Clifford Forrester, Kathleen Fox, Mashundra Maclin, Courtney Smith, and Joyce Woznica

TEam 73  
Homework #4

Loan Analysis

Contents

[Introduction 1](#_Toc10057978)

[Universal Bank Data Fields 1](#_Toc10057979)

[Background 1](#_Toc10057980)

[#1 – Logit and Probit Analysis 2](#_Toc10057981)

[Logit Analysis 2](#_Toc10057982)

[Probit Analysis 4](#_Toc10057983)

[Conclusions 5](#_Toc10057984)

[#2 – Moderating Effects 5](#_Toc10057985)

[#3 – Final Regression Model 6](#_Toc10057986)

[#4 – Neural Network Analysis 6](#_Toc10057987)

[#5 – Prediction Model of Neural Network 7](#_Toc10057988)

List of Figures

[Figure 1: Descriptive Statistics on Bank Dataset 1](#_Toc10057967)

[Figure 2: Tests for Linearity 2](#_Toc10057968)

[Figure 3: Logit run with RStudio and Same Code from Rcmdr 2](#_Toc10057969)

[Figure 4: AIC Step for the Best Logit Model Script and Results 3](#_Toc10057970)

[Figure 5: Best Logit Model without Experience Variable Summary 4](#_Toc10057971)

[Figure 6: Best Probit Model R-Code 4](#_Toc10057972)

[Figure 7: Best Probit Model without Experience Variable Step Results 4](#_Toc10057973)

[Figure 8: Best Logit Model Summary 5](#_Toc10057974)

[Figure 9: Neural Network R-code without Binary Observations 6](#_Toc10057975)

[Figure 10: Neural Network Diagram without Binary Observations 7](#_Toc10057976)

[Figure 11: Neural Network Results Matrix from “neuralnet” R-command without binary variables 7](#_Toc10057977)

# Introduction

We will be using data from Universal Bank to determine the factors that influence whether a customer will take out a loan from the bank. For this assignment we will be submitting a *logit*, *probit* and neural network analysis of loan acquisition behavior.

## Universal Bank Data Fields

The following fields are in the comma-separated-value file provided for this exercise.

* **ID** – unique identifier
* **PersonalLoan** – did the customer accept the personal load offered (1=Yes, 0=No)
* **Age** – customer’s age
* **Experience** – number of years of profession experience
* **Income** – annual income of the customer ($000)
* **ZIP.code** – home address zip code
* **Family** – family size of customer
* **CCAvg** – average spending on credit cards per month ($000)
* **Education** – education level (1) undergraduate, (2) graduate, (3) advanced/professional
* **Mortgage** – value of house mortgage ($000)
* **Securities** – does the customer have a securities account with the bank? (1=Yes, 0=No)
* **CDAccount** – does the customer have a certificate of deposit with the bank? (1=Yes, 0=No)
* **Online** – does the customer use Internet banking facilities (1=Yes, 0=No)
* **CreditCard** – does the customer use a credit card issued by Universal Bank? (1=Yes, 0=No)

## Background

Prior to doing any analysis on the data, we ran some general descriptive statistics on the data in the dataset provided.

A screenshot of a cell phone

Description automatically generated

Figure 1: Descriptive Statistics on Bank Dataset

We also performed various tests to determine if correlation and regression were applicable to our dataset. In this analysis, we found the following:

* No multicollinearity
* Single outlier

A screenshot of a social media post

Description automatically generated

Figure 2: Tests for Linearity

Our tests confirmed that we could use general linear correlation for this dataset, but we did have a single outlier. However, we decided to leave the outlier in for our testing, but wanted to note that we ran these tests.

# #1 – Logit and Probit Analysis

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%)

Clifford to provide response, Joyce put in what she had done

We performed a *logit* and *probit* analysis using ‘Rcmdr’ in the RGui and scripting in R with RStudio both showing the same results.

## Logit Analysis

A screenshot of a cell phone

Description automatically generated

Figure 3: Logit run with RStudio and Same Code from Rcmdr

Using the AIC Step method (the same that it used in Rcmdr), we found the best model as shown in *Figure 3: AIC Step for the Best Logit Model Script and Results*.

A screenshot of a cell phone

Description automatically generatedA screenshot of a cell phone

Description automatically generated  
A screenshot of text

Description automatically generated

Figure 4: AIC Step for the Best Logit Model Script and Results

We determined that we could drop Experience since the *p-value* for this variable is actual > 0.05 (0.09458).

A screenshot of a cell phone

Description automatically generated

Figure 5: Best Logit Model without Experience Variable Summary

## Probit Analysis

\*this is just some ideas on how to do this section\*

A screenshot of a cell phone

Description automatically generated

Figure 6: Best Probit Model R-Code

A screenshot of a cell phone

Description automatically generated

Figure 7: Best Probit Model without Experience Variable Step Results

However, we dropped out CustomerID because it is simply a row number or arbitrary ID and should not be included.

A screenshot of a cell phone

Description automatically generated

Figure 8: Best Logit Model Summary

## Conclusions

\*\* Do we use the spreadsheets he used in class here?

Using these two methods, we determined that the following variable were significant because of their below 0.05 *p-value* as shown in the descriptions in the figures provided in this section.

* Income – The higher the income, the more likelihood of taking out a loan. In fact, for a $1 of income, the likelihood goes up by X%.
* Family – finish each of these
* CCAvg
* Education
* SecuritiesAccount
* CDAccount
* Online
* CreditCard

As shown in *Figure 5: Best Logit Model without Experience Variable* and above, the variables have different effects on the likelihood of taking out a loan. We have indicated this information in the bullets above.

# #2 – Moderating Effects

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%)

Kathleen to provide response

# #3 – Final Regression Model

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%)

Courtney to provide response

# #4 – Neural Network Analysis

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%)

Mashundra to provide response, Joyce put in what she had done.

Using the variables that we found significant in both our *logit* and *probit* analysis performed in the section entitled #1 – Logit and Probit Analysis beginning on page 2. These variables were as follows:

* Income
* Family
* CCAvg
* Education
* SecuritiesAccount
* CDAccount
* Online
* CreditCard

It was indicated that the binomial observations can often cause problems in regression and neural networks, so we dropped these and reran the neural network as shown in *Figure 9: Neural Network R-code without Binary Observations*.

A screenshot of a cell phone

Description automatically generated

Figure 9: Neural Network R-code without Binary Observations

The resulting plotted neural network is shown in *Figure 10: Neural Network Diagram without Binary Observations*.

A close up of a map

Description automatically generated

Figure 10: Neural Network Diagram without Binary Observations

# #5 – Prediction Model of Neural Network

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%)

Justify your answers. Provide a snapshot of output from your analysis in your final paper.

Using the information provided in our neural network model, the following prediction model was used (reflected in *Figure 11: Neural Network Results Matrix from “neuralnet” R-command*).

A screenshot of a cell phone

Description automatically generated

Figure 11: Neural Network Results Matrix from “neuralnet” R-command without binary variables

Since binary variables can cause issues with neural networks, we provided the results matrix for our neural network without these variables, but including all variables that were determine significant in the *probit* and *logit* models.

Using this information, we created the XXX – Joyce to finish.

Joyce to finish this after Prof Harter shows in class on Wednesday night.