Ally Fornino

Influenza Analysis Interim Report

**[Tableau Presentation Link](https://public.tableau.com/shared/7XNJQXPRN?:display_count=n&:origin=viz_share_link)**

**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 a state has a