

PREPARING FOR INFLUENZA SEASON: INTERIM REPORT

March 5, 2024

1. 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.

2. Research Hypothesis

Hypothesis: *“If a citizen is 65 years of age or older, then they are more at risk of developing serious complications, including death, from the flu”.*

3. Data Overview

Influenza Deaths by Geography (CDC)

- This data shows the distribution of influenza-related deaths each month (Jan-2009 to Dec-2017) for all states in the US subdivided by age groups.

Population Data by Geography, Time, Age and Gender (US Census Bureau)

- This data shows population figures each year (2009-2017) for all US counties. Subdivisions are included for male and female as well as age groups.

4. Data Limitations

Influenza Deaths by Geography (CDC)

- **Missing data:** The death variable for many records was recorded as “suppressed”. From the data, it is clear “suppressed” applies to records where deaths were less than 10. This means that the analysis will slightly understate the impact of influenza.
- **Reporting:** The method of obtaining the cause of death from death certificates may obscure the impact of influenza. Influenza may not be listed as the cause of death but may have played a role in the mortality of some vulnerable individuals.

Population Data by Geography, Time, Age and Gender (US Census Bureau)

- **Estimations:** Although rigorous estimation methods were applied by the US Census Bureau, estimations may not be fully representative of the population.
- **Human error:** Whilst a high level of rigor can be assumed of this data due to its source, multiple collection methods and the need for manual entry present some likelihood of errors.

Overall Limitations

- **Timeframe:** Both data sets only cover from 2009 to 2017. Insights regarding trends and relationships should be looked at cautiously as they may have become outdated by the 2024 influenza season. For example, recent world events such as the Covid-19 pandemic caused well-known shifts in the public health sector and may have also caused significant shifts to population sizes across numerous states.
- **Indirect connection with project:** Neither dataset directly addresses hospital and clinic staffing levels or needs. Population and mortality data, whilst useful, does not account for patients who go to the hospital because of the flu and live. These patients also impact the workload of medical professionals.

5. Descriptive Analysis

Summary statistics for core variables are provided in the table below. The data covers each US state for each year from 2009 to 2017:

Variable Name	Mean	Standard Deviation	Outlier Percentage
Deaths (65+ years old)	826	1,014	3.92%
Deaths Total	905	1,154	4.14%
Population (65+ years old)	807,370	887,070	6.32%
Population Total	5,972,686	6,806,511	4.79%
MR (65+ years old)	0.0008%	0.0010%	0.44%
MR (<65 years old)	0.0847%	0.0435%	4.58%

Table 1: Summary statistics

A very strong correlation (0.94) was found between the 'Deaths' and 'Population' variables for the 65+ age group. This shows that states with a larger aged population (65 and older) will have a higher number of their aged population die from influenza.

Variables	Deaths (65+ years old) & Population (<65 years old)
Correlation Coefficient	0.94
Strength of Correlation	Strong

Table 2: Correlation Summary

6. Results and Insights

Research hypothesis:	If a citizen is 65 years of age or older, then they are more at risk of developing serious complications, including death, from the flu.
Independent variable:	The age of an individual (Age Groups)
Dependent variable:	Risk of developing serious complications or death (Mortality Rate)
Null hypothesis:	"The mortality rate of the 65 or older group will be less than or equal to the under 65 group" $H_0: \mu_{65 \text{ or older}} \leq \mu_{\text{under 65}}$
Alternate hypothesis:	"The mortality rate of the 65 or older group will be greater than the under 65 group" $H_A: \mu_{65 \text{ or older}} > \mu_{\text{under 65}}$
T-test:	One-tailed test
Significance level:	$\alpha = 0.05$
Results:	The mortality rate for the 65 and over population was higher (M = 0.0847%, SD = 0.0435%) than the mortality rate for the under 65 population (M = 0.0008%, SD = 0.0010%). At a p-value of 5.5605E-157, which is extremely close to 0, the t-test is significant (given the significance level = 0.05) and allows us to reject the null hypothesis.

Table 3: Statistical hypothesis testing summary

7. Remaining Analysis and Next Steps

Further analysis should be undertaken in the following areas:

- Comparisons should be made between the size of vulnerable populations in each state and hospital staffing levels to determine where shortages of medical professionals are likely to occur.
- A deeper temporal analysis of hospital visits or influenza mortality for each state throughout the months or weeks of the year is advised to identify the beginning and end of influenza season in each state. This will aid the development of staffing timetables.

After further analysis:

- Project deliverables will be created and selected for use in a final presentation. These include visualisations made in Tableau (bar chart, column chart, tree map, time forecast, histogram, bubble charts, scatter plots, spatial visualisation, word cloud)
- Final presentation for stakeholders where recommendations for where available frontline staff should be sent in preparation for flu season across the US.

8. Appendix

A) Project Brief

[Link to project brief](#)

B) Original Data Sets

Influenza deaths by geography (CDC) - [Download Data Set](#)

Population data by geography, time, age, and gender (US Census Bureau) - [Download Data Set](#)

C) Cleaned and Integrate Data Set

[Link to cleaned and integrated data set](#)