Correlation between hospital charges and local poverty rates

Hsuan Lee, Chen Yang and Yanyan Zhang
University of Utah
May 2nd, 2017

Background and motivation

Hospital charges vary in a surprisingly large range. Reasons for that remain largely unexplained [1]. Our hypothesis in this study is that higher local poverty rates are associated with lower hospital charges. We would like to see whether it is true by analyzing these two datasets:

- A provider level summary of Inpatient Prospective Payment System (IPPS) discharges, average charges and average Medicare payments for the Top 100 Diagnosis-Related Groups (DRG)
- Data on poverty rates by county

Key ideas

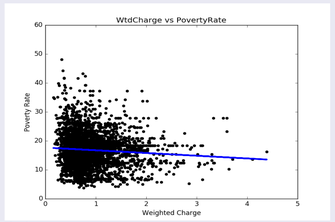
- We took into account all the counties whose centers are within 50 miles of the hospital, and generated weighted poverty rates. The weight of the poverty rates was based on the distance between each county center and the hospital.
- To get better linear regression models, we separated hospitals into five groups based on the locations: west, midwest, northeast, southwest and southeast. Based on this and an approach similar to the 5-means clustering, we obtained the five 'best fit' linear regression models.

Methods

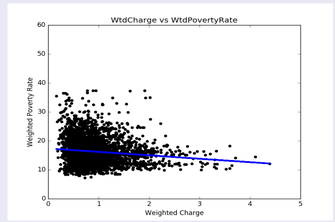
- We used the least squares method to generate linear regression models.
- We applied the 5-means clustering method to get five 'best fit' linear regression models.

Main results

It shows in Figure 1 that the linear regression model fit the data better after the original poverty rates were updated to weighted poverty rates. The mean squared error dropped from 34.06 to 21.47.



(a) Hospital charge levels v.s. Original poverty rates

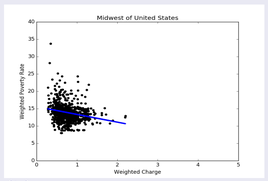


(b) Hospital charge levels v.s. Weighted poverty rates

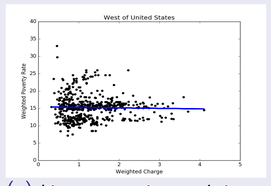
Figure 1: Linear regression results for nationwide data

To improve the models, we separated hospitals to five groups according to their locations following the grouping in Figure 2a. The linear regression models are shown in Figure 2b - 2f.

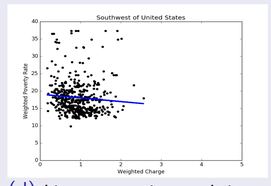
(a) Five region groups



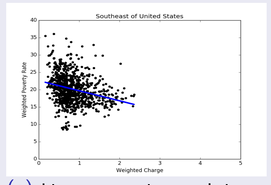
(b) Linear regression result in the midwest



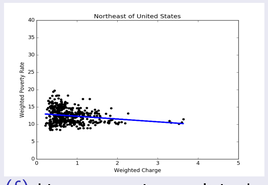
(c) Linear regression result in the west



(d) Linear regression result in the soutwest



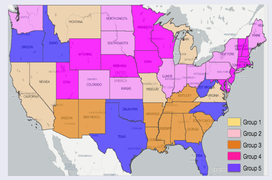
(e) Linear regression result in the southeast



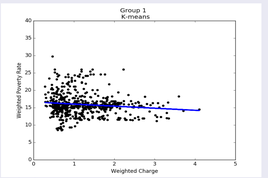
(f) Linear regression result in the northeast

Figure 2: Linear regression models for five region groups

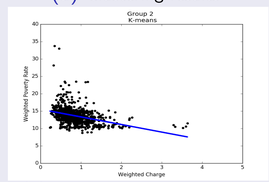
After applying the 5-means clustering method, we obtained the five 'best fit' linear regression models, shown in Figure 3b - 3f. States corresponding to these five models are colored differently in Figure 3a.



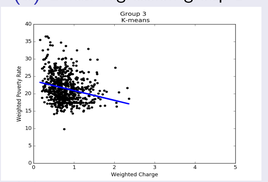
(a) Clustering results



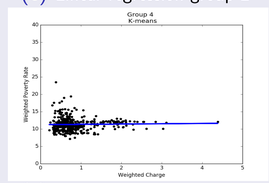
(b) Linear regression group 1



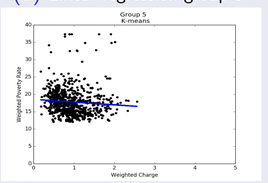
(c) Linear regression group 2



(d) Linear regression group 3



(e) Linear regression group 4



(f) Linear regression group 5

Figure 3: Five 'best fit' linear regression models

Conclusions and lessons

- There was negative correlation between hospital charges and local poverty rates. But the correlation was not very strong.
- Getting data ready for analysis can take a long time.
- It can be tricky to choose right starting parameters.
- One should try different approaches to apply on the same data.
- Real data might not be as beautiful as we want them to be.

Acknowledgement

We would like to thank Jeff Phillips for discussions and advice.

Reference

[1] James D. Park, Edward Kim, and Rachel M. Werner, *Inpatient hospital charge variability of U.S. hospitals*. Journal of General Internal Medicine, 30(11): 1627-1632, November 2015.