

**D207**

Brandon Gillins

Western Governors University

D207- Task 1

June 02, 2024

## Table of Contents

<i>A1. Relevant Question:</i>	3
<i>A2. Stakeholder Benefits:</i>	3
<i>A3. Relevant Dataset Description:</i>	3
<i>B1. Chi-Square Test:</i>	4
<i>B2. Results from Chi-Square Test:</i>	4
<i>B3. Technique Justification:</i>	5
<i>C. Univariate Statistics:</i>	5
Continuous Variables:	5
Categorical Variables:	6
Findings:	7
<i>D. Bivariate Statistics:</i>	8
Findings:	8
<i>E1. Results of the Hypothesis Test:</i>	10
<i>E2. Discussion of the Limitations of the Data Analysis:</i>	10
<i>E3. Recommended Course of Action Based on the Results:</i>	10
<i>F. Video Recording</i>	10
<i>G. Coding Sources</i>	11
<i>H. Sources</i>	11

## A1. Relevant Question:

**Question:** What are the key health indicators that contribute to patient readmissions through emergency services?

## A2. Stakeholder Benefits:

Emergency services are exceedingly costly for both patients and health organizations. Providers that operate under a value-based care system track member health information to identify conditions such as diabetes and compliance with critical health markers like A1c or CKD (Chronic Kidney Disease). By focusing our data analysis on patients who have utilized emergency services, we aim to uncover patterns that can inform the health team, enabling primary care providers (PCPs) and emergency care teams to collaborate effectively. This collaboration is pivotal in reducing the number of emergency admissions through proactive, preventative care strategies. The following stakeholders stand to benefit from such data analysis:

- **Hospital Administration:** By comprehending the primary causes of readmissions, administrators can develop targeted interventions to reduce readmission rates. This proactive approach helps avoid penalties from the Complaints to Medicare (CTM) and improves star ratings, which in turn positively impacts enrollment numbers.
- **Healthcare Providers:** Doctors and nurses can concentrate on specific risk factors identified during patient treatment and post-discharge care. This focused approach enhances patient outcomes and overall satisfaction by ensuring more personalized and effective care protocols.
- **Patients:** Improved care protocols, informed by detailed data analysis, reduce the likelihood of patients being readmitted. This leads to better health outcomes and lower medical costs for patients, contributing to their overall well-being and satisfaction with the healthcare system.
- **Insurance Companies:** Insurers can leverage this data to design better coverage plans that encourage preventive care, thus reducing overall healthcare costs. By directing up to 80% of the plan's budget towards patient care, insurers can enhance the value returned to members. In areas with relatively healthy populations, insurers may provide additional benefits, such as gym memberships or other health services not typically covered by insurance, thereby fostering a healthier member base and reducing long-term costs.

## A3. Relevant Dataset Description:

The data relevant to answering the question about key health indicators contributing to patient readmissions through emergency services includes various health markers. Here is a concise overview of the data selection and cleaning process:

## Data Cleaning and Preparation:

### Health Markers (Categorical Data):

- **Columns:** ReAdmis, HighBlood, Stroke, Diabetes, Overweight, Arthritis, Hyperlipidemia, BackPain, Anxiety, Allergic\_rhinitis, Reflux\_esophagitis, Asthma.
- **Description:** These binary indicators (0 or 1) represent the presence or absence of specific health conditions. Null values were filled with 0, assuming that the absence of data implies the absence of the condition.

### 2. Initial Admin (Categorical Data):

- **Description:** How the patient was initially admitted into the hospital (emergency admission, elective admission, observation). We are exploring the question of how to reduce the number of emergency admissions. The data was filtered by carving out only those that were admitted through emergency services. Which left a remaining 5060 rows of data.
- This approach will reduce the degrees of freedom needed for the Chi Square test and keep our data population specific for the question we are trying to answer.

## B1. Chi-Square Test:

In this analysis, we performed chi-square tests to investigate the association between various health indicators and patient readmissions through emergency services. For each health indicator, we formulated the following hypotheses:

- **Null Hypothesis (H0):** There is no significant association between the health indicator and patient readmissions.
- **Alternative Hypothesis (H1):** There is a significant association between the health indicator and patient readmissions.

