Patient Readmission Analysis (Chi-Square Test)

Ashley Traore

Question:

Is there a statistically significant difference between the service the patient receives and the readmission status?

Benefit:

An organization can benefit from this test by using the results to make business decisions. By understanding which factors have an impact on patient readmission, the company can properly allocate resources more effectively. They can also use the information from this analysis to evaluate if additional training to employees is necessary to decrease the readmission rate of certain patients.

Relevant data:

The relevant data for this test are the Services column, which contains the patient's primary service that was received while in the hospital, and the ReAdmis column, which contains a yes/no value for whether the patient was readmitted within a month of release.

Justification for choosing this analysis technique:

I chose to use the Ch-square test for this analysis because it allows the comparison between two categorical variables. Therefore, this is the best test to perform when determining if there is a statistically significant difference between services and readmission status.

Identify the Distribution of Categorical and Continuous Variables using Univariate Statistics:

Continuous Variables: Total Charge and Initial Days

Categorical Variables: Services and Back Pain

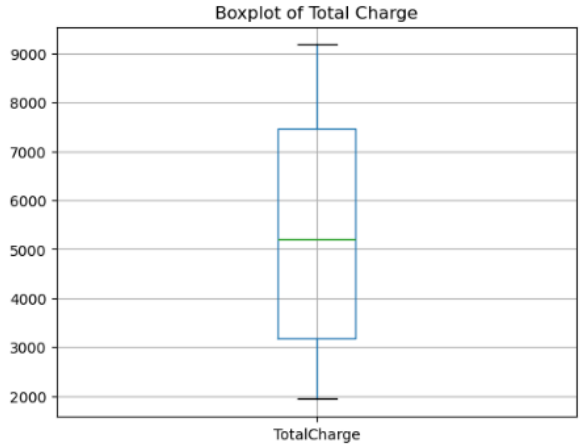
Visual Findings of Categorical and Continuous Variables using Univariate Statistics:

Continuous Variables:

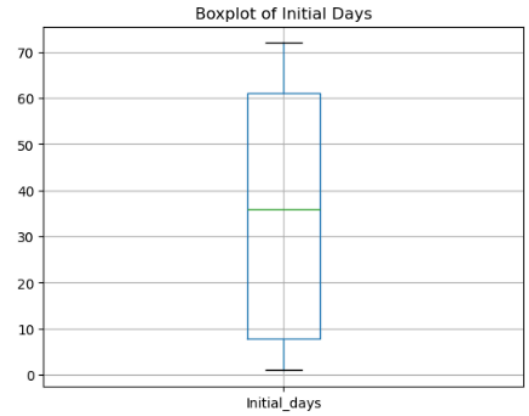
Univariate Statistics

Creating a boxplot to show the spread of 2 continuous variables using univariate statistics.

```
[22]: #Creating a boxplot to show the spread of TotalCharge
medicalData.boxplot(column=['TotalCharge'])
plt.title('Boxplot of Total Charge')
plt.show()
```



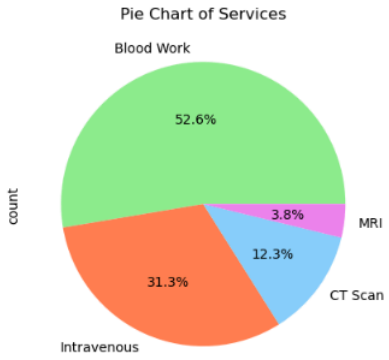

```
[24]: #Creating a boxplot to show the spread of Initial_days
medicalData.boxplot(column=['Initial_days'])
plt.title('Boxplot of Initial Days')
plt.show()
```



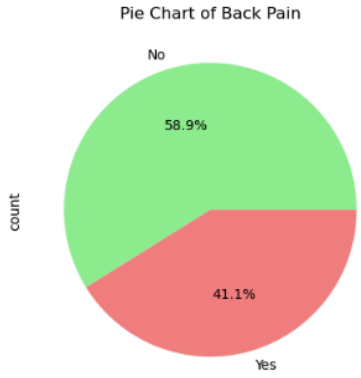
Categorical Variables:

Creating a pie chart to show the spread of 2 categorical variables using univariate statistics.

```
[27]: #Creating a pie chart to show the spread of Services
medicalData['Services'].value_counts().plot(kind='pie', autopct='%1.1f%%', colors=['#90ee90','#ff7f50', '#87cefa', '#ee82ee'])
plt.title('Pie Chart of Services')
plt.show()
```




```
[19]: #Creating a pie chart to show the spread of BackPain
medicalData['BackPain'].value_counts().plot(kind='pie', autopct='%1.1f%%', colors=['#90ee90','#f08080'])
plt.title('Pie Chart of Back Pain')
plt.show()
```



Identify the Distribution of Categorical and Continuous Variables using Bivariate Statistics:

Continuous Variables: Income and TotalCharge, Income and Additional_charges

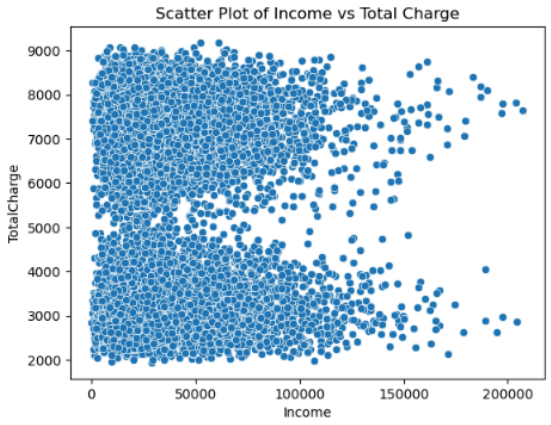
Categorical Variables: Anxiety and Overweight, BackPain and Overweight

Visual Findings of Categorical and Continuous Variables using Bivariate Statistics:

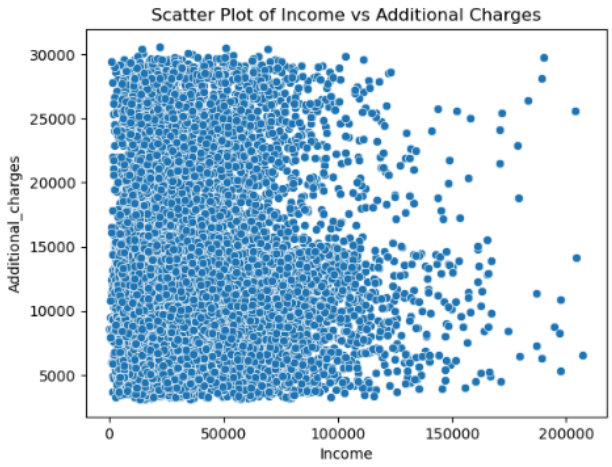
Continuous Variables:

Bivariate Analysis

[56]:
```
# Shows the distribution of Income vs TotalCharge using bivariate statistics
sns.scatterplot(x='Income', y='TotalCharge', data=medicalData)
plt.title('Scatter Plot of Income vs Total Charge')
plt.show()
```

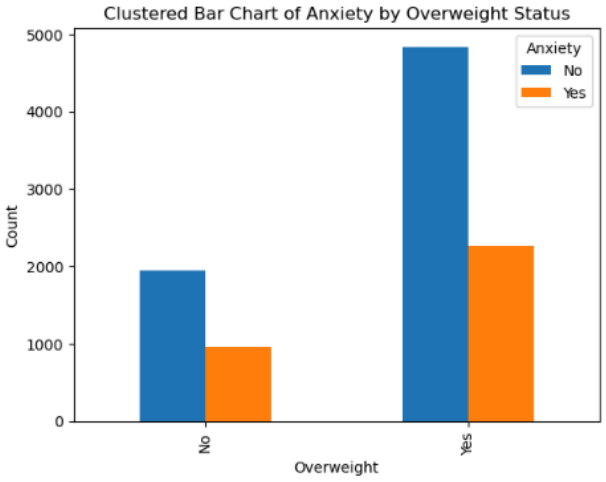


[54]:
```
# Shows the distribution of Income vs Additional_charges using bivariate statistics
sns.scatterplot(x='Income', y='Additional_charges', data=medicalData)
plt.title('Scatter Plot of Income vs Additional Charges')
plt.show()
```

