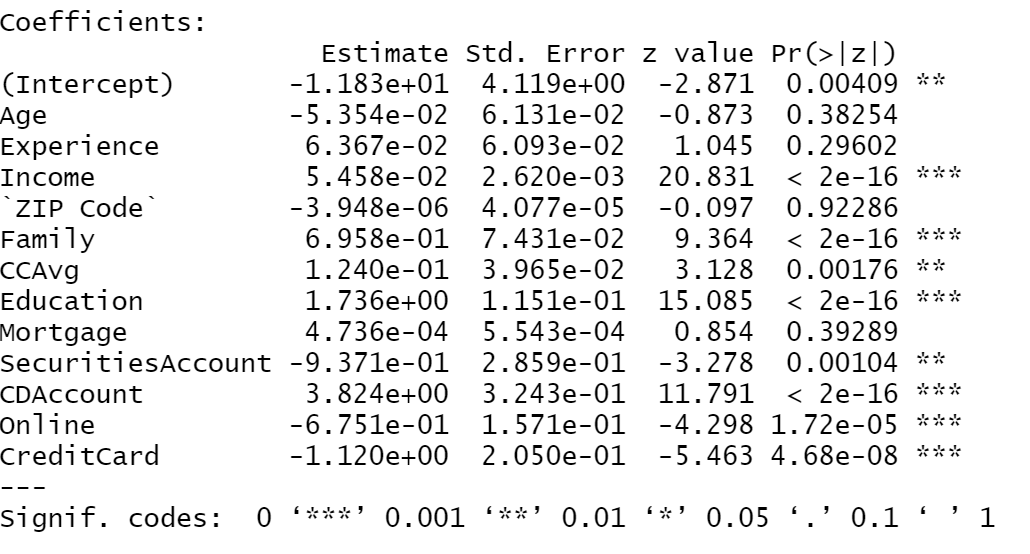
SCM 651

Homework #4

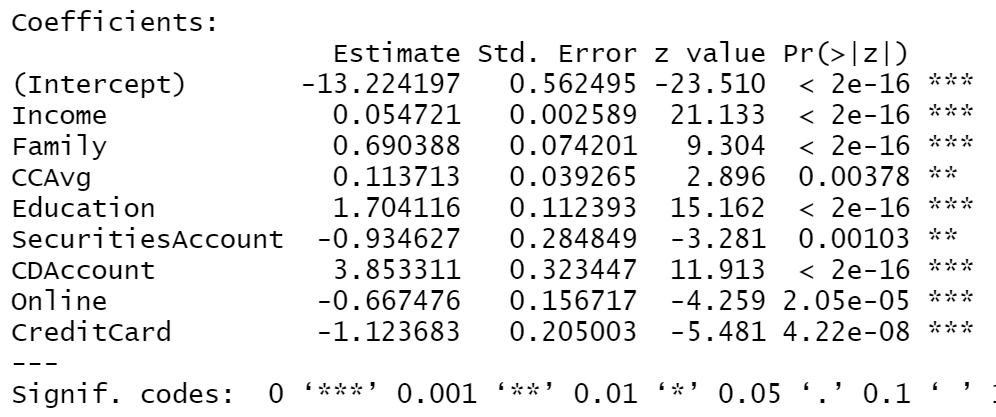
Team 4: Phoebe Sheahan, Emilio Ramos Monzalvo, James Eakins, Shaun Hall, J.R. Slouffman

1. Perform a logit and probit analysis of the variables that affect whether a customer takes out a loan. Consider only the main 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.

First, we ran the logit model with all variables included.



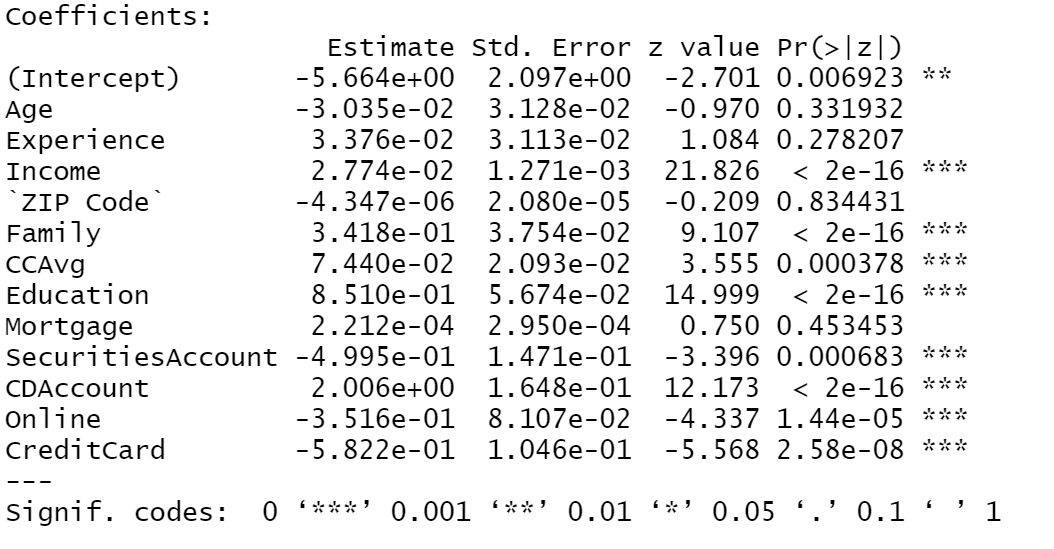
After determining the most significant variables based on their probability we reran the logit and the outcome showed that all variables we determined to be significant were in fact significant.



