Laycee Glass

OEM2 Task 1

Exploratory Data Analysis

7/10/24

Western Governors University

## OEM2 Performance Assessment

## Student Information

Student Name: Laycee Glass

ID Number: 001896026

Date: 7/10/24

## A1. Question

The question I chose to answer for the medical dataset is the following: Is there a significant difference in Vitamin D levels between patients who were readmitted and those who were not readmitted?

## A2. Analysis Benefit

Stakeholders of the organization could benefit from this analysis because of improved patient outcomes and satisfaction. If there is a significant difference found in Vitamin D levels, healthcare providers could implement targeted interventions to monitor and manage these levels. Educating patients about how to maintain proper Vitamin D levels and how it relates to their overall health can lead to higher patient engagement. These patients, in turn, are more likely to experience better health outcomes, which enhances their perception of quality of care.

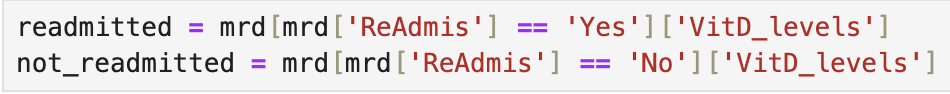
Another proposed benefit would be cost reduction. Providers would be able to reduce the number of readmissions based on Vitamin D levels, leading to significant cost savings. Readmissions are costly and require more resources to be used.

### A3. Relevant Data

* VitD\_levels – the levels measured in ng/mL
* ReAdmis – if the patient was readmitted within a month of release

### B1. T-test Code

To run the t-test, I used the following code after importing the csv file (as seen in .ipynb document):



A close up of a sign

Description automatically generatedA close-up of a white background

Description automatically generatedA screenshot of a computer code

Description automatically generated

### B2. Results

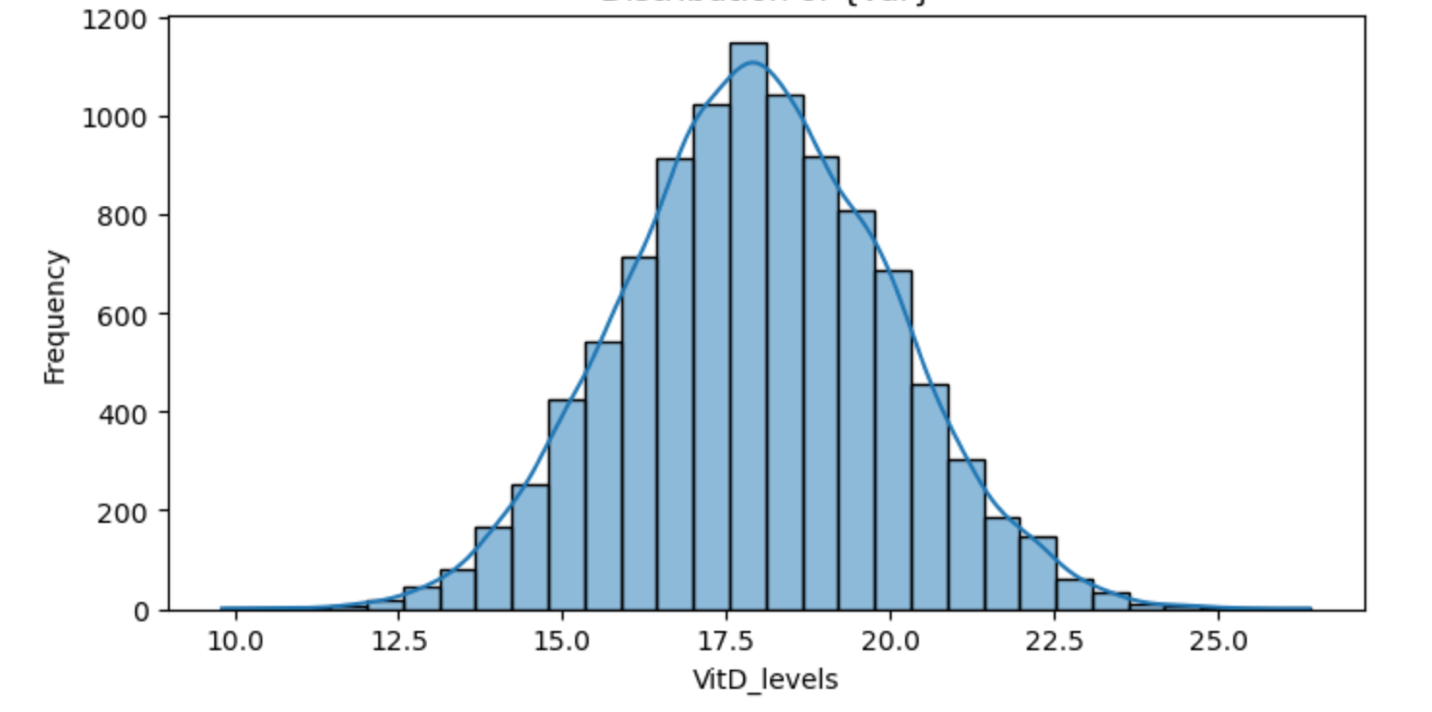
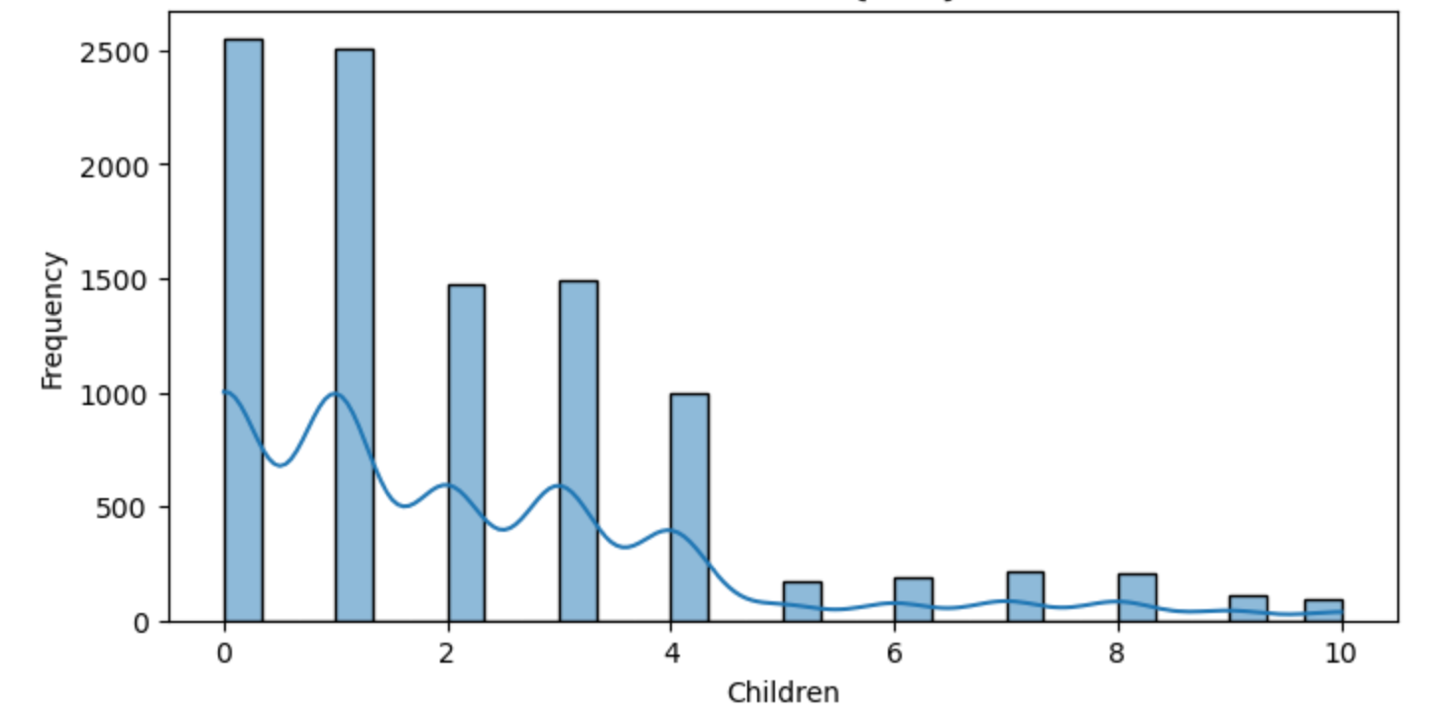
The t-test produced a t-statistic of 0.40822135318600616 and a p-value of 0.683119914739662, with 9998 degrees of freedom. Based on the code I wrote, it also produced the sentence “There is no significant difference in Vitamin D levels."

### B3. Justification

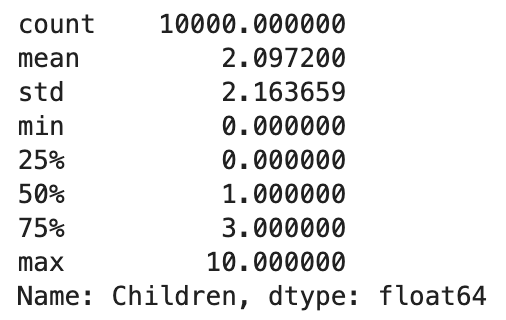
A t-test was ideal for this analysis because it is specifically designed to compare the means of two groups; in our case, the groups were Vitamin D levels and readmission status. It also worked due to the type of data present. It was well-suited for one continuous dependent variable (Vitamin D levels) and one categorical independent variable (readmission status).

### C1. Univariate Statistics

I identified the distribution of two continuous variables in the dataset, which were VitD\_levels and Children. This is how the distributions looked:



Based on these graphs, the distribution of Vitamin D levels was normal, and the distribution of the number of children had a positive skew. Using the .describe( ) command, the following summary statistics were generated:

A number of numbers and letters

Description automatically generated with medium confidence

A graph of a number of people

Description automatically generatedA graph with blue squares

Description automatically generatedI also identified the frequency counts of two categorical variables, ReAdmis and Marital. Their counts are shown by the following bar charts:

These bar charts show that there was a greater distribution of no samples for the ReAdmis variable, and the greatest distribution of Marital was widowed but all seemed fairly equal. The specific statistics of each variable are shown below:

A black text on a white background

Description automatically generatedA close-up of a number

Description automatically generated

### D1. Bivariate Statistics

For this section, I compared two continuous variables, VitD\_levels and Children, against each other. For these two variables, I was able to calculate correlation and plot a scatter plot with a regression line. It is shown here:

A graph of blue dots

Description automatically generated The correlation here was calculated to be 0.009487013110847639. This indicated a poor correlation, which we can also see from the scatter plot. For example, for patients that had 0 children, Vitamin D levels ranged from as low as 12.0 to 25.0. This trend continued across the board, even to the opposite end where patients who had 10 children had Vitamin D levels between 13.0 and 23.0. There was no steady increase or decrease in levels for those who had children versus those who did not.

For the categorical variables of ReAdmis and Marital, I created a contingency table, performed a chi-square test, and plotted a heatmap to compare. The contingency table and statistics are shown here:

A white paper with numbers and text

Description automatically generated

Per the calculated p value, there was no significant association between readmission status and marital status. For visualization purposes, I created the following heatmap for these two variables:

A blue and yellow chart

Description automatically generated

As the map shows, there is no clear association. Looking at patients who were divorced, 1283 were not admitted and 678 were admitted. Each category of marital status had similar values for being readmitted and not being readmitted. No specific marital status was more likely to be readmitted.

### E1. Results

For the initial t-test I ran, the null hypothesis of the question was there is no significant difference in Vitamin D levels between readmitted and non-readmitted patients. Based on the calculated p-value being greater than alpha (0.05), I would fail to reject this null hypothesis.

### E2. Limitations

Conducting a t-test was appropriate in this scenario, but there are limitations to this test as well. For my test specifically, it assumed the variances of the Vitamin D levels were equal for both groups (readmitted, not readmitted). If the variances were not equal, it could lead to inaccurate results.

Another limitation to consider for a t-test is that the data should be a simple random sample, meaning it is representative of a total population. Without knowing how/where/when the data was collected means this dataset could possibly not fit this description, which could again lead to inaccurate results or introduction of bias (Maverick, n.d.).

### E3. Recommended Action

Because Vitamin D levels did not seem to have any correlation with readmission status, my next step would be to investigate other variables such as age, income, doctor visits, etc. to identify possible influences on readmission status.

### G. Code Sources

I did not use any outside sources to complete the code for this assessment.

### H. Content Sources

Maverick, J. B. (n.d.). *What assumptions are made when conducting a T-test?*. Investopedia. https://www.investopedia.com/ask/answers/073115/what-assumptions-are-made-when-conducting-ttest.asp