Interim Report: Effective Medical Staff Deployment for Influenza Season

Project Overview

Motivation: A count of the historical influenza deaths gives an indication of the severity of flu in an area. Deaths can be prevented with flu shots and adequate medical staff. In the United States, each state has a different population composition, meaning that some states will have more vulnerable populations.

Objective: Examine influenza trends across the United States and provide recommendations to the medical staffing agency on how to best allocate their resources (staff) in each state.

Scope: The agency covers all hospitals in each of the 50 states of the United States, and the project will plan for the upcoming influenza season.

Research Hypothesis

If the state has a large proportion of their population 65 years and older (vulnerable population subset)[1], then the influenza rates and subsequent mortality rates are likely to be higher.

Data Overview

US Census Data: Population data from 2009 to 2017 collected by the US Census Bureau on an annual basis. The dataset contains the county and respective state, the year the data was collected in, the population of that respective year, biological sex information, and age group categories in five-year increments up until 85 where it is just "85 years and over."

Influenza Deaths Data: Influenza-related deaths data from 2009 to 2017 collected by the CDC monthly. This dataset contains the state, month, year, age group categories in ten-year increments, and their respective mortality rates.

Data Limitations

US Census Data: The dataset is dependent on a variety of methods for its collection, so these are as accurate as possible, but are still an estimate of reality. The dataset should not be bias as it is simply a general demographic breakdown; however, the collection method may be considered biased, which in turn make the data itself subsequently bias. The Bureau is dependent on people responding, which not only causes a lag, but also manual errors.

Influenza Deaths Data: Those 65 years and older may have a myriad of health concerns that might lead to death; however, on their death certificate, only one cause of death may be listed, and influenza may not be selected if other (more serious) concerns were present (e.g., diabetes, AIDS, old age, etc.). While influenza may have been a contributing or aggravating factor in these deaths, if it is not listed, the data cannot be captured appropriately. Additionally, the dataset suppressed any death count of 10 or less, leading to five (5) as a stand-in number for the calculations.

Descriptive Analysis

- The average or mean population for those 65 years and older is 806,989, with a standard deviation of 887,017.
  - 11% were outliers, meaning at least 68% of values are within one standard deviation of the mean, therefore is considered a normal distribution.[2]
- The average influenza deaths for those 65 years and older is 897, with a standard deviation of 972.
  - 13% were outliers, meaning at least 68% of values are within one standard deviation of the mean, therefore is considered a normal distribution.
- In statistical analysis, the closer a correlation coefficient is to 1, the stronger the relationship of the two variables is. In this case, the correlation coefficient is 0.9, making the relationship between older age and influenza mortality rates strong.

Comparing Population & Influenza Death Rates: 65+ Year Old vs < 65-Year-Old						
	65+ Population	65+ Deaths	Mortality %	< 65 Population	< 65 Deaths	Mortality %
Minimum	47,809	180	**0.38%**	417,527	480	0.11%
Maximum	5,078,704	5,694	**0.11%**	33,617,095	1339	0.00%
Mean (Average)	806,989	897	**0.11%**	5,167,038	537	0.01%
Std Deviation	887,017	972	**0.11%**	5,944,643	119	0.00%

*** There is significantly less people 65+ years and older in the United States, but they have over ten times the mortality rate than those younger than 65 years old.**

	Variable 1	Variable 2
Dataset Name	US Census Data, > 65 years old	Influenza Deaths Data, > 65 years old
Sample or Population?	Sample	Sample
Normal Distribution?	Yes	Yes
Variance	7.9E+11	944,307
Standard Deviation (1 Std Dev)	887,017	972
Standard Deviation (2 Std Dev)	1,774,034	1944
Mean	806,989	897
Outlier Percentage (1 Std Dev)	11%	13%
Outlier Percentage (2 Std Dev)	6%	4%
Upper Limit (1 Std Dev)	1,694,006	1,869
Upper Limit (2 Std Dev)	2,581,023	2,840
Lower Limit (1 Std Dev)	-80,028	-75
Lower Limit (2 Std Dev)	-967,045	-1,047
Total Outliers (Upper Limit, 1 Std Dev)	49	58
Total Outliers (Upper Limit, 2 Std Dev)	29	18
Total Outliers (Lower Limit, 1 Std Dev)	N/A	N/A
Total Outliers (Lower Limit, 2 Std Dev)	N/A	N/A

*** Since there can't be a negative population or death values, these lower limit numbers are irrelevant/unusable.**

Results and Insights

Research Hypothesis: If the resident is 65+ years old, then they are at a higher risk for influenza death.
Null Hypothesis: The mortality rate of people over 65 years old is greater than or equal to people under 65 years old.
Alternative Hypothesis: The mortality rate of people over 65 years old is greater than people under 65 years old.

