NPPES Analysis

Healthcare Data Analytics and Data Mining

Group 3

Jinman Rong
Runxua Cao
Ting Xiao
Xinlan Wu
Yunyi Wang
Ziyu Tang
(Audit) Peihan Tian
(Audit) Fangzhou Xiang

Introduction

This report analyzes the status about the US healthcare providers using the National Plan and Provider Enumeration System (NPPES) database.

Main topics are discussed as follows:

- Healthcare providers and their first licensed states
- Gender difference in practicing as a "Sole Proprietor"
- Gender difference in low and high risk-reward categories
- Density of MRI centers per population by states and reasons behind

Through this report, you will have a simple understanding of the supply side of the healthcare market in United States.

Healthcare providers and their first licensed states

The information of the doctor for each member is presented as follows：

Table 1 provider license state of the doctor for each member

Provider License State	Group Member (Last Name)
GA	Cao
NH	Rong
GA	Tang
FL	Tian
GA	Wang
TN	Wu
FL	Xiang
NC	Xiao

Gender difference in practicing as a "Sole Proprietor"

The report explores the gender difference in practicing as a "Sole Proprietor" in 9 states, including California, Ohio, Tennessee, Wisconsin, Kentucky, Iowa, New Mexico, Maine and Wyoming. The 2x2 cross table for gender by Sole Proprietor is presented below:

Table 2 Gender difference in practicing as a "Sole Proprietor"

	Female	Male
SP	226805	135606
Non_SP	458823	257113

The result is statistically significant (p-value < 2.2e-16). Thus, practicing as a "Sole Proprietor" is related to gender and females are less likely to establish solo practice offices compared with males.

Gender difference in low and high risk-reward categories

Are male healthcare individual providers more likely than their female peers to choose the practices that are associated with higher risk for a higher reward? The answer is yes. According to the Salary.com website, the Obstetrics & Gynecology and Pediatrics have the lowest average salary, while Surgery and Orthopaedic Surgery have the highest one. Assumed that high reward always means high risk, we class “Obstetrics & Gynecology” and “Pediatrics” as low risk/reward category and “Surgery” and “Orthopaedic Surgery” as high risk/reward one. The 2x2 cross table for gender by low and high risk-reward categories in 9 states is presented below:

Table 3 Gender difference in low and high risk-reward categories

	High risk high reward	Low risk low reward
Female	1837	14494
Male	10079	9403

The result is statistically significant (p-value < 2.2e-16). Thus, it can be concluded that the gender difference may affect healthcare individual provider’s choice among high or low reward/risk practices, and male healthcare individual providers are more likely than their female peers to choose the practices that are associated with higher risk for a higher reward.

Density MRI centers per population by states and reasons behind

The density of MRI centers by state population varies from state to state. The heat map is presented below:

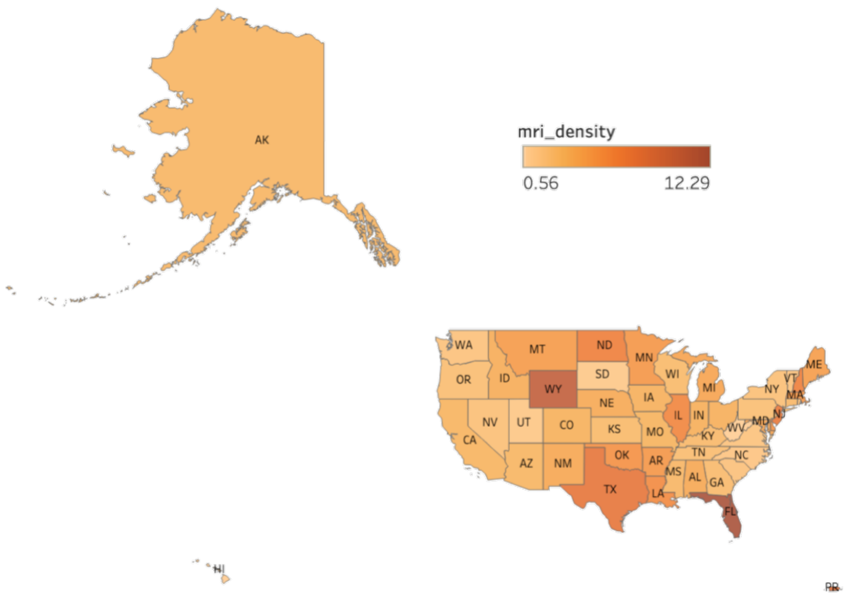

Figure 1 National heat map of MRI centers

The population statistics about the US population by states is referred from the World Population Review: *US States - Ranked by Population 2019*[1].

It can be seen that the density of MRI centers per population in Florida is the highest, which is about 12.29 MRI centers per 1000,000 population. While West Virginia has the least whose number is about 0.56.

