Lab_2_Submission_Jasmol_Dhesi_Suhas_Panthari.rmd

2024-07-21

Lab 2: Describing a Bivariate Relationship

Introduction

The United States healthcare system, notorious for its exorbitant costs, consistently ranks as the most expensive globally. To curb these soaring costs while enhancing the quality of care, home health services have become increasingly prevalent. The rising interest in home health services has spurred significant research into their impact on healthcare outcomes. This study investigates the following question:

Do home health services affect the frequency of emergency room visits for Medicare beneficiaries?

By investigating whether home health services can affect the number of inpatient-covered stays, the research addresses a crucial question with implications for cost containment and quality of care. This inquiry is relevant for healthcare providers, policymakers, and patients alike, as it seeks to optimize healthcare resources and improve patient outcomes. By employing a detailed Ordinary Least Squares (OLS) regression analysis, the study aims to provide robust insights into this relationship using a t-test on regression coefficients with $\alpha = 0.05$.

Data

To analyze the relationship between home health care and inpatient-covered stays, we sought the most comprehensive and stratified dataset available. We identified the Chronic Conditions Data Warehouse from the Centers for Medicare and Medicaid Services (CMS) "Fee-for-Service Geographic Variation Public Use File," as suitable data to conduct this research study.

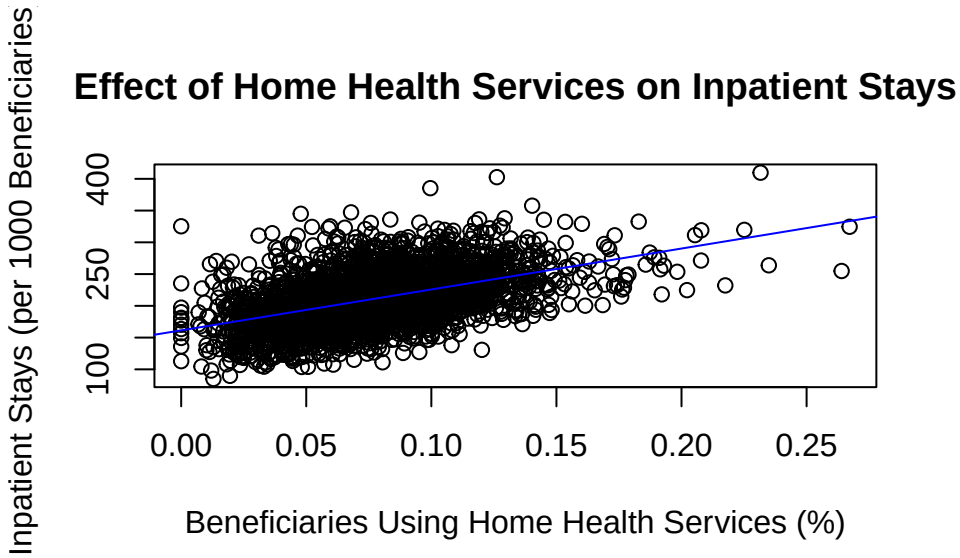
Within Medicare and Medicaid beneficiaries, a subset of beneficiaries are enrolled in the "Fee-for-Service" (FFS) program. This dataset includes many parameters across several regions, capturing data on FFS program users. Each observation corresponds to a specific county or state, or is labeled "national".

In examining the CMS dataset, we concentrated on the most recent **county** data from the year 2022 (3198 observations). Our first step involved meticulously cleaning the data by removing null values (183 observations) in the fields related to home health services (BENES_HH_PCT) and inpatient stays covered by Medicare beneficiaries (IP_CVRD_STAYS_PER_1000_BENES). These parameters were of primary interest for our study, as they are crucial for analyzing the relationship between home health services utilization and the frequency of emergency room visits.

Having acquired the necessary data and identified the critical parameters, we were ready to test the linear regression model against our null hypothesis:

Null Hypothesis: *"Home health services usage by Medicare beneficiaries does not affect the inpatient stays covered by Medicare."*

The results of our regression can be seen in the visualization below:



The visualization above shows the data points centered around the regression line, indicating that the regression model accurately captures the relationship between the variables.

Model Evaluation

The Ordinary Least Squares (OLS) method was used to develop our regression linear model. The OLS approach proved effective, as demonstrated by its fit to the data, with the best-fit line accurately aligning with the scatter plot. This level-level model provided a clear initial evaluation of the relationship between the variables.

To further optimize the model, we incorporated a transformation by introducing the age parameter. This adjustment was necessary to account for the declining health levels in the older population, who are known to require more medical attention. Including age as a variable enhances the model's accuracy by acknowledging the increased healthcare needs of older individuals, thereby refining the relationship between home health services and emergency room visits. Additionally, we restricted the upper age limit to 72 to prevent outliers from skewing the data, ensuring a more robust and reliable analysis.

