# **MAT 303 Module Five Problem Set Report**

Logistic Regression

[Your Full Name]

[Your SNHU Email]

Southern New Hampshire University

Note: Replace the bracketed text on page one (the cover page) with your personal information.

## **1. Introduction**

*Discuss the statement of the problem with regard to the statistical analyses that are being performed. Address the following questions in your analysis:*

* *What is the data set that you are exploring?*
* *How might your results be used?*
* *What type of analyses will you be running in this problem set?*

Caution sign icon Answer the questions in a paragraph response. Remove all questions and this note before submitting! Do not include R code in your report.

## **2. Data Preparation**

*There are some important variables that you have been asked to analyze in this problem set. Identify and explain these variables. Address the following questions in your analysis:*

* *What are the important variables in this data set?*
* *How many rows and columns are present in this data set?*

Caution sign icon Answer the questions in a paragraph response. Remove all questions and this note before submitting! Do not include R code in your report.

## **3. First Logistic Regression Model**

### **Reporting Results**

*Report the results of the regression model. Address the following questions in your analysis:*

* *Write the general form* *and the prediction equation of a logistic regression model for defaulting on credit as response variable and credit utilization, education as predictors.*
* *Now write the prediction equation of this model in terms of the natural log of odds to express the beta terms in linear form.*
* *What do the following terms from the model above mean in terms of an individual defaulting on their credit?*
  1. pi

     
  2. Equation

     A fraction with pi in the numerator, and one minus pi in the denominator
* *Create this logistic regression model. Write the prediction model equation (in terms of the natural log of odds) using outputs obtained from your R script. Round all figures to four decimal places.*
* *Interpret the estimated coefficient of credit utilization.*
* *Obtain the confusion matrix and report the counts for true positives, true negatives, false positives, and false negatives.*
* *Report the following:*
  1. *Accuracy*
  2. *Precision*
  3. *Recall*

Caution sign icon Answer the questions in a paragraph response. Remove all questions and this note before submitting! Do not include R code in your report.

### **Evaluating Model Significance**

*Evaluate model significance for the regression model. Address the following questions in your analysis:*

* *Perform the Hosmer-Lemeshow goodness of fit test to assess whether the model is appropriate for the data set. Identify the null and alternative hypotheses, the test statistic, and the P-value. Use a 5% level of significance.*
* *Which terms are significant in the model based on Wald’s test? Use a 5% level of significance.*
* *Obtain the Receiver Operating Characteristic (ROC) curve. Interpret the graph and explain what it illustrates.*
* *What is the value of AUC? Interpret what this value represents.*

Caution sign icon Answer the questions in a paragraph response. Remove all questions and this note before submitting! Do not include R code in your report.

### **Making Predictions Using Model**

*Make predictions using the regression model. Address the following questions in your analysis:*

* *What is the probability of an individual defaulting on credit who has a credit utilization of 35% and has high school education? Find the odds of this event occurring. Comment on these outputs.*
* *What is the probability of an individual defaulting on credit who has a credit utilization of 35% and has post graduate education? Find the odds of this event occurring. Comment on these outputs.*

Caution sign icon Answer the questions in a paragraph response. Remove all questions and this note before submitting! Do not include R code in your report.

## **4. Second Logistic Regression Model**

### **Reporting Results**

*Report the results of the regression model. Address the following questions in your analysis:*

* *Write the general form* *and the prediction equation of a logistic regression model for defaulting on credit as the response variable and credit utilization, assets, and missed payment as predictors. Note that this general form should be written in terms of E(y) and exponents.*
* *Now write the prediction equation of this model in terms of the natural log of odds to express the beta terms in linear form.*
* *Create this logistic regression model. Write the prediction model equation (in terms of the natural log of odds) using outputs obtained from your R script. Round all figures to four decimal places*
* *Obtain the confusion matrix and report the counts for true positives, true negatives, false positives, and false negatives.*
* *Report the following:*
  1. *Accuracy*
  2. *Precision*
  3. *Recall*

Caution sign icon Answer the questions in a paragraph response. Remove all questions and this note before submitting! Do not include R code in your report.

### **Evaluating Model Significance**

*Evaluate model significance for the regression model. Address the following questions in your analysis:*

* *Perform the Hosmer-Lemeshow goodness of fit test to assess whether the model is appropriate for the data set. Identify the null and alternative hypotheses, the test statistic, and the P-value. Use a 5% level of significance.*
* *Which terms are significant in the model based on Wald’s test? Use a 5% level of significance.*
* *Obtain the ROC curve. Interpret the graph and explain what it illustrates.*
* *What is the value of AUC? Interpret what this value represents.*

Caution sign icon Answer the questions in a paragraph response. Remove all questions and this note before submitting! Do not include R code in your report.

### **Making Predictions Using Model**

*Make predictions using the regression model. Address the following questions in your analysis:*

* *What is the probability of an individual who has a credit utilization of 35%, owns only a car, and has missed payments in the last 3 months. Find the odds of this event occurring. Comment on these outputs.*
* *What is the probability of an individual who has a credit utilization of 35%, owns a car and a house, and has not missed payments in the last 3 months. Find the odds of this event occurring. Comment on these outputs.*

Caution sign icon Answer the questions in a paragraph response. Remove all questions and this note before submitting! Do not include R code in your report.

## **5. Conclusion**

*Describe the results of the statistical analyses and address the following questions:*

* *Based on the analysis that you have performed and assuming that the sample size is sufficiently large, would you recommend using this model? Why or why not?*
* *Fully describe what these results mean in your scenario using proper statistical terms and concepts.*
* *What is the practical importance of the analyses that were performed?*

Caution sign icon Answer the questions in a paragraph response. Remove all questions and this note before submitting! Do not include R code in your report.

## **6. Citations**

*You are not required to use external resources for this report. If none were used, remove this entire section. However, if you used any resources to help you with your interpretation, you must cite them. Use proper APA format for citations.*

*Insert references here in the following format:*

Author's Last Name, First Initial. Middle Initial. (Year of Publication). Title of book: Subtitle of book, edition. Place of Publication: Publisher.