Determinants of Outpatient Doctor Visits: Analyzing the Impact of Demographic, Health, and Socioeconomic Factors

Stanislau Papko

2024-10-23

1 Introduction

In this project, I will be working with the DoctorContacts data set available in the Ecdat package. The original source is "The Structure of Demand for Medical Care: Latent Class versus Two-Part Models" (2002) by Deb, P. and P.K. Trivedi from *Journal of Health Economics*. It is a cross-section from 1977–1978. The DoctorContacts data set provides information on the number of outpatient visits individuals make to a medical doctor over a year. It also contains some demographic and socioeconomic factors that may influence utilisation of health services, such as age, sex, income, education. This data set consists of 20186 observations and following 15 variables:

mdu (double) - number of outpatient visits to a medical doctor;

lc (double) - log(coinsrate+1) where coinsurance rate is 0 to 100;

idp (logical) - individual deductible plan?;

lpi (double) - log(annual participation incentive payment) or 0 if no payment;

fmde (double) - log(max(medical deductible expenditure)) if IDP=1 and MDE>1 or 0 otherwise;

physlim (logical) - physical limitation?;

ndisease (double) - number of chronic diseases;

health (factor) - self–rate health (excellent,good,fair,poor);

linc (double) - log of annual family income (in $);

lfam (double) - log of family size;

educdec (double) - years of schooling of household head;

age (double) - exact age;

sex (factor) - sex (male,female);

child (logical) - age less than 18?;

black (logical) - is household head black?

Research Question I have formulated as follows: "How do demographic, health and socioeconomic factors, such as age, sex, family income, health status, presence of a physical limitation and the number of chronic diseases, affect the number of outpatient visits to a medical doctor?"

This question will help explore which factors have a significant impact on number of outpatient visits to a medical doctor and which of them lead to higher or lower healthcare utilization. To do this, I will focus on the mdu variable as the primary dependent variable.

For one of the plots I have to modify the dependent variable. I need to categorize it. To do this, I use the code below and create new categorical variable mdu_category.

```
breaks_mdu <- c(0, 1, seq(6, 11, by = 5), Inf)
labels_mdu <- c("0", "1-5", "6-10", "11+")
DoctorContacts$mdu_category <- cut(DoctorContacts$mdu,
                                   breaks = breaks_mdu,
                                   include.lowest = TRUE,
                                   right = FALSE,
                                   labels = labels_mdu)
```

In the future, I will also need a new categorical variable derived from the linc variable, which will divide the observed individuals into those with low, mid or high income. To do this, I create the variable income_category using the code below.

```
quantiles <- quantile(DoctorContacts$linc, probs = c(0, 1/3, 2/3, 1),
                      na.rm = TRUE)
DoctorContacts$income_category <- cut(DoctorContacts$linc,
                                      breaks = quantiles,
                                      labels = c("Low income",
                                                 "Mid income",
                                                 "High income"),
                                      include.lowest = TRUE,
                                      right = FALSE)
```

2 Analysis

Since the main response variable is mdu,the number of outpatient visits to a medical doctor, it is proper to begin our analysis by showing the distribution of the values of this variable. For this purpose I will use the histogram. It is important to note that the binwidth was chosen to be 1 in order to show all observed values on the plot. As you will notice below some of them are very important (e.g. 0), which was definitely something to consider. The resulting histogram can be found in Figure 1.

As we can see below the mdu variable can take quite a wide range of values, outpatient visits to a medical doctor, from 0 to almost 80. However, only a few are quite common, those taking values less than 10. So for example we see that individuals, who haven't got outpatient visits to a medical doctor over a year, are more than 6000 out of 20186 observations and those who've only been to the doctor once are almost 4,000. People who've been to the doctor more than 10 times over a year are the exception to the rule.

