

Ale

There is a new version of this application that will load in 115.068 seconds. [Reload page now](#)
Attempt 1 of 3

[Ask for Help](#)

[Print](#)

NBM3 — NBM3 TASK 1: LINEAR REGRESSION MODELING

PREDICTIVE MODELING — D208
PRFA — NBM3

TASK OVERVIEW	SUBMISSIONS	EVALUATION REPORT
---------------	-------------	-------------------

COMPETENCIES

4030.5.2 : Multiple Regression

The graduate employs multiple regression algorithms with categorical and numerical predictors in describing phenomena.

4030.5.3 : Regression Implications

The graduate makes assertions based on regression modeling.

INTRODUCTION

As a data analyst, you will assess data sources for their relevance to specific research questions throughout your career. In your previous coursework, you have performed data cleaning and exploratory data analysis on your data. You have seen basic trends and patterns and can now start building more sophisticated statistical models. In this course, you will use regression models. You will explore both linear regression and logistic regression models and their assumptions.

For this task, you will select **one** of the provided data files from the “Data Sets and Associated Data Dictionaries” found in the Web Links section.

You will then review the data dictionary related to the raw data file you have chosen and prepare the data set file for linear regression modeling. The organizations connected with the given data sets for this task seek to analyze their operations and have collected variables of possible use to support the decision-making processes. You will analyze your chosen data set using linear regression modeling, create visualizations, and deliver the results of your analysis.

REQUIREMENTS

Your submission must be your original work. No more than a combined total of 30% of the submission and no more than a 10% match to any one individual source can be directly quoted or closely paraphrased from

sources, even if cited correctly. The similarity report that is provided when you submit your task can be used as a guide.

You must use the rubric to direct the creation of your submission because it provides detailed criteria that will be used to evaluate your work. Each requirement below may be evaluated by more than one rubric aspect. The rubric aspect titles may contain hyperlinks to relevant portions of the course.

*Tasks may **not** be submitted as cloud links, such as links to Google Docs, Google Slides, OneDrive, etc., unless specified in the task requirements. All other submissions must be file types that are uploaded and submitted as attachments (e.g., .docx, .pdf, .ppt).*

Part I: Research Question

- A. Describe the purpose of this data analysis by doing the following:
1. Summarize **one** research question that is relevant to a real-world organizational situation captured in the data set you have selected and that you will answer using multiple linear regression in the initial model.
 2. Define the goals of the data analysis.

Note: Ensure that your goals are within the scope of your research question and are represented in the available data.

Part II: Method Justification

- B. Describe multiple linear regression methods by doing the following:
1. Summarize **four** assumptions of a multiple linear regression model.
 2. Describe **two** benefits of using Python or R in support of various phases of the analysis.
 3. Explain why multiple linear regression is an appropriate technique to use for analyzing the research question summarized in part I.

Part III: Data Preparation

- C. Summarize the data preparation process for multiple linear regression analysis by doing the following:
1. Describe your data cleaning goals and the steps used to clean the data to achieve the goals that align with your research question including your annotated code.
 2. Describe the dependent variable and *all* independent variables using summary statistics that are required to answer the research question, including a screenshot of the summary statistics output for each of these variables.
 3. Generate univariate and bivariate visualizations of the distributions of the dependent and independent variables, including the dependent variable in your bivariate visualizations.
 4. Describe your data transformation goals that align with your research question and the steps used to transform the data to achieve the goals, including the annotated code.
 5. Provide the prepared data set as a CSV file.

Part IV: Model Comparison and Analysis

- D. Compare an initial and a reduced linear regression model by doing the following:
1. Construct an initial multiple linear regression model from *all* independent variables that were identified in part C2.
 2. Justify a statistically based feature selection procedure or a model evaluation metric to reduce the initial model in a way that aligns with the research question.
 3. Provide a reduced linear regression model that follows the feature selection or model evaluation process in part D2, including a screenshot of the output for each model.
- E. Analyze the data set using your reduced linear regression model by doing the following:
1. Explain your data analysis process by comparing the initial multiple linear regression model and reduced

linear regression model, including the following element:

- a model evaluation metric
2. Provide the output and *all* calculations of the analysis you performed, including the following elements for your reduced linear regression model:
 - a residual plot
 - the model's residual standard error
 3. Provide an executable error-free copy of the code used to support the implementation of the linear regression models using a Python or R file.

