D207 Performance Assessment

Adrienne Stafford

01/07/24

A1. I chose the medical data set to analyze for the D207 Performance assessment. Readmission of patients is a top concern for hospitals in any area. Understanding the correlation between the patients readmitted to the hospital and the underlying causes that have brought them back could help doctors better assess patients during their initial visit in hopes that they will not need to return. During this analysis, I will answer if there is a correlation between readmission, high blood pressure, reliability, and respectful response.

A2. Stakeholders can use this information to improve patient care during initial intake. Understanding what brings patients back to the hospital will give care doctors the knowledge to order tests when patients present certain symptoms. This may also lead to forming a general understanding that with certain symptoms, hospital staff can expect to see the patient return to the hospital just due to the nature of the issue.

A3. Variables

|  |  |
| --- | --- |
| Data Used in Analysis | |
| ReAdmis | yes, no |
| Highblood | yes, no |
| Item4 (Reliability) | 1 to 8 (1 = most important,  8 = least important) |
| Item8 (Evidence of active listening from doctor) | 1 to 8 (1 = most important,  8 = least important) |

B1. Code used to run the analysis:

# Downloading packages:

# Standard data science imports

import numpy as np

import pandas as pd

from pandas import DataFrame

# Visualization libraries

import seaborn as sns

import matplotlib.pyplot as plt

%matplotlib inline

# Statistics packages

import pylab

import statsmodels.api as sm

import statistics

from scipy import stats

# Import chisquare

from scipy.stats import chisquare

from scipy.stats import chi2\_contingency

# File import

df=pd.read\_csv('medical\_clean.csv', index\_col='CaseOrder')

#Reviewing Variables

print(df)

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

#Rename Variables

df.rename (columns = {

'Item1':'TimelyResponse',

'Item2':'TimelyFixes',

'Item3':'TimelyReplacements',

'Item4':'Reliability',

'Item5':'Options',

'Item6':'RespectfulResponse',

'Item7':'Courteous',

'Item8':'Listening'},

inplace=True)

print(df)

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

#Contigency Table

contigency = pd.crosstab(df['ReAdmis'], df['RespectfulResponse'],

normalize='index')

contingency

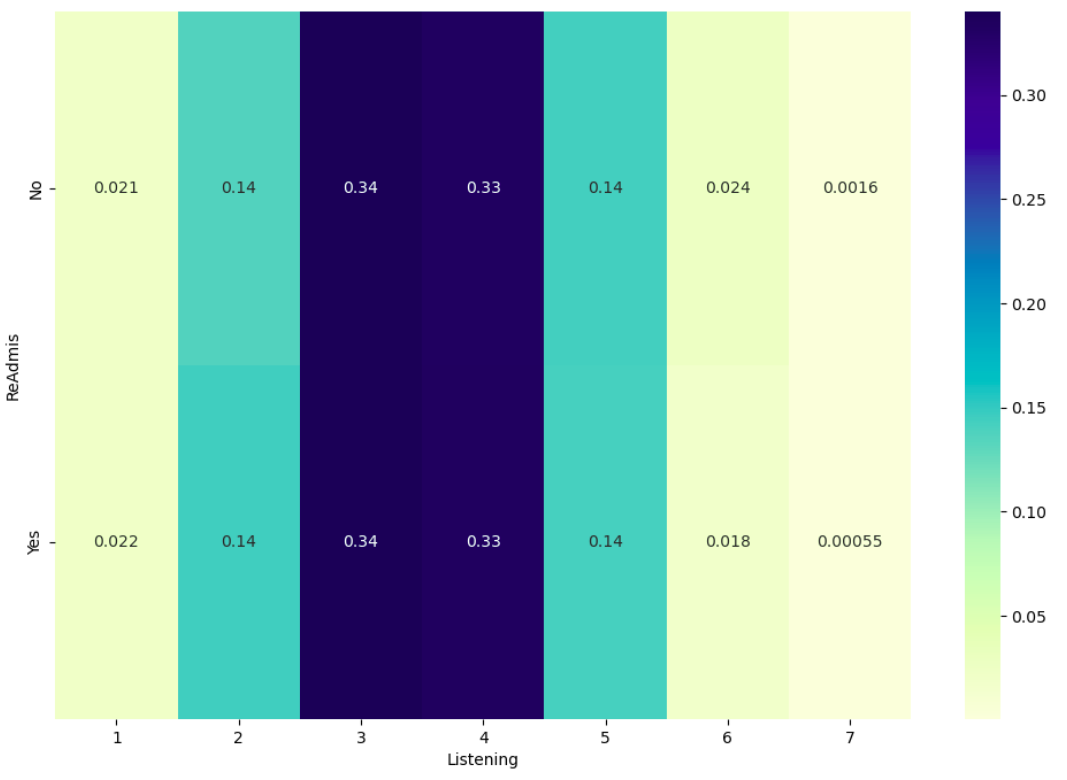
c,p, dof, expected = chi2\_contingency(contigency)

