**Healthcare Costs and Demographics: Do Population**

**Demographics Affect Healthcare Costs?**

Christina Rawls, Eduardo Panneflek, Kristian Pavlock, Taylor Galloway

**Project Description:**

The purpose of this project is to use Python, Pandas, and Matplotlib to explore whether there is a relationship between healthcare costs in California and median age, income per capita and population.

**Scope of Research:**

This was carried out by accessing the both the Census API as well as Centers for Medicare and Medicaid Services. The Census provided data in regard to median age, income per capita and population, this data was filtered to only look at zip codes specifically in California. Centers for Medicare and Medicaid Services provided a dataset which included specific medical procedures, where they were conducted and the average total payments. The top 5 procedures across the state of California were used to test the following hypotheses, see Figure 1. for reference.

Procedure Definitions:

* 871: Septicemia or Severe Sepsis w/o MV >96 Hours w/ MCC
* 291: Heart Failure & Shock w/ MCC
* 872: Septicemia or Severe Sepsis w/0 MV >96 Hours w/o MCC
* 470: Major Joint Replacement or Reattachment of Lower Extremity w/o MCC
* 392: Esophagitis, Gastroent & Misc. Digest Disorders w/o MCC

MCC: Major Complications and Comorbidities

**Figure 1. Top 5 Most Frequent Diagnoses**

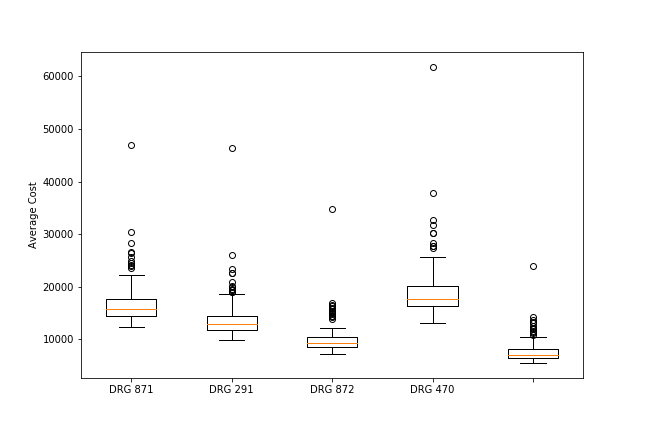
A screenshot of a cell phone

Description automatically generated

**Data Cleaning Procedures**

Prior to beginning the analysis data points that existed above the upper quartile of each of the procedures were dropped, in order to more appropriately represent the data.

**Figure 2. Boxplot of the top 5 most frequent procedures, prior to dropping outliers**



**Figure 3. Boxplot of the top 5 most frequent procedures, after dropping outliers**

A screenshot of a video game

Description automatically generated

**Hypotheses:**

1. If there is a relationship between cost of healthcare and population, there will be higher healthcare costs in areas with higher population.
2. If there is a relationship between cost of healthcare and income per capita, there will be higher healthcare costs in areas with higher income per capita.
3. If there is a relationship between cost of healthcare and median age, there will be higher healthcare costs in areas with higher median age.

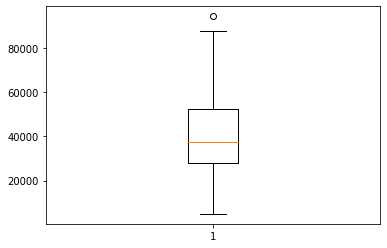
**Hypothesis 1**

*If there is a relationship between cost of healthcare and population, there will be higher healthcare costs in areas with higher population.*

From our boxplot you can see that the population in California, from the sample we took is skewed slightly to the left. Therefore, 50% of the zip codes included in this sample has a population that is greater than ~40,000. A simple linear regression of the data showed that there is not a strong relationship between population and the average cost of medical procedures at hospitals within the corresponding zip code (r = 0.01, p = 0.92). When looking at the p-values, we fail to reject our null hypothesis.

* 291 - Heart Failure & Shock w/ MCC did not show a relationship between population and average cost (r = 0.04, p = 0.62)
* 392 - Esophagitis, Gastroent & Misc Digest Disorders w/o MCC did not show a relationship between population and average cost (r = -0.01 p = 0.91)
* 470 - Major Joint Replacement or Reattachment of Lower Extremity w/o MCC did not show a relationship between population and average cost (r = 0.0, p = 0.99)
* 871 - Septicemia or Severe Sepsis w/o MV >96 Hours w/ MCC did not show a relationship between population and average cost (r = -0.03, p = 0.75)
* 872 - Septicemia or Severe Sepsis w/o MV >96 Hours w/o MCC did not show a relationship between population and average cost (r = 0.5, p = 0.56)

