

NBM2 – NBM2 TASK 2: LOGISTIC REGRESSION FOR PREDICTIVE MODELING

PREDICTIVE MODELING – D208

PRFA – NBM2

TASK OVERVIEW

SUBMISSIONS

EVALUATION REPORT

EVALUATION REPORT – ATTEMPT 3 – REVISION NEEDED

Overall Evaluator Comments

EVALUATOR COMMENTS

To support you in demonstrating proficiency, you'll be unable to resubmit your task until you've consulted with a Course Instructor. The comments below provide details about needed revisions.

A1: Research Question.

Competent The summary includes 1 research question that is relevant to a realistic organizational situation and can be addressed using the selected data set and logistic regression.

There are no comments for this aspect.

A2: Objectives and Goals.

Competent The submission clearly defines the objectives or goals of the data analysis, and the objectives or goals are reasonable for the scope of the scenario and are represented in the available data.

There are no comments for this aspect.

B1: Summary of Assumptions.

Approaching Competence The submission summarizes the assumptions of a logistic regression model, but 1 or more of the assumptions contain inaccuracies.

EVALUATOR COMMENTS: ATTEMPT 3

The response contains assumptions of logistic regression. The response contains an incorrect statement.

B2: Tool Benefits.

Approaching Competence The submission describes the tool(s) chosen in support of various phases of the analysis, but the description does not include the benefits of using the tool(s) for logistic regression, or the benefits do not logically align with the goal of the analysis.

EVALUATOR COMMENTS: ATTEMPT 3

The response discusses the packages used and the various phases of Logistic Regression. The response is incomplete as a description of which packages are used in which specific phases of the Logistic Regression analysis

B3: Appropriate Technique.

Competent The submission accurately explains why logistic regression is an appropriate technique to analyze the research question from Part I.

EVALUATOR COMMENTS: ATTEMPT 3

NA

C1: Data Goals.

Approaching Competence The submission describes the data preparation goals but not the data manipulations that will be used to achieve the goals. Or the manipulations are not in alignment with the data preparation goals, logistic regression analysis, or the research question.

EVALUATOR COMMENTS: ATTEMPT 3

The response describes some data preparation goals with corresponding manipulations. The response contains some statements that are incorrect/nonsensical; the response contains an incorrect method of data preparation.

C2: Summary Statistics.

Approaching Competence The submission discusses summary statistics but does not discuss the target variables or *all* predictor variables that need to be gathered from the data set to answer the research question, or the discussion contains inaccuracies.

EVALUATOR COMMENTS: ATTEMPT 3

The response contains the number of observations of some variables. The response is incomplete as it does not describe the type of each variable; the response contains an incorrect measure of central tendency for most variables; the response incorrectly is incomplete as it does not describe the number of variables; the response shows that a few variables were incorrectly manipulated; the response does not describe the number of observations for all the variables used in the initial model.

C3: Steps to Prepare the Data.

Approaching Competence The submission explains the steps used to prepare the data for the analysis, but is missing 1 or more necessary steps, or the steps do not include the annotated code, or are not related to a logistic regression analysis.

EVALUATOR COMMENTS: ATTEMPT 3

The response provides the code with the steps to prepare the data. The response contains incorrect steps of preparing the data; the code does not run due to errors.

C4: Visualizations.

Not Evident The submission does not include *both* univariate and bivariate visualizations of the distributions of variables in the cleaned data set.

EVALUATOR COMMENTS: ATTEMPT 3

The aspect will be assessed once revisions to other task prompts are in place.

C5: Prepared Data Set.

Not Evident The submission does not provide a data set.

EVALUATOR COMMENTS: ATTEMPT 3

The aspect will be assessed once revisions to other task prompts are in place.

D1: Initial Model.

Approaching Competence The submission provides an initial logistic regression model from *some*, but not *all*, predictors identified in Part C2, or the model contains inaccuracies.

