Unit 1 Theory

You may answer the questions by either using LaTex or inserting pictures.

## Pictures

To insert a picture: 1. Copy an image to the image directory in your local Unit 1 Assignment folder. 2. Link to the image with the following code:

![sample picture](images/sample.png)



sample picture

## LaTex

To use LaTex, use

$equation$ for inline equations in the text  
and $$equation$$ to display an equation on a new line.

Example: This is an inline equation .

This is to display an equation:

## Question 1:

True/False Explain

A high p-value associated with a test statistic indicates that there is a high probability that we reject the null in favor of the alternative.

### Your response

## Question 2:

1. Is linear regression an example of supervised or unsupervised learning? Explain.
2. Is linear regression an example of a parametric or nonparametric approach? Explain.

### Your response

## Question 3:

### ISLR Chapter 3 Question 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 = 50, = 20, = 0.07, = 35, = 0.01, = −10.

1. Which answer is correct, and why?

* For a fixed value of IQ and GPA, males earn more on average than females.
* For a fixed value of IQ and GPA, females earn more on average than males.
* For a fixed value of IQ and GPA, males earn more on average than females provided that the GPA is high enough.
* For a fixed value of IQ and GPA, females earn more on average than males provided that the GPA is high enough.

1. Predict the salary of a female with IQ of 110 and a GPA of 4.0.
2. 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.

### Your response

## Question 4:

### ISLR Chapter 3 Question 4:

Suppose 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.

1. Suppose that the true relationship between X and Y is linear, i.e. . 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.
2. Answer (a) using test rather than training RSS.
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 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.
4. Answer part 3 using test rather than training RSS.

### Your response

## Question 5:

Suppose that we want to estimate the model:

Show that .

### Your response