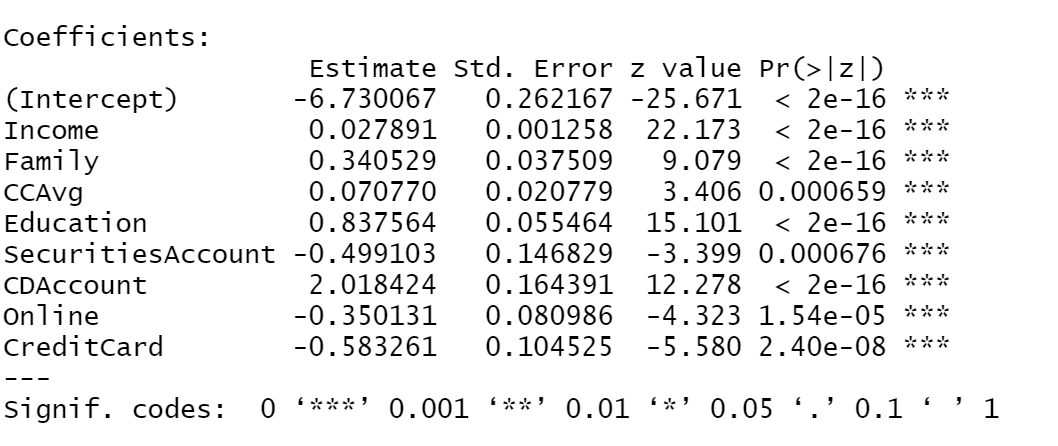
The most significant variables are Income, Family, CCAvg, Education,SecuritiesAccount, CDAccount, Online and CreditCard. Income, Family, CCAvg, Education, and CDAccount all positively increases the probability of someone taking out a personal loan. The higher your income the more likely you are to take out a personal loan. The more members in your family the more likely you will take out a loan. Your CCAvg being higher created a greater likelihood in someone taking out a personal loan. The more educated you are the more likely you will be to take out a loan. If you do have a CDAccount you are more likely to take out a loan. The variable is more significant because it is a binomial outcome, therefore if the person does have an CDAccount it very significantly increases the probability that they will take out a loan .

The variables that will have a negative effect on the likelihood of someone taking out a loan are Online, CreditCard, and Securities accounts. If you bank online you are less likely to take out a loan. If you have a credit card you are less likely to take out a loan, this is significant because it has a greater negative effect based on the fact that this is a binomial variable. Securities account also negatively affects the likelihood of someone taking out a personal loan.

Next, we ran the probit analysis with all the variables included.



After running the probit we could see that the most important variables were the same for both models.

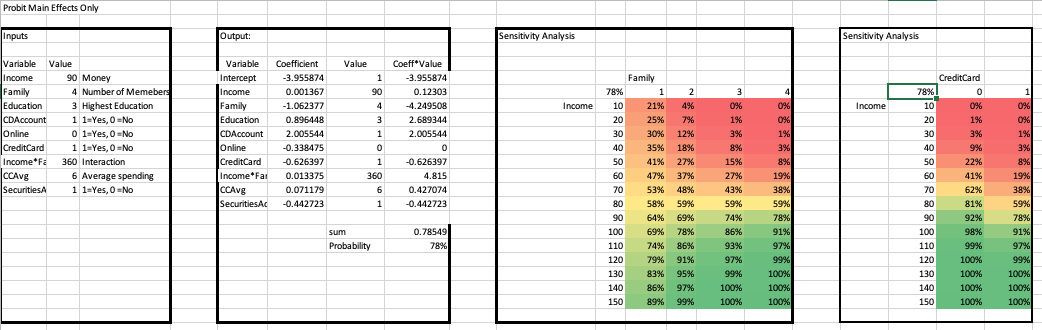
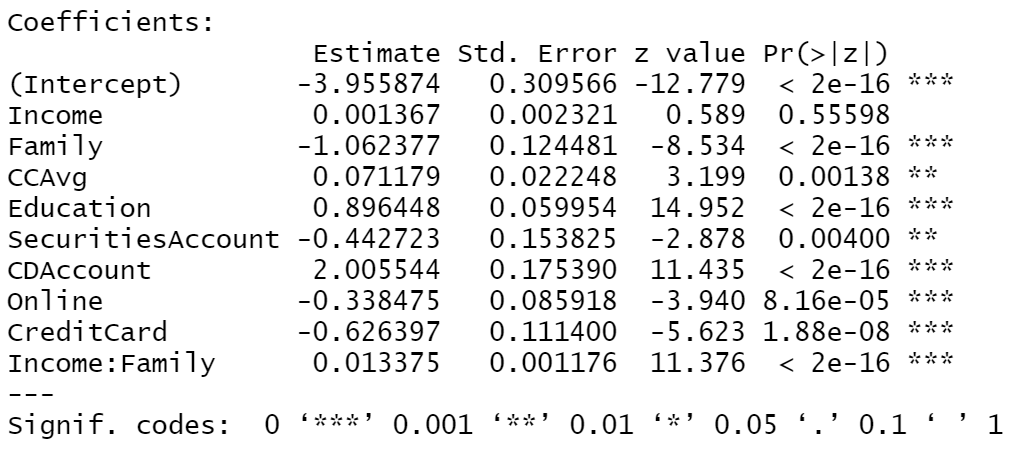


The probit analysis shows that Income, Family, CCAvg, Education, and CDAccount all positively affect the probability of someone having a personal loan. The more income that you have the more likely you are to take out a loan. The larger your family size the more likely to take out a loan. CCAvg of a higher value increases the While SecuritiesAccount, Online, and CreditCard all negatively affect the probability of someone taking out a loan.