Part V: Data Summary and Implications

- F. Summarize your findings and assumptions by doing the following:
 1. Discuss the results of your data analysis, including the following elements:
 - a regression equation for the reduced model
 - an interpretation of the coefficients of the reduced model
 - the statistical and practical significance of the reduced model
 - the limitations of the data analysis
 2. Recommend a course of action based on your results.

Part VI: Demonstration

- G. Provide a Panopto video recording that includes the presenter and a vocalized demonstration of the functionality of the code used for the analysis of the programming environment, including the following elements:
 - an identification of the version of the programming environment
 - a comparison of the initial multiple linear regression model you used and the reduced linear regression model you used in your analysis
 - an interpretation of the coefficients of the reduced model

Note: The audiovisual recording should feature you visibly presenting the material (i.e., not in voiceover or embedded video) and should simultaneously capture both you and your multimedia presentation.

Note: For instructions on how to access and use Panopto, use the "Panopto How-To Videos" web link provided below. To access Panopto's website, navigate to the web link titled "Panopto Access," and then choose to log in using the "WGU" option. If prompted, log in using your WGU student portal credentials, and then it will forward you to Panopto's website.

To submit your recording, upload it to the Panopto drop box titled "Regression Modeling – NBM3 | D208." Once the recording has been uploaded and processed in Panopto's system, retrieve the URL of the recording from Panopto and copy and paste it into the Links option. Upload the remaining task requirements using the Attachments option.

- H. List the web sources used to acquire data or segments of third-party code to support the application. Ensure the web sources are reliable.
- I. Acknowledge sources, using in-text citations and references, for content that is quoted, paraphrased, or summarized.
- J. Demonstrate professional communication in the content and presentation of your submission.

File Restrictions

File name may contain only letters, numbers, spaces, and these symbols: ! - _ . * ' ()

File size limit: 200 MB

File types allowed: doc, docx, rtf, xls, xlsx, ppt, pptx, odt, pdf, txt, qt, mov, mpg, avi, mp3, wav, mp4, wma, flv, asf, mpeg, wmv, m4v, svg, tif, tiff, jpeg, jpg, gif, png, zip, rar, tar, 7z

RUBRIC

A1:RESEARCH QUESTION

NOT EVIDENT

A summary of 1 research question is not provided.

APPROACHING COMPETENCE

The summary includes 1 research question, but the research question is not relevant to a realistic organizational situation or cannot be addressed using relevant data from the selected data set and multiple linear regression in the initial model.

COMPETENT

The summary includes 1 research question that is relevant to a realistic organizational situation and can be addressed using relevant data from the selected data set and multiple linear regression in the initial model.

A2:GOALS

NOT EVIDENT

The submission does not define the goals of the data analysis.

APPROACHING COMPETENCE

The submission defines the goals of the data analysis, but 1 or more of the goals are not relevant.

COMPETENT

The submission defines the goals of the data analysis, and the goals are relevant.

B1:SUMMARY OF ASSUMPTIONS

NOT EVIDENT

The submission does not summarize *any* assumptions of a multiple linear regression model.

APPROACHING COMPETENCE

The submission summarizes 4 assumptions of a multiple linear regression model, but 1 or more of the assumptions contain inaccuracies. Or the submission only includes 1–3 assumptions of a multiple linear regression model.

COMPETENT

The submission accurately summarizes 4 assumptions of a multiple linear regression model.

B2:TOOL BENEFITS

NOT EVIDENT

The submission does not

APPROACHING COMPETENCE

The submission describes *only* 1

COMPETENT

The submission describes 2

describe *any* benefits of using Python or R in support of various phases of the analysis.

benefit of using Python or R in support of various phases of the analysis. Or the description includes 2 benefits of using the tool, but 1 or both benefits do not logically align with the goal of the analysis.

benefits of using Python or R in support of various phases of the analysis, and the benefits logically align with the goal of the analysis.

B3: APPROPRIATE TECHNIQUE

NOT EVIDENT

The submission does not explain why multiple linear regression is an appropriate technique to analyze a research question.

APPROACHING COMPETENCE

The submission explains why multiple linear regression is an appropriate technique, but the explanation does not relate to the research question from part I. Or the explanation contains inaccuracies.

