SCM 651

Homework 4

Group Members: Brenon Tate, Yacine Ide, Melissa Mosier, Nicholas Nguyen

Due: 6/12/2022

Submitted: 6/12/2022

Table of Contents

[1 Introduction 3](#_Toc105948785)

[2 Logit and Probit Analysis 4](#_Toc105948786)

[3 Moderating Effects 6](#_Toc105948787)

[4 Regression Model 8](#_Toc105948788)

[5 Neural Network Analysis 10](#_Toc105948789)

[6 Prediction Model 12](#_Toc105948790)

# Introduction

Using the Universal Bank Data set that was provided, we performed a Logit and Probit analysis to determine which attributes were significant when determining a person’s probability to take out a loan. We then analyzed different combinations of moderating effects to determine which interactions were significant. Using the significant attributes and interactions, we created a regression model and performed a sensitivity analysis.

Lastly, we performed a neural network analysis using the variables we found significant. From this neural network analysis, we created a prediction model and associated sensitivity analysis.

# Logit and Probit Analysis

**Perform a logit and probit analysis of the variables that affect whether a customer takes out a loan. Consider only main effects (main variables, no moderating effects). Which variables are significant? How do the significant variables influence the likelihood of taking out a loan? Copy screen snapshots of your analysis in R to your report. (20%)**

Figure 1 and Figure 2 shown below are screen shots from the Logit and Probit analysis performed in R. An inspection of the P values shows that only some of the attributes are significant when it comes to determining if a person will be given a loan or not.

A picture containing text, plaque

Description automatically generated

Figure 1 Logit Output

Commentary on Fig.1 output: Based on the above Logit model, CCAvg, CDAccount, CreditCard, Education, Family, Income, Online and SecuritiesAccount were all significant variables. This makes sense as we know that the banks investigate this personal info when you apply for a loan and they impact how much you are approved for, and whether you get approved at all. CCAvg, CDAccount, Education, Income (higher the income), Family (the larger the family), Experience, Mortgage all have a positive coefficient and more than likely will lead to these groups accepting a personal loan. CreditCard, Online, and Securities Account are all negative, meaning they are associated with a lower likelihood of accepting a personal loan.

Our significant variables (Lowest P-values) are CreditCard, CDAccount, SecuritiesAccount, CCAvg, Education, Family, Income, Online.

The variables that negatively affect the decision (Negative coefficient) are Age, CreditCard, Online, SecuritiesAccount, ZipCode.

A picture containing text, plaque

Description automatically generated

Figure 2 Probit Output

Commentary on Fig.2 output: Based on the above Probit model, CCAvg, CDAccount, Education, Income (higher the income), Family (the larger the family), Experience, Mortgage all have a positive coefficient and make one more likely to accept a personal loan. CreditCard, Online and Securities Account are all negative and mean a client will be less likely to accept a personal loan.

Our significant variables (Lowest P-values) are the following: CreditCard, CDAccount, SecuritiesAccount, CCAvg, Education, Family, Income, online.

The variables that negatively affect the decision (Negative coefficient) are Age, CreditCard, Online, SecuritiesAccount, ZipCode.

# Moderating Effects

**Add moderating effects (interactions of variables). Which interactions make sense conceptually? Which interactions are statistically significant? How do you interpret the coefficients on these variables? Copy screen snapshots of your analysis in R to your report. (20%)**

**\*For this question, we chose to use only attributes which were most significant from the Logit and Probit output, all other combinations were omitted.**

* Based on the output above, we can see that Family, Income and Education, would be considered significant variables and would have the most impact on a decision to accept a personal loan.
* In our logit model, we calculated the following coefficients for Education, Income and Family:
  + Family: 0.695839386
  + Education: 1.735968471
  + Income: 0.054581665

Those 3 coefficients make sense because people with a higher degree of education usually require financial assistance to pay off things such as student loans. Family also would make sense because the more people living in a household, the higher the probability of needing more money and therefore personal loans. Income is the lowest coefficient, which means that as income increases, people are less likely to seek personal loans. An interesting interaction we found was between education and income. The low coefficient for income could also be because people with high income also typically have advanced degrees, as they usually are doctors, lawyers, engineers, etc.

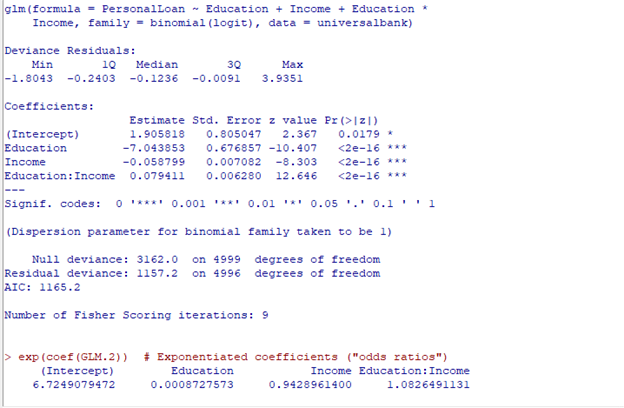


Figure 3 Logit Output with Moderating Effects (Education:Income)

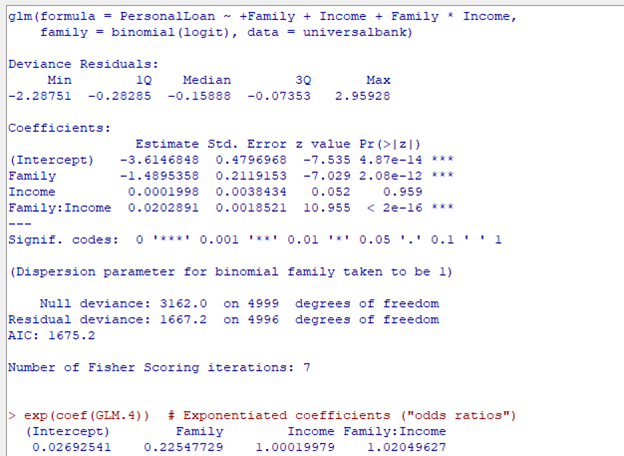


Figure 4 Logit Output with Moderating Effects (Family:Income)

A picture containing text

Description automatically generated

Figure 5 Logit Output with Moderating Effects (Family:Education)

# Regression Model

**Create a final regression model with the variables that you feel are important (both main effects and interaction terms). Use the moderating effect that was significant and its two individual main effects. Create a spreadsheet prediction of the model. Perform a sensitivity analysis as seen earlier in the semester. Which variables have the greatest influence on the customers’ loan behavior (combined main effects and interaction effects)? Copy screen snapshots of your analysis in R to your report. (20%)**