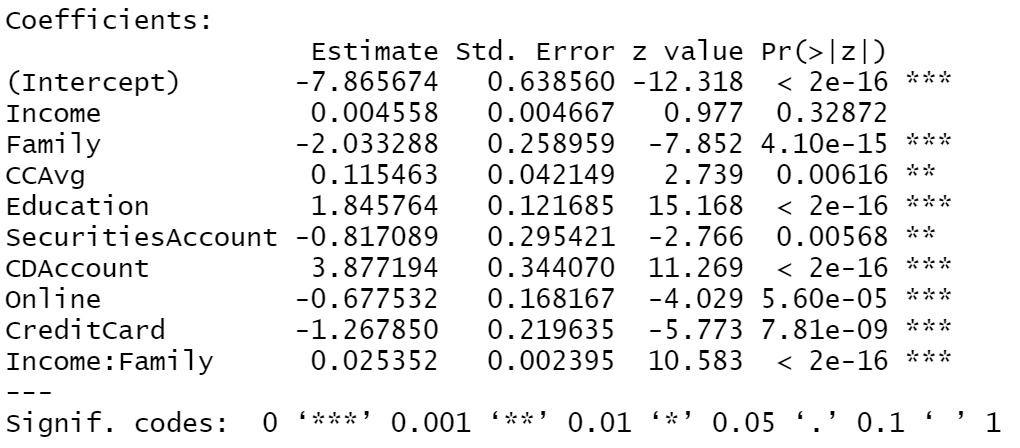
2. Add moderating effects (interactions of variables). Which interactions make sense conceptually? Which interactions are statistically significant? How do you interpret the coefficients on these variables? Copy screen snapshots of your analysis in R to your report.

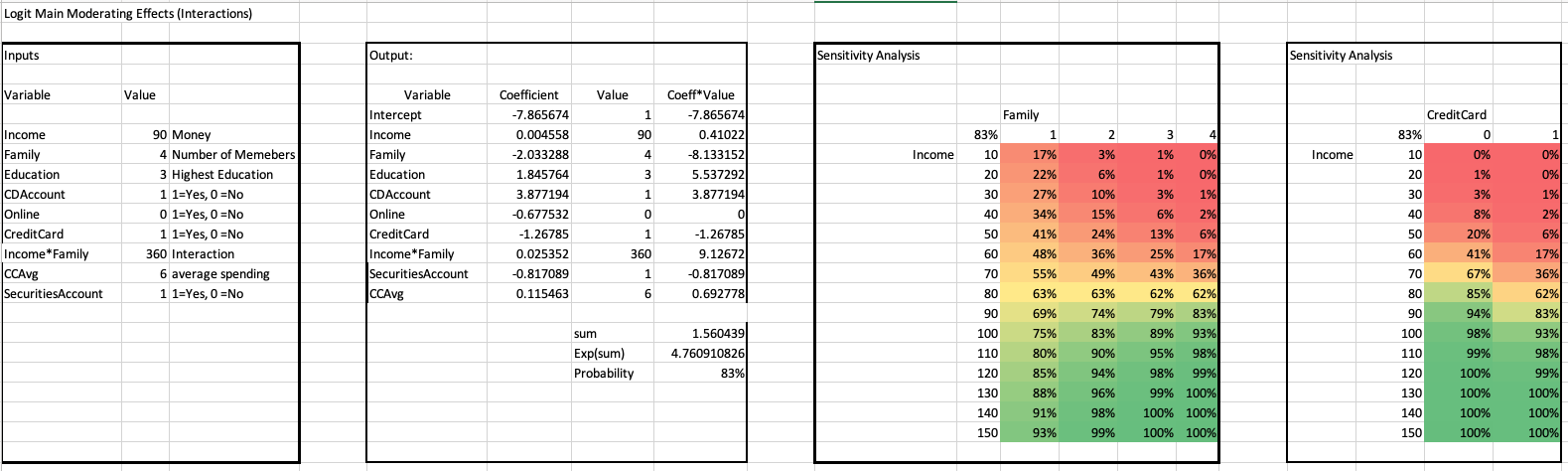
The two variables that would have the most interactions conceptually would be family and income. It makes sense that a larger family would have the most income. Both of these variables would influence a person to take out a personal loan. In contrast, if there is a lower class family with many family members, then the income per family member would be lower which should make the loan harder to get. Therefore, we performed the moderating effect analysis using both the Probit and the Logit Function which can be seen below.

Probit with moderating factors of Family and Income :



Logit moderating effects of income and family:

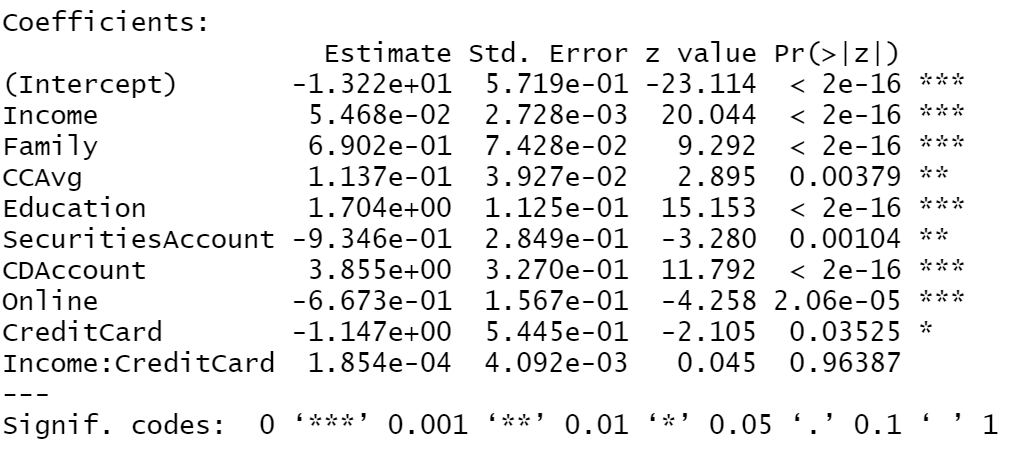


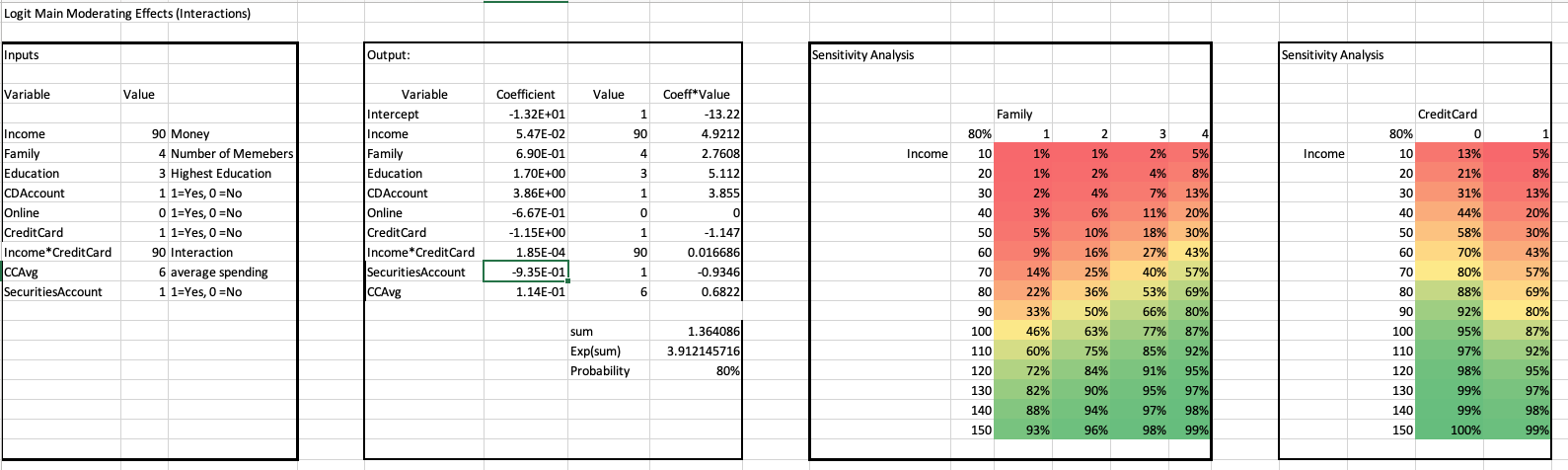


The most interesting effect of the moderating effects is that the Family parameter seems to be insignificant using either the Probit or the Logit functions. We also see that you are less likely to get the loan as the size of your family increases. As we look at the coefficients, the bigger the number, the more effect the parameter will have on the result. If the coefficient is negative, then it lowers the probability of the person getting the loan. This can be visualized in the sensitivity analysis of the Family versus Income.

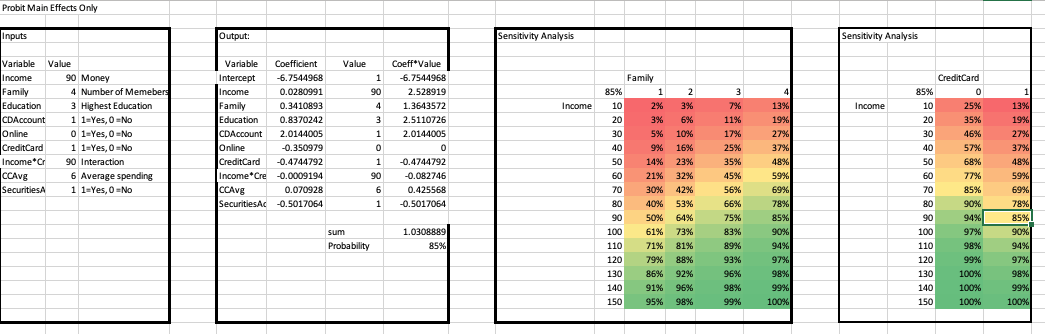
The other combination of parameters that makes conceptual sense to combine is the Income and Credit Card parameters. If you have a history with the bank and a high salary, then it should be very likely to get a loan. On the other hand, since the variables are positively correlated, then putting them into the model will not give us promising results as we can see below. The interaction between Credit Card and Income turned out to be an insignificant parameter.

Logit moderating effects of Income and Credit Card:

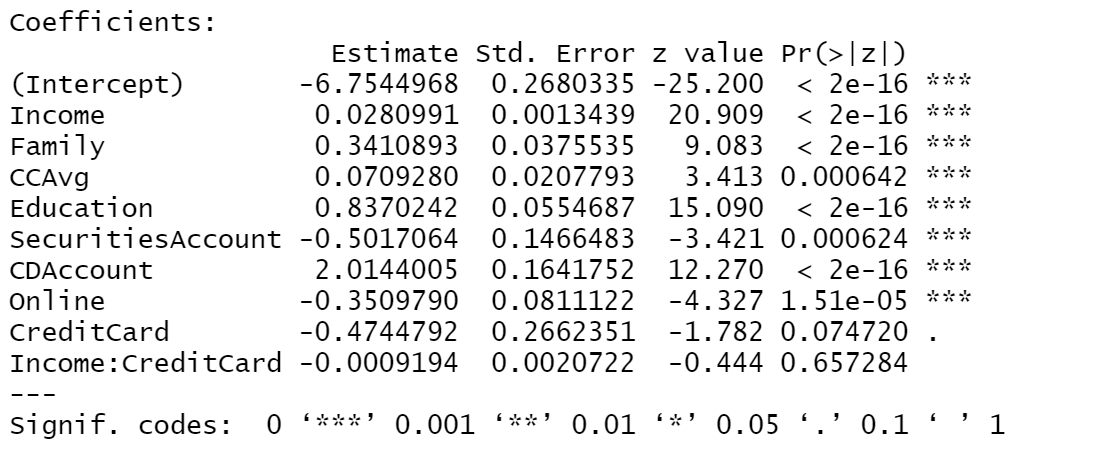


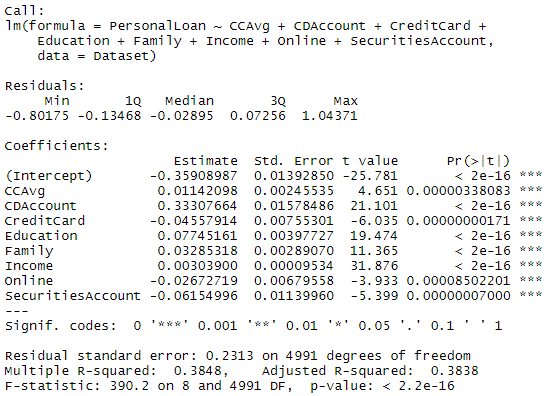


Probit with moderating factors of Income and Credit Card:



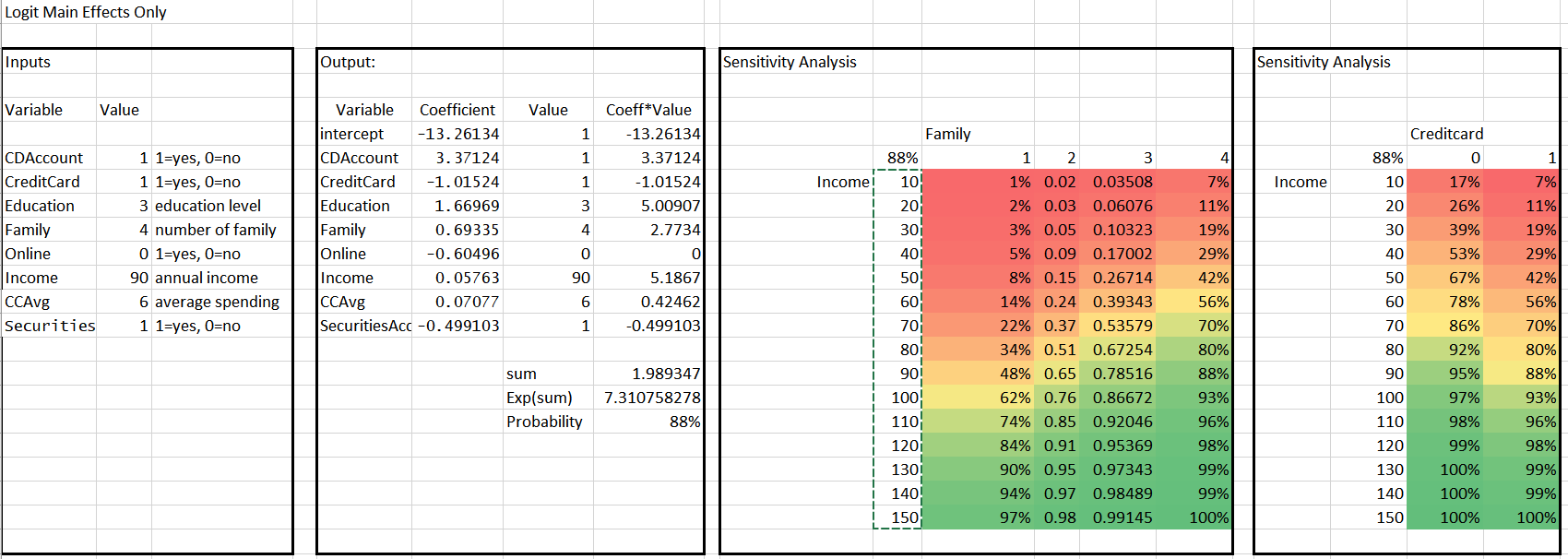
3. Create a final regression model with the variables that you feel are important (both main effects and interaction terms). Create a spreadsheet prediction of the model. Which variables have the greatest influence on the customers’ loan behavior (combined main effects and interaction effects)? Perform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

