CS 422 Homework2

Xingli Li, Illinois Institute of Technology

1. Exercises

Question 1.1

Section 3.7 (Exercises), page 123

Exercise 6.

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

Answer:

The least square line equation:

$$y_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

we obtain:

$$y = \hat{\beta}_0 + \hat{\beta}_1 \bar{x}$$

$$y = \bar{y} - \hat{\beta}_1 + \hat{\beta}_1 \bar{x} = \bar{y}$$

So the least squares line always passes through the point ($\bar{x}$,$\bar{y}$).

Question 1.2

Section 3.7 (Exercises), page 120

Exercises 1, 3, 4-a.

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:
$$ H_0^{(1)}: \beta _1=0$$
$$ H_0^{(2)}: \beta _2=0$$
$$ H_0^{(3)}: \beta _3=0$$
 The null hypotheses associated with table 3.4 don't have an effect on sales.
their respective fields are TV, radio ,newspaper,   the TV, radio corresponding p-values are highly significant , the newspaper p-value is not significant. So I think that newspaper advertising budget do not affect sales.

3 . Suppose we have a data set with five predictors...

  (a) Which answer is correct, and why?

i. For a fixed value of IQ and GPA, males earn more on average than females.
ii. For a fixed value of IQ and GPA, females earn more on average than males.
iii. For a fixed value of IQ and GPA, males earn more on averag than females provided that the GPA is high enough.
iv. For a fixed value of IQ and GPA, females earn more on average than males provided that the GPA is high enough.

Answer:

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

For the males:

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

For the females:

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

when GPA>=3.5, males earn more on average than females provided that the GPA is high enough. Therefore I choose (iii).

(b) Predict the salary of a female with IQ of 110 and a GPA of 4.0.

Answer:
We obtain :

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

The salary of a female is 137100$.

(c) 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:
False. We need to test the hypothesis H_0: $\beta^{4}=0$ and look at the p-value associated with the F statistic.

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 + β1 X + β2 X 2 + β3 X 3 + ε.

(a) Suppose that the true relationship between X and Y is linear,i.e. Y = β0 + β1 X + ε. 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.

Answer:

$$RSS = \sum_{i=i}(y_i - \hat{y}_i)^2$$

From the formula, I think, without knowing the detailed data, the comparison between the two models is uncertain.