2024

Preparing for Influenza Season

Surabhi Seth

Interim Report

5/8/2024

Contents

Project Overview

Goal:

To help a medical staffing agency that provides temporary workers to clinics and hospitals on an as-needed basis. The analysis will help plan for influenza season, a time when additional staff are in high demand. The final results will examine trends in influenza and how they can be used to proactively plan for staffing needs across the country.

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:

The objective of the project is to determine when to send staff, and how many, to each state. This process needs to be considered in terms of efficiency and overall attempt to lower the death rate due to Influenza and hence save as many human lives as possible.

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.

Stakeholder Identification

The below groups are the key stakeholders in the project:

- Medical agency frontline staff (nurses, physician assistants, and doctors)
- Hospitals and clinics using the staffing agency's services
- Influenza patients
- Staffing agency administrators

Hypothesis

Upon analysis of available data, we can observe that the influenza death vulnerability increases with age and beyond 75 years, it increases sharply. Hence, vulnerable population of 75 years or more are more prone to death due to influenza.

Data Overview

The following data sets covering influenza in the United States will be used during the project:

1. Influenza deaths by geography

Data Sourcing: This is an external data source. The data provided by the various healthcare providers and is provided to CDC. Hence the data can be considered from reliable source.

Data limitations: As the data is collected by different medical providers and testing agencies, the data is not exhaustive as it does not account for tests done at other agencies. The data might have an element of bias, depending upon the testing and payment for each sample at the agencies.

2. Population data by geography, time, age, and gender

Data Sourcing: This is an external data source. The data provided by the US Census Bureau, is a government data, we can classify this as a trustworthy data source.

Data limitations: As the data collected by different agencies and through surveys, manual errors could be possible. If the initial census shows higher male female ratio, a projection bias might adjust it in the same ratio, in spite of actual female population change.

