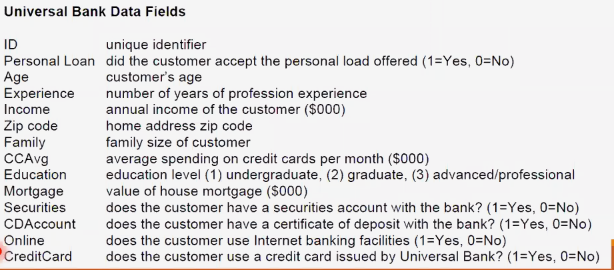
# SCM 651 Homework 4 – Loan Analysis

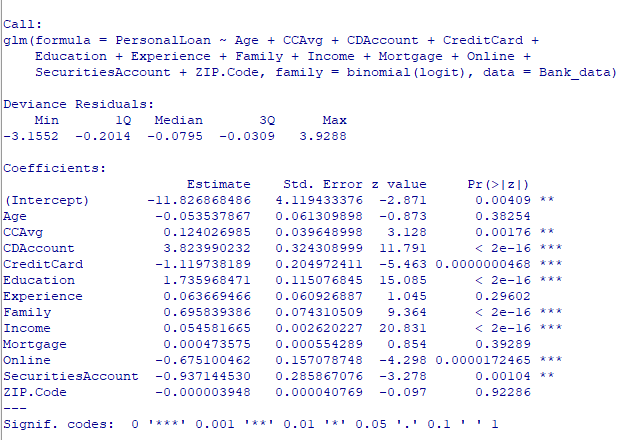
## Perform a logit and probit analysis of the variables that affect whether a customer takes out a loan. Consider only main effects. Which variables are significant? How do the significant variables influence the likelihood of taking out a loan?

Logit Regression

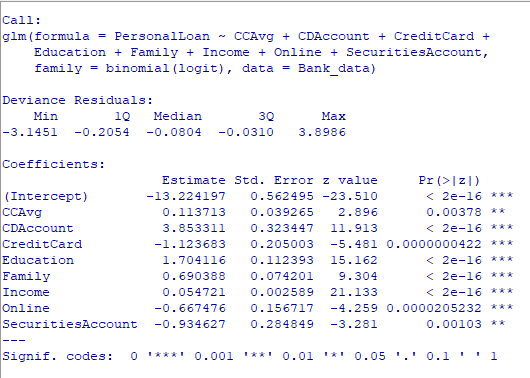
The Universal bank dataset was used to perform a logit regression to predict which customers will take out a personal loan. The dataset contains 14 variables, and a description of each variable is shown below:



We will start with all of the variables except for CustomerID (ID) as this is just an identifier and will not be correlated with the PersonalLoan variable.

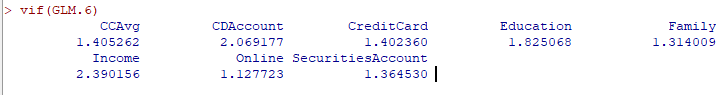


Most of the variables in the dataset are statistically significant in predicting whether someone accepts a personal loan, though Age, Experience, Mortgage, and Zip.Code are not. We will remove these variables and run the regression analysis again, until only statistically significant variables are left.



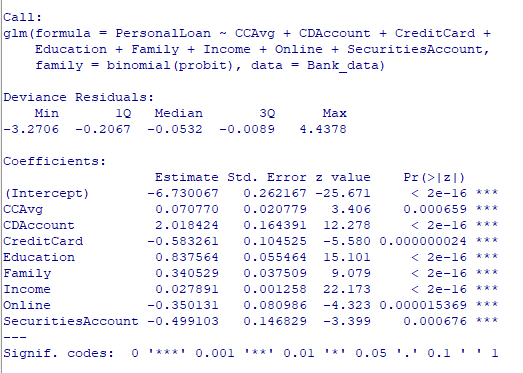
After removing all variables that are not statistically significant to α = 0.05 we are left with the logit model above. The remaining significant variables are CCAvg, CDAcount, CreditCard, Education, Family, Income, Online, and SecuritiesAccount. Higher (CCAvg), having a CDAccount with Universal Bank, higher Education, having a larger family Family and higher Income increase the odds that a customer accepts a PersonalLoan. However, having a Credit Card with Universal Bank, banking Online and having a Securities Account decrease the odds of accepting a PersonalLoan.

Checking the variance inflation factor of the remaining predictor variables, we see that none of them are collinear since the factor for each is less than 10.



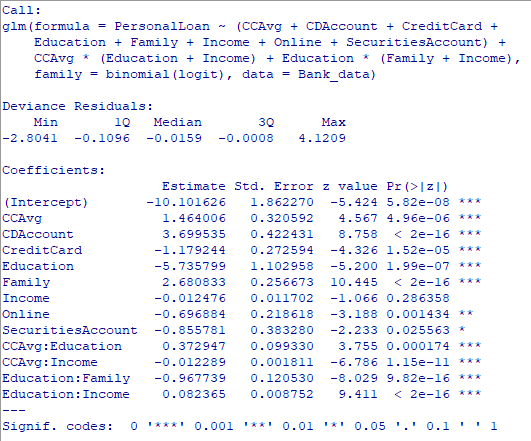
# Probit Regression

Performing a probit regression on the same dataset yields similar results, with the same set of variables being statistically significant as in the logit regression.



## 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?

After adding moderating effects, a logit regression model was created in R Commander with the following variables:



Keeping all of the main effects from the first logit model, we found that CCAvg:Education, CCAvg:Income, Education:Family, and Education:Income are statistically significant moderating effects. Even though Income has become statistically insignificant, we must keep it in the model because its interaction with CCAvg is significant.

