Applied Statistic Modeling Assignment 2

Jiaming Deng 22302794

1)

1. β0 is the expected value of yi when xi1 is equal to 0, that is, when the user is in group A.
2. β1 is the difference in the expected value of time y which is the time that users use the product between groups B and A when all other variables are held constant.
3. εi is the random error term.

2)

1. β0 is the expected value of yi when all other variables are 0.
2. β1 is the influence of xi1 on yi when xi2 is 1.
3. β2 is the effect of xi2 equal to 2 on yi when xi1 is 0.
4. β3 is the effect of xi2 equal to 3 on yi when xi1 is 0.
5. β4 is the effect of the interaction between xi1 and xi2=2 on yi.
6. β5 is the effect of the interaction between xi1 and xii2=3 on yi.

Both β2 and β3 are used to obtain the influence of xi2 on yi, because xi2 has three categories, namely 1, 2, and 3; according to the formula, it can be seen that xi2=1 is the control category, and β2 means that the category of xi2=2 and the control category are in the Yi gap. In the same way, β2 represents the difference in yi between the category with xi2=3 and the control category.

β4 and β5 are interaction terms that represent the effect of the combination of xi1 and xi2 on yi. Specifically, as above, xi2=1 is the control category. β4 represents the difference in the effect of x1 on y between the class with x2=2 and the control class, while β5 represents the difference in the effect of x1 on y between the class with x2=3 and the control class.

The interpretation of the coefficients in Model 2 differs from that in Model 1 in that Model 2 includes an interaction term that allows the effect of xi1 on yi to vary depending on the value of xi2 whereas Model 1 includes only a single coefficient for the effect of xi1 on yi, assuming all Values are all affected the same.

3)

1. The coefficient of group B is estimated to be 0.004632, but Pr(>|t|) = 0.993 means that it is not statistically significant, indicating that users do not spend significantly more time using the new interface than the old interface.
2. Adjusted R-squared is -0.00505, indicating that the model has almost no change in the data, and F-statistic is 6.662e-05, indicating that the model has no significant statistical significance. All in all, according to Model 1, there is no significant evidence that users' usage time has changed significantly after switching to the new interface.

4)

1. The coefficient of groupB:user_levelmedium is -1.2829, which means that under the premise of keeping all other variables constant, the product use time y of users with medium level in groupB is 1.2829 lower than that of users with the same level in groupA. Pr(>|t|) =0.07303 shows that the statistics are interesting.
2. The coefficient of groupB:user_levelhigh is -3.2593, which means that under the premise of keeping all other variables constant, the product usage time y of the high-level users in groupB is 3.2593 lower than that of the same-level users in groupA. Pr(>|t|) =8.41e-06 shows statistical significance.
Overall, the model is statistically significant. Negative and different coefficients for both interaction terms indicate that the new interface is less effective for users with intermediate or advanced experience, and the statistical significance is different for different levels of users.

5)

Model 1 concludes that there is no significant difference in streaming time between users in Group A and Group B. The conclusion drawn from model 2 is that according to different user levels, after updating to the new interface, the usage time of medium-level users does not change significantly, but for your high-level users, the usage time has a slight negative impact. These two conclusions are not contradictory, because model 1 does not consider user levels, while model 2 considers user levels, so the conclusions drawn are more detailed.
We recommend that the streaming service continue to use the new user interface, and focus on the experience of high-level and mid-level users, and collect feedback from these users to improve the interface.