An Apple a Day Keeps the Doctor Away

Analyzing Annual Visits to the Physician

STAT 476/776 | Spring 2020

I. Background

Motivating Question

In the 1990s, the German government began a campaign to reform its healthcare system, which led to the first-ever national health survey in Germany conducted in 1998. One particular question the government investigated was the rate at which adults visited their doctors each year. The legislators wanted to ensure that the healthcare system matched the needs of the population. It was important to know whether the population consistently visited the doctor or whether there were specific groups that tended to visit the doctor more often. It was not affordable to cover all visits; yet, the system needed to finance any typical medical expenses. Therefore, we want to better understand and explain the underlying latent behaviors of individuals in regard to their propensity toward doctor visitations.

Data

This paper explores the number of times German adults visit the doctor in a given year, using results from the 1998 German Health Survey. The dataset includes a population sample of 1,127 adults and their answers to the following questions:

1. number of visits to the doctor in 1998 [*numvisit*]
2. healthy or not healthy [*badh*] (self-reported)
3. age [*age*]

It is important to note that the sampling method in this survey is unclear and therefore raises uncertainty in regard to our ability to extend the following insights from this dataset to the population as a whole.

The visit counts range from 0 (no check-ups) to 40. Since the survey results consist of each respondent's answers, this dataset provides us the opportunity to analyze the disaggregated individual behavior. However, in order to visualize the dataset in a histogram and conduct chi square tests, we group the counts in bins from 0 to 10+, which we will use consistently throughout this paper. The actual raw data can be seen below.

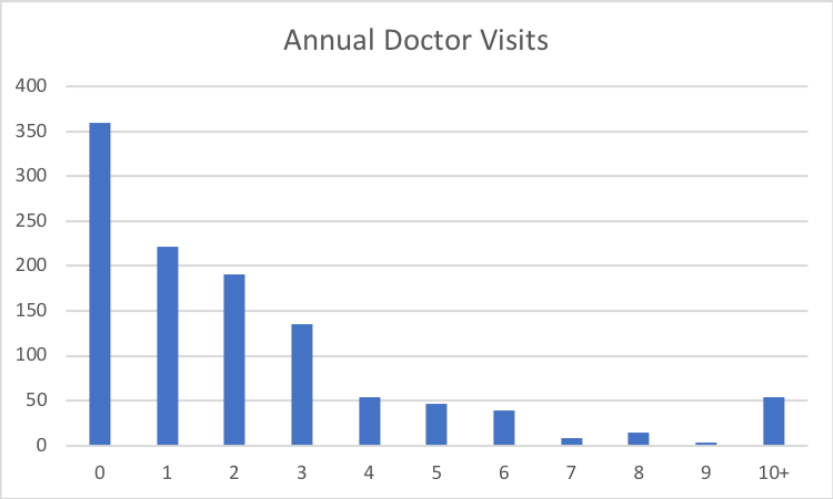


Figure 1: Number of doctor visits per year - raw survey data

The histogram exhibits that there is a large number of 0 visits and small-digit count observations, as well as a little bump at the right-censored bin.

Expectations

In this paper, we will be investigating the propensity of individuals to visit the doctor. Within the overall population, we expect slight heterogeneity since people have different health concerns to attend to or are more cautious when it comes to their health. However, we also know there is a common societal tendency to visit the doctor a standard amount of times per year (barring any emergencies), steering the distribution toward homogeneity. For example, in Germany, medical professionals and insurance companies suggest that an average healthy person visit their doctor 1-2 times a year, depending on the individual's age.

Within the healthy and young population, we assume that individuals are more homogenous since there are not many reasons for people to have an extreme tendency to or not to visit the doctor. However, within the healthy and old population, it is more likely for individuals to be heterogeneous since some people would be more careful about their health as they age. For the unhealthy population, it would make sense for individuals to be homogenous since each person is consistently and constantly monitoring his or her health.