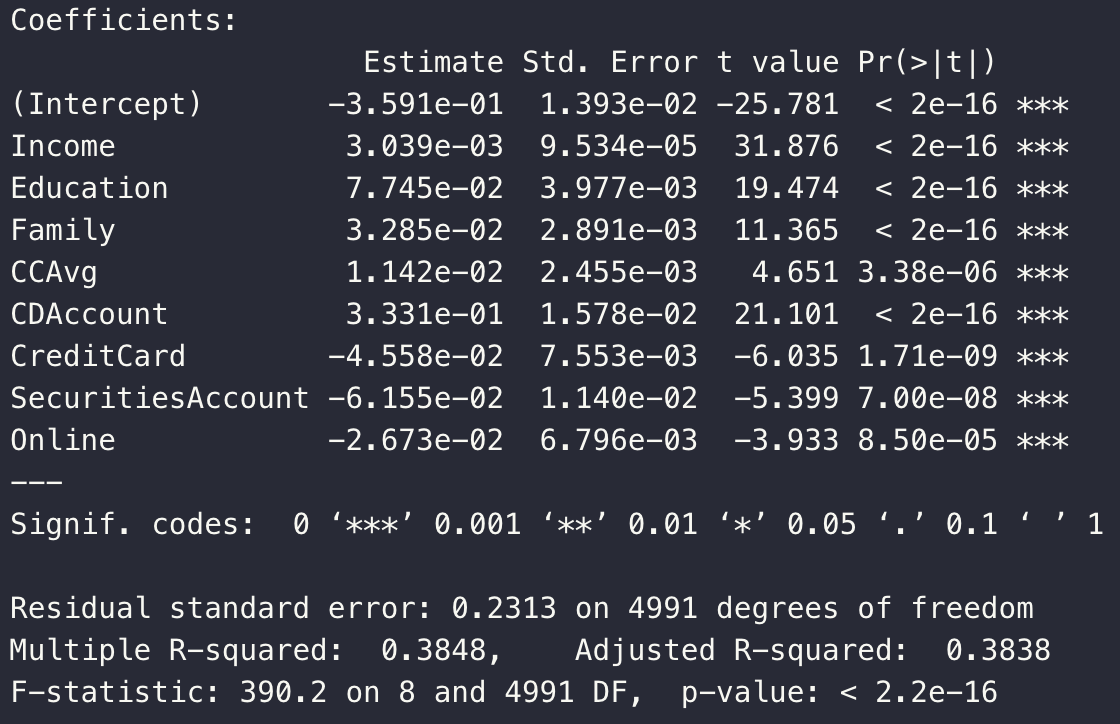


Figure 6 Regression Model with Significant Attributes

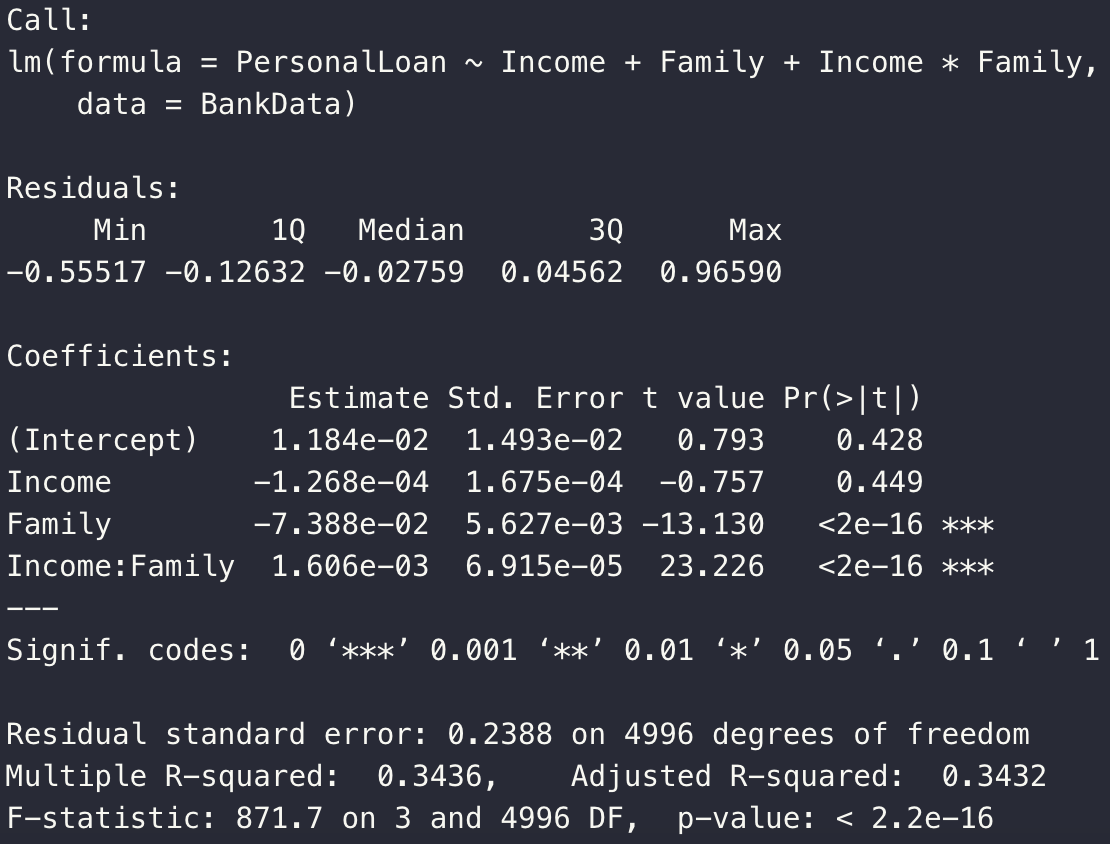


Figure 7 Regression Model with Moderating Effect

Based on our regression above, we decided to use a Regression Model using family and income with moderating effects because of the impactful coefficient and significance (Low p value).

The family and income analysis shows an interesting reaction between the variables. Having both a high income results in the lowest probability of taking out a loan. The interaction here is interesting because, despite Family having a stronger coefficient than Income, the effect when combined with Income is low. It seems that income overpowers family size when considering whether someone will accept a personal loan.

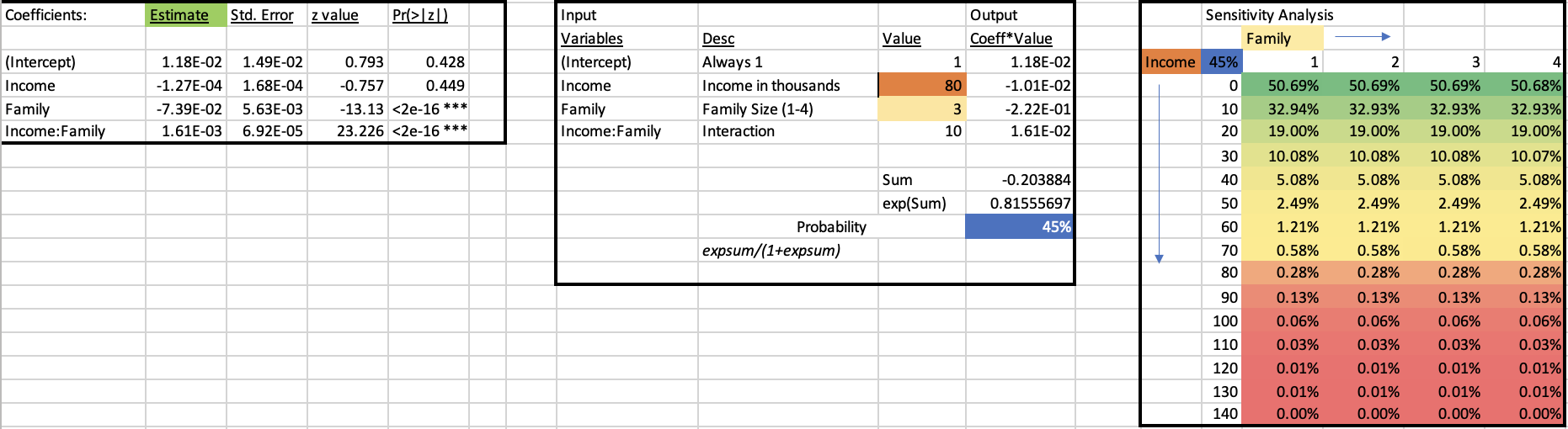


Figure 8 Education. & Income Sensitivity Analysis

# Neural Network Analysis

**Perform a neural network analysis of the variables in part 3 above. Copy screen snapshots of your final neural network model in R to your report. (20%)**

Chart

Description automatically generated

Figure 9 Neural Network

Text

Description automatically generated

Figure 10 Neural Network Matrix Output

Our final neural network model in R had the following values for coefficients:

* Y = Personal Loan
* X1 (Significant variable 1) = Family Size
* X2 (Significant variable 2) = Income
* Node 1: Intercept is 5.85, Income is -.028 Family is -.769
* Node 2: Intercept is 4.69, Income is 3.73 and Family is 2.20
* To Y (Personal Loan): 1.75, -7.59, -1.10

We decided to analyze Family Size and Income because they had the most significant impact on the decision making to take out a loan or not. The neural network ran for a few steps but had a high number of errors. Adding more significant variables would have lowered the number of errors.

# Prediction Model

**Create a prediction model of the neural network in part 4. Using the prediction model, perform a sensitivity analysis for the neural network model like the Logit and Probit sensitivity analysis. (20%)**

Figure 10 shown below details the prediction and sensitivity analysis using the provided template and based on our neural network output. The What-If table shows that income and family size, as expected, play a large part in whether someone will accept a loan.

The interaction between family and income is very clear here: the more money a family makes and the larger the family, the more likely this family will take a personal loan. It does still appear, as we found earlier, that income has a stronger effect than family – we can see that an income below 70k generally makes one unlikely to take on a loan. However, we do see that large families are more likely to take out loans, even if income is over 100k. Our recommendation is to push personal loans with large families, especially families that can anticipate paying down such a loan because they have a sustainably large income.

Graphical user interface, application, table, Excel

Description automatically generated

Figure 11 Neural Network Sensitivity Analysis