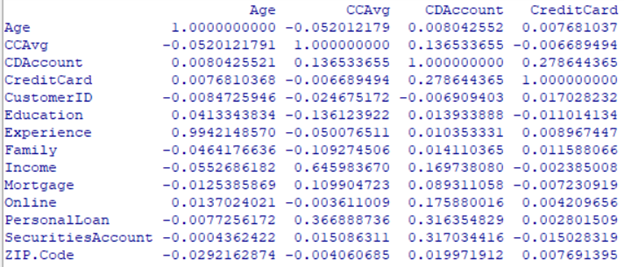
David Doman, Alexa Mowbray, Ryan Cathcart, Jason Taylor

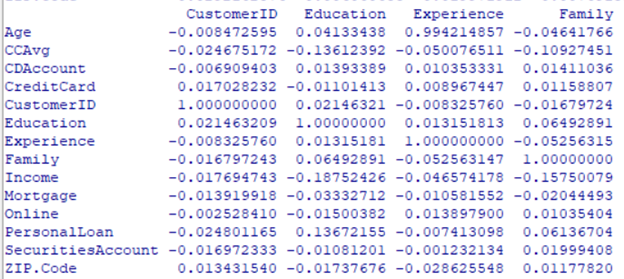
**Loan Analysis**

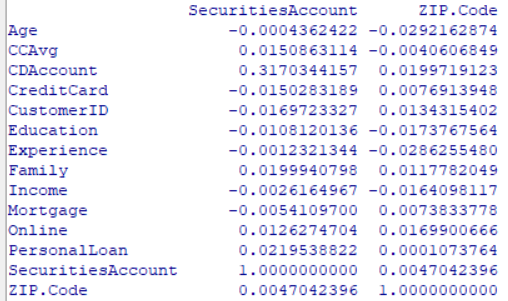
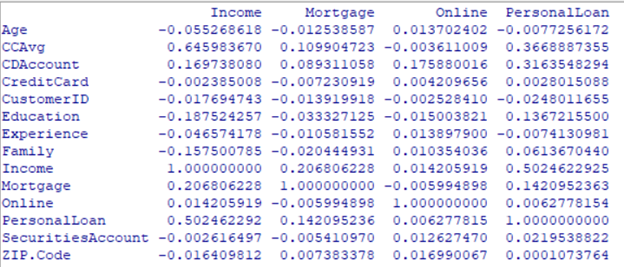
An analysis was done to determine the various variables that went into determining which factors ultimately influenced whether a customer took out a loan at Universal Bank.

*1) Perform a logit and probit analysis of the variables that affect whether a customer takes out a loan. Consider only 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. (20%)*

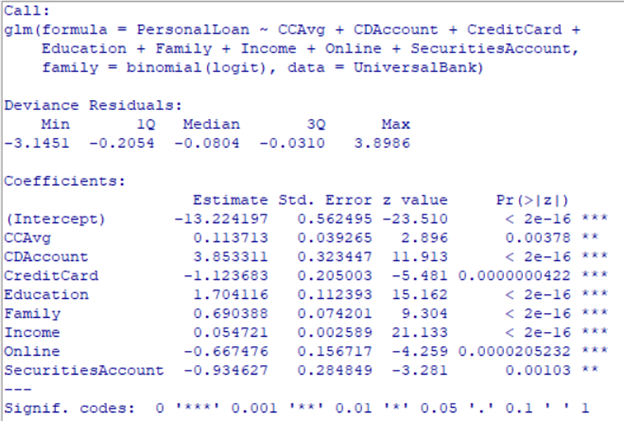
A logit analysis of the different factors was completed to glean which variables were significant and whether those variables influenced whether a loan was taken or not. Our output (personal loan) was not in the logit analysis with the other variables. The correlation matrix was used to identify variables that were highly correlated. Below is a screenshot of the analysis from R:





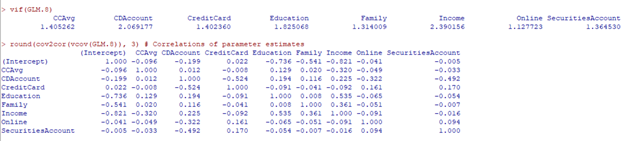


We noticed an extremely high correlation with a value of almost 1 between Age and Experience so we decided to drop one of these variables. This happened to be Age since it was the less significant of the two. We then reran the logit analysis and identified those variables that were not significant. Zip Code, Mortgage, Customer ID, and Experience were insignificant so they were removed from the run. Below are the results for the logit analysis with the remaining significant variables:

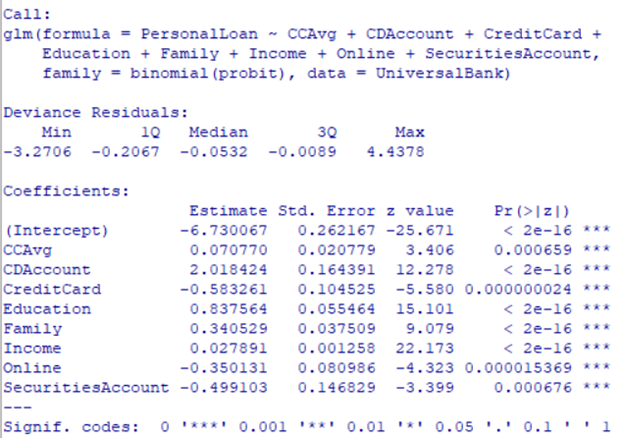


CCAvg, CDAccount, Credit Card, Education, Family, Income, Online and SecurititesAccount were the remaining significant variables used for the logit analysis. CCAvg, CDAccount, CreditCard, Education, Family, Income, and Online were the most significant, while CCAvg and SecuritiesAccount were fairly significant.

We then decided to run the variation inflation analysis to double check that none of our variables had a VIF over 5. If any were to have been over 5, it would have presented some concern. All variables had exceptionally low VIF’s so our analysis looks good in that aspect.



After logit analysis, we ran a probit analysis. While logit relies on logistic distribution, probit focuses on the normal distribution. The biggest difference between the two is that the coefficients end up being different. While logit is more sensitive to extreme values, probit is more sensitive to values closer to the mean. Below are the results for the probit analysis:

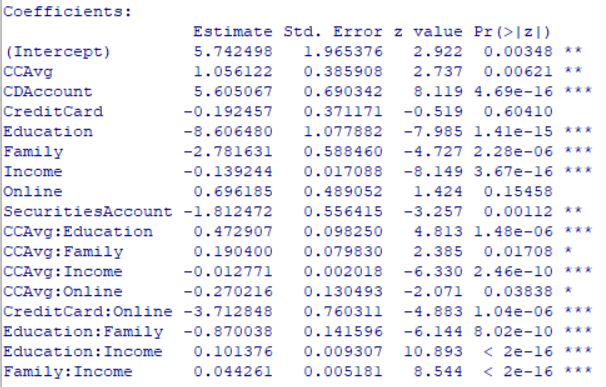


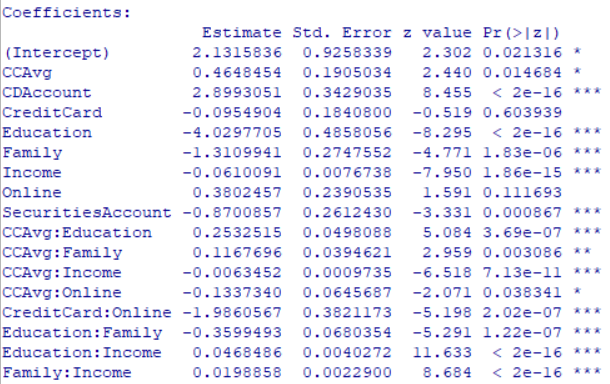
The coefficients that increase the likelihood of customers taking out a loan include: CCAvg, CDAccount, Education, Family, and Income. CDAccount had the highest value of the previously mentioned variables. The coefficients that decrease the likelihood of a customer taking out a loan include: CreditCard, Online and SecuritiesAccount.

*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. (20%)*

We first ran the logit analysis including all the possible interactions between our significant variables used in our main effect logit run. Since there were many interactions, it was beneficial to use parentheses around our main effects and squaring it. This allowed us to save a lot of time typing out each possible interaction manually. After the initial run, we then continued to run the analysis multiple times, each time removing those interactions that were insignificant. Although some of our main effects would become insignificant during these runs, we did not remove any of them since they were found to be significant in our main effect runs from question 1. Below are the results for both the logit and probit analyses:

**LOGIT with Moderating Effects**

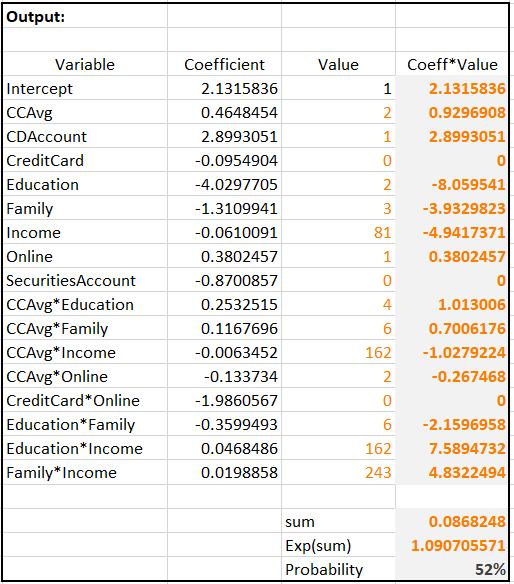
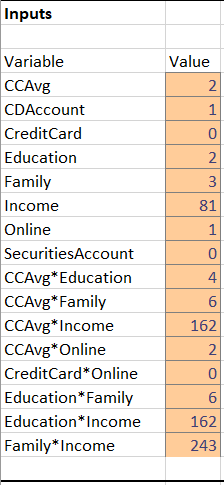


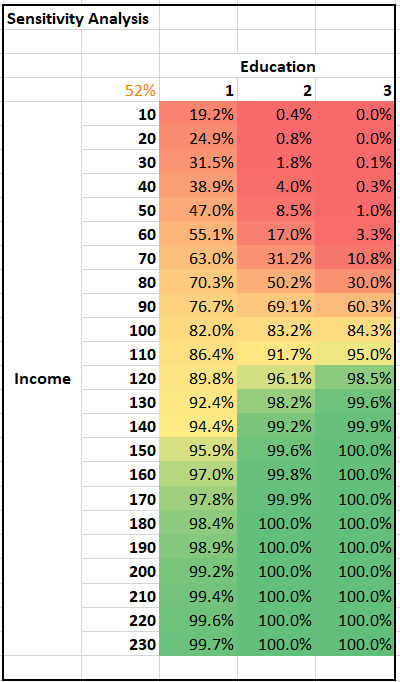
**Probit with Moderating Effects**

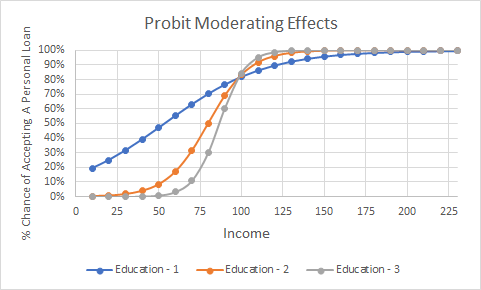
Our results showed that CCAvg is moderated by Education, Family, Income and Online. CreditCard was moderated by Online, Education is moderated by Family and Income and Family is moderated by Income. The CCAvg interactions tell us that if you use credit cards a lot and are highly educated or have a large family there is an increased likelihood of taking out a loan. The interaction between having a credit card and online told us that if you have a credit card with Universal Bank and do your banking online then you are less likely to accept a personal loan. Higher education and higher income increase the likelihood of accepting a personal loan.

The coefficients of the education and income variables are both less than 0. To arrive at the probability, we take the logarithm of the sum of exponents which will always return a positive number. We can therefore conclude that the more volatile of the two factors is income.