**Figure 4. Population in California by Zip Code**

****

**Figure 5. Population vs. Overall Hospital Costs**

**A screenshot of a cell phone

Description automatically generated**

**Figure 6. Population vs. Cost of Procedure 291**

**A close up of a map

Description automatically generated**

**Figure 7. Population vs. Cost of Procedure 392**

**A screenshot of a cell phone

Description automatically generated**

**Figure 8. Population vs. Cost of Procedure 470**

**A close up of a map

Description automatically generated**

**Figure 9. Population vs. Cost of Procedure 871A screenshot of a cell phone

Description automatically generated**

**Figure 10. Population vs. Cost of Procedure 872**

**A close up of a map

Description automatically generated**

**Hypothesis 2**

*If there is a relationship between cost of healthcare and income per capita, there will be higher healthcare costs in areas with higher income per capita.*

A simple linear regression of the data showed that there is not a strong relationship between per capita income and the average cost of medical procedures at hospitals within the corresponding zip code (r = 0.11, p = 0.16). When looking at the p-values, we fail to reject our null hypothesis.

* 291 - Heart Failure & Shock w/ MCC did not show a relationship between per capita income and average cost (r = 0.04, p = 0.62)
* 392 - Esophagitis, Gastroent & Misc Digest Disorders w/o MCC did not show a relationship between per capita income and average cost (r = -0.05 p = 0.54)
* 470 - Major Joint Replacement or Reattachment of Lower Extremity w/o MCC did not show a relationship between per capita income and average cost (r = 0.11, p = 0.17)
* 871 - Septicemia or Severe Sepsis w/o MV >96 Hours w/ MCC did not show a relationship between per capita income and average cost (r = 0.24, p = 0.0)
* 872 - Septicemia or Severe Sepsis w/o MV >96 Hours w/o MCC did not show a relationship between per capita income and average cost (r = 0.0, p = 0.97)

**Figure 11. Per Capita Income vs. Overall Hospital Costs**

**A screenshot of a cell phone

Description automatically generated**

**Figure 12. Per Capita Income vs. Cost of Procedure 291**

**A close up of a map

Description automatically generated**

**Figure 13. Per Capita Income vs. Cost of Procedure 392**

**A close up of a map

Description automatically generated**

**Figure 14. Per Capita Income vs. Cost of Procedure 470**

**A close up of a map

Description automatically generated**

**Figure 15. Per Capita Income vs. Cost of Procedure 871**

**A close up of a map

Description automatically generated**

**Figure 16. Per Capita Income vs. Cost of Procedure 872**

**A close up of a map

Description automatically generated**

**Hypothesis 3**

*If there is a relationship between cost of healthcare and median age, there will be higher healthcare costs in areas with higher median age.*

A simple linear regression of the data showed that there is not a strong relationship between median age of the residents living in a zip code and the average cost of medical procedures at hospitals in that zip code (r = -0.07, p = 0.4). The linear regression showed, for each of the following procedures:

* 291 - Heart Failure & Shock w/ MCC did not show a relationship between median age and average cost (r = -0.06, p = 0.49)
* 392 - Esophagitis, Gastroent & Misc Digest Disorders w/o MCC did not show a relationship between median age and average cost (r = -0.16, p = 0.04)
* 470 - Major Joint Replacement or Reattachment of Lower Extremity w/o MCC did not show a relationship between median age and average cost (r = -0.05, p = 0.53)
* 871 - Septicemia or Severe Sepsis w/o MV >96 Hours w/ MCC did not show a relationship between median age and average cost (r = -0.0, p = 0.98)
* 872 - Septicemia or Severe Sepsis w/o MV >96 Hours w/o MCC did not show a relationship between median age and average cost (r = -0.11, p = 0.16)

**Figure 17. Median Age vs. Average Cost of all Procedures**

A screenshot of a cell phone

Description automatically generated

**Figure 18. Median Age vs. Cost of Procedure 291**

A screenshot of a cell phone

Description automatically generated

**Figure 19. Median Age vs. Cost of Procedure 392**

A screenshot of a cell phone

Description automatically generated

**Figure 20. Median Age vs. Cost of Procedure 470**

A screenshot of a cell phone

Description automatically generated

**Figure 21. Median Age vs. Cost of Procedure 871**

A screenshot of a cell phone

Description automatically generated

**Figure 22. Median Age vs. Cost of Procedure 872**

A close up of a map

Description automatically generated

**Conclusion**

All in all, we did not have enough evidence to reject the null hypothesis. Nor did we see a correlation within the data parsed for each of the 5 most performed medical procedures in California. Through the data gathered, we have come to the conclusion that there is no correlation found between the demographic data (Mean Age, Population, and Per Capita Income) and the average cost of the top 5 medical procedures in California. Limitations we found were the fact that all records pulled related to medical costs were procedures that occurred with customers who have Medicare. As well as only using the top 5 medical procedures were investigated as a part of this report.