|  |
| --- |
| **Question 1: 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). Copy screen snapshots of your analysis in R to your report. (20%)**  Probit model and summary:    **Which variables are significant?**  *The intercept is significant along with the variables Income, Family, CCAvg, Education, SecuritiesAccount, CDAccount, Online, and CreditCard.*  **How do the significant variables influence the likelihood of taking out a loan?**  *The coeffcients for Income, Family, CCAvg, and Education are all positive sloping. As a result, when they increase in magnitude the likelihood of a loan increases. Additionally, as SecuritiesAccount, Online, and CreditCard go from 0 to 1, the likelihood of a personal loan decreases as their coefficients are negative. And lastly, as CDAccount goes from 0 to 1 the likelihood of a personal loan increases as its coefficient is positive.*  Logit model and summary:    ***Which variables are significant?***  *The intercept is significant as are the variables CCAvg, CDAccount, CreditCard. Education, Family, Income, Online, and SecuritiesAccount all with p-values less than our .05 threshold.*  ***How do the significant variables influence the likelihood of taking out a loan?***  *The coefficients for CCAvg, Family, Education, and Income are all positive so as they increase the likelihood of a loan increases. As CDAccount go from 0 to 1, the likelihood of a personal loan increases. As CreditCard, Online and SecuritiesAcounts go from 0 to 1 the likelihood of a personal loan decreases.* |
| **Question 2:**  **Add moderating effects (interactions of variables). Copy screen snapshots of your analysis in R to your report. (20%)**  **Which interactions make sense conceptually?**  *It conceptually makes sense to combine Education and Income as people with high income probably feel more confident to repay a loan if they have the education to increase income over time or more easily get another job if they lose their current source of income.*   |  |  | | --- | --- | | Logit: Education\*Income | *The intercept, Education, Income and the moderating effect of Education\*Income are all statistically significant with p-values less than 0.05.*  *Taken individually the coefficients of Income and Education show a decreased likelihood of taking out a loan. Education itself has a negative impact on taking out a loan. A potential reason for this could be that the client already had a large loan burden due to college loans. When combined, the interaction of these two variables increases the likelihood of taking out a loan. It’s possible this is the result of high-income earners who have already completed college or a master’s degree looking for a potential home loan. Or that it is easier for high earners to qualify for a loan.* | |
| **Question 3:**  **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. Copy screen snapshots of your analysis in R to your report. (20%)**     1. **Create a spreadsheet prediction of the model.** 2. **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)?**  The income data ranges from $8,000 to $224,000. Education level was divided into 3 buckets – 1 for at least a bachelor’s degree, 2 for at least a graduate degree and 3 for those with more advanced degrees and professionals. We ran a sensitivity analysis on these two variables which showed increasing likelihood of a personal loan being approved as education and income both increased, as expected. However, we expected income to contribute more than education but that does not bear out in the analysis. Despite very high levels of income, when a person only had a bachelor’s degree the highest predicted likelihood of a loan only reached 37.3%. For education levels 2 and 3 the effect was more pronounced. The likelihood of a loan reached 100% at $209k and above for level 2 graduate degrees and 100% for about $162k for level 3 education levels. |
| **Question 4:**  **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%)**  **result.matrix**    **plot(net)**    We ran the neural network model multiple times before we felt we had a useable solution. The hazard with the model is that the solution could get stuck in a local optima. The main issue is the coefficient for the Income variable. If it is too large then it blows up the logistic math and Excel cannot handle this in the sensitivity in Question 5. |
| **Question 5:**  **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%)**      The sensitivity model is consistent with our earlier sensitivity model generated from the logistical regression. The model provides additional evidence in support of the relationship established in the logistical models. In particular, the Education level is a significant driver of loan approval. One anomaly worth noting occurs when Income reaches $116,000. At this level Education 3 lags level 2, 56% to 67%, but then rises to 100% before Education level 2 does. This is something worth investigating further.  One key piece of information that we felt would have helped put these results into context further was the size of the loans applied for and the purpose for the loans. Perhaps we would have been able to draw more distinct subgroups and reach stronger conclusions about what determines loan approval had we had access to this information. |