II. Basic NBD Model

We first fit the basic NBD model to the entire dataset of annual doctor visits using the disaggregated observations, calculating the probability of occurrence for each individual respondent and optimizing the Maximum Likelihood Estimation (MLE). We present the resulting histogram and model parameters below.

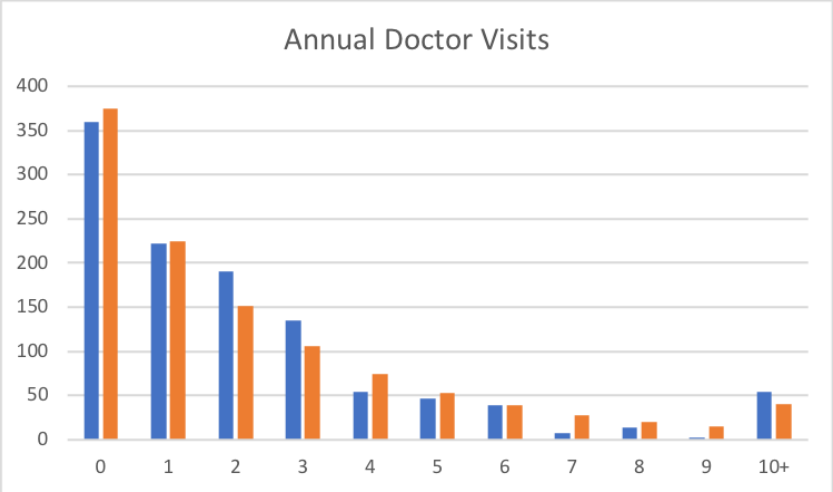


Figure 2: Actual vs. Expected (from the NBD model) of annual doctor visits

r	***alpha***	***LL***	χ^2	***p-value***
0.806	0.343	-2297.113	55.825	3.052E-09

The *r* parameter for this model (0.806) is less than but close to 1, implying that there is slight heterogeneity within the overall population. This result makes sense since there is a large proportion of individuals who visited the doctor 0-3 times and then another bump at the right-censored bin. Additionally, the model's parameters are consistent with our original expectation that there would be slight heterogeneity to account for individuals who need more medical attention or who are more cautious about their health as well as those who tend to avoid the doctor.

However, using aggregated binned data, we conduct a chi square test and obtain a very low p-value, which implies that the basic NBD model performs poorly. This is consistent with a visual inspection of

the histogram above, which shows a bad fit of the data. Taking a closer look, we see that there are too many 2 and 3 observations than expected, as well as too few larger-digit observations on the right tail.

Although the chi square test and visual inspection reveal that the basic NBD model fits poorly, we can check the robustness of the NBD model by comparing the results from the disaggregated MLE approach to both a basic NBD using aggregated data and a NBD model using the "Means and Zeroes" method.

Model Method	***r***	***alpha***
Disaggregated MLE	0.806	0.343
Aggregated MLE	0.874	0.384
Means and Zeroes	0.873	0.371

We see model parameters above are relatively consistent; the slight difference can be explained by the loss of granularity when rolling up the data. We can now be more confident that a NBD model will fit the dataset.

III. Effect of "Doctor's Orders" — Adding a Spike

As mentioned above, healthcare professionals in Germany suggest that individuals visit the doctor biannually. Based on this information, we hypothesize that a portion of the population are adults that religiously follow the professional advice (barring any medical emergency). If so, we should see in the data a group of "hard-core 2-time visitors". We therefore add a spike at 2 to the original NBD model to account for those people. The results are summarized below.

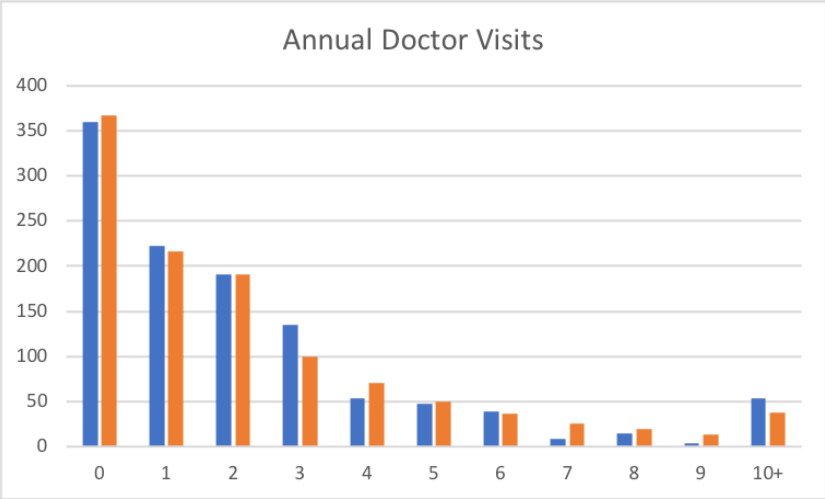


Figure 3: Actual vs. Expected (from the NBD with spike at 2) of annual doctor visits

r	*alpha*	*Spike at 2*	*LL*	χ^2	*p-value*
0.791	0.344	0.042	-2151.199	46.556	6.815E-08

The *r* parameter for this model (0.791) is less than 1 and smaller than the *r* from the basic NBD model, illustrating that there is slightly more heterogeneity within the overall population. We also see that about 4% of the population visits the doctor consistently twice a year. The small spike parameter, however, implies that either the "hard-core 2-time visitors" group is negligible or our original hypothesis about the existence of the "2-time" tendency may be false.

Again, we compute the chi square test, which yields a low p-value. The visual fit of the histogram, however, has improved slightly from the original model. We therefore conclude once more that the dataset as a whole does not fit this basic NBD model, even with a spike.

IV. Differences Among Individuals with Good vs. Bad Health

After fitting the NBD model for the entire dataset using two approaches and still not obtained good models, we want to look at differences in propensity toward visiting the doctor between individuals with good health and those with bad health. We hypothesized above that there will be more people with bad health who are particularly cautious and tend to visit the doctor more often. We therefore split the data

based on respondents' self-reported health status (0 = good; 1 = bad) to investigate this claim. These two NBD models are summarized below.

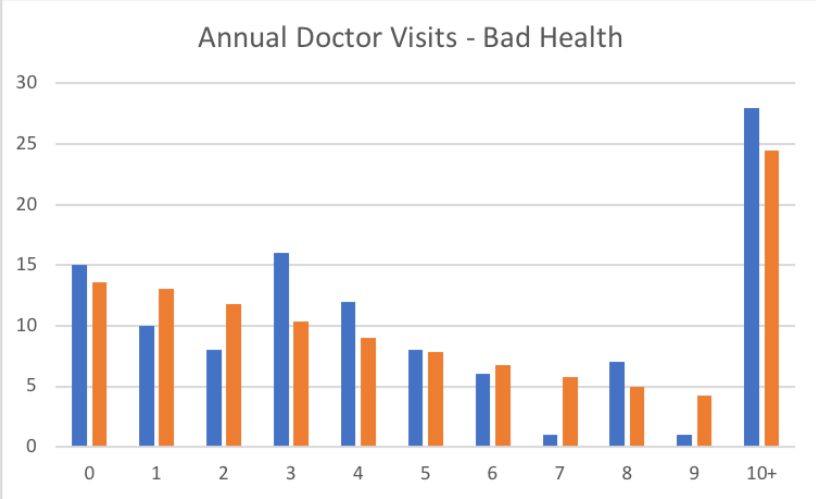


Figure 4: Actual vs. Expected (from the NBD) of annual doctor visits of individuals with bad health

r	***alpha***	***Spike at 2***	***LL***	χ^2	***p-value***
1.136	0.186	0.042	-323.219	13.976	0.082

First, we look at the segment of individuals who self-reported that they had bad health. The *r* parameter for this model (1.136) is above 1, implying that there is homogeneity within this segment of the population. This result makes sense since once someone has bad health, their tendency to or not to visit the doctor would not be particularly extreme; all the people in the segment have to be generally more cautious about their health. The proportionally large amount of 10+ visit counts (compared to the "healthy" model below) illustrates that the propensity to visit the doctor in this population segment is higher than that of a healthy adult. The *r* parameter in this model is also a lot higher than our result from the models on the entire dataset. We could therefore hypothesize that the segment of healthy individuals must have more heterogeneity (which we will investigate below).

This time, the chi square test gives us positive results. Given the p-value of 0.082, we can say that the model's fit is significant. While an inspection of the histogram, on the other hand, could lead to a different conclusion due to the seemingly poor fit, it is important to note that the graph is slightly skewed since the dataset is smaller and each additional difference between expected and actual is therefore magnified.

We then look at the segment of individuals who self-reported that they were in good health.

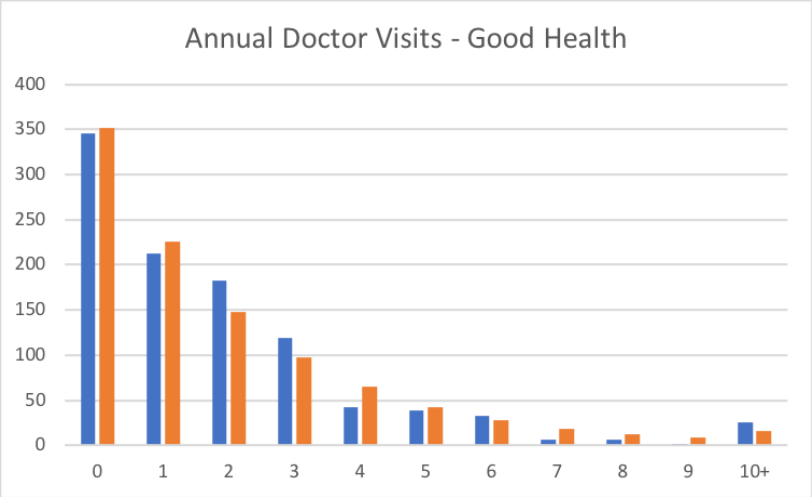


Figure 5: Actual vs. Expected (from the NBD) of annual doctor visits of individuals with good health

r	***alpha***	***LL***	χ^2	***p-value***
0.963	0.497	-1912.184	42.898	9.183E-07

The *r* parameter for this model (0.963) is slightly below 1, demonstrating that there is some heterogeneity within this segment of the population. This result is consistent with our original assumptions that certain healthy individuals could have different tendencies in regard to visiting the doctor based on personal factors while the average healthy person will just go for normal check-ups. Proportionally there are a lot less right-skewed observations among the healthy segment, which makes sense from a logical standpoint that healthy individuals do not need to visit the doctor as often. Interestingly though, the *r* parameter in this model is also higher than our result from the models on the entire dataset. A possible conclusion could be that the splitting of the segments separated a lot of the observations on the extremes (0 visits vs. 10+ visits) and therefore the individual models no longer need to explain as wide of heterogeneous behavior.

However, we compute the chi square test and yield a low p-value. The fit of the histogram also is not great visually. We therefore conclude that the healthy segment does not fit this NBD model. While we

could try to fit a spike at 2 according to our aforementioned hypothesis, we will explore another factor instead that could better explain the different behaviors within the healthy segment.

To conclude this section, we want to compare this model to the basic NBD to find out whether the behaviors of the segments are in fact different. We can first make the observation heuristically by looking at the difference between means across the segments, as well as the entire dataset.

	Mean
Full Dataset	2.35
Segment: Good Health	1.94
Segment: Bad Health	6.11

From the means, we can see a clear difference between the tendencies of visiting the doctor of individuals in the healthy segment and the unhealthy one. To statistically prove the difference between the models, we conduct a Likelihood Ratio Test to reveal whether we can reject the null hypothesis that the models are the same.

LL Large	***LL Small***	***Difference***	**df**	***p-value***
-2297.113	-2235.403	123.419	3	1.416E-26

Since the p-value is extremely low, we can conclude that the two NBD models are different, and therefore the good health and bad health segments do not follow the same distribution. This conclusion is consistent with our hypotheses above about the latent differences among individuals based on their health statuses.

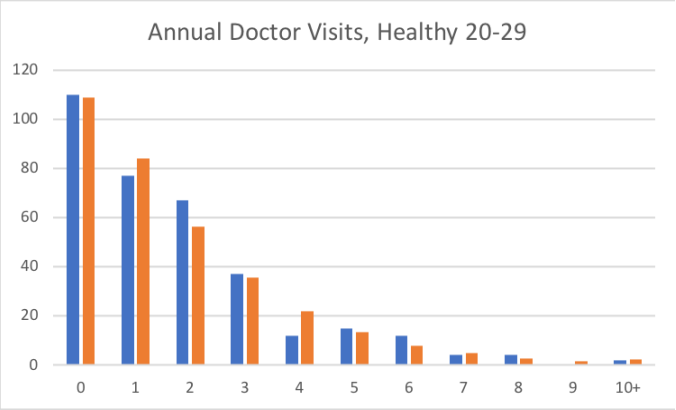
V. Comparing Behaviors of Different Age Groups within the Healthy Population Segment

Now that we know there is a difference between the good health and bad health segments, we want to improve the model for the healthy population by digging into other factors that could explain their latent behaviors. As noted in our expectations in the beginning, a healthy young person would not have much

reason to visit the doctor other than for a check-up; however, once a person gets older, even if they are healthy, they could start to be more cautious and more vigilant in visiting the doctor. To see if age explains some of the propensity, we split the healthy segment into 4 age groups: 20-29, 30-39, 40-49, and 50-60 and fit four separate NBD models.

Younger (ages 20-29 and 30-39)

The histograms and resulting parameters are shown below.



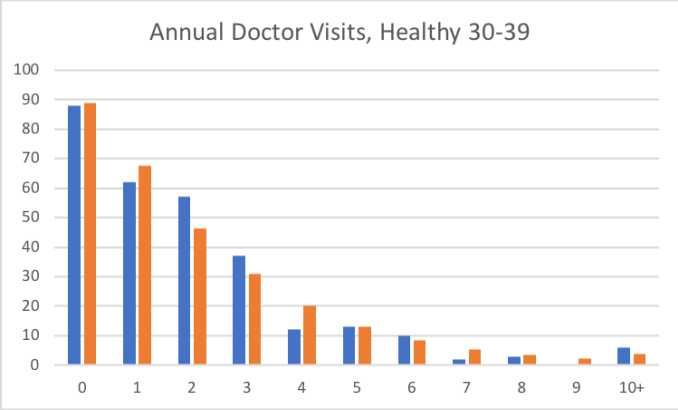


Figure 6 & 7: Actual vs. Expected (from the NBD) of annual doctor visits of individuals with good health, ages 20-29 and 30-39

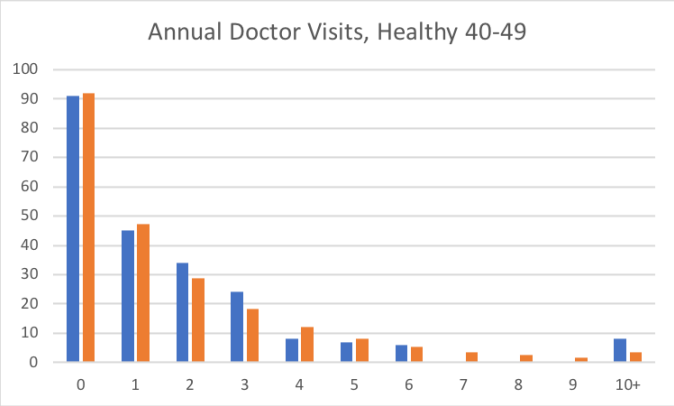
Segment	***r***	***alpha***	***LL***	χ^2	***p-value***
20-29	1.363	0.767	-614.760	11.783	0.161
30-39	1.228	0.618	-551.173	13.455	0.097

The *r* parameters for these models (1.363 and 1.228 respectively) are above 1, implying that there is homogeneity within the young, healthy population. This result makes sense since there is a large proportion of individuals who visited the doctor 0-3 times and few observations on the tail. Additionally, the model's parameters are consistent with our original expectation that there would be homogeneity since young and healthy individuals do not have much of a tendency, and even more so, any extreme tendencies, to visit the doctor.

The chi square tests give us positive results. Given the high p-values, we can say that the models' fits are significant. Again, although by inspection the histogram may look unaligned, the smaller sample sizes magnify even the smallest differences.

Older (ages 40-49 and 50-60)

The histograms and resulting parameters are shown below.



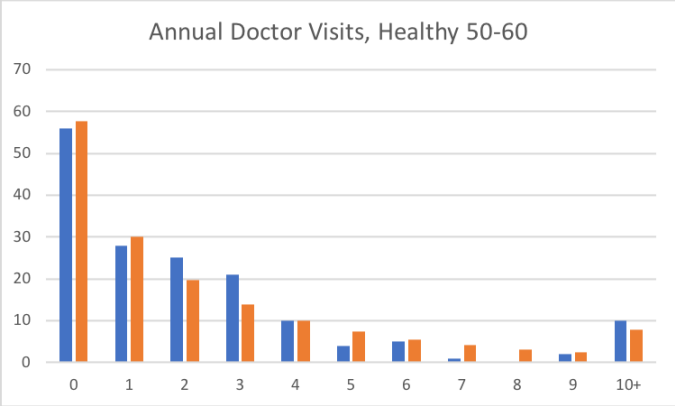


Figure 8 & 9: Actual vs. Expected (from the NBD) of annual doctor visits of individuals with good health, ages 40-49 and 50-60

Segment	***r***	***alpha***	***LL***	χ^2	***p-value***
40-49	0.744	0.426	-397.353	17.357	0.027
50-60	0.661	0.266	-335.253	12.975	0.113

The *r* parameters for these models (0.744 and 0.661 respectively) are below 1, implying that there is heterogeneity within the older, healthy population. This result also makes sense since there is a large proportion of individuals who visited the doctor 0-3 times as well as a bump in observations on the tail. Additionally, the model's parameters are consistent with our original expectation that there would be heterogeneity since as people get older, some tend to be more prudent about their health.

The chi square tests also give us positive results. Given the high p-values, we can say that the models' fits are significant. By inspection, the histogram looks well aligned relative to the small dataset and short axes.

Model Comparison

In order to know whether the age segmentation does a better job than a monolith healthy population subset, we again conduct a Likelihood Ratio Test to reveal whether the models have statistically different distributions. We see the results below.

LL Large	*LL Small*	*Difference*	**df**	*p-value*
-1912.184	-1898.539	27.290	6	.000128

Since the p-value is extremely low, we can conclude that the two NBD models are different, and therefore the breakdown of ages is a better model at explaining the behaviors and propensities of healthy individuals than the larger single model.

VI. Conclusion

Using the NBD models, we find that it is difficult to explain the propensity for doctor visits of the overall population. Instead, we are able to break down the population into more comprehensible segmentations: good health and bad health. We conclud that individuals with bad health tend to be more homogenous in their propensity to visit the doctor, presumably since the entire segment is dealing with medical issues and are not visiting the doctor on their "own accord". On the other hand, the healthy segment could be split further by age, in which we see that younger individuals are more homogenous while older individuals are more heterogeneous. With this information, the German government should write its healthcare policies based on age groups since people's tendencies differ across those segments. (They should not base the legislation on people with bad health since that is a small fraction of the population according to this survey).

Moving forward, the government should compare results across other sample populations as well as multiple different years in order to conclude whether this data can serve as a basis for their models. Additionally, they should find a better way at collecting health status data (instead of self-reporting) in order to verify that it is accurate.