What factors are associated with the average payment of clinic visiting

Jie Lian

November 2019

1   Introduction

It is interested to understand what factors are associated with the average payment of clinic visiting. There are some factor candidates, including operation costs of clinic, taxes rate, government engagement, and individual health condition. In this simply study, I want to examine whether operation costs and government engagement can explain the variance of the average payment of clinic visiting.

2   Data

In this short study, I collect data for three sets of variables: average payment of clinic visiting, operation costs, and government engagement. Their data resources and collection method are explained as follows.

- Average payment of clinic visiting. The data resource is Google BigQuery public data. There are four tables focusing outpatient charge from 2011 to 2014. In order to make the study simple, I only use the 2011 table. I use google official API to collect data.

- Operation costs. I use nursing hourly salary as the measurement of this variable. The data come from the average registered nurses salaries by state table on www.nursingce.com. The data is for 2019. I assume the salaries have not changed from 2011 to 2019. I use web scrapping to collect data.

- Government engagement. I use number of new enacted health care policies in 2011 to proximate this variable. There are three topics of policies,

nutrition, physical activity, and obesity. I consider them as three different independent variables. And also, I add their sum to the model to be another independent variable. I used the official designated API to collect data.

The dependent variable in this mini study is the average payment of clinic visiting. The distribution of this variable is showed in Figure 1.

The scatter plot with fitted line between nurses' hourly salary and average payment of clinic visiting is showed in Figure 2. Unfortunately, the attempt to make same plot for other independent variable was not successful because Python can't converge the fit line correctly, showing that the low correlation between those independent variables and the dependent variable.

3   Model and Result

I run a linear regression model to better understand the relationships between independent variables and dependent variables.

The model is as follows:

average total payments $= \beta_1 + \beta_2 * Nutrition + \beta_3 * Obesity + \beta_4 * PhysicalActivity + \beta_5 * totalPolicy + \beta_6 * hourlySalary + u$

Unfortunately, the p-values of government engagement variables are all much greater than 0.05. Therefore, the hypothesis that the government engagement is associated with average total payment of clinic visiting is not supported. But the relationship between hourly nursing salary and the average payment are supported by the model result, which is showed as Figure 3.
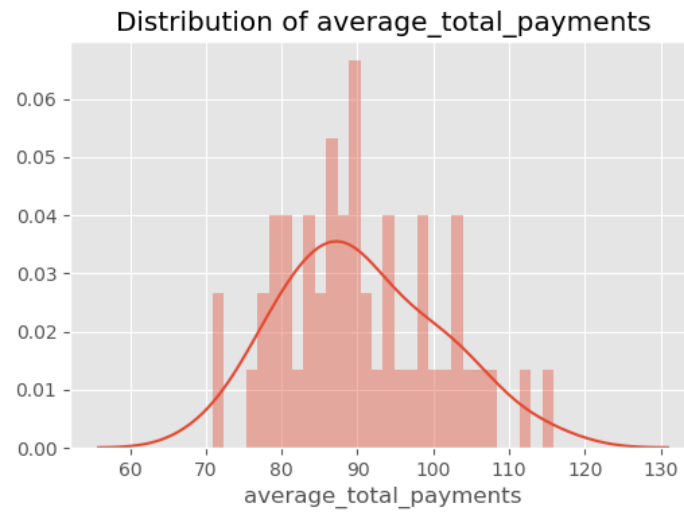
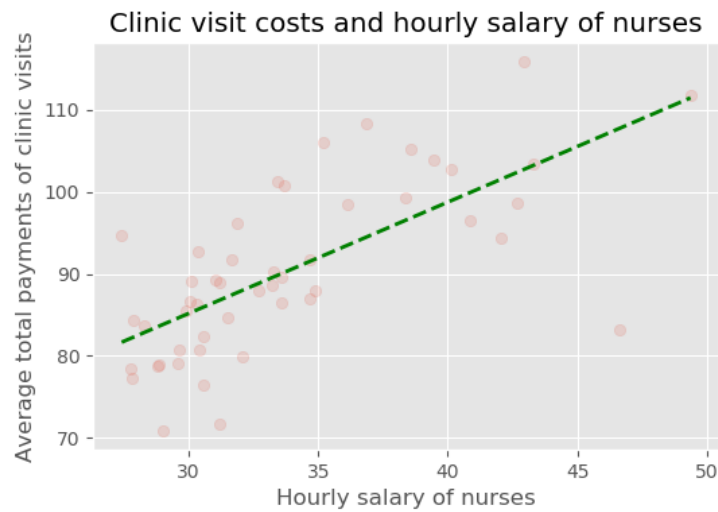Figure 1: Distribution of the average payment of clinic visiting in 2011

Figure 2: The relationship between average payment of clinic visiting and nurses' hourly salary

```
                         OLS Regression Results
==============================================================================
Dep. Variable:     average_total_payments   R-squared:                       0.531
Model:                                OLS   Adj. R-squared:                  0.479
Method:                     Least Squares   F-statistic:                     10.19
Date:                    Tue, 12 Nov 2019   Prob (F-statistic):           1.28e-05
Time:                            02:48:12   Log-Likelihood:                 -139.06
No. Observations:                      41   AIC:                             288.1
Df Residuals:                          36   BIC:                             296.7
Df Model:                               4
Covariance Type:                nonrobust
==============================================================================
                     coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const              41.4377      7.967      5.201      0.000      25.279      57.596
Nutrition          -0.0805      0.143     -0.562      0.578      -0.371       0.210
Obesity             0.2255      0.310      0.728      0.471      -0.403       0.854
Physical Activity  -0.1405      0.171     -0.821      0.417      -0.487       0.206
totalPolicy         0.0045      0.032      0.141      0.889      -0.060       0.069
hourlySalary        1.4502      0.235      6.182      0.000       0.974       1.926
==============================================================================
Omnibus:                       10.202   Durbin-Watson:                   1.884
Prob(Omnibus):                  0.006   Jarque-Bera (JB):               12.937
Skew:                          -0.709   Prob(JB):                      0.00155
Kurtosis:                       5.358   Cond. No.                     1.14e+16
==============================================================================
```

Figure 3: OLS resut

4