Renee Probetts

IEMS 308

1/26/18

Clustering Assignment

**Medicare Payment Amounts**

**Executive Summary**

Today there are approximately 44 million beneficiaries of Medicare in the United States. It is useful to know where that money is going and who it is going to. Knowledge where the majority of the money is going and if it is more heavily spent on certain people can be used to better fund research initiatives and where to focus preventative efforts. For example, if the providers with the highest percent of beneficiaries with strokes are the providers with the highest amount Medicare is paying, one can look into ways to prevent strokes in beneficiaries with that chronic condition. Preventative measures can end up saving a significant amount of money in the long run.

In order to segment the Medicare providers into similar categories, clustering analysis can be performed to determine what similarities are found in groups with different payment amounts.

Clustering analysis uses a specific algorithm, k-means in this case, to group the data together based on similarity measures. In this case the measures included the amount Medicare paid the provider; the ages, genders, and chronic conditions of the beneficiaries; and the total services provided. By doing so, the Center for Medicare and Medicaid Services can easily examine what providers who costing more money have in common.

Upon first inspection of the data, and as expected, we saw that the amount of money Medicare spent was what ranged the most between providers, and would likely drive the analysis. After running the clustering analysis, we saw that this was in fact the case. Chronic conditions; however, had significantly no variance within clusters. The number of services provided and number of beneficiaries receiving these services did vary between clusters.

This information is not useful when it comes to chronic conditions; nevertheless, it is practical to know that the money is being spent in a rational way… with more services and more beneficiaries, more money is paid.

**Problem Statement**

An average of 91864.99 dollars are paid by Medicare for each provider’s services. We were tasked with determining whether certain chronic conditions (beneficiaries meeting the CCW chronic condition criteria) were costing the Medicare system more money. And if so, which ones were costing the most. With this information, Medicare can invest in specific preventative programs to reduce costs in the long run.

**Methodology**

The first step in clustering analysis is determining which variables are relevant. We began by exploring the data to determine what was going on. We found that the locations and place of service were irrelevant to the question posed, as was the race of beneficiaries. We created new data frames, which only included the pertinent variables. Those being ‘total\_medicare\_payment\_amt’, ‘total\_med\_medicare\_payment\_amt’, ‘total\_drug\_medicare\_payment\_amt’, ‘beneficiary\_average\_age’, ‘beneficiary\_female\_count’, ‘beneficiary\_male\_count’, ‘total\_services’, ‘total\_unique\_benes’, as well as all the ones beginning with ‘beneficiary\_cc’.

For the purpose of reducing calculation times, this data set was split into samples and the same actions were performed six different times to determine the best algorithm, the optimal number of clusters. We then calculated the within sum of squares (wss) from the k-means clustering algorithm with varying numbers of clusters. By plotting the wss, we were able to see the optimal k-value. Over the six samples, the best number of clusters was 10. Using R software, we did a k-means clustering analysis. We calculated the centers of each cluster and calculated the variance seen in Table 1. This gives an idea of how different each feature is between clusters.

|  |  |
| --- | --- |
| Feature | Variance |
| afib | 9.192040003 |
| alzrdsd | 36.17168697 |
| asthma | 5.526267108 |
| cancer | 37.02612864 |
| chf | 37.62346125 |
| ckd | 35.23572279 |
| depr | 11.10396049 |
| diab | 25.83326773 |
| hyperl | 8.338838943 |
| hypert | 7.233789182 |
| ihd | 2.526959651 |
| ost | 34.91332038 |
| schio | 1.435211523 |
| strk | 8.509804108 |
| payment | 3.12899E+15 |
| med payment | 3.1484E+15 |
| drug medicare | 4.99417E+11 |
| average age | 5.469731914 |
| female count | 29218912811 |
| male count | 13460482779 |
| total services | 1.13559E+13 |
| unique benes | 82337046492 |

Table : Center Variance

**Analysis**

Conclusions