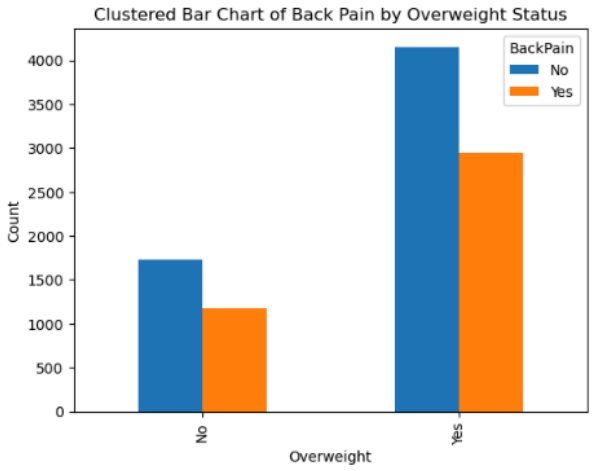


Categorical Variables:

[62]:
```
# Cluster bar chart to show the distribution of Anxiety to Overweight status using bivariate statistics
overweight_anxiety = pd.crosstab(medicalData['Overweight'], medicalData['Anxiety'])
ax = overweight_anxiety.plot(kind='bar')
ax.set_title('Clustered Bar Chart of Anxiety by Overweight Status')
ax.set_xlabel('Overweight')
ax.set_ylabel('Count')
plt.show()
```



[64]:
```
# Cluster bar chart to show the distribution of Back pain to Overweight status using bivariate statistics
overweight_anxiety = pd.crosstab(medicalData['Overweight'], medicalData['BackPain'])
ax = overweight_anxiety.plot(kind='bar')
ax.set_title('Clustered Bar Chart of Back Pain by Overweight Status')
ax.set_xlabel('Overweight')
ax.set_ylabel('Count')
plt.show()
```



Results of Hypothesis Test:

The results of the Ch-square test show a p-value of 0.03. Since this value is less than the common significance level of 0.05, there is a statistically significant difference between the type of service the patient receives and their readmission status. In other words, the difference in readmission rates is likely influenced by the type of service provided during the initial admission. The Chi-square value of 8.89 with 3 degrees of freedom is greater than the critical value of 7.81. Therefore, it supports this finding by indicating a noticeable discrepancy between observed and expected frequencies. (Bobbitt, 2020)

Limitations of the Analysis

Chi-square analysis has several limitations. For instance, it is sensitive to sample size. A small sample size can result in unreliable results. While Chi-square tests can identify if there's a relationship between variables, they cannot determine if one variable causes change in another. Additionally, these tests are restricted to categorical data and cannot be used with continuous or numerical variables. (*Advantages and disadvantages of Chi-Square Test, 2025)*

Recommended Course of Action:

Based on the test results, it is recommended that the company evaluate both the services they provide and the personnel delivering them. By identifying services with a higher readmission rate, the company can determine whether additional training is necessary or if more resources should be allocated to that area.

H. Sources:

Bobbitt, Z. (2020, July 14). *How to find the chi-square critical value in python*. Statology. https://www.statology.org/chi-square-critical-value-python/

GeeksforGeeks. (2025, January 2). *Python - Pearson's Chi-Square Test*. https://www.geeksforgeeks.org/python-pearsons-chi-square-test/

Advantages and disadvantages of Chi Square Test. AspiringYouths. (n.d.). https://aspiringyouths.com/advantages-disadvantages/chi-square-test/