Statistics Learning, Module 2, 2021

Syndicate Problem Set # 2

All responses are to be neatly typed up. This includes any equations that you choose to present. The page setup should have at least 2cm margins on all sides, with all texts formatted to 12pt font size and at least 1.5 spacing. Present your numerical results neatly in tables and plots. R screenshots are not acceptable unless explicitly requested.

The submitted report should not exceed 6 pages. Graphs and tables may be presented in an Appendix, with the length of the appendix not exceeding 4 pages.

The assignment is due on Monday June 14th at 9PM. You are expected to work through these problems over the weeks prior, as indicated below.

Fishing Mode (To be completed during Week 4)

The data file "fishing.csv" contains cross sectional data on the choice of fishing mode. The data also contains the price associated with each mode, the catch rate (average number of fish per hour) for each mode, and the monthly income of each individual.

Ordered Logit Model

Assume that there is a preferential order to the choices:

$$Beach < Pier < Boat < Charter$$

Make sure that the factor levels are reordered to reflect this preferential order. For the purpose of this exercise, assume that all predictors affect the latent utility in a linear manner. Construct an ordered logit model to explain the choice of fishing mode. Present your model in latent utility representation. Discuss how the prices, catch rates and individual level income affect the propensity of selecting each choice based on this model.

Your response to this should not exceed 1 page.

Multinomial Logit Model

Now assume that there is no preferential order to the choices.

1. Construct the multinomial logit model to explain the choice of fishing mode. Assume that the prices and catch rates of all three modes to impact the utility of each of the possible choice. Present your model using an appropriate table format. Discuss how the prices, catch rates and individual level income affect the choice based on this model.

   Your response to this should not exceed 1 page.

2. Discuss how the ordered logit and the multinomial logit models are different. For the purpose of modelling fishing mode choice, which model do you think is more appropriate? Justify your answer. (Maximum 200 words.)

Credit Card Balances - Revisited (To be completed during Week 5)

Revisit the credit card balance data that you investigated in Week 1.

Tobit Regression

You may have already noticed that the lower limit of balance is 0 and the upper limit is the credit limit of each customer. Convert the balance data into a ratio:

$$BalanceRatio_i = \frac{Balance_i}{Limit_i}$$

This ratio measures the bank customer's level of credit card account utilization relative to the available limit. This measure is contained within the interval $[0,1]$.

Revisit the final model you constructed in Week 1. Fit the Tobit model on $BalanceRatio_i$ with the same set of predictors that you have chosen in Week 1. Note that since $Limit_i$ is used to construct the dependent variable, it should not be included in your model here. Present the estimated Tobit model. From the Tobit model, what are the drivers of the level of credit card account utilization? Discuss whether the results from this model contradict with what you have found in Week 1. Your discussion should highlight how the two models are different.

Your response to this question should not exceed 1 page.

Poisson Regression

***For the purpose of the following exercises, exclude the "Limit" and "Balance" variables from the consideration set.

Suppose that you would like to construct a model to explain the number of credit cards held by a bank customer. The response variable of interest is now "Cards".

1. Use a Poisson regression model to assess how the customer demographic characteristics and credit rating influence the number of credit cards they hold. Conduct a model selection exercise to select an appropriate Poisson regression model. Present your choice of Poisson regression model and discuss the results.

   Your response to this question should not exceed 1 page.

2. Estimate a linear regression model on the "Cards" data with the exact same set of predictors that you have chosen for the Poisson regression model above. Present your model. Compare this result to your Poisson regression model. What is the difference between the two models? Which of the two models do you think is the more appropriate model to use in this case? Justify your answer.

   Your response to this question should not exceed 1 page.