CH_03_LINEAR_REGRESSION_CONCEPTUAL

Tuesday, January 31, 2017
5:24 PM

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.**

	Coefficient	Std. error	t-statistic	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

TABLE 3.4. *For the Advertising data, least squares coefficient estimates of the multiple linear regression of number of units sold on radio, TV, and newspaper advertising budgets.*

ANSWER:

- Null Hypothesis $H_0 =$ $Coeffiencet\ of\ TV, Radio\ \&\ Newspaper = 0$ ***OR***
- Null Hypothesis $H_0 =$ $Advertisement\ spen\ on\ TV, Radio\ \&\ Newspaperhave\ NO\ effect\ on\ predicting\ the\ Sales$
- **Conclusion based on p-Values:**
  - Coefficients of TV and Radio & Intercept term estimated (based on the sample training data) has a p-Value < 0.00001 which means:
    - A statistically significant evidence exist that Sales is dependent on advertisement spend on TV and Radio
    - It indicates that the coefficient values of TV, Radio and Intercept are highly unlikely to be 0 as the 95% confidence interval for both does NOT include 0
    - Coefficients of TV and Radio both have +VE signs - indicating a positive effect on Sales with increasing value of both variables
    - For $1000 spent in TV advertisement - Sales is likely to go up by **average 46 units** when all other variables stay same
    - For $1000 spent in Radio advertisement - Sales is likely to go up by **average 189 units** when all other variables stay same
  - Coefficients of Newspaper term estimated (based on the sample training data) has a p-Value of 0.8599 which means:
    - A statistically significant evidence exist that Sales is unlikely to depend on advertisement spend on Newspaper
    - It indicates that the, during repeated estimation, coefficient values of Newspaper is highly likely to be 0 as the 95% confidence interval for both includes 0

2. **Carefully explain the differences between the KNN classifier and KNN regression methods.**
   - KNN classifier and KNN regression methods are closely related in formula. However, the final result of KNN classifier is the classification output for Y (qualitative), where as the output for a KNN regression predicts the quantitative value for f(X)
   - In a default working mode - the output of KNN Classifier is decided based on the "majority votes" a class receives. The output of KNN-Regression is the average of Y-value of K nearest neighbors

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 ˆβ0 = 50, ˆβ1 = 20, ˆβ2 = 0.07, ˆβ3 = 35, ˆβ4 = 0.01, ˆβ5 = −10.**
   - **(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 average 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.**
   - **(b) Predict the salary of a female with IQ of 110 and a GPA of 4.0.**
   - **(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:

(a) >> iii

$$Y = 50 + 20*GPA + 0.07*IQ + 35*(Gender) + 0.01*(IQ*GPA) - 10*(Gender*GPA)$$

For Male:

$$Y = 50 + 20*GPA + 0.07*IQ + 0.01*(IQ*GPA)$$

For Female:

$$Y = 50 + 20*GPA + 0.07*IQ + 35 + 0.01*(IQ*GPA) - 10GPA$$

$$Y = 85 + 10*GPA + 0.07*IQ + 0.01*(IQ*GPA)$$

Male Salary > Female Salary WHEN:

$$50 + 20*GPA + 0.07*IQ + 0.01*(IQ*GPA) > 85 + 10*GPA + 0.07*IQ + 0.01*(IQ*GPA)$$

$$>> 50 + 20*GPA > 85 + 10*GPA$$

$$>> 10*GPA > 35$$

$$>> GPA > 3.5$$

When GPA is high enough (> 3.5) Males earns more than female…

(b) >> 137.1 units

$$Y = 50 + 20*4 + 0.07*110 + 35 + 0.01*(110*4) - 10*(4)$$

$$= 50 + 80 + 7.7 + 35 + 4.4 - 40$$

$$= 137.1$$

(c) >> FALSE

To verify if the GPA/IQ has an impact on the quality of the model we need to test the hypothesis H0:β4^=0H0:β4^=0 and look at the p-value associated with the tt or the FF statistic to draw a conclusion.

4. Q4, Q5, Q6, Q7

URL: https://rpubs.com/ppaquay/65559