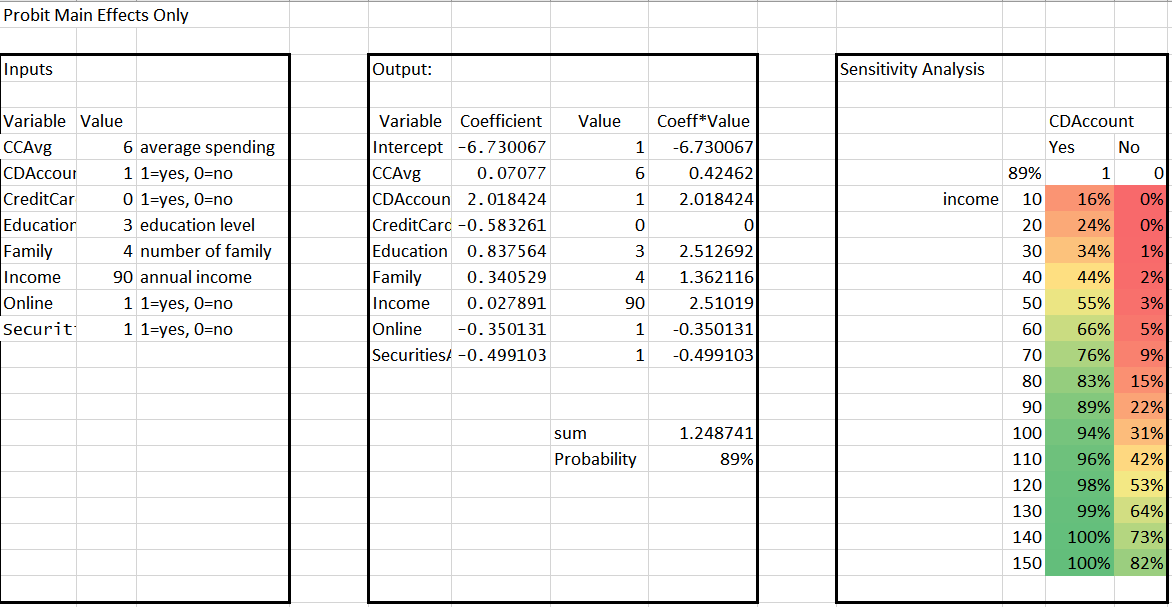
Most important variables from previous analysis are CDAccount, CreditCard, Education, Family, income and Online. These are all common significant variables from both logit and probit analysis. After running the final regression model with these variables, all eight variables showed as significant variables. CCAvg, CDAccount, Education, Family and Income positively affect the model, which means these will increase the probability of getting a personal loan. And CreditCard, SecuritiesAccount and Online will negatively affect the probability. The main effects would be the CDAccount since it has the highest coefficient number. We have another regression model from the Excel that shows similar results as the one from the R which is attached on the appendix.



For sensitive analysis, we tested on the number of family members and income by using variables from logit analysis. We observed that when there are more family members with higher income, it increases the probability of getting a loan. Also, when the person has a credit card, no matter how many numbers, they would have a higher probability of getting loans. Another graph is the one when we used the variables from the probit analysis. It indicates that if you have the CDAccount, the probability will significantly increase to getting a loan than when you don’t have the CDAccount. The person who has a high income, if they don’t have a CDAccount, then he will have much less chance to get a loan than the person who has a low income.

We selected the values of variables by the logic that we found from the earlier analysis. The person who has the most family members and highest degree would get the higher probabilities so we assigned 4 for the family member and 3 for the education. According to the rscript we ran, 90 was the most occurring number for the income for the family of 4 with the highest education. Most of these people do have a CDaccount, SecurityAccount and CreditCard too. So we choose value 1 for those inputs. Lastly, they don’t use online.





4. Perform a neural network analysis of the variables found to be significant in the logit and probit analysis above. Copy screen snapshots of your final neural network model in R to your report.

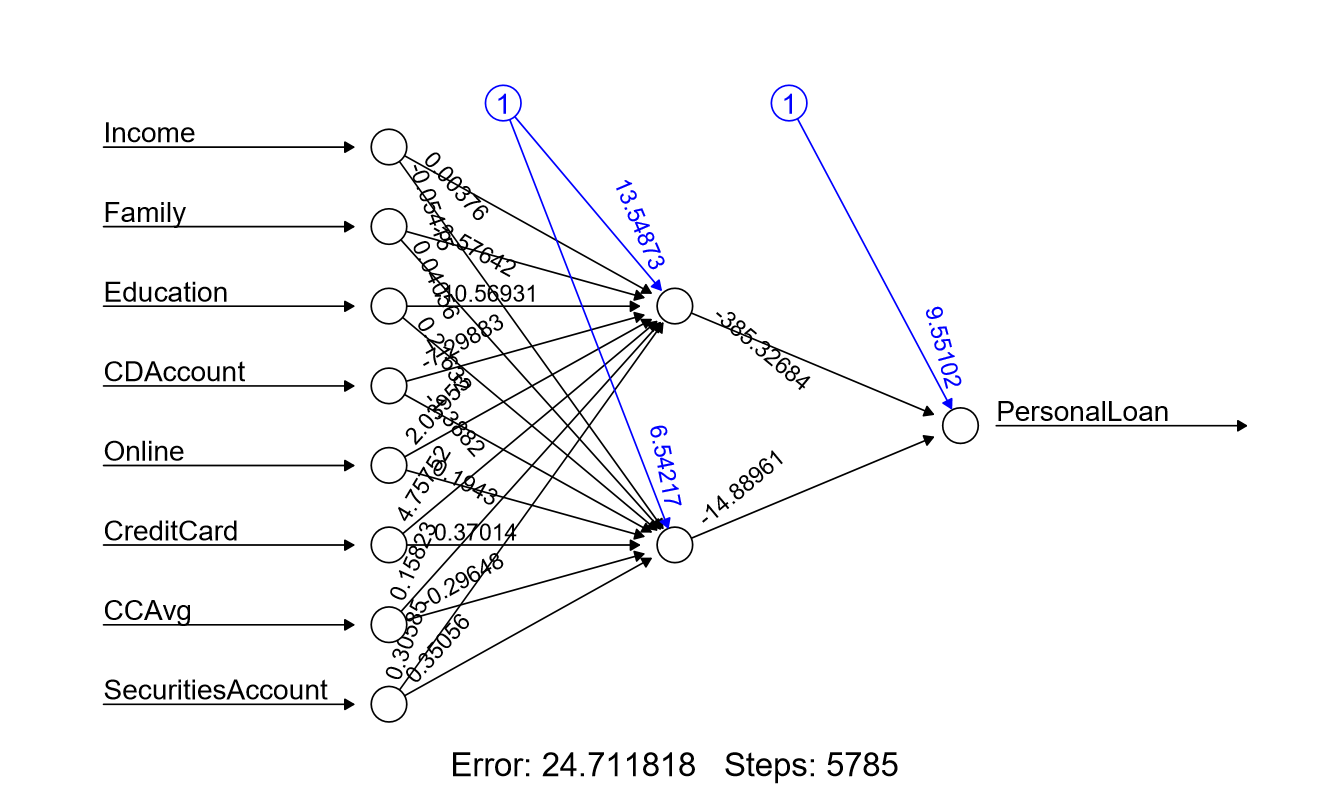
Based off the logit analysis and the significant variables that we identified the neural network output:

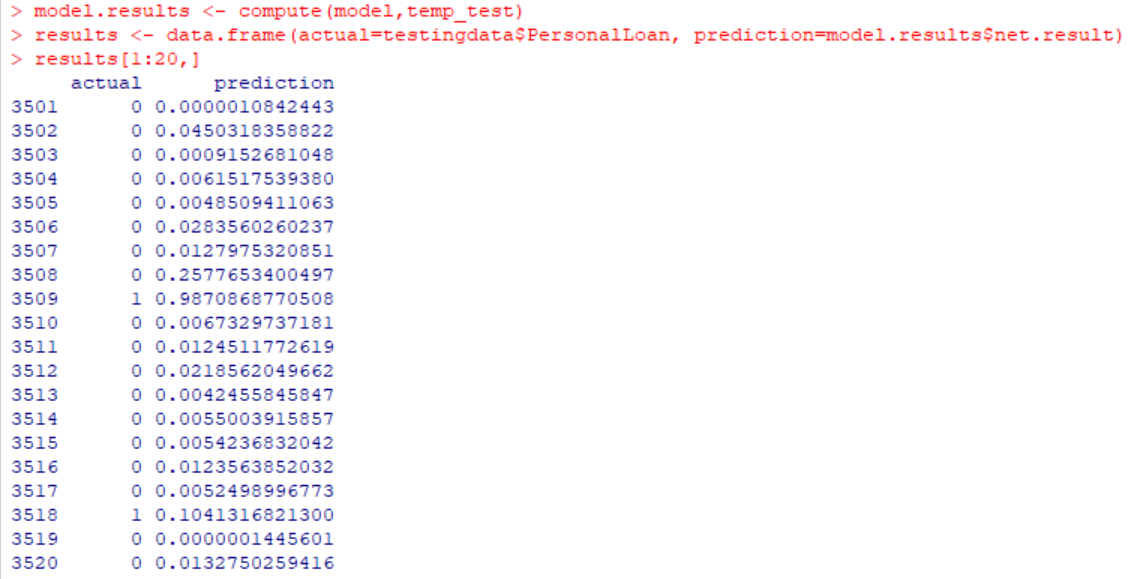
> trainingdata <- HMK4[1:3500,]

> testingdata <- HMK4[3501,5000,]

> model <- neuralnet(PersonalLoan ~ Income + Family + Education + CDAccount + Online + CreditCard + CCAvg + SecuritiesAccount, trainingdata, hidden=2, lifesign="minimal", linear.output=FALSE, threshold=0.1)

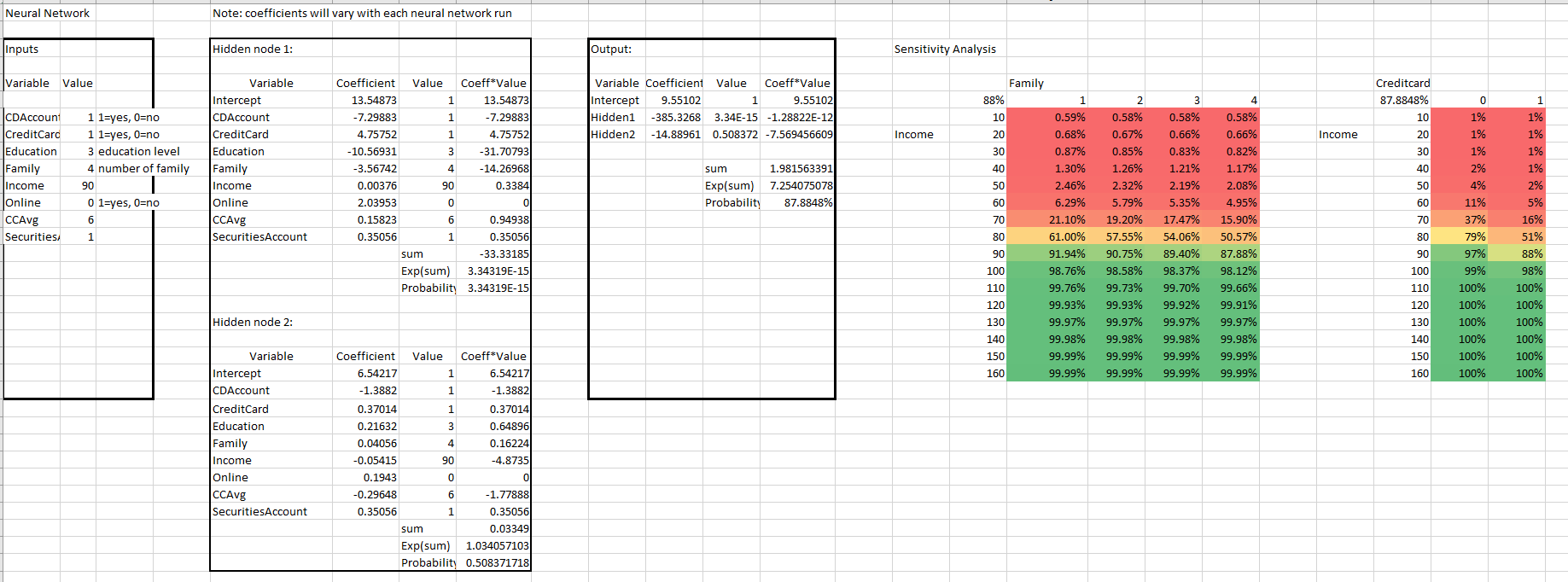
hidden: 2 thresh: 0.1 rep: 1/1 steps: 5785 error: 24.71182 time: 12.6 secs



