Jared Yu

Module 5 Discussion

Question:

Let be a continuous random variable, be a categorical variable that has 3 levels (L1, L2, L3), be a categorical variable that has 2 levels (“yes” or “no”), be a continuous variable. Construct a multiple linear regression model such that we can study the effect of on and study whether the effect of on is equal across all levels of and .

Ans:

For convenience, let the index for two of the regressor variables be interchanged so that the model can be more easily understood. Here, let become and become , while remains the same. Therefore, the question becomes a matter of studying the effect of on and studying whether the effect of on is equal across all levels of and . The reason is that the original is a continuous variable, so no indicator variables are required. However, and the original are both categorical variables which require indicator variables to represent properly in the linear regression model. In the case of , it is a binary categorical variable, therefore it will require only a single () coefficient. However, in the original , it has three levels which requires two () coefficients to represent correctly using indicator variables.

We can represent the first indicator variable, , as follows,

Then, we can represent the second indicator variable, (the new) , by expanding it to and . The way that these two can then represent the original can be seen below in Table 1.

|  |  |  |
| --- | --- | --- |
| Original |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

Table The above table shows the original categorical variable broken into two separate indicator variables, and . The table shows that is represented by , is represented by , and is represented by .

Therefore, assuming then that a first-order model is appropriate for the given problem, we have

In equation (1), the intercept of the regression line depends on the factors in an additive fashion. So, none of the regressor variables depend on each other. It is possible to look at interaction effects by combining variables together, such as , but that will not be explored for simplicity.

The first question asks us to study the effect of on . Here, has been changed to , so we are the effect of the continuous variable (new) on . We can analyze this by using the following hypothesis test,

If is not rejected, then it indicates that the original regressor can be deleted from the model. We can perform this test using the partial test, where the test statistic is as follows,

In the above formula, , and is the design matrix with a first column of ones while also including the variables . It can also be thought of as measuring the contribution of as if it were the last variable added to the model. If , we reject , concluding that contributes significantly to the regression model at the confidence level . In other words, if is larger than , then we can conclude that the original has a significant impact on at confidence level .

The next question is asking us to study whether the effect of (the original) on is equal across all levels of (the original) and . To check whether or not the effect of on is equal across all levels of the two categorical variables, we would need to examine all the different possible models involving each of the combinations of levels amongst and . This can be seen below in Table 2.

|  |  |  |
| --- | --- | --- |
| (new) |  | Regression Model |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

Table The above table shows the six different models that could be designed to test whether or not the impact of (the original) on is equal across all levels of (the original) and .

Like before, we can construct a hypothesis test for each of these models, utilizing the partial test to determine the results. In each of these hypotheses’ tests, we would again be examining vs. . The difference however is that in , the full model (FM) and reduced model (RM) would correspond to different sets of the coefficients. For example, in the first row of Table 2, the FM would consist of , , and , while the RM would consist of and .