# Preparing for Influenza Season: Interim Report

**Felicia Andersen November 29,2023**

## 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.

## Hypothesis

If specific age groups demonstrate elevated influenza mortality rates, then an increased allocation of medical frontline staff support may be warranted based on geographic location.

## Data Overview

**CDC Influenza Mortality Data:** The data shows the geographic and monthly spread of influenza across the United States over multiple years and age groups.

**US Census Bureau Population Data:** The data shows US population counts over multiple years by location and age groups.

## Data Limitations

**CDC Influenza Mortality Data:** The primary limitation of this dataset is due to incompleteness, particularly when mortality counts fall below a total of 10. Instances where counts reach this threshold are denoted as “suppressed” based on privacy policy. This restriction does not allow for a comprehensive analysis across all age groups, hindering the identification of potentially vulnerable populations. This limitation diminishes the dataset in informing strategies for optimal medical staffing during the influenza season, as it may obscure insights into specific age groups that could benefit from targeted interventions.

**US Census Bureau Population Data:** The limitation of this dataset is that it primarily captures population counts, lacking direct correlations to influenza deaths. The reliability of the data is contingent on the accuracy of its collection and reporting methods, which could be caused by potential administrative errors.

## Descriptive Analysis

The following variables based on the datasets were reviewed. Conducting a descriptive analysis, helped to identify relationships between the variables and whether those relationships are meaningful.

| MORTALITY RATE PER 100,000 | AVERAGE | STANDARD DEVIATION |
|---|---|---|
| 65 years and younger | 1.47 | 0.92 |
| 65 years and older | 91.99 | 37.29 |

The correlation between mortality rates and the 65 and older age group is notably strong, with a coefficient of (0.96), compared to (0.51) for the 65 years and younger age group. This underscores the heightened vulnerability of the 65 and older age group to influenza related deaths. Recognizing this correlation is crucial for strategic planning in terms of allocating medical resources, considering factors such as when, where, and how many medical staff are needed, particularly in states where this age group faces a higher mortality rate.

## Results & Insights

To confirm the relationship between age and influenza mortality I formulated the following hypotheses:

**Null Hypothesis:** The mean influenza mortality rate of individuals younger than 65 and older than 65 years old is the same.

**Alternative Hypothesis:** The mean influenza mortality rate of individuals younger than 65 and older than 65 years old is NOT the same.

t-Test: Two-Sample Assuming Unequal Variances

| *Influenza Mortality Rate per 100,000* | *<65 years* *Total* | *>65 years* *Total* |
|---|---|---|
| Mean | 0.793585412 | 84.77915238 |
| Variance | 0.997978855 | 1894.598453 |
| Observations | 459 | 459 |
| Hypothesized Mean Difference | 0 | |
| df | 458 | |
| t Stat | -41.32739773 | |
| | | |
| P(T<=t) one-tail | 6.2532E-157 | |
| t Critical one-tail | 1.648187415 | |
| P(T<=t) two-tail | 1.25E-156 | |
| t Critical two-tail | 1.965157098 | |

A two-tailed t-test revealed that the p-value of 1.25E-156 is lower than alpha of 0.05, leading to the rejection of the null hypothesis. Consequently, the alternative hypothesis is supported.

The interpretation of this testing has concluded that with a 95% confidence level, the findings indicate individuals aged 65 years and older have a higher mortality rate compared to other age groups. This underscores the significance of age as a factor in influenza mortality.

## Remaining Analysis & Next Steps

**The following is what remains for analysis:**

- Composition and Comparison Charts
- Temporal Visualizations and Forecasting
- Statistical Visualizations: Histograms, Box Plots, Scatter Plots, and Bubble Charts
- Spatial Analysis
- Textual Analysis

**Next Steps:**

Completing the final deliverable, which includes a Tableau video presentation to stakeholders with key insights for medical staffing needs for the upcoming influenza season. This includes any feedback from stakeholders to ensure the final deliverable meets expectations and provides the necessary information for decision making.

## APPENDIX

**Project Overview:**

Project Management Plan

**Hypothesis:**

Business Requirements and Hypothesis Development

**Data Overview**:

- Software used: Microsoft Excel to handle data cleaning, profiling, and statistical analyses.
- Statistical methods used: Descriptive Statistics, Inferential Statistics which includes completing a t-Test. I used a one-tailed/sample t-Test to compare one group (aged 65+) against the total number of influenza deaths.
- Datasets used: CDC Influenza deaths by geography and US Census Bureau Population data by geography, time, age, and gender, CDC Influenza Visits Data Set, and Survey of flu shot rates in children.

**Results and Insights:**

Further insights about identifying all vulnerable populations could be incomplete due to having "suppressed" information from the CDC Influenza deaths by geography dataset. The reason why there is suppressed mortality data can be further explained here.