Hospital Readmissions Data Analysis and Recommendations for Reduction

Background

In October 2012, the US government's Center for Medicare and Medicaid Services (CMS) began reducing Medicare payments for Inpatient Prospective Payment System hospitals with excess readmissions. Excess readmissions are measured by a ratio, by dividing a hospital's number of "predicted" 30-day readmissions for heart attack, heart failure, and pneumonia by the number that would be "expected," based on an average hospital with similar patients. A ratio greater than 1 indicates excess readmissions.

Exercise Directions

In this exercise, you will:

- critique a preliminary analysis of readmissions data and recommendations (provided below) for reducing the readmissions rate
- construct a statistically sound analysis and make recommendations of your own

More instructions provided below. Include your work **in this notebook and submit to your Github account**.

Resources

- Data source: https://data.medicare.gov/Hospital-Compare/Hospital-Readmission-Reduction/9n3s-kdb3 (https://data.medicare.gov/Hospital-Compare/Hospital-Readmission-Reduction/9n3s-kdb3)
- More information: http://www.cms.gov/Medicare/medicare-fee-for-service-payment/acuteinpatientPPS/readmissions-reduction-program.html (http://www.cms.gov/Medicare/medicare-fee-for-service-payment/acuteinpatientPPS/readmissions-reduction-program.html)
- Markdown syntax: http://nestacms.com/docs/creating-content/markdown-cheat-sheet (http://nestacms.com/docs/creating-content/markdown-cheat-sheet)

In [32]:
```
%matplotlib inline

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import bokeh.plotting as bkp
from mpl_toolkits.axes_grid1 import make_axes_locatable
from scipy import stats
```

In [3]:
```
# read in readmissions data provided
hospital_read_df = pd.read_csv('data/cms_hospital_readmissions.csv')
```

Preliminary Analysis

In [4]:
```
# deal with missing and inconvenient portions of data
clean_hospital_read_df = hospital_read_df[hospital_read_df['Number of Discharg
es'] != 'Not Available']
clean_hospital_read_df.loc[:, 'Number of Discharges'] =
clean_hospital_read_df['Number of Discharges'].astype(int)
clean_hospital_read_df = clean_hospital_read_df.sort_values('Number of Dischar
ges')
```

```
C:\Users\amungale\AppData\Local\Continuum\Anaconda3\lib\site-packages\pandas
\core\indexing.py:477: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/st
able/indexing.html#indexing-view-versus-copy
  self.obj[item] = s
```

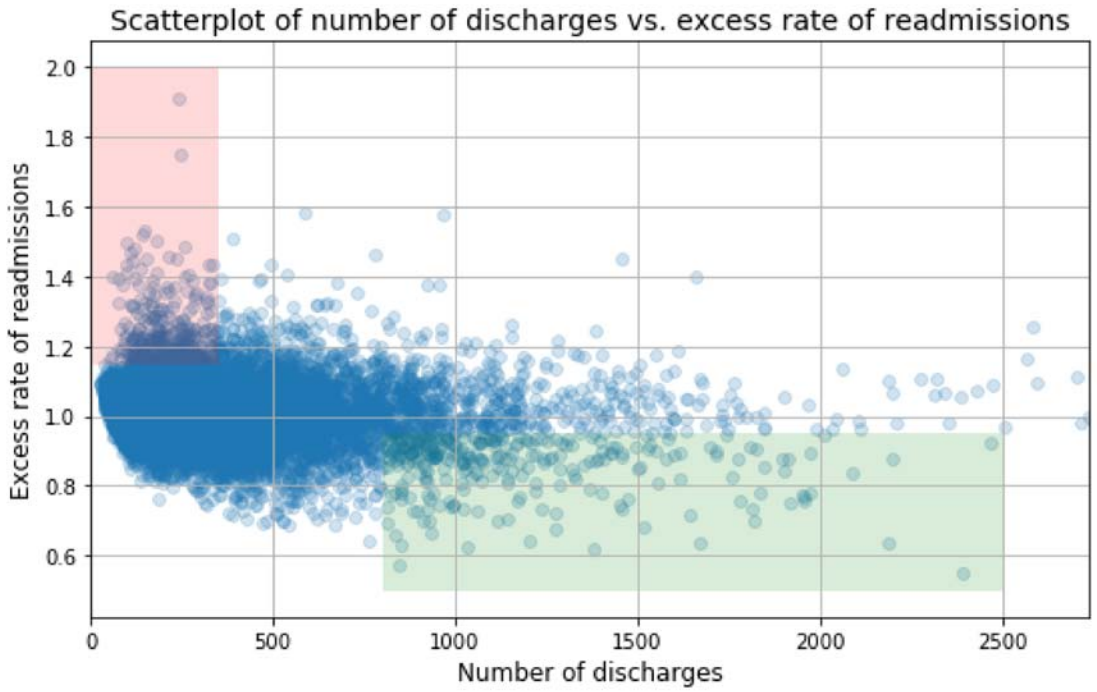
In [5]:
```
# generate a scatterplot for number of discharges vs. excess rate of readmissi
ons
# lists work better with matplotlib scatterplot function
x = [a for a in clean_hospital_read_df['Number of Discharges'][81:-3]]
y = list(clean_hospital_read_df['Excess Readmission Ratio'][81:-3])

fig, ax = plt.subplots(figsize=(8,5))
ax.scatter(x, y,alpha=0.2)

ax.fill_between([0,350], 1.15, 2, facecolor='red', alpha = .15, interpolate=Tr
ue)
ax.fill_between([800,2500], .5, .95, facecolor='green', alpha = .15, interpola
te=True)

ax.set_xlim([0, max(x)])
ax.set_xlabel('Number of discharges', fontsize=12)
ax.set_ylabel('Excess rate of readmissions', fontsize=12)
ax.set_title('Scatterplot of number of discharges vs. excess rate of readmissi
ons', fontsize=14)

ax.grid(True)
fig.tight_layout()
```