% Table created by stargazer v.5.2.3 by Marek Hlavac, Social Policy Institute. E-mail: marek.hlavac at gmail.com % Date and time: Tue, Jul 23, 2024 - 1:10:16 AM

Despite its effectiveness, the OLS model is contingent upon the satisfaction of key assumptions. These assumptions are:

1. I.I.D Observations: The data must be independent and identically distributed.
2. Existence of a Unique Best Linear Predictor (BLP): The model must not exhibit issues such as heavy tails or perfect collinearity.

The I.I.D. assumption may be challenged if low home health service rates in one county influence neighboring counties' healthcare systems, potentially leading to geographic clustering. Nevertheless, significant differences among adjacent counties suggest that observations can often be considered independent.

To validate the second assumption, our data must be free from heavy tails and perfect collinearity. Our analysis indicates that the data does not exhibit heavy tails, which supports the existence of a BLP. This finding is visually supported by the Inpatient Stays versus Home Health Beneficiaries graph. Additionally, to address potential collinearity issues, we restricted our analysis to county-level data. By excluding state and national data, we mitigate concerns of perfect collinearity, ensuring that the value of one observation does not directly linearly affect another.

Table 1:

	Dependent variable:
	Inpatient Stays per 1000 Beneficiaries
Home-Health Beneficiary Percentage	640.130***
	(22.951)
Age Below 72 (Indicator)	10.247**
	(4.362)
HH-Beneficiary Percent : Age Below 72 Indicator	3.584
	(51.421)
Constant	159.487***
	(1.922)
Observations	3,015
R^2	0.254
Adjusted R^2	0.254
Residual Std. Error	35.976 (df = 3011)
F Statistic	342.413*** (df = 3; 3011)
Note:	*p<0.1; **p<0.05; ***p<0.01

Results and Interpretation

```
##
## t test of coefficients:
##
##                                Estimate Std. Error t value Pr(>|t|)
## (Intercept)                    159.4871     1.9225 82.9596  < 2e-16 ***
## BENES_HH_PCT                   640.1300    22.9515 27.8906  < 2e-16 ***
## AGE_INDICATORTRUE               10.2472     4.3624  2.3490  0.01889 *
## BENES_HH_PCT:AGE_INDICATORTRUE   3.5835    51.4211  0.0697  0.94445
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

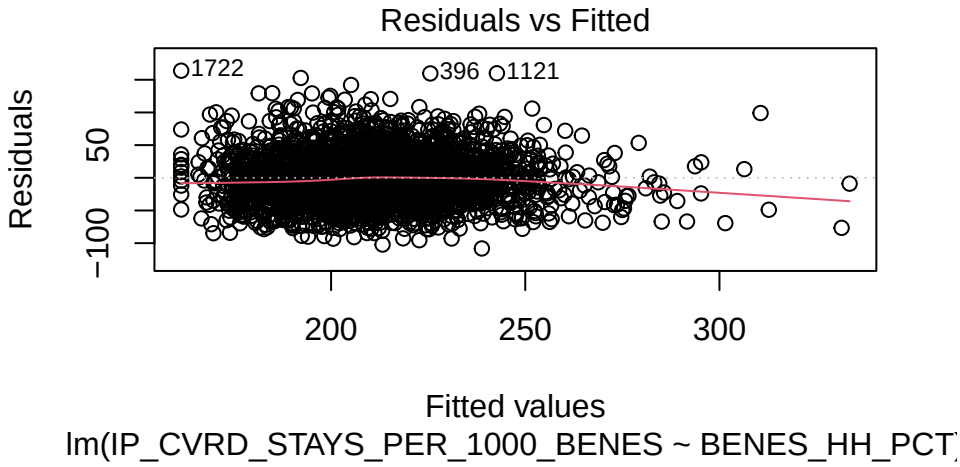
The model's intercept term, the coefficient for 'BENES_HH_PCT', and the indicator for age 'AGE_INDICATORTRUE' are all statistically significant with p<0.05. The interaction term between the percentage of home health beneficiaries and age indicator 'BENES_HH_PCT:AGE_INDICATORTRUE', however, is not statistically significant. These results reject our null hypothesis as home health services usage is found to affect inpatient stays.

The model intercept term and coefficient for ('BENES_HH_PCT'), are practically significant due to their large scale in the regression model. A 1% increase in the percentage of home health beneficiaries is associated with an increase of 6.40 inpatient stays per 1,000 beneficiaries. The small positive effect of age indicates that older populations do require more inpatient care, but this is not a substantial moderating factor in the home health services relationship.

Given these findings, healthcare policymakers should consider these dynamics when designing and allocating resources for home health services. Future research might explore additional variables or different geographic levels to further understand these relationships.

Appendix

Link to data set: Medicare Geographic Variation - by National, State & County | CMS Data



Other attempts to optimize the model included log and polynomial transformations, which did not result in any appreciable increase in accuracy. Consequently, we used an indicator variable representing whether the average beneficiary age is less than 72. This approach would distinguish between age groups while maintaining the model's integrity.