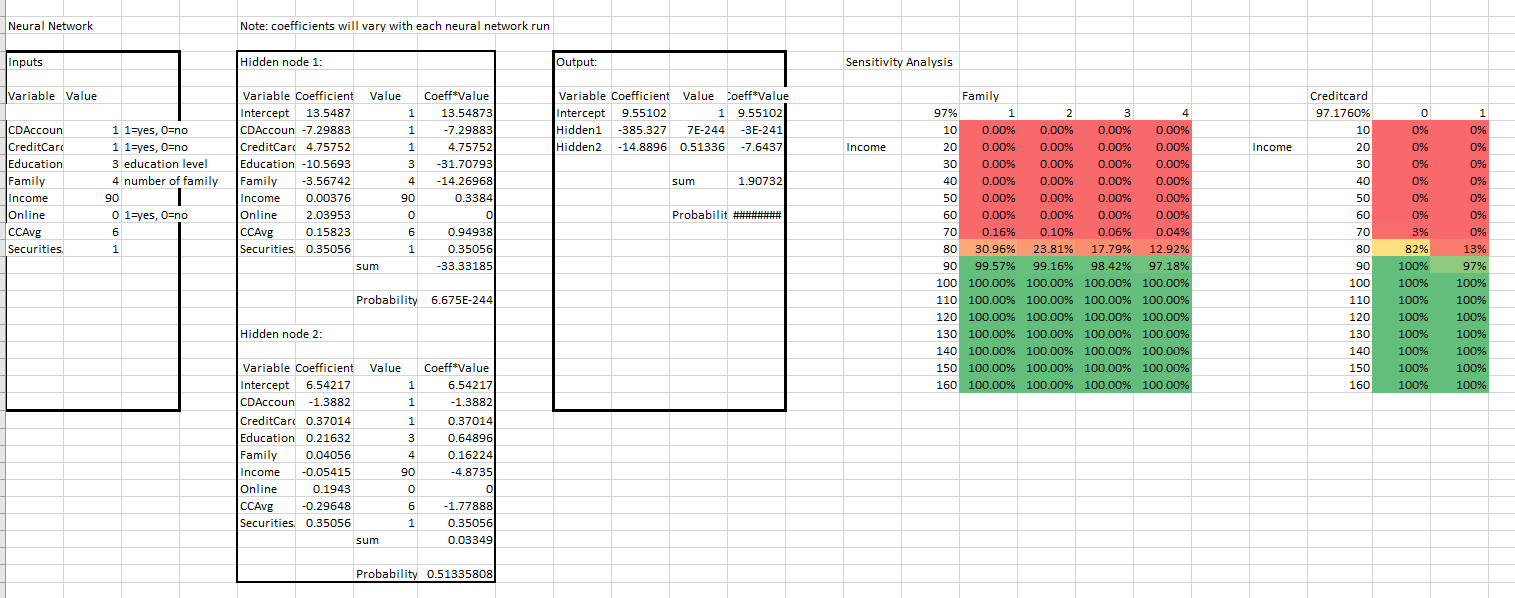
****

5. Create a prediction model of the neural network. Using the prediction model, perform a sensitivity analysis for the neural network model similar to the logit and probit sensitivity analysis.

Logit Sensitivity



Probit Sensitivity



The Neural Network (NN) Logit Sensitivity analysis was conducted using the same variables with the Logit Analysis and the moderating effects. The first comparison was conducted between Income and Family. These numbers depicted an increase in probability at the greater than 80K€ income. At this level of income an individual is more likely to go for a personal loan. This is similar to the Logit Sensitivity analysis in results, but is an opposite effect with respect to logical thinking as the family size increases. The NN sensitivity analysis depicts a smaller family size as a higher probability for a personal loan on a range from 1 to 4 members (61%-51%). On the other hand the Logit Sensitivity Analysis shows the probability of a personal loan to increase with an increase in family size. From a logical standpoint, this scenario makes sense compared to the NN Logit Sensitivity Analysis. A person with higher family numbers would potentially have more reasons to obtain a personal loan than say a person with fewer family members. This dichotomy (NN Logit Sensitivity versus Logit Sensitivity) is interesting from the Personal Loan standpoint. A bank may choose a different method of analysis when researching Personal Loan probabilities depending on the criteria it sets for the loan.

A different reasoning follows for the sensitivity analysis relative to the moderating effects of Income and Credit Card comparison. A person without any credit cards and a higher income is more likely to get a personal loan. From a logical stand this may not make sense. The NN Logit Sensitivity analysis depicts a person with no card from Universal Bank to be more than likely to get a loan, but it may be tough since they “may” not have any credit to their name from Universal Bank. This could only make sense if there was a deeper analysis like conducting more moderating effect analyses such as taking into account if they have credit cards with other banks.

The NN Probit Analysis produced similar results as the NN Logit but with lower probabilities. For instance, a smaller family size has a higher probability for a personal loan when the range is from 1 to 4 but the values are from 31%-12%, respectively. The lower numbers may be attributed to Probit being centered on the mean. With the large data set, one could also surmise that there are many extremes which is why the NN Probit Sensitivity Analysis may be a bit more reflective of the true data set. Of interesting note, the similar results for the Income to Credit Card NN Probit Sensitivity Analysis were realized as in the NN Logit Sensitivity Analysis. This may be a point about the interactions of these two variables. Their dependencies may have a relationship and can possibly be a good indicator when determining if a person will want a Personal Loan.

**Appendix**