Preliminary Report

Read the following results/report. While you are reading it, think about if the conclusions are correct, incorrect, misleading or unfounded. Think about what you would change or what additional analyses you would perform.

A. Initial observations based on the plot above

- Overall, rate of readmissions is trending down with increasing number of discharges
- With lower number of discharges, there is a greater incidence of excess rate of readmissions (area shaded red)
- With higher number of discharges, there is a greater incidence of lower rates of readmissions (area shaded green)

B. Statistics

- In hospitals/facilities with number of discharges < 100, mean excess readmission rate is 1.023 and 63% have excess readmission rate greater than 1
- In hospitals/facilities with number of discharges > 1000, mean excess readmission rate is 0.978 and 44% have excess readmission rate greater than 1

C. Conclusions

- There is a significant correlation between hospital capacity (number of discharges) and readmission rates.
- Smaller hospitals/facilities may be lacking necessary resources to ensure quality care and prevent complications that lead to readmissions.

D. Regulatory policy recommendations

- Hospitals/facilties with small capacity (< 300) should be required to demonstrate upgraded resource allocation for quality care to continue operation.
- Directives and incentives should be provided for consolidation of hospitals and facilities to have a smaller number of them with higher capacity and number of discharges.

Setup an appropriate hypothesis test.

We are testing if smaller hospitals have higher excess readmission rates compared to larger hospitals. We can measure hospital size by the number of discharges. Lets split our data set into 2 groups, hospitals with greater than 280 discharges and hospitals with less. This roughly divides our data set into two equal parts.

Let p_1 be the proportion of excess readmissions (readmissions rates higher than 1) amongs smaller hospitals, and p_2 be the proportion of excess readmissions amongst larger hospitals.

Null hypothesis: $p_1 - p_2 = 0$

Alternative Hypothesis: $p_1 - p_2 > 0$

In [18]:
```
total_len = len(clean_hospital_read_df['Number of Discharges'])
sum(clean_hospital_read_df['Number of Discharges']>280)/total_len
```

Out[18]: 0.49835895664190705

In [34]:
```
small_hosp = clean_hospital_read_df[clean_hospital_read_df['Number of Discharg
es']<=280]
large_hosp = clean_hospital_read_df[clean_hospital_read_df['Number of Discharg
es']>280]
p1 = sum(small_hosp['Excess Readmission Ratio']>1)/len(small_hosp)
p2 = sum(large_hosp['Excess Readmission Ratio']>1)/len(large_hosp)
p = (p1*len(small_hosp) + p2*len(large_hosp)) / (len(small_hosp)+len(large_hos
p))
p,p1,p2
```

Out[34]: (0.51295560545862839, 0.53064738292011016, 0.49514731369150777)

Compute and report the observed significance value (or p-value).