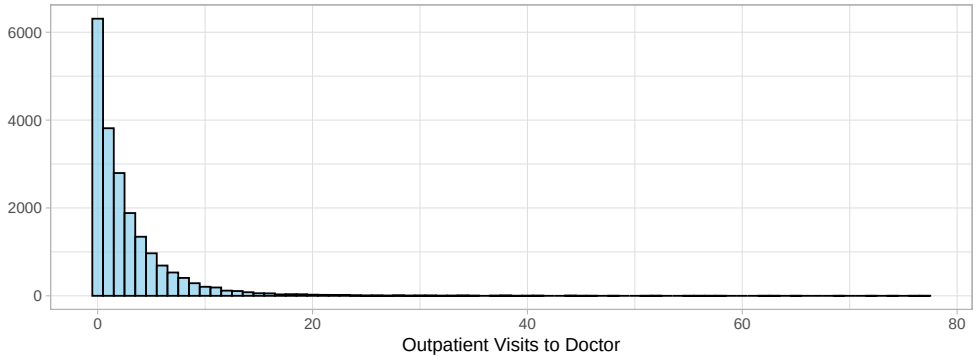

Figure 1: Distribution of outpatient doctor visits

Next, we look at how the number of outpatient visits to a medical doctor changes with increasing age of individuals and whether this number depends on the sex of individuals. Here I will show the explained variable mdu on the y-axis, the variable age on the x-axis and for the variable sex I will make 2 facets for males and females. As you can see I am dealing with a large number of observations, so I have to deal with the overlap problem. For this purpose I decided to use hex geom. I also decided to use the smooth curve with LOESS (locally estimated scatterplot smoothing) method to demonstrate the overall trend. You can see the result in Figure 2.

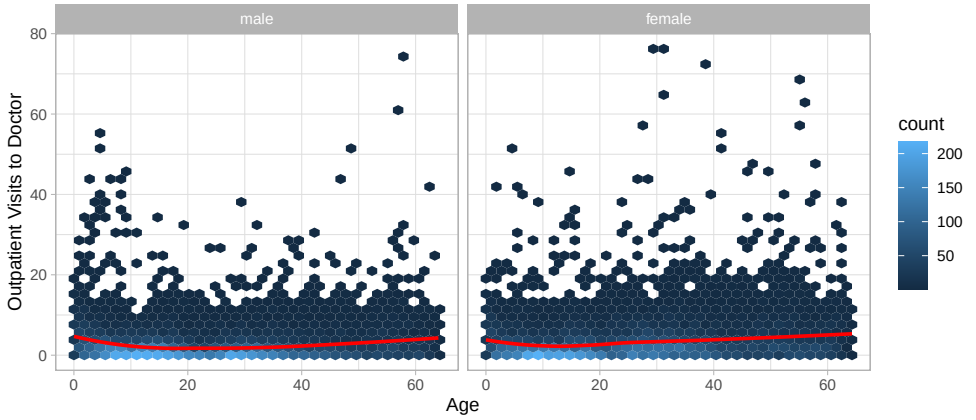

Figure 2: Number of outpatient doctor visits by age and sex

First of all we see that there is no significant difference between men and women, as clearly evidenced by the similar smooth curves for both genders. We also see that the relationship between the age of individuals and the number of outpatient visits to the doctor is small. The average number of doctor visits varies slightly with increasing age of individuals and is on a fairly narrow range, roughly between 2 and 5 visits. However, some interesting things can be observed. For example, we see that the average number of visits is slightly lower in the case of teenagers than in the case of very young children and adults. The lighter lower part of the graph is characteristic of just the age of teenagers, which may mean that teenagers can often have no outpatient visits to the doctor at all during the year. Small differences between males and females are seen after adolescence. So we see that after the age of 20, women's doctor visits

begin to increase smoothly with increasing age. For men, however, this growth begins much later, after the age of 30. These differences may be due, for example, to women taking a more responsible approach to their health or to potential pregnancies that may be the reason for preventive visits. However, as I noticed earlier, these patterns are rather insignificant. I would also like to point out that the above would be better seen if I changed the scale of the y-axis, but I have abandoned this idea to not reduce the interpretability of the plot.

Now we move on to something that is theoretically directly related to doctor visits - health. I start by introducing the correlation between self–rate health and number of outpatient visits to a medical doctor. Here I decided to use boxplot. Note that I used transparency to avoid overlap when presenting outliers on the plot. The result is shown in Figure 3.