print('p-value = ' + str(p))



plt.figure(figsize=(12,8))

sns.heatmap(contigency, annot=True, cmap="YlGnBu")



df\_stats = df.describe()

print(df\_stats)

# Import label encoder

from sklearn import preprocessing

# label\_encoder object

label\_encoder = preprocessing.LabelEncoder()

# Encode labels

df['ReAdmis'].unique()

df\_num ={"ReAdmis\_numeric":{"No":0, "Yes": 1, "nan":np.NAN}}

df['ReAdmis\_numeric'] = df['ReAdmis']

df.replace(df\_num, inplace=True)

plt.hist(df['ReAdmis\_numeric'])

plt.show()

df[['Listening', 'Reliability','TimelyFixes','Courteous']].hist()

plt.savefig('df\_pyplot.jpg')

plt.tight\_layout()

df['HighBlood'].unique()

df['HighBlood\_numeric'] = df['HighBlood']

df.replace(df\_num, inplace=True)

plt.hist(df['HighBlood\_numeric'])

plt.show()

df.boxplot(['ReAdmis\_numeric'])

plt.savefig('df\_boxplots.jpg')

df.boxplot(['HighBlood\_numeric'])

plt.savefig('df\_boxplots.jpg')

df.boxplot(['Reliability'])

plt.savefig('df\_boxplots.jpg')

bivariate = df[['Reliability', 'ReAdmis\_numeric','RespectfulResponse','HighBlood\_numeric']]

sns.heatmap(bivariate.corr(), annot=True)

plt.show()

plt.scatter(df['ReAdmis\_numeric'], df['RespectfulResponse'])

plt.show()

sns.lmplot(x = 'RespectfulResponse', y = 'TimelyFixes', data = df);

data\_one=df[['ReAdmis\_numeric', 'HighBlood\_numeric']]

sns.heatmap(data\_one.corr(),annot=True )

alpha=0.05

def chi\_test(col\_1, col\_2):

cont= pd. crosstab (col\_1, col\_2, margins=False)

print(cont)

c, p, dof, expected = chi2\_contingency(cont)

print('p-value = %.2f' %(p))

print('dof value= %d' % (dof))

print('expected= %s' %(expected))

if p <= alpha:print('Reject the null hypothesis, the two variables are dependent')

else:print('Two variables are independent')

result=chi\_test(df['ReAdmis\_numeric'], df['Listening'])

print(result)

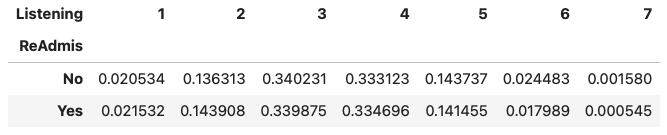
result=chi\_test(df['ReAdmis\_numeric'], df['HighBlood\_numeric'])

print(result)

result=chi\_test(df['ReAdmis\_numeric'], df['RespectfulResponse'])

print(result)

B2. Outputs:





A screenshot of a computer

Description automatically generated

A close up of numbers

Description automatically generated

A close-up of numbers

Description automatically generated

B3. Analysis Technique Justification:

The chi-square technique is best used in the analyses that look for correlations. The method looks for differences in categorical variables and any correlation in two or more variables. ReAdmis has a Bernoulli distribution and is categorical. The Chi-square method will show any differences among other categorical variables.

C1. Univariate Statistics:

The below output gives me a summary of statistics for each of the variables.

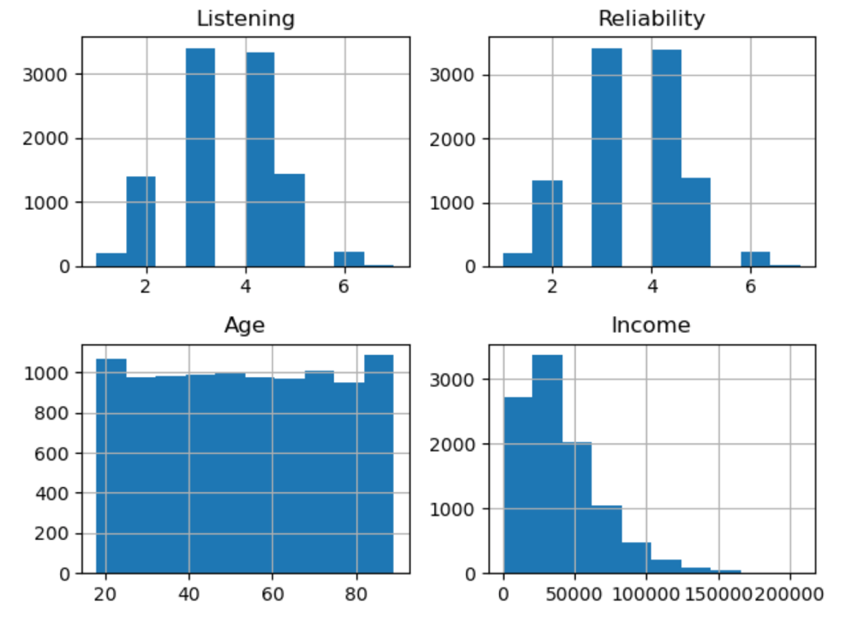
A screenshot of a computer

Description automatically generated

A screenshot of a computer code

Description automatically generated

Continuous and Categorical variables



Listening has distribution of Normal.

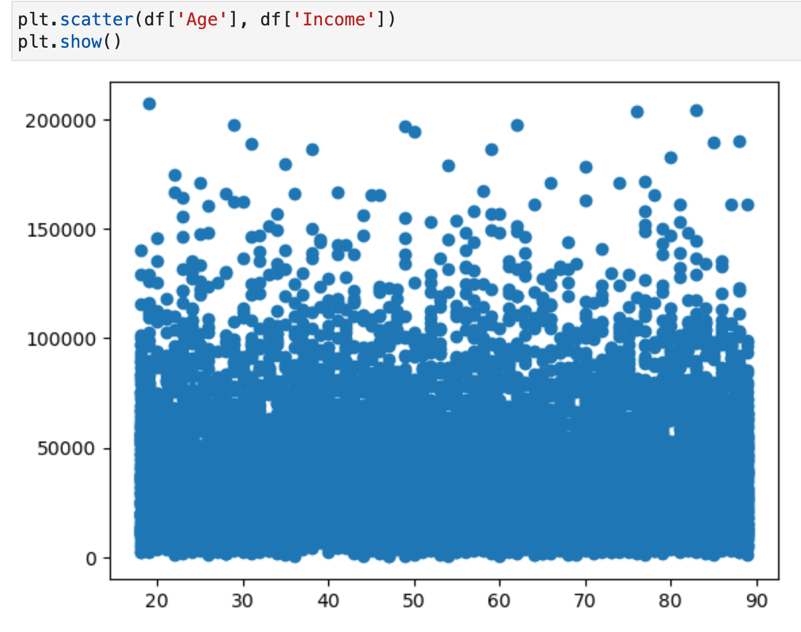
Reliability has distribution of Normal.

Age has distribution of uniform.

Income has distribution of Right Skewed or Chi-Squared.

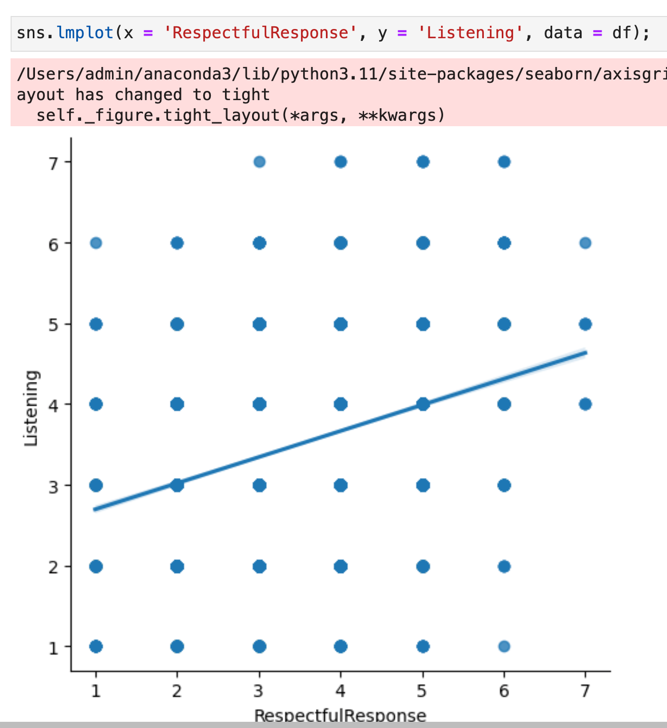
D1. Bivariate Statistics

Continuous variables



Has a normal distribution.

Categorical variables



Has a Uniform distribution.

E1. Hypothesis Test Summary

During this analysis, I looked for any correlation between Admis, Highblood, Item4(RespectfulResponse), and Item8(Listening). I compare two variables by using Hypothesis Testing. When a P-value Exceeds 0.05, it is considered statistically insignificant and fails to reject the null hypothesis. A P-value less than 0.05 is considered statistically significant and rejects the null hypothesis. With P-values of 0.26 (ReAdmis, Listening), 0.84 (ReAdmis, Highblood), and 0.41 (ReAdmis, RespectfulResponse), the variables show no correlation and fails to reject the null hypothesis.

E2. Limitations

Some can assume that when a patient is readmitted to the hospital, high blood pressure, patient perception of a respectful response, and the hospital personnel actively listening to the patient play a role in them being readmitted to the hospital within 30 days. I have discovered that there is no statistical significance between the variables in my research question. The data set is limited as it does not identify why the patient was admitted to the hospital for either visit. Using this data set could lead to inaccurate conclusions due to the general data set.

E3. Recommended COA

My recommendation is to gather data that shows the reason for any admission. This data can be compared to pre-existing conditions and readmissions within 30 days to find any correlations. Updating the data to include reasons for admission will provide a more adequate analysis of patient readmission and produce a better solution to reduce readmissions.

F. Panopto

See Link

G. References:

Dahiru, T. (2008, June). *P - value, a true test of statistical significance? A cautionary note*. Annals of Ibadan postgraduate medicine. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4111019/#:~:text=Being%20a%20probability%2C%20P%20can,other%20than%20due%20to%20chance.

*How to fix: Typeerror: Lmplot() got multiple values for argument “data.”* Stack Overflow. (1968, November 1). https://stackoverflow.com/questions/74342437/how-to-fix-typeerror-lmplot-got-multiple-values-for-argument-data

*Not getting expected results in chi-squared test using Python*. Stack Overflow. (2022, July 1). https://stackoverflow.com/questions/62925425/not-getting-expected-results-in-chi-squared-test-using-python

H. Sources:

None