The coefficient for CCAvg changed from 0.07 to 1.46 after adding moderating effects, indicating that it has a strong positive effect on whether a customer accepts a PersonalLoan. CCAvg is moderated by Education to increase the odds of accepting a PersonalLoan, and Income to slightly decrease the odds of accepting a PersonalLoan. Therefore, highly educated people who also spend a lot on their credit cards every month are more likely to accept a PersonalLoan than less educated people who also don’t spend as much on credit cards. This offsets the new coefficient for Education alone of -5.74, while it was 0.84 without moderating effects. Education is also moderated by Family and Income. The coefficient on Family increased from 0.34 to 2.68 when adding moderating effects, while Income only had a small change from 0.02 to -0.01.

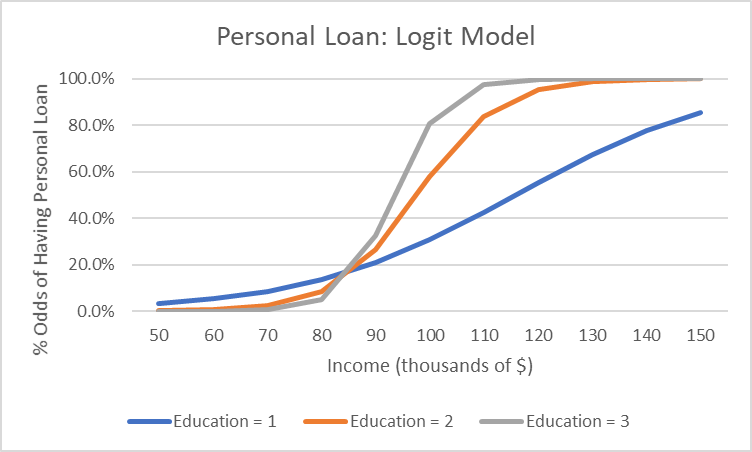
Overall, this model indicates that larger families with higher household incomes and education levels that are comfortable spending on credit cards every month are more likely to have a personal loan.

## 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.

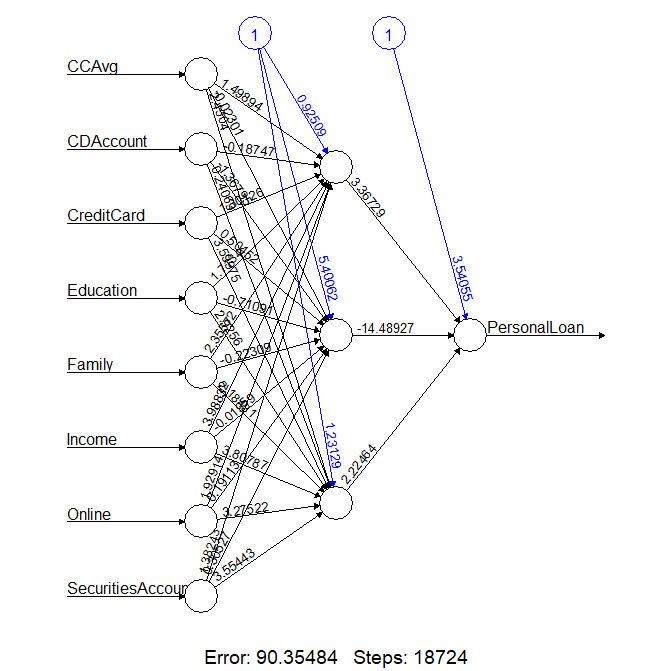


The spreadsheet model shown above indicates that Education:Income is the one of the stronger positive effects on whether a customer accepts a PersonalLoan, along with CCAvg, CDAccount and Family. The variables with the strongest negative impact to the likelihood of a customer having a Personal loan is Education, along with the baseline level from the Intercept. The large negative contribution from Education alone is largely moderated out by Education:Income and CCAvg:Education.

A sensitivity analysis done with this model highlights the impact that Income and Education have on the probability. Over all, income lower than $85,000 has a smaller effect on the likelihood a person accepts a personal loan, (less than 20%), but the possibility increases dramatically when income is over $85,000. When we also consider Education, we find that at low to middle income levels, less than about $85,000, someone with lower Education is more likely to accept a Personal loan. At Income greater than $85,000 however, this inverts rather quickly and people with more Education are more likely to accept a PersonalLoan.

## Perform a neural network analysis of the variables found to be significant in the logit and probit analysis above.



The neural net shown above was created based on the Universal Bank data in R, using only the main effect variables from the initial Logit model. There are three hidden nodes, and a threshold level of 0.04 was used to get the model to converge to reasonably low error in the default number of steps.

## 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.

A sensitivity analaysis of yhe neural net model for Education equal to 1 is very similar to the Logit model, with a 50% probability of having a PersonalLoan at an Income of about $120,000.

At Education levels 2 and 3 however, the neural net shows considerably higher probability of having a PersonalLoan at a given Income level than the Logit model. In the neural net, the probability curves versus Income are very smooth and similar, with the 50% probability level shifted to lower Income levels for higher Education. It should be noted however that this neural net model was failing to calculate probabilities for any Education level when Income was greater than $170,000, while the Logit model was able to calculate a probability. While $170,000 was within the bounds of the available data for Income, which ranged from $8,000 to $224,000 with the 50th percentile being at $64,000, there may not have been enough data at the higher Income range to properly develop the neural net. However, since the probabiliity of having a PersonalLoan at an Income of $170,000 is greater than 95% for any Education level, this neural net model is probably adequate for our purposes.

# Academic Integrity