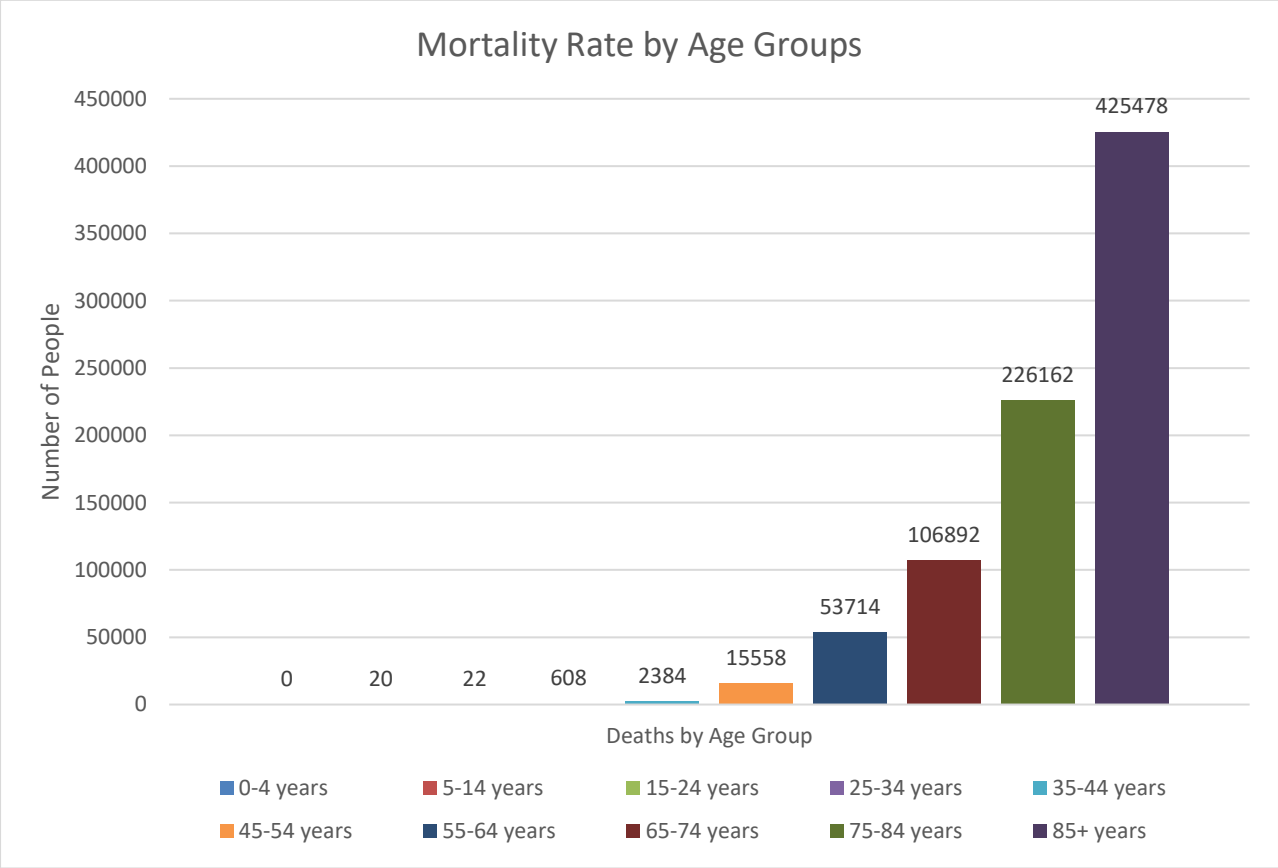
3. Integrated Data Set

The influenza Mortality and US population data sets were combined using the concatenated variable State-Year as a key id [Combined key]. This was further combined into age groups for population and mortality rates.

The age groups were divided into 2 groups: High Risk (75+ Years of Age), Low Risk (Under 75 Years of Age)

Mortality rate trends across Age groups for US

In analysing the number of deaths due to Influenza, as the age groups progress from 0-4 year to 85+ years, the mortality rate increases significantly. The below chart provides the perspective on the deaths by age groups, which is skewed on the higher age groups.



Descriptive Analysis

The research hypothesis proposes that a higher population in the 75+ age group will result in an overall higher number of influenza deaths in the same age group. A descriptive analysis was conducted, which helped to determine and improve upon data quality and to understand the correlation between various variables. I determined the Mean, Standard deviation and correlation coefficient for population and mortality rates of 75years and above to arrive at my hypothesis validations and eventual conclusions.

	Mortality 75+	Population 75+	Total Population 75+
Standard Deviation	853	17546	313640
Mean	710	5990	97842
Correlation Coefficient	0.94		
Strength of Correlation	Strong Correlation		

The correlation coefficient strongly supports the hypothesis that individuals aged 75 and above are more prone to severe influenza complications and a higher risk of death. The other factors that could potentially further impact be like vaccination rates, access to healthcare, general health, and various demographic and environmental elements.

Results and Insights

To obtain further insights from the sample data sets, inferential statistics were used to translate the research hypothesis into conclusions about the population as a whole.

❖ Null Hypothesis: The influenza death rate of patients 75 years or older is less than or equal to the influenza death rate of patients younger than 75 years old.

❖ Alternative Hypothesis: The influenza death rate of patients 75 years or older is greater than the influenza death rate of patients younger than 75 years old

T Test: Two-Sample Assuming Unequal Variances	75+ rate	upto 75 rate
	0.002092107	0.00024294
Mean	0.001606089	0.000162503
Variance	6.32765E-07	2.53541E-08
Observations	458	458
Hypothesized Mean Difference	0	
df	493	
t Stat	38.08230469	
P(T<=t) one-tail	3.029E-149	
t Critical one-tail	1.647950228	
P(T<=t) two-tail	6.058E-149	
t Critical two-tail	1.964787445	

Remaining Analysis and Next Steps

After confirming that high risk populations (75+ years of age) do experience higher influenza deaths, our analysis will begin to look for states where populations older than 75 years of age is higher. The key next steps suggested are:

a. Analyse the Influenza death rate trends by States and years (2009 -2017)
b. Analyse the population density of 75+ (65+ if the sample size is low) across states and their variability from 2009 – 2017
c. Estimation of the population density change across states (county level for detailed planning) for the upcoming Influenza period.
d. Staffing predictions and logistics based on the historical trends of patients handled by each worker, capacity of each hospital, tests and vaccinations etc.

These analytics shall then be put into more visualizations with several possible trends and outcome driven analysis.