1 http://worldpopulationreview.com/states/

The reasons why Florida has the highest density of MRI centers per population is related with four entities: the state government, the patients, the healthcare providers and the medical schools

State government:

Among the ranking of 50 states government health expenditure per capita, Florida government ranked 42 with the spend of $4,256 per capita on all programs, which reveals that Florida government may not pay much attention to healthcare to some extent. Therefore, it is not surprised that Florida has the 5th lowest health insurance coverage among all states[2]. The government's less health expenditure leads to more patients' self-payment, which may affect patients' preference for healthcare treatment, that is, patients will prefer lower cost healthcare services.

Patients:

Florida has the highest percentage of aged 65 and over population (19.8%) and the 22nd highest population with a disability (13.6%), showing that the market size of healthcare, especially MRI, is quite huge, and the industry is very promising given the large demand[3].

Given the unwillingness of Florida government spending on healthcare, many patients have to pay by themselves to see a doctor. When it comes to MRI scanning, they may choose imaging centers instead of hospitals given the huge price gap between these two providers[4], making the MRI centers under huge demands in Florida.

Table 4 Cost gap of MRI between imaging center and hospital in Florida [5]

MRI scan cost in Florida by insurance type		
Insurance	Hospital MRI costs	Imaging center MRI costs
Insured patient	$160 – $1,643 (25% co-pay)	$80 – $986 (25% co-pay)
High-deductible patient	$641 – $6,570	$321 – $3,942
Medicare/Medicaid patient	$54 – $124 (20% co-pay)	$36 – $83 (20% co-pay)
Cash patient	318– $725	$272 – $518

Healthcare providers:

There is a “healthy” loop for MRI imaging centers in the market: the promising healthcare market springs up many imaging centers, which may eventually lower the MRI scanning prices because of price competition. In return, the lower MRI scan price provided by imaging centers than that provided by hospitals attracts more patients, which keeps the loop moving forward.

2 https://www.usatoday.com/story/money/economy/2018/08/20/what-your-state-spends-health-care-per-capita/37121541/
3 https://www.usatoday.com/story/money/economy/2018/08/20/what-your-state-spends-health-care-per-capita/37121541/
4 https://www.businessinsider.com/how-much-an-mri-costs-by-state-2017-3
5 https://affordablescan.com/blog/mri-cost-in-florida/

Medical schools:

MRI technology in Florida leads the industry. Latest development in technology have led to the introduction of superconducting magnets with field strength above 20T. The imaging of neuro disease using High Field MR (INUMAC) project in University of Florida has been developing a powerful 11.75 Tesla magnet that will dwarf others in terms of size and quality of imaging.[6] The edging MRI technology may lower the MRI facilities costs and promote the quality to some extent, and finally boost the MRI market in Florida.

Conclusion

For the individual healthcare suppliers, the statistically significant differences are found between females and males in practicing as a "Sole Proprietor" and low and high risk-reward categories. Females are less likely to establish solo practice offices or choose the practices that are associated with higher risk for a higher reward. This may because of the different risk preferences between gender.

For the organizational healthcare suppliers, we found that the density of MRI centers by population varies from state to state, and Florida has the highest density of MRI centers per population. Four entities in Florida contributes to this situation: the state government, the patients, the healthcare providers and the medical schools. However, we can't say that our conclusion is robust because we don't have detailed information about how many MRI machines are in each facility by states. More data and information should be dug into to get reliable insights.

6 https://healthmanagement.org/c/imaging/news/north-america-continues-to-drive-mri-technologies-market

Appendix

Healthcare Analysis Assignment-1 code

group 3

11/5/2019

```r
rm(list=ls())
library(data.table)
library(sandwich)
library(tidyverse)
library(lmtest)
library(ggplot2)
library(knitr)
library(psych)
library(dplyr)
```

Question 2

```r
# Import data
df <- fread('~/Downloads/analyze_data.csv')

#Replace null with NA
df[df$`Provider Gender Code`==""]=NA
df[df$`Is Sole Proprietor`==""|df$`Is Sole Proprietor`=="X"]=NA

#Select the three colomns we needed to create a new dataframe for further analysing
df1=df[,c('Provider Gender Code','Is Sole Proprietor','Provider Business Practice Location Address State Name')]

#Omit NA
df1=na.omit(df1)

#Filter the states
df1=df1[df1$`Provider Business Practice Location Address State Name`=='CA'|df1$`Provider Business Practice Location Address State Name`=='OH'
       |df1$`Provider Business Practice Location Address State Name`=='TN'|df1$`Provider Business Practice Location Address State Name`=='WI'
       |df1$`Provider Business Practice Location Address State Name`=='KY'|df1$`Provider Business Practice Location Address State Name`=='IA'
       |df1$`Provider Business Practice Location Address State Name`=='NM'|df1$`Provider Business Practice Location Address State Name`=='ME'
       |df1$`Provider Business Practice Location Address State Name`=='WY']

#Create dataframe male/female for Gender by Yes/No for Sole Proprietor
FY=df1[which(df1$`Provider Gender Code`  == 'F' & df1$`Is Sole Proprietor`=='Y'), ]
FN=df1[which(df1$`Provider Gender Code`  == 'F' & df1$`Is Sole Proprietor`=='N'), ]
```

```r
MY=df1[which(df1$`Provider Gender Code` == 'M' & df1$`Is Sole Proprietor`=='Y'), ]
MN=df1[which(df1$`Provider Gender Code` == 'M' & df1$`Is Sole Proprietor`=='N'), ]

#Create the 2*2 matrix
gen_prop=matrix(c(nrow(FY),nrow(MY),nrow(FN),nrow(MN)),nrow=2,ncol=2,byrow=T,dimnames=list(c('Y','N'),c('F','
M')))

#Fisher test
fisher.test(gen_prop)
##
##  Fisher's Exact Test for Count Data
##
## data:  gen_prop
## p-value < 2.2e-16
## alternative hypothesis: true odds ratio is not equal to 1
## 95 percent confidence interval:
##  0.9294875 0.9450474
## sample estimates:
## odds ratio
##  0.9372444
```

