PREPARING FOR INFLUENZA SEASON:

An Interim Report

Renu Balaji / 14.02.2024

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 some states have a higher overall population age (>55 years) (vulnerable population), then they tend to have higher influenza-related mortality rates.

DATA OVERVIEW

a) Population Data (Source: US Census Bureau)

This is a census data collected for the entire United States which gives information about the total population, gender-wise, state-wise and age-wise. The variables in this data set include State, Year, Total population, Male population, Female population and population from <5 years till >85 years.

b) Influenza deaths (Source: Centre for Disease Control)

This is the data collected by CDC where there is information provided for the number of deaths due to influenza in a particular year in different states. The data also highlights the different age groups. The variables in this data-set include State, Year, State code, Month, Month code, age groups and death count.

DATA LIMITATIONS

a) Population Data (Source: US Census Bureau)

The main limitation of this data is the ***time lag*** due to the annual collection of data. There is also a possibility of ***manual errors*** in cases where the data has been entered manually. The data is collected in regular intervals (annual) by the Census Bureau and hence should be bias-free.

b) Influenza deaths (Source: Centre for Disease Control)

Manual errors by patients while entering their data may occur. The data is collected and updated frequently by CDC so there should not be any biases.

DESCRIPTIVE ANALYSIS

	Standard Deviation	Variance	Mean	Outlier count	Outlier Percentage
Death count (>55 years)	730.83	534112.71	468.92	18	1%
Population (>55 years)	534201.74	2.9E+11	386851.728	84	5%

Correlation		
Variable	Death count from >55 years and the sum of population >55 years	Total population (all age groups) vs the total death count (all age groups)
Proposed Relationship	As the population age increases, the death count also increases	As the total population increases, the death count increases
Correlation Coefficient	**0.893**	**0.954**
Strength of Correlation	**Strong**	**Strong**
Usefulness / Interpretation	In different states, as the age of the population increases, more and more people fall into a higher risk of death due to influenza. This perfectly aligns with the proposed hypothesis.	As the population number increases, the death count also increases.

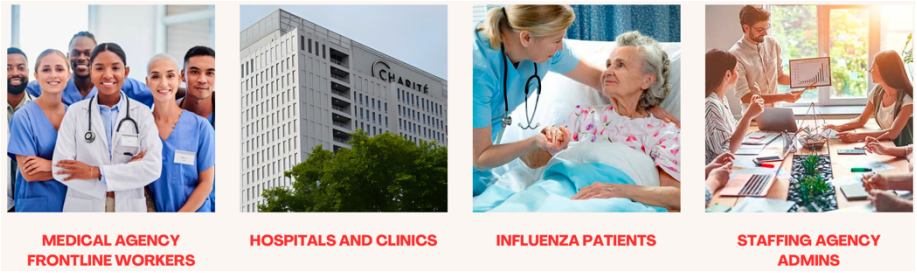
RESULT AND ANALYSIS

Research Hypothesis	If some states have a higher overall population age (>55 years), then they tend to have higher influenza-related mortality rates
Independent Variable	Population of age group >55 years
Dependent Variable	Death count for age group >55 years
Null Hypothesis (H0)	The death ratio of population lesser than or equal to 54 years of age is higher.
Alternative Hypothesis (Ha)	The death ratio of population greater than 55 years of age is higher.
Type of test	One tailed test (one direction since we want to know the death count of the population greater than 55 years of age)
Significance (α)	**0.05**
p-value	The **p-value is 1.4927E-158** which is extremely low when compared to the significance (α=0.05) hence the null hypothesis has been rejected. Therefore, the states with the death ratio of age group of 54 years and below is lower than the death ratio of states having age group of greater than 55 years. The confidence level of this result is 95%.

REMAINING ANALYSIS AND NEXT STEPS

- The resulting data can be presented to the stakeholders to indicate the vulnerable groups. The analysis rejected the null-hypothesis, therefore this information would also enable them to divide their staff members state-wise depending on which state has a total population with an age > 55 years. Those states with a population > 55 years will require more staffing facilities.
- Another analysis of the vaccination rate and it's effectiveness amongst the population with an age greater than 55 years can be studied. This can provide information if more vaccination campaigns must be brought in before the influenza season.
- The death count per month can be analysed to understand which months have the highest influenza-related mortality.

APPENDIX

<table>
<tr><td>Data overview</td><td><a>Population Data (Source: US Census Bureau)
Data set analysed from the year 2009-2017</td></tr>
<tr><td></td><td><a>Influenza deaths (Source: Centre for Disease Control)
Data set analysed from the year 2009-2017</td></tr>
<tr><td>Stakeholders</td><td> 
</td></tr>
<tr><td>Schedule and Milestones</td><td>
<table>
<tr><th></th><th>MILESTONES</th></tr>
<tr><td>Week 1-3 (Prior to flu season)</td><td>-Collect data + cleaning
-Design the project
-Kickstart with vaccination drives</td></tr>
<tr><td>Week 3-6 (before flu season)</td><td>-Data analysing
-Profiling data
-Allocating required staff for each hospital
-Analyse flu shots to hospitalization ratios</td></tr>
<tr><td>Week 6-9 (before flu season)</td><td>-Data transformation and final statistical analysis
-Testing the hypothesis
-Data visualisation</td></tr>
<tr><td>Week 9-12 (during and after flu season)</td><td>-Preparing the visualisations to present them to the non-technical audience
-Concluding the hypotheses tested</td></tr>
</table>
</td></tr>
</table>

Results and Insights

	Standard Deviation	Variance	Mean	Outlier count	Outlier Percentage
Death count (>55 years)	730.83	534112.71	468.92	18	1%
Population (>55 years)	534201.74	2.9E+11	386851.728	84	5%

Data Wishlist

- Data indicating influenza case-load across different states in different months of the year
- Data indicating influenza related hospitalizations and mortality rates in terms of population age across different states
- Data indicating vaccinations to influenza-related hospitalizations and across regions
- Data indicating influenza related hospitalization rate in terms of number of people in the vulnerable population