In [39]:
```
#P Value
z_num = p1-p2
se = np.sqrt(p*(1-p)*(1/len(small_hosp) + 1/len(large_hosp)))
z = z_num/se
z
#1-stats.norm.cdf(z)
```

Out[39]: 3.8211094103433183

Report statistical significance for α = .01.

We calculate a z value of 3.8211. Not only is this test statistically significant at $\alpha = 0.01$, this is significant for $\alpha = 0.0001$

Discuss statistical significance and practical significance. Do they differ here? How does this change your recommendation to the client?

We have statistical evidence that the proportions of excess readmissions is truly higher among smaller hospitals. This can be subjective in a practical setting when discussing medium sized hospitals and which category the fall into, but we know that given how we divided the hospitals (number of discharges being less than 280 vs grater), the smaller hospitals have higher readmission rates. I would agree with the premliminary reccomemdation of either moving towards consolidating smaller hospitals and their resources with larger ones, or demanding some sort of evidenceo higher levels of quality of care amongst smaller hospitals. The one exception I would make would be to further investigate medium sized hospitals, with discharge numbers between 500 and 1000, to see if they the quality of care differs from the hospitals with >1000 discharges and hospitals with <300 discharges.

Look at the scatterplot above.

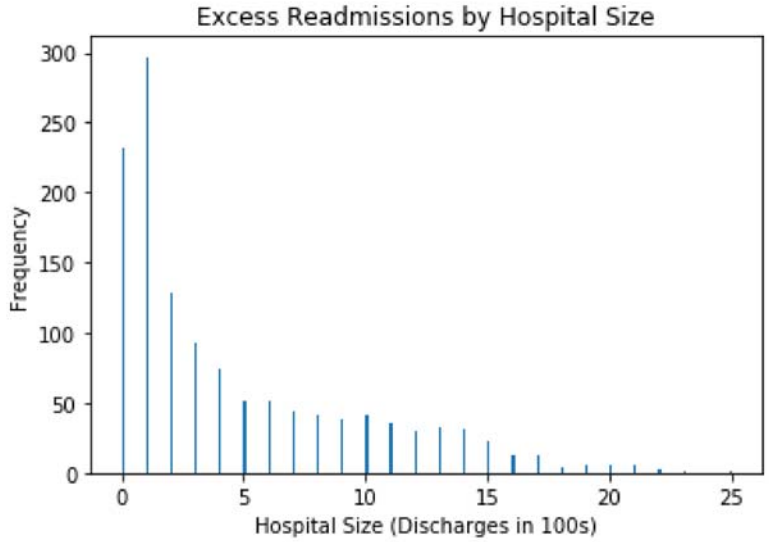
```
  - What are the advantages and disadvantages of using this plot to convey informat
ion?
  - Construct another plot that conveys the same information in a more direct manne
r.
```

The scatterplot above is useful in identifying the general distribution of our data points in relation to the two variables we are exploring. We can clearly see where the majority of our data points fall, and a loose trend of lower readmission rates correlating with higher discharge numbers, or hospital size. One disadvantage to this plot is that many of our data points are cluttered into the mass of points among hospitals with discharge sizes between 0 and 1000. We cannot distinguish the individual data points, and more importantly, we cannot distinguish any potential trends within this data. A better wat to present this would be to plot number of discharges vs the number of exccess readmissions. Shown below, the trend is more clearly visible.

In [ ]:

In [93]:
```
clean_hospital_read_df['HighReadmission'] = clean_hospital_read_df['Excess Rea
dmission Ratio']>1
scatter_data = clean_hospital_read_df.groupby('Number of Discharges').sum()
#x=scatter_data.index.values
#y=scatter_data.HighReadmission.values
plt.hist(scatter_data.HighReadmission, bins=250)
plt.title("Excess Readmissions by Hospital Size")
plt.xlabel("Hospital Size (Discharges in 100s)")
plt.ylabel("Frequency")
plt.show()

#fig, ax = plt.subplots(figsize=(20,20))
#ax.scatter(x, y,alpha=0.9)
```



In [ ]: