**Performance Assessment: Exploratory Data Analysis (OEM2)**

Bader Ale

Department of Information Technology, Western Governors University

D207 Exploratory Data Analysis

Professor William Sewell

Table of Contents

[Part A: 3](#_Toc137201946)

[Research Question 3](#_Toc137201947)

[Question Importance 3](#_Toc137201948)

[Data Identification 3](#_Toc137201949)

[Part B: 3](#_Toc137201950)

[Technique Used 3](#_Toc137201951)

[Part C: 4](#_Toc137201952)

# Part A:

## Research Question

For this performance assessment, my research question is: Is there a significant difference in readmission based on the number of office visits?

## Question Importance

This research question is particularly valuable because it can help the hospital administration, investors and stakeholders determine if the amount of visits the physician makes to the patient affects the readmission rate. In this way, readmission rates can be decreased, potentially saving the facility money and resources.

## Data Identification

In order to answer the research question, the right variables have to be chosen. In this case *ReAdmis* and *Doc\_visits.* Per the Data Dictionary, *ReAdmis* is defined as whether or not the patient was readmitted back to the hospital within thirty days following their original discharge while *Doc\_visits*  is defined as the number of times the primary physician visited the patient in the hospital during their admittance.

# Part B:

## Technique Used

The technique chosen for this analysis was the t-test and more specifically a two-sample t-test. This technique was chosen because the research question concentrated on two variables, *ReAdmis* and *Doc\_visits; ReAdmis* is a categorical variable with two levels (Yes and No) while *Doc\_visits* is a discrete quantitative variable. Two groups will be further created – *ReAdmis\_yes* and *ReAdmis\_no*. The independent samples (two-sample) t-test will calculate the means of the two groups, *Readmis\_yes* and *Readmis\_no* and using the formula will generate both a t-statistic and a p-value.

A screenshot of a computer program

Description automatically generated with low confidence

Figure 1: Identifying t-test Variables.

A screenshot of the analysis output is shown below:

A picture containing text, font, line, white

Description automatically generated

Figure 2: t-test Output

# Part C: