Hillel Krief 15-03-2022

Project Overview- Preparing for Influenza Season

Project Overview

Motivation

The United States has an influenza season where more people than usual suffer from the flu. Some people, particularly those in vulnerable populations, develop serious complications and end up in the hospital. Hospitals and clinics need additional staff to adequately treat these extra patients. The medical staffing agency provides this temporary staff.

Objective

Determine when to send staff, and how many, to 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 a state has a high population of ages 65+, then a higher rate of influenza deaths will occur.

Data Overview

Census Population

This is an external data source for the United States population sorted by county, year, total population, male population, female population, and population by age group. This data is owned by the US Census Bureau which is a government organization and is therefore trustworthy.

CDC Influenza Deaths

This is an external data source for deaths caused sorted by influenza by gender and age group. This is administrative data collected by the CDC from places like: hospital, Public Health providers, labs, and other healthcare offices. This data is manually collected. Data is updated once per week.

Data Limitations

Census Population

This data is from 2017, and therefore is not up to date with today's population. Especially since our analysis focuses on vulnerable populations, census data can change pretty dramatically.

CDC Influenza Deaths

There is a lot of incomplete data, as many of the death numbers are suppressed because of data privacy limitations. This does not invalidate the data, however, it is important to take into account the missing data.

Descriptive Analysis

The standard deviation of vulnerable aged deaths by state and year is **1015**. Similarly, the standard deviation of population by state and year is **1327686**. This shows us the magnitude of geographical implications on our project, that we must take location into account.

The correlation between vulnerable deaths and vulnerable parts of different populations is **22%** which shows a very weak relationship.

Data Spread

	Variable One	Variable Two
Dataset Name	Death Data	Population Data
Variable	Vulnerable Deaths	Vulnerable Ages
Sample or Population	Sample	Sample
Normal Distribution		
Variance	1029237.011	1.76275E+12
Standard Deviation	1014.513189	1327686.584
Mean	828.0917031	1195657.809
Outliar Percentage	4%	
One SD Above	1843	2523344.393
Two SD Above	2858	3851030.977
One SD Below	187	0
Correlation		
Outlier Count Higher	18	0
Outlier Count Lower	0	0

Correlations

Variables	Total Flu Deaths	Vulnerable Population
	22%	

Results and Insights

The results of a one tailed T- test between the Bottom and Top states that contained the most vulnerable deaths and the grand total deaths in each state is the following:

Null Hypothesis- *A state with a high population of citizens ages 65+ will be less than or equal to a state with a low population of citizens 65+*

Alternative Hypothesis*- A state with a high population of citizens ages 65+ will have a higher death rate than a state with a low population of citizens 65+*

	Bottom Vuln Deaths	*Top Vuln Deaths*
Mean	8.76715E-05	0.000140089
Variance	4.07756E-09	2.32962E-09
Observations	10	10
Hypothesized Mean Difference	0	
df	17	
t Stat	-2.070819371	
P(T<=t) one-tail	0.026956705	

Hypothesis testing showed that our null hypothesis can be disproved which shows that there must be a relationship between the vulnerable populations and death rate.

Next Steps

As described in my descriptive analysis, we have a weak correlation between vulnerable deaths and vulnerable parts of populations, which contradicts our inferential statistical calculations that did show strong relationship. Our next step would be to dive into hospitalization numbers due to influenza or do to some more testing to confirm our hypothesis to be true. Just because we have a weak correlation doesn't mean there is no relationship.