**State Poverty Levels against SNF Medicare Payments: A deep dive into SNF Payments and Lengths of Stay for 2015 by Medicare**

**By Andrea Fox**

For my final project I wanted to look at Medicare payments for SNF (skilled nursing facilities) to see if payments were affected by the state poverty level. I found a dataset that outlines all SNF payments by Medicare for 2015 by State (1). My initial thought was that I expected states with a higher poverty level to have a higher ALOS (average length of stay) and higher payments by Medicare.

I started with putting together my data frames. I had 3 specific data frames: 1 that contained the full csv file of data, 1 that specified all of the data by the top 10 richest states, and 1 that specified all of the data by the top 10 poorest states. These states are broken into the following (2):

Top 10 poorest states

* MS – Mississippi
* NM – New Mexico
* LA – Louisiana
* KY – Kentucky
* WV – West Virginia
* AR – Arkansas
* AL – Alabama
* TN – Tennessee
* AZ – Arizona
* GA – Georgia

Top 10 richest states

* NH – New Hampshire
* MN – Minnesota
* HI – Hawaii
* ND – North Dakota
* MD – Maryland
* UT – Utah
* VA – Virginia
* MA – Massachusetts
* WY – Wyoming
* VT - Vermont

For the most part I used the smaller two data sets, but I did use the full one for some analyzing. I then created histograms for all my major variables by all 50 states, this includes: total stays, ALOS, total charge amount, total Medicare allowed amount, and total Medicare payment amount. From observing these histograms, it looks like my theory could be incorrect because the highest bars aligned with states that had the highest populations, which most do not fall within either top 10 poorest or richest states. I then created two histograms to specifically look at total SNF Medicare payments by each top 10, which again looks to disprove my theory as Massachusetts had a higher number of payments than the highest within my poverty states.

After the histograms I started to delve into the data specifically for my top 10 states. By looking at the mean for ALOS I was correct in my theory that the states that are higher in poverty had longer stays, but total stays and total Medicare payments were higher for my top 10 richest states. The variance and standard deviation was also higher for ALOS for the poorer states, but much higher for all other variables for the richer states. When I attempted to plot my PMF comparing payments by each dataset the graph originally was blank. It took me awhile to figure out that my PMF is incredibly low (between 0-0.0007), but my payment amounts were in the high thousands to low millions. Once I adjusted the axis numbers, I was able to see more, but the bars are so close together that it looks like one giant bar. The CDF or cumulative distribution function looks pretty much the same for both total stays by top 10 states and total payments by top 10 states. Each graph shoots to 1 quickly, which if I remember correctly shows a high relationship between them. For correlation I plotted total stays by total payments for all states and ALOS by total charge amounts for all states. Total stays by payments had a very high correlation at 0.9, but ALOS for charge amounts had a very weak correlation at 0.08.

I finished out the project by running hypothesis tests as well as regression analysis. I ran three different hypothesis tests with diffmeanspermute and correlationpermute giving me what looked like accurate answers and the t-test looking incorrect as it had p-value higher than 1. The difference in means gave me 0.0 which means it is statistically significant. When testing correlation, I got 0.65 so less statistically significant. The t-test I’m not sure I did something incorrect here or if my p-value really is 1.18, but it does not seem accurate as most of the time p-value should never exceed 1. For regression analysis I had some difficulty understanding everything that came back, but total payment by state looks to be significant looking at the p-value, while payments by state for total stays was not as significant. I think with regression analysis I could spend more time in reading on to understand the significance of my results.

Overall, I would have to say my original hypothesis was incorrect in that the top 10 poorest states received higher payments than the top 10 richest states. However, I was correct when I said that the states with higher poverty levels had longer ALOS. I think I needed to look at states with higher population as those states probably had some effect on my numbers. This was definitely a very helpful exercise in really making me think outside the box on my hypothesis as well as evaluating the data. I’m sure there are plenty of things that I missed, but I felt like I had the right variables to answer my questions. I also think my limited knowledge in python held me back some as I didn’t feel confident in coding everything.

References:

1. Centers for Medicare & Medicaid Services (n.d.). Medicare Skilled Nursing Facility Provider Reports. Retrieved January 15, 2020, from <https://www.kaggle.com/cms/medicare-skilled-nursing-facility-provider-reports#medicare-skilled-nursing-facility-snf-provider-aggregate-report-cy-2015.csv>
2. World Population Review (n.d.). Poverty Rate by State 2020. Retrieved January 16, 2020, from <http://worldpopulationreview.com/states/poverty-rate-by-state/>