The chi-square test results include the chi-square statistic, p-value, and degrees of freedom. A common threshold to determine statistical significance is a p-value of 0.05. If the p-value is less than 0.05, we reject the null hypothesis, indicating a significant association. If the p-value is greater than or equal to 0.05, we fail to reject the null hypothesis, indicating no significant association. We are trying to determine how much a surprise factor we see from our given dataset and the smaller the value the greater that surprise. (Kozyrkov, 2022). (Please see attached code in .ipynb file)

## B2. Results from Chi-Square Test:

Health Indicator	P-value	Result
High Blood Pressure	0.285	Failed to reject the null hypothesis
Stroke	0.695	Failed to reject the null hypothesis
Diabetes	0.653	Failed to reject the null hypothesis
Overweight	0.301	Failed to reject the null hypothesis
Arthritis	0.973	Failed to reject the null hypothesis
Hyperlipidemia	0.97	Failed to reject the null hypothesis
BackPain	0.608	Failed to reject the null hypothesis
Anxiety	0.472	Failed to reject the null hypothesis
Allergic_rhinitis	0.655	Failed to reject the null hypothesis
Reflux_esophagitis	0.399	Failed to reject the null hypothesis
Asthma	0.923	Failed to reject the null hypothesis

### B3. Technique Justification:

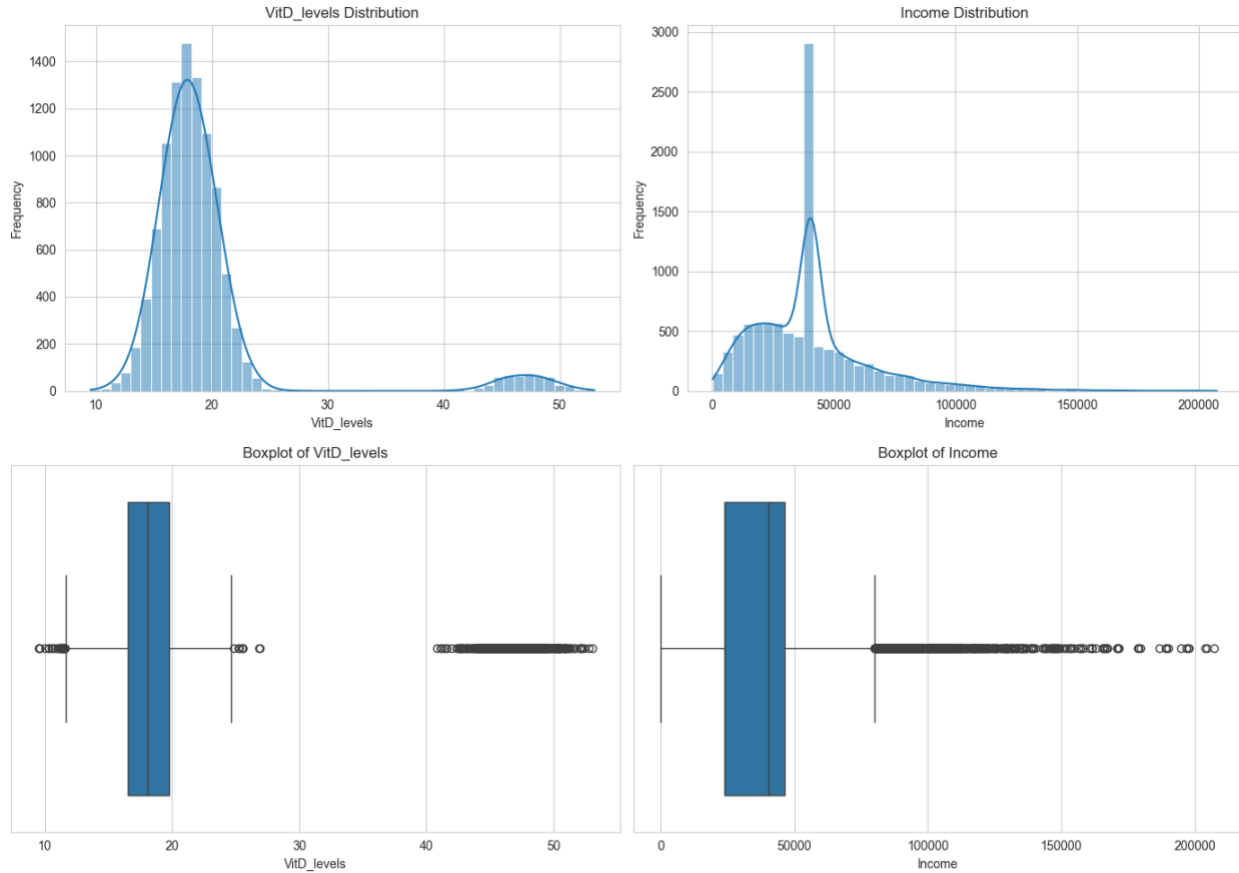
The chi-square test was chosen because both the ReAdmis column and the health indicator columns are categorical data, specifically binary (true/false). The chi-square test is ideal for determining if there is a significant association between two categorical variables. It compares observed frequencies of occurrences to expected frequencies under the assumption of independence. This method is well-suited for the data, as it allows for the assessment of whether the presence of certain health conditions is associated with patient readmissions through emergency services.

