Preparing for the Upcoming Influenza Season: Interim Report

Project overview

Motivation: The United States has a time every year where more people than normal are getting sick from the influenza. Certain populations in certain areas can develop severe symptoms and need to get treatment in urgent care clinics and in the worst case, hospitals. For this time of the year, more staff are needed to treat patients in clinics and hospitals (which are provided by the medical staffing agency on a temporary basis), but other interventions, such as influenza shots, can reduce the load on clinics and hospitals as well.

Objective: Determine where and how many additional medical staff to send to various clinics and hospitals in states around the country.

Scope: Since influenza is a national issue, the agency can cover hospitals in all 50 states. This project will help plan for the upcoming influenza season nationwide using historical data.

Hypothesis

If an individual is less than 5 years old and older than 65 years old, then they are more likely to die from influenza.

Data overview

1. **Influenza deaths**: monthly data collected between 2009-2017 by the Center for Disease Control (CDC) listing the number of influenza deaths by US state and 10-year age groups (except a <1 year and 1-4 year old group).
2. **Census data**: yearly estimates of the total, male, female, and 5-year age increment populations between 2009-2017 collected by the U.S Census Bureau in every county in each state.

Data limitations

1. **Influenza deaths**: Influenza death data are compiled by a US government organization, so the quality of the data is likely to be reliable and mostly free of any bias, despite the collection happening across multiple institutions in every state. These data are compiled by month, which is still coarse for our purposes, but will be useful for historical analysis and monthly updates during the intervention. Deaths are only recorded with a corresponding death certificate which could lower the accuracy of the data for those who did not seek medical treatment and died somewhere other than a hospital or clinic. Unlike the census data, our influenza deaths dataset is only summarized by state and we therefore do not have the finer scale county information. Fewer than 10 deaths are suppressed in the final dataset and make up a majority of records so comparison of different geographical locations can be difficult.
2. **Census**: Census data are collected by a US government organization, so the quality of the data is likely to be reliable and mostly free of any bias, especially since censuses are conducted to represent all individuals in a population. The data, however, are finalized as one number (by population demographic) every year so there is a significant time gap between new records. This means the census will not be very useful for estimating populations in real time during the influenza medical personnel intervention, but can be used for historical analysis. Population records are conducted and entered manually, so there could be errors while collecting or entering the data. Other characteristics that may affect influenza vulnerability such as race, education level, average annual income, etc. are not included in this dataset so this limits the accuracy of determining areas with higher percentages of vulnerable populations.

Descriptive analysis

	Variable 1	**Variable 2**
	Population <5,65+ years combination	**Deaths <5,65+ years combination**
Mean	1223782	998.0108932
Standard deviation	1335106.233	974.8327333
Linear model slope	0.000690364	
Linear model intercept	153.155621	
Correlation coefficient	0.945505314	

The correlation between the population and influenza deaths of the combined young (<5 years old) and old (65+ years old) age categories (proposed vulnerable population) was a highly correlated positive relationship. This suggests that as the vulnerable population increases, so does the number of influenza deaths.

Results and Insights

To test the research hypothesis that extremely young and older people are more likely to die from influenza, we can use our statistical hypotheses based on mortality rates:

The null hypothesis/H_0: the rate of influenza deaths of young/old people are not higher than middle age groups.

The alternative hypothesis/H_a: the rate of influenza deaths of young/old people are higher than middle age groups.

Summary stat	Young/old age groups	Middle age groups
Mean	0.001060504	0.00018958
Variance	2.4666E-07	3.5907E-08
Observations	459	459
Hypothesized Mean Difference	0	
df	589	
t Stat	35.10153851	
alpha	0.05	
P(T<=t) one-tail	8.50E-147	

We have several indications to reject the null hypothesis (means of the variables are equal) in favor of the alternative hypothesis (mean of the young/old group is higher than the middle age groups). Based on the t-test statistics, we see that the t-statistic is very high and the accompanying p-value is very small. This shows that there is close to a 0% chance that the results happened by chance/the variables are very likely statistically significantly different.

Next steps

Now that we have established that very young children and old adults are at a higher risk of dying from influenza, we need to determine where the highest percentage of those populations are located to assign risk categories to different states. We can do this with bar charts in Excel or another kind of visual programming software such as Tableau. This process will help us determine where to focus resources for medical personnel assignments. To narrow down when flu season generally occurs, we can use the influenza deaths dataset and plot the number of deaths over time or based on monthly averages for all

years in the time series. We can then concentrate on sending resources in these months or as need be during the intervention.

The final report will be included in a PowerPoint presentation or Tableau storyboard presented to the medical professional, staffing agency, and patient stakeholders. It will contain historical analysis of influenza deaths by population demographics, geographical location, and time of the year using data visualizations and statistical analysis. Using insights from this historical analysis, I will present a plan of/make recommendations for allocating staffing agency resources based on state risk categories at times of year when influenza deaths are at their worst.