1. Describe the null hypotheses to which the p-values, given in the following table, correspond. Explain what conclusions can be drawn based on the p-values. Your explanation should be phrased in terms of sales, TV, radio, and newspaper.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Coefficient | Std. error | t-stat | p-value |
| Intercept | 2.939 | 0.3119 | 9.42 | < 0.0001 |
| TV | 0.046 | 0.0014 | 32.81 | < 0.0001 |
| radio | 0.189 | 0.0086 | 21.89 | < 0.0001 |
| newspaper | -0.001 | 0.0059 | -0.18 | 0.8599 |

**Answer:** The null hypothesis is testing if every estimated coefficient is equal to 0, given the data set used. We reject the null if any coefficient’s confidence interval does not contain 0. We can conclude that the intercept, TV, and radio are all not 0 with reasonable confidence.

1. Carefully explain the differences between KNN classifier and KNN regression methods

**Answer:** classification tries to predict which group/class a predictor belongs to by calculating the local (size K) probability, whereas regression predicts a numeric value by calculating the local average.

1. Suppose we have a data set with five predictors, GPA, IQ, Gender (1 for female), Interaction between GPA and IQ, and Interaction between GPA and Gender. The response is starting salary after graduation (in 1000’s). Suppose we have a least squares and get B0 = 50, B1=20, 0.07, 35, 0.01, -10 resp.
   1. Which answer is correct and why?
      1. For a fixed value of IQ and GPA, males earn more on average than females.
      2. For a fixed value of IQ and GPA, females earn more on average than males
      3. For a fixed value of IQ and GPA, males earn more on average than females provided that the GPA is high enough
      4. For a fixed value of IQ and GPA, females earn more on average than males provided that the GPA is high enough.

**Answer:** iii

Formula for a female:

Formula for a male:

Because of the 20 term in the male’s GPA, sufficiently high GPA will increase a male’s GPA much more.

* 1. Predict the salary of a female with IQ of 110 and GPA of 4.0

**Answer:** = 85 + 4(10 + (0.01\*110)) + (0.07\*110) = 137.1

* 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 and justify.

**Answer:** False, the small coefficient value only implies there is a small change in the model, it does not mean the interaction term is not important. P-value is used to determine whether a parameter should be used in the model.

1. A set of data (n=100) with p=1 and Y quantitative. A linear regression model and a cubic regression is fit.
   1. Suppose the true relationship between X and Y is linear. Consider the training RSS for both models. Would we expect one to be lower than the other, the same, or not enough information?

**Answer:** We would expect the RSS of the linear model to be lower because it more accurately fits the true relationship.

* 1. Same as a, but test instead of training.

**Answer:** The cubic model likely overfit the training data, therefore, we would expect the RSS of the linear model to be lower.

* 1. Suppose the true relationship is not linear, but unsure how far from linear, same as a.

**Answer:** The RSS of a nonlinear model would be lower than the RSS of a nonlinear model due to its flexibility, and fitting the training set better.

* 1. Same as c, but with the test set instead of training.

**Answer:** There is not enough information to say.

1. Using 3.4, argue that in the case of simple linear regression, the least squares line always passes through the point .

**Answer:** is the linear regression formula and because , we can substitute for and get the following equation . With canceling we can show that .