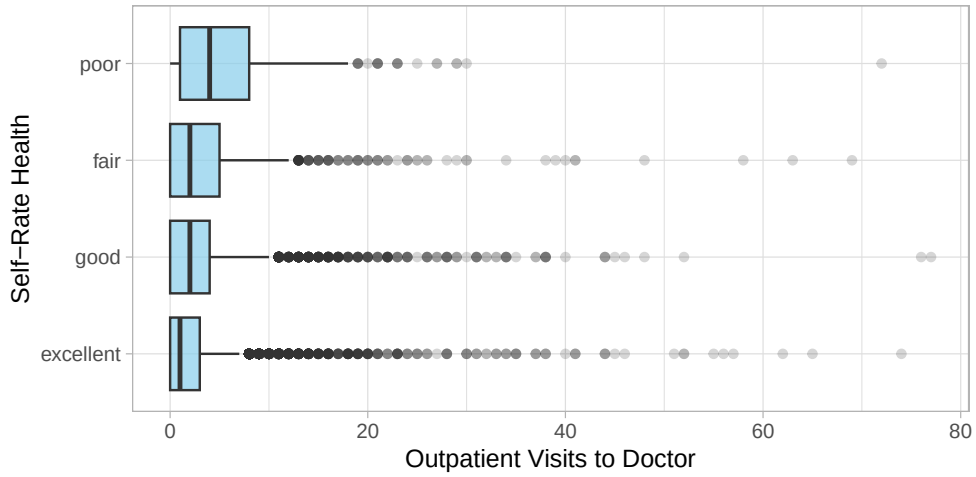

Figure 3: Number of outpatient doctor visits by self-rated health status

From the plot above, I conclude that the worse is self–rate health, the greater the number of outpatient visits to the doctor. This is well illustrated by the increase in the corresponding medians on the boxplot. So the median for "Excellent" rate is about 1, while the median for "Poor" rate is about 4. I also note the large variability within categories and the big number of outliers, shown in the graph by the dots. For example, there was an individual who rated his health as "Excellent", but had been to the doctor more than 70 times. Surprisingly, I note only few extreme outliers for the "Poor" category and the lowest number of outliers overall. I speculate that this may be due to the smallest representation of this group.

We're staying on the topic of health status. This time we'll look how having physical limitations and the number of chronic diseases affect the number of outpatient visits to a medical doctor. Now I will use the scatter plot. On the y-axis I place the variable mdu, the variable ndisease on the x-axis and for the variable physlim I will make 2 facets. Once again, I have to face the problem of overlap. This time I'll use a few counteracting methods: I'll again use transparency, reduce the size of the dots on the plot and use the jitter position adjustment. I'll also use a smooth curve with LOESS method to see the general trend. The result is shown in Figure 4.

In the plot, we can see that both explanatory variables affect the explained variable. Firstly, we see that as the number of chronic diseases increases, the number of visits

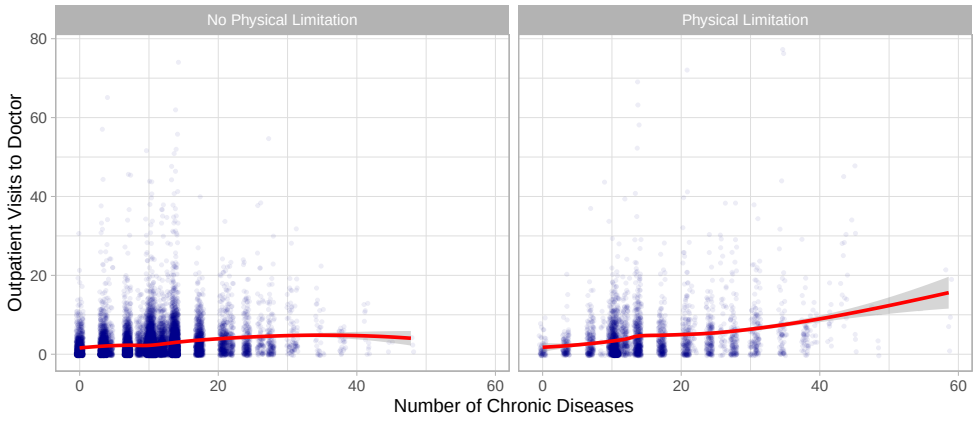

Figure 4: Number of outpatient doctor visits by number of chronic diseases and physical limitations

to the doctor also increases. We also see that this growth is not the same for people with and without physical limitations. We see that in the absence of chronic diseases, the number of doctor visits is close to 0, about 1-2. As chronic diseases increase and in the absence of physical limitation, we have seen a smooth and small increase in the number of doctor visits. So with an extreme number of chronic diseases, the average number of visits is approximately 5. This increase is more dramatic in the presence of a physical limitation. In this case, with an extreme number of chronic diseases, the number of visits on average exceeds the value of 10. Additionally, I also note the relationship between the explanatory variables. The graph shows that in the presence of a physical limitation, the number of chronic diseases is generally higher than in the absence of such a limitation.