EVALUATOR COMMENTS: ATTEMPT 3

The response creates an initial model. The initial model is inaccurate due to improper treatment of variables.

D2: Justification of Model Reduction.

Approaching Competence The submission justifies a variable selection procedure or a model evaluation metric to reduce the initial model but does not justify *both*, or the justification is not in alignment with the research question, or the variable selection procedure is not statistically based.

EVALUATOR COMMENTS: ATTEMPT 3

The response describes a method of variable selection. The variable selection is not statistically based as it is creating a new model after removing all the variables with a p-value $> .05$ at once; the response is incomplete as it does not utilize a model evaluation metric to know when to stop reducing the initial model, it instead utilizes the p-values of variables in the model to reduce the initial model.

D3: Reduced Logistic Regression Model.

Not Evident The submission does not provide a reduced logistic regression model.

EVALUATOR COMMENTS: ATTEMPT 3

The aspect will be assessed once revisions to other task prompts are in place.

E1: Model Comparison.

Approaching Competence The submission explains the data analysis process by comparing the initial and reduced logistic regression models but does not include *all* of the given elements, or the explanation contains inaccuracies.

EVALUATOR COMMENTS: ATTEMPT 3

The response describes model evaluation metrics such as Pseudo-R-squared, Confusion matrices, and ROC curves. The response is incomplete as a summary of the logic of the variable selection method is not provided; the response incorrectly analyzes an ROC curve.

E2: Output and Calculations.

Not Evident The submission does not provide the output or calculations of the analysis performed.

EVALUATOR COMMENTS: ATTEMPT 3

The aspect will be assessed once revisions to other task prompts are in place.

E3: Code.

Approaching Competence The submission provides the code used to support the implementation of the logistic regression models, but the code is incomplete or contains inaccuracies.

EVALUATOR COMMENTS: ATTEMPT 3

The response provides Python code. The code doesn't run due to errors.

F1: Results.

Approaching Competence The submission discusses the results of the data analysis, but the discussion does not address *all* of the given elements, or the discussion contains inaccuracies.

EVALUATOR COMMENTS: ATTEMPT 3

The response provides the statistical and practical significance of the model. The response incorrectly truncates one of the coefficients of the reduced model; the response is incomplete as a discussion of the limitations of the data analysis is not provided as it only describes the limitations of the data and not of the data analysis (and stating that logistic regression has a tendency to overfit is not sufficient); the interpretation of the values of the specific coefficients is not provided.

F2: Recommendations.

Not Evident The submission does not recommend a course of action based on results.

EVALUATOR COMMENTS: ATTEMPT 3

The aspect will be assessed once revisions to other task prompts are in place.

G: Panopto Demonstration.

Approaching Competence A Panopto video recording is provided, but it does not include *all* of the given elements, or the video does not capture *both* the presenter and the functioning code for the duration of the video.

EVALUATOR COMMENTS: ATTEMPT 3

A Panopto video is provided that demonstrates the functioning code, compares the full and reduced models, and has the presenter visible for the duration of the video. The interpretation of the specific coefficient values is not provided; a summary of the programming environment used is not provided.

H: Sources of Third-Party Code.

Not Evident The submission does not record *any* web sources.

EVALUATOR COMMENTS: ATTEMPT 3

The aspect will be assessed once revisions to other task prompts are in place.

I: Sources.

Not Evident The submission does not include both in-text citations and a reference list for sources that are quoted, paraphrased, or summarized.

EVALUATOR COMMENTS: ATTEMPT 3

The aspect will be assessed once revisions to other task prompts are in place.

J: Professional Communication.

Not Evident Content is unstructured, is disjointed, or contains pervasive errors in mechanics, usage, or grammar. Vocabulary or tone is unprofessional or distracts from the topic.

EVALUATOR COMMENTS: ATTEMPT 3

The aspect will be assessed once revisions to other task prompts are in place.