*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. (20%)*





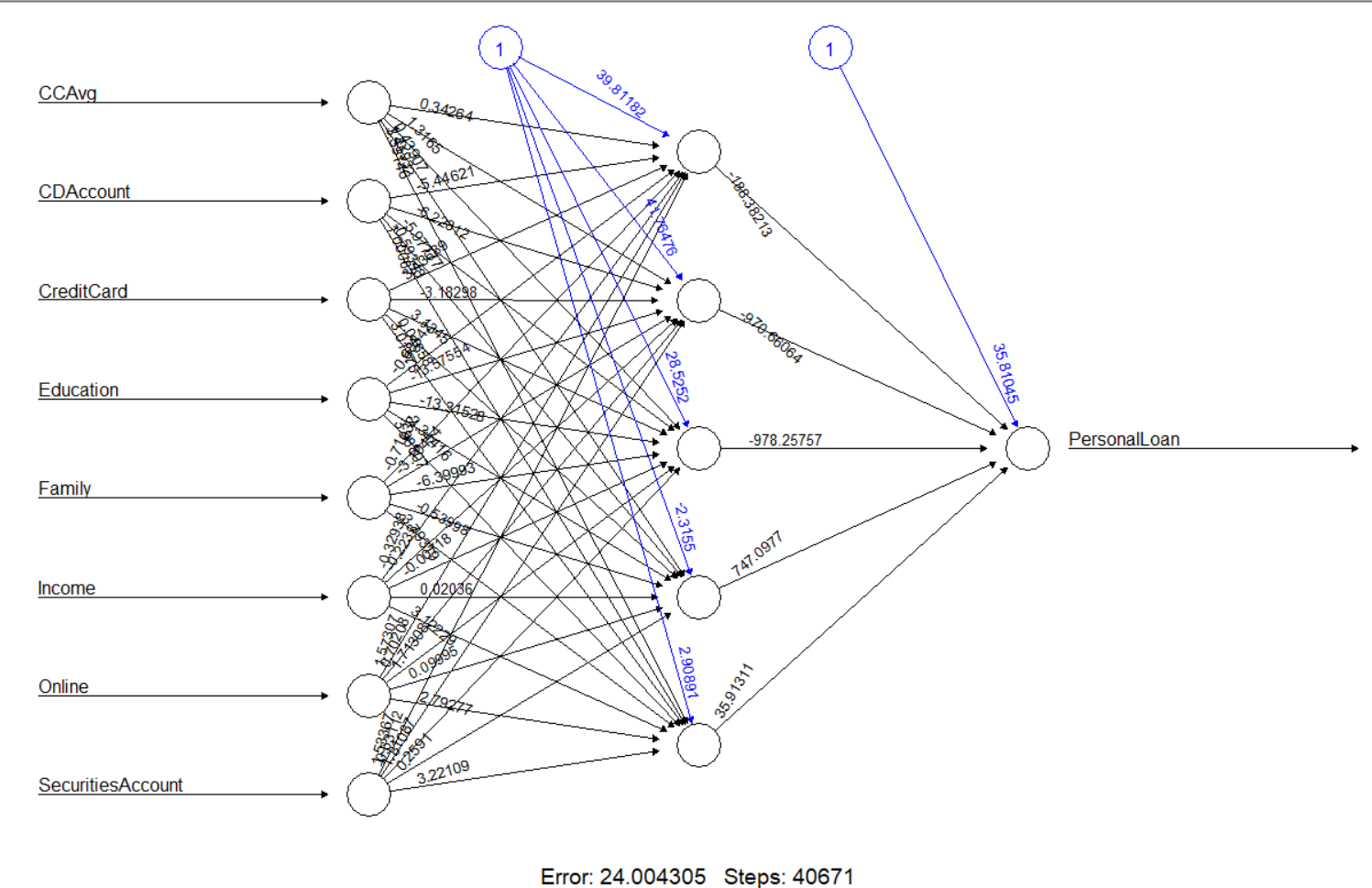


CCAvg, CDAccount and Online all have positive influences on the likelihood that someone will accept a personal loan with CDAccount having the largest impact. Education, Family, and Income had the largest negative coefficients. CreditCard and SecuritiesAccount both have negative impact on if a loan will be accepted as well. The large negative coefficient for standalone Education, Family and Income is moderated/offset by the high positive coefficient of the Education:Income and Family:Income interactions. Family: Income also had an extremely high coefficient, which increases the likelihood of accepting a loan.

*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. (20%)*

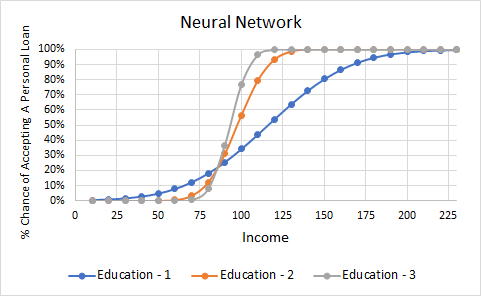


Since the hidden nodes should be about two-thirds the amount of your input nodes, we decided to use 5 hidden nodes. We were able to achieve an error value of 24, which is quite good compared to running the analysis with 3 or 4 hidden nodes.

*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. (20%)*

**

The Neural Network analysis resulted in similar findings to that of the logit and probit analyses. We see those less educated and have incomes lower than about $90,000 are more likely to accept a loan than those with the similar incomes and higher educations. Those with level 2 & 3 Education levels seem to have no chance of accepting a loan until income exceeds about $60,000. We can also see that those with Education levels 2 & 3 and incomes greater than $125,000 that it is almost a certainty that they will accept a personal loan. Furthermore, at an income level of $125,000, those with Education level 1 are half as likely to accept a loan than Education 2 & 3 individuals.