Question 3

```r
npi=fread(file='~/Downloads/npidata_pfile_20050523-20191013.csv',
          col.names=c('NPI','Entity Type Code','State','Gender','Taxonomy'),
          select = c('NPI','Entity Type Code','Provider Business Practice Location Address State
Name','Provider Gender Code','Healthcare Provider Taxonomy Code_1') )
npi_gender=npi[Gender=='M'|Gender=='F']
npi_state=npi_gender[State=='CA'|State=='OH'|State=='TN'|State=='WI'|State=='KY'|State=='IA'|State=='NM'|Stat
e=='ME'|State=='WY']
npi_taxo=npi_state[Taxonomy=='208000000X'|Taxonomy=='207V00000X'|Taxonomy=='208600000X'|Taxonomy=='207X00000X
']
npi_taxo[(Taxonomy=='208000000X'|Taxonomy=='207V00000X'),':='(risk='low')]
npi_taxo[(Taxonomy=='208600000X'|Taxonomy=='207X00000X'),':='(risk='high')]
npi=npi_taxo
head(npi)
##           NPI Entity Type Code State Gender   Taxonomy risk
## 1: 1841293891                1    TN      M 208600000X high
## 2: 1417950460                1    NM      M 207X00000X high
## 3: 1376546317                1    KY      M 208000000X  low
## 4: 1922001957                1    TN      M 207V00000X  low
## 5: 1568465599                1    ME      M 208600000X high
## 6: 1417950569                1    IA      M 207X00000X high
a=npi[Gender=='F'&risk=='high',sum(`Entity Type Code`)]
b=npi[Gender=='F'&risk=='low',sum(`Entity Type Code`)]
c=npi[Gender=='M'&risk=='high',sum(`Entity Type Code`)]
d=npi[Gender=='M'&risk=='low',sum(`Entity Type Code`)]
```

```
data=matrix(c(a,b,c,d),ncol=2)
fisher.test(data)
##
##  Fisher's Exact Test for Count Data
##
## data:  data
## p-value < 2.2e-16
## alternative hypothesis: true odds ratio is not equal to 1
## 95 percent confidence interval:
##  0.1118644 0.1252415
## sample estimates:
## odds ratio
##  0.1183885
```

Question 4

```
dt             = fread("~/Downloads/analyze_data.csv")
state_data     = fread("~/Downloads/state_data.csv")
dt_org=dt[dt$"Entity Type Code"==2,]
dt_org_mri=dt_org[dt_org$"Healthcare Provider Taxonomy Code_1"=="261QM1200X",]
dt_org_mri_state            =   data.frame(dt_org_mri   %>%
                                  group_by(state=dt_org_mri$"Provider Business Practice Location Address State
Name") %>%
                                  summarise("Healthcare Provider Taxonomy Code_1"= n()))
state_pop_mri               =   merge(dt_org_mri_state, state_data, by = "state")
state_pop_mri$Pop_new       =   state_pop_mri$Pop/1000000
state_pop_mri$mri_density   =   state_pop_mri$Healthcare.Provider.Taxonomy.Code_1/state_pop_mri$Pop_new
head(state_pop_mri)
##   state Healthcare.Provider.Taxonomy.Code_1      State      Pop    Pop_new
## 1    AK                                   2     Alaska   735720   0.735720
## 2    AL                                  18    Alabama  4898246   4.898246
## 3    AR                                  15   Arkansas  3026412   3.026412
## 4    AZ                                  20    Arizona  7275070   7.275070
## 5    CA                                 113 California 39747267 39.747267
## 6    CO                                  18   Colorado  5770545   5.770545
##   mri_density
## 1    2.718425
## 2    3.674785
## 3    4.956364
## 4    2.749114
## 5    2.842963
## 6    3.119289
write.csv(state_pop_mri, file="~/Downloads/state_pop_mri.csv")
```