Q11

Rahul Atre

2023-10-09

Q11 [Regression Modeling]

Suppose we have a dataset with five predictors: X_1 = GPA, X_2 = IQ, X_3 = Gender (1 for Female and 0 for Male, other genders are not represented in the data), X_4 = Interaction between GPA and IQ, X_5 = Interaction between GPA and Gender. The response is the starting salary after graduation (in thousands of dollars). We use least squares to fit the model, and obtain the coefficients: $\beta_0 = 50$, $\beta_1 = 20$, $\beta_2 = 0.07$, $\beta_3 = 35$, $\beta_4 = 0.01$, $\beta_5 = -10$.

1. Which of the following statement is valid? 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.

For this case, the parameters of the fitted regression model have been provided using the least squares method:

$\hat{y} = 50 + 20X_1 + 0.07X_2 + 35X_3 + 0.01X_4 - 10X_5$.

Let's consider the two cases given to us, if gender is male ($X_3 = 0$, $X_5 = 0$). Note that X_5 is affected here since it is the interaction of gender with GPA.

For female, it is when ($X_3 = 1$), and X_5 becomes X_1 since gender = 1, so the only "interactable" variable left is GPA. So,

$E[\hat{y}|X_3 = 0] = 50 + 20X_1 + 0.07X_2 + 0.01X_4$

$E[\hat{y}|X_3 = 1] = 50 + 20X_1 + 0.07X_2 + 35(1) + 0.01X_4 - 10X_1 = 85 + 10X_1 + 0.07X_2 + 0.01X_4$

Let's check when male's earn more salary than females:

$E[\hat{y}|X_3 = 0] > E[\hat{y}|X_3 = 1]$

$50 + 20X_1 + 0.07X_2 + 0.01X_4 > 85 + 10X_1 + 0.07X_2 + 0.01X_4$

$(20 - 10)X_1 + (0.07 - 0.07)X_2 + (0.01 - 0.01)X_4 > (85 - 50)$

$10X_1 > 35$

$X_1 > 3.5$

Therefore, when a male's GPA is above 3.5, he would earn more than the average female. So (iii) is the only true statement.

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

Given information: $X_1 = 4.0$, $X_2 = 110$, $X_3 = 1, X_4 = (X_1)(X_2) = (4.0)(110) = (440)$, $X_5 = (X_1)(X_3) = (4.0)(1) = 4.0$

So, plugging these values into our salary equation:

$\hat{y} = 50 + 20(4.0) + 0.07(110) + 35(1) + 0.01(440) - 10(4) = 137.1$ (in thousands).

Therefore, the predicted salary of a female with an IQ of 110, and a GPA of 4.0 is $137,100$.

3. 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.

This statement is **false** because the beta parameter for GPA/IQ (X_4) has nothing to do with the interaction effect between the two terms. We can verify the interaction effect between GPA and IQ by performing a hypothesis test using the t-statistic.

The hypothesis would be defined as follows:

$H_0 : \beta_4 = 0$ $H_1 : \beta_4 \neq 0$