### C. Univariate Statistics:

#### Continuous Variables:

**VitD\_levels:** Distribution of the number of days the patient stayed in the hospital during the initial visit.

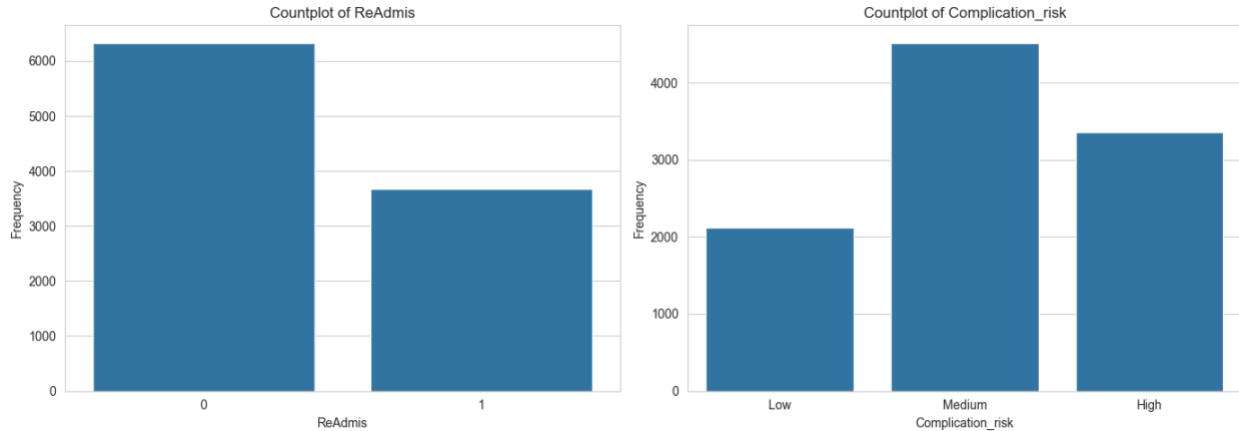
**Income:** Distribution of the report income of the patients.



Statistics	VitD_levels	Income
count	10000	10000
mean	19.41	40484.64
std	6.72	24883.79
min	9.52	154
25%	16.51	23955.5
50%	18.08	40481
75%	19.79	46466
max	53.02	207249

### Categorical Variables:

1. **ReAdmis:** Distribution of readmission status (yes, no).
2. **Complication\_risk:** Level of complication risk the patient was identified from primary patient assessment.



### ReAdmis:

Values	Counts	% of Total
0	6331	63.31%
1	3669	36.69%

### Complication\_risk:

Values	Counts	% of Total
Low	2125	21.25%
Medium	4517	45.17%
High	3358	33.58%

### Findings:

**Income Distribution:** The income distribution exhibits a noticeable spike around \$40,000 due to filling missing values with the mean. Additionally, the data is skewed towards lower to medium income groups, with most individuals earning below \$40,000. This suggests a varied population, prompting further analysis of the clinic locations and their demographics. Understanding the selection process of the initial dataset, which spans multiple states and cities, could provide valuable insights into the population being served and observed.