Finally, I will use the variables I created at the beginning. Now I want to show the correlation between family income and number of outpatient visits to a medical doctor. In the plot, both of these variables will be represented by the new categorical variables income_category and mdu_category. This time I decided to use a mosaic plot, which was chosen for its compactness and interpretability. The resulting plot can be found in Figure 5.

The plot demonstrates 4 categories, which show the number of outpatient doctor visits. However, we can see that if we look at the difference between the categories, their number can be reduced to two: 0 and 1+. So we see that people from low-income families are the most common among people who have not been to a doctor. People from high-income families, on the contrary, are the least common. Among individuals, who have visited a doctor at least once, the three income categories are roughly equally prevalent. I also note that as family income increases, the number of visits to the doctor also increases, but for the last 3 categories this increase is hardly noticeable. Additionally, the mosaic plot demonstrates the relative sizes of the categories by outpatient doctor visits. So the largest category is 1-5 visits. I also note the large size of the category where individuals did not visit a doctor, despite the fact that it takes only one value. The situation is opposite in the case of the last category, "11+". It is on the contrary small, despite the large number of accepted values.

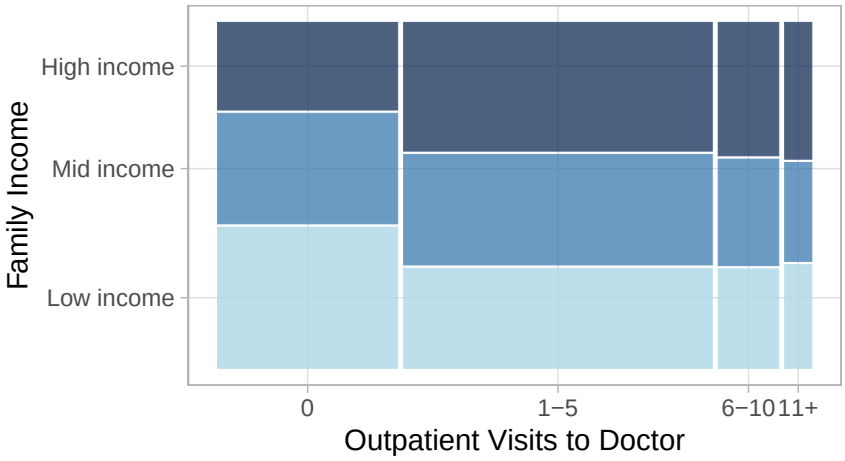

Figure 5: Relationship between family income and outpatient doctor visits

3 Conclusion

The analysis shows that demographic, health and socioeconomic factors all play a role in determining the number of outpatient visits to a medical doctor, though their influence varies in significance. The most influential factors, which affect outpatient visits are self-rate health status, physical limitation and chronic diseases. Individuals with poorer health or with more chronic diseases tend to visit doctors more frequently, especially if they also have physical limitation.

Family income also affects the number of outpatient visits to a medical doctor. Low-income families are more likely to avoid visiting a doctor. Differences in the frequency of visits among income groups are less pronounced for individuals, who visit a doctor at least once.

Age and sex have only a minimal impact on the number of outpatient visits to a doctor.

Overall, the data suggests that while socioeconomic and demographic factors matter, health-related variables have a more significant influence on the frequency of doctor visits.

4 References

Croissant, Y. (2023). Ecdat: Data Sets for Econometrics. R package version 0.4-4. Available at: https://cran.r-project.org/web/packages/Ecdat/index.html

Deb, P., & Trivedi, P. K. (2002). The structure of demand for medical care: Latent class versus two-part models. Journal of Health Economics, 21(4), 601–625.