COMPETENT

The submission accurately explains why multiple linear regression is an appropriate technique to analyze the research question from part I.

C1: DATA CLEANING

NOT EVIDENT

The submission includes neither a description of the data cleaning goals nor the steps used to clean the data to achieve the goals.

APPROACHING COMPETENCE

The submission describes either the data cleaning goals, or the steps used to clean the data to achieve the goals but not *both*. Or the steps are not in alignment with the data cleaning goals or multiple linear regression analysis. Or the annotated code is incomplete or not included.

COMPETENT

The submission describes the data cleaning goals and the steps used to clean the data to achieve the goals. The goals and steps align with each other and with multiple linear regression analysis. The annotated code is complete.

C2: SUMMARY STATISTICS

NOT EVIDENT

The submission does not use summary statistics to describe the dependent variable and *all* independent variables required to answer the selected research question. Or a screenshot of the summary statistics output is not provided for *all* dependent and independent variables.

APPROACHING COMPETENCE

The submission inaccurately uses summary statistics to describe the dependent variable or *all* independent variables. Or the description of variables is incomplete. Or a screenshot of the summary statistics output is provided for some but not *all* dependent and independent variables.

COMPETENT

The submission accurately uses summary statistics to describe the dependent variable and *all* independent variables required to answer the selected research question and provides a screenshot of the summary statistics output for *all* dependent and independent variables.

C3:VISUALIZATIONS

NOT EVIDENT

The submission does not include *both* univariate and bivariate visualizations of the distributions of the dependent and independent variables.

APPROACHING COMPETENCE

The submission generates *both* univariate and bivariate visualizations of the distributions of the dependent and independent variables, but *some* visualizations contain inaccuracies. Or the bivariate visualizations do not include the dependent variable.

COMPETENT

The submission accurately generates *both* univariate and bivariate visualizations of the distributions of the dependent and independent variables. The bivariate visualizations include the dependent variable.

C4:DATA TRANSFORMATION

NOT EVIDENT

The submission includes neither a description of the data transformation goals nor the steps used to transform the data to achieve the goals.

APPROACHING COMPETENCE

The submission describes either the data transformation goals or the steps used to transform the data to achieve the goals but not *both*. Or the steps are not in alignment with the data transformation goals, multiple linear regression analysis, or the research question. Or the annotated code is incomplete or not included.

COMPETENT

The submission describes the data transformation goals and the steps used to transform the data to achieve the goals. The goals and steps align with each other, multiple linear regression, and the research question. The annotated code is complete.

C5:PREPARED DATA SET

NOT EVIDENT

The submission does not provide a data set.

APPROACHING COMPETENCE

The submission provides the prepared data set, but the data set is not fully prepared or is incomplete. Or the submission is not provided as a CSV file.

COMPETENT

The submission provides the fully prepared data set in the form of a CSV file.

D1:INITIAL MODEL

NOT EVIDENT

The submission does not provide an initial multiple linear regression model.

APPROACHING COMPETENCE

The submission provides an initial multiple linear regression model from some, but not *all*,

COMPETENT

The submission provides an accurate initial multiple linear regression model from *all*

independent variables identified in part C2, or the model contains inaccuracies.

independent variables identified in part C2.

D2:JUSTIFICATION OF MODEL REDUCTION

NOT EVIDENT

The submission justifies neither a feature selection procedure nor a model evaluation metric to reduce the initial model.

APPROACHING COMPETENCE

The submission justifies either a feature selection procedure or a model evaluation metric to reduce the initial model, but the justification is not in alignment with the research question. Or the feature selection procedure or model evaluation metric is not statistically based.

COMPETENT

The submission justifies a statistically based feature selection procedure or a model evaluation metric to reduce the initial model. The justification is in alignment with the research question.

D3:REDUCED LINEAR REGRESSION MODEL

NOT EVIDENT

The submission does not provide a reduced linear regression model.

APPROACHING COMPETENCE

The submission provides a reduced linear regression model, but it is not in alignment with the justification from part D2. Or a screenshot is not provided for each model.

COMPETENT

The submission provides a reduced linear regression model that is in alignment with the justification from part D2, and a screenshot is provided for each model.