A one-tailed test was conducted to examine the relationship between age and influenza mortality rates.

- There is a positive relationship[3] between the two values observed: as age increases, so does the risk for influenza deaths.
- The p-value[4] is 5.8871E-175 – well below the 0.05 threshold to reject the null hypothesis.
  - With 95% confidence, it can be concluded that people over 65 years old are at a higher risk for influenza death.

Remaining Analysis and Next Steps

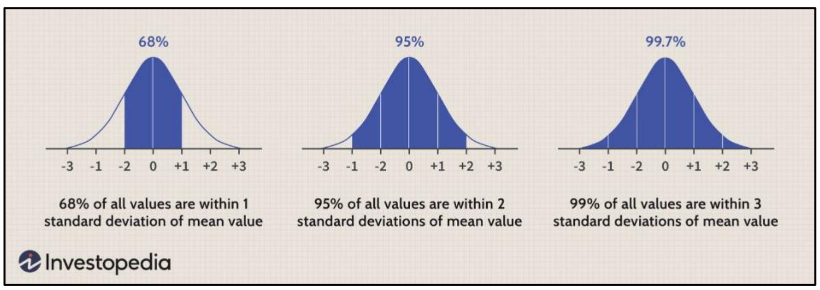
- Review which states have the highest population of 65+ years residents *and* which have the highest death rate, then recommend more medical staff be deployed to those states.
  - It is important to consider both scenarios (high pop. of 65+ residents and high 65+ death rate) to best allocate our limited resources. A state can have a high pop. of 65+ residents but may have a low death rate due to already having enough medical staff and resources.
  - Effort should be focused on states that would benefit the most (e.g., reduced death rate) from the deployment.
- Review the current medical staff availability for each state and include it in analysis moving forward.
- Additional deliverables:
  - Tableau storyboard with temporal, statistical, and spatial visualizations
  - Video presentation with visualizations, conclusions, and recommendations

Appendix

t-Test: Two-Sample Assuming Unequal Variances		
	% of Death by Pop: < 65 Years	% of Death by Pop: 65+ Years
	0.000755062	*0.003765004*
Mean	0.000268142	0.001311181
Variance	7.57225E-08	2.61871E-07
Observations	458	458
Hypothesized Mean Difference	0	
df	701	
t Stat	-38.41816448	
P(T<=t) one-tail	5.8871E-175	
t Critical one-tail	1.647030228	
P(T<=t) two-tail	1.1774E-174	
t Critical two-tail	1.96335386	

[1] Vulnerable Population Subset: The CDC identified multiple groups as being part of the population most vulnerable to influenza deaths, including people 65 years and older, children 5 years old and under, pregnant women, people with compromised immune systems (e.g., autoimmune diseases, cancer, chronic health conditions, etc.), and certain racial and ethnic groups. To provide a more concentrated approach to a solution, this project opted to focus on one subset of the multitude of vulnerable populations to not spread resources too thin. (Source: CDC)

[2] *"For all normal distributions, 68.2% of the observations will appear within plus or minus one standard deviation of the mean; 95.4% of the observations will fall within +/- two standard deviations; and 99.7% within +/- three standard deviations. This fact is sometimes referred to as the "empirical rule," a heuristic that describes where most of the data in a normal distribution will appear."* See below graph for more detail. (Source: Investopedia)



[3] *"Negative correlation is sometimes described as inverse correlation. In statistics, positive correlation describes the relationship between two variables that change together, while an inverse correlation describes the relationship between two variables which change in opposing directions."* In other words, in a positive relationship, when one value increases, so does the other; in a negative relationship, when one value increase, the other decreases. (Source: Investopedia)

[4] *"A p-value is the probability of the data-generating mechanism corresponding to a specified null hypothesis to produce an outcome as extreme or more extreme than the one observed."* See below graph for more detail. (Source: GIGA Calculator)