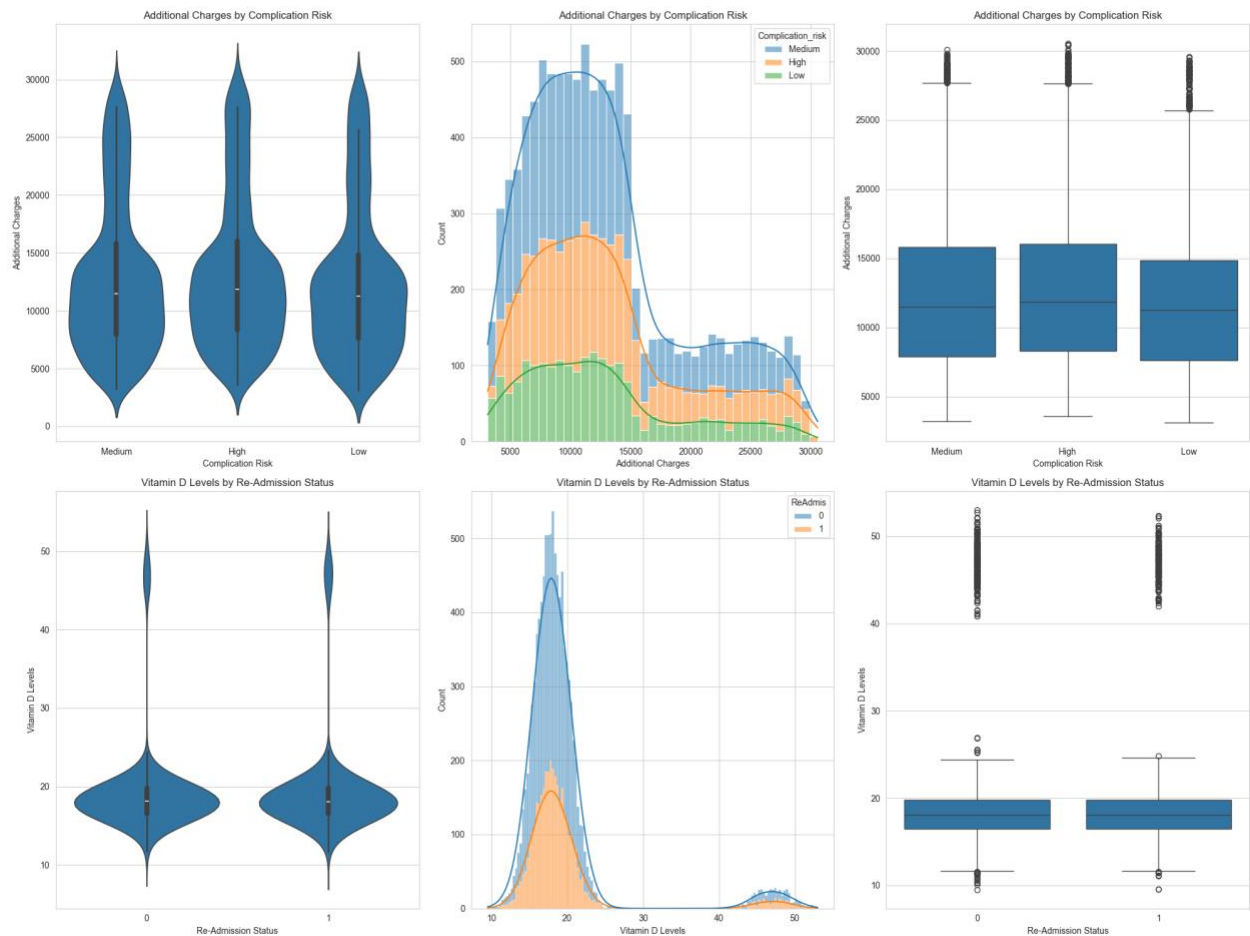
**Vitamin D Distribution:** The Vitamin D levels exhibit a bimodal distribution, with a primary peak around 20 and a smaller secondary peak around 45. This secondary peak warrants further investigation to understand potential errors or anomalies in the data. Consulting an expert could help determine the cause and implications of this spike, especially in relation to readmission rates, as Vitamin D levels are linked to immune health.

**Readmission Distribution:** The analysis of readmission status reveals that a significant portion of patients were readmitted, with a readmission rate of 38%, which is notably high for any healthcare organization.

**Complication Risk Distribution:** The complication risk data shows a majority of patients falling into the medium-risk category. However, a substantial number are classified as

high risk. Investigating the readmission rates within each complication risk category could provide insights into the accuracy and predictive power of the complication risk metric, aiding in focusing efforts on specific patient populations. The situation for high readmissions is likely complex, involving multiple factors. A detailed examination of the complication risk data could help identify new approaches to manage and reduce readmission rates effectively.

## D. Bivariate Statistics:



## Findings:

**Additional Charges by Complication Risk:** The violin plot depicts the distribution of additional charges across different complication risk categories: Low, Medium, and High. I found that the violin plot minimized the mean difference, so additional graphs were designed to understand the shape and mean difference. The standard deviation between all groups are similar.



Complication Risk	count	mean	std	min	25%	50%	75%	max
Low	2125.0	12490.777581	6489.168777	3125.70	7631.94	11270.390	14864.7300	29611.39
Medium	4517.0	12866.641913	6539.973676	3213.08	7918.38	11485.000	15828.1100	30087.65
High	3358.0	13306.659321	6561.141954	3589.08	8334.01	11865.155	16049.5875	30566.07

To further

understand the difference an ANOVA test would be appropriate to understand the difference of means from each category. The ANOVA test results reveal a significant difference in the means of additional charges among these categories. Specifically, the ANOVA test yields an F-statistic of 10.58 and a p-value of less than 0.0001, indicating that at least one group mean is significantly different from the others.

To further investigate which specific groups differ from each other, a post-hoc Tukey HSD test was conducted. The results show that additional charges for the High complication risk group are significantly higher than those for the Low and Medium risk groups. However, there is no significant difference between the Low and Medium risk groups. This suggests that patients categorized under High complication risk incur notably higher additional charges compared to those in Low and Medium risk categories.

**Vitamin D Levels by Re-Admission Status:** The distribution of Vitamin D levels is quite balanced between readmitted and non-readmitted patients, with both groups showing similar variabilities. This finding suggests that there is likely no significant relationship between Vitamin D levels and readmission rates within this population. The means of both values are very similar, and both show similar distribution and shape of values. To verify the means are expected we can perform an ANOVA test.

Re-Admission	count	mean	std	min	25%	50%	75%	max
No	6331.0	19.326818	6.531744	9.519012	16.513829	18.085481	19.789091	53.019124
Yes	3669.0	19.560825	7.039983	9.553535	16.511677	18.064019	19.793537	52.370764

The ANOVA test yields an F-statistic of 2.81 and a p-value of less than 0.093. This means we would not reject the Null hypothesis stating there is not enough of a surprise factor between members that were re admitted and those that are not by comparing there means.

This outcome is somewhat discouraging, as it indicates that Vitamin D levels might not be a useful predictor for readmission risk. To further understand this relationship, it would be beneficial to investigate how the population was selected and consider the timing of the Vitamin D tests relative to admission times. Examining these factors might reveal a better correlation and help refine the analysis.

## E1. Results of the Hypothesis Test:

The chi-square tests performed indicate that there is no significant association between the analyzed health indicators and patient readmissions through emergency services. This conclusion is drawn from the fact that all p-values exceeded the threshold of 0.05. Therefore, we fail to reject the null hypothesis, implying there is no significant association between the health indicators and patient readmissions. In mathematical terms, it's often easier to prove that something doesn't exist than to prove it does. Thus, by failing to reject the null hypothesis, we are asserting that there is no significant association.

## E2. Discussion of the Limitations of the Data Analysis:

The size of the dataset could influence the results, and larger datasets might yield different findings. However, based on my experience in the healthcare system, healthcare issues are rarely caused by a single factor. They tend to be complex, involving multiple interrelated factors. This complexity could mean that the health indicators analyzed in this study are not the primary factors influencing patient readmissions, or that their effects are moderated by other variables not included in the analysis.

## E3. Recommended Course of Action Based on the Results:

Given the results, it would be cautious to explore additional variables that could impact patient readmissions. Specifically, investigating the compounding effects of related diseases, such as heart disease and diabetes or heart disease and stroke, could provide more insights. Understanding how these diseases interact and contribute to readmissions might help identify more effective interventions. Further research is needed to identify these factors and develop strategies aimed at reducing patient readmissions, ultimately improving patient outcomes, and reducing healthcare costs.

## F. Video Recording.

### Recording Link

<https://wgu.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=b294c1fc-b5c9-4f4a-978e-b183015eb0a5>

## G. Coding Sources

None.

## H. Sources

Kozyrkov, C. (2022, April 4). Retrieved from YouTube:  
<https://www.youtube.com/watch?v=gjF4RKJ-m6s&t=1s>