E1:MODEL COMPARISON

NOT EVIDENT

The submission does not explain the data analysis process by comparing the initial multiple linear regression model and reduced linear regression models.

APPROACHING COMPETENCE

The submission explains the data analysis process by comparing the initial multiple linear regression model and reduced linear regression models but does not include the given element. Or the explanation contains inaccuracies.

COMPETENT

The submission accurately explains the data analysis process by comparing the initial multiple linear regression model and reduced linear regression models, including the given element.

E2:OUTPUT AND CALCULATIONS

NOT EVIDENT

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

APPROACHING COMPETENCE

The submission provides the output and calculations of the analysis performed, but not all of

COMPETENT

The submission provides the accurate output and calculations of the analysis performed,

the output is included, or 1 or more of the calculations are missing or contain inaccuracies. Or the submission does not include *all* the given elements.

including *all* the given elements.

E3:CODE

NOT EVIDENT

The submission does not provide the code used to support the implementation of the linear regression models.

APPROACHING COMPETENCE

The submission provides an executable copy of the code used to support the implementation of the linear regression models, but the code is incomplete or contains inaccuracies. Or the submission is not provided in the form of a Python or R file.

COMPETENT

The submission provides an executable copy of the code used to support the implementation of the linear regression models, and the code is complete and accurate in the form of a Python or R file.

F1:RESULTS

NOT EVIDENT

The submission does not discuss the results of the data analysis, or the discussion is not in alignment with the research question and the data analysis.

APPROACHING COMPETENCE

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

COMPETENT

The submission accurately discusses the results of the data analysis, and the discussion addresses *all* the given elements and is in alignment with the research question and the data analysis.

F2:RECOMMENDATIONS

NOT EVIDENT

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

APPROACHING COMPETENCE

The submission recommends a course of action based on results, but the recommendation is not appropriate based on the research question or the results of the data.

COMPETENT

The submission recommends an appropriate course of action based on the results as they relate to the research question.

G:PANOPTO DEMONSTRATION

NOT EVIDENT

The submission does not provide a Panopto video recording.

APPROACHING COMPETENCE

A Panopto video recording is provided that includes a

COMPETENT

A Panopto video recording is provided that includes a

demonstration of the functionality of the code used for the analysis, but *all* the listed elements are not discussed. Or the video does not capture both the presenter and the vocalized presentation of the functioning code in a Panopto video recording.

demonstration of the functionality of the code used for the analysis. The demonstration includes *all* the listed elements. For the duration of the presentation, the video captures both the presenter and the vocalized presentation of the functioning code in a Panopto video recording.

H:SOURCES OF THIRD-PARTY CODE

NOT EVIDENT

The submission does not list *any* web sources.

APPROACHING COMPETENCE

The submission lists only *some* of the web sources used to acquire data or segments of third-party code. Or the web sources are not reliable.

COMPETENT

The submission lists *all* web sources used to acquire data or segments of third-party code, and the web sources are reliable.

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.

APPROACHING COMPETENCE

The submission includes in-text citations for sources that are quoted, paraphrased, or summarized and a reference list; however, the citations and/or reference list is incomplete or inaccurate.

COMPETENT

The submission includes in-text citations for sources that are properly quoted, paraphrased, or summarized and a reference list that accurately identifies the author, date, title, and source location as available.

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.

APPROACHING COMPETENCE

Content is poorly organized, is difficult to follow, or contains errors in mechanics, usage, or grammar that cause confusion. Terminology is misused or ineffective.

COMPETENT

Content reflects attention to detail, is organized, and focuses on the main ideas as prescribed in the task or chosen by the candidate. Terminology is pertinent, is used correctly, and effectively conveys the intended meaning. Mechanics, usage, and grammar promote accurate interpretation and understanding.

WEB LINKS

[Data Sets and Associated Data Dictionaries](#)

If you have trouble with the link, copy and paste the link directly into your web browser.

[Panopto Access](#)

Sign in using the "WGU" option. If prompted, log in with your WGU student portal credentials, which should forward you to Panopto's website. If you have any problems accessing Panopto, please contact Assessment Services at assessmentservices@wgu.edu. It may take up to two business days to receive your WGU Panopto recording permissions once you have begun the course.

[Panopto FAQs](#)

[Panopto How-To Videos](#)

