear regression. Thus, we can expect RSS for cubic regression to be lower.*
   4. **Answer (c) using test rather than training RSS.**  
      *Answer: As the distance from linear is not known we cannot compare the value for test RSS. If its near to cubic, cubic regression will produce lesser RSS and if its near to linear, it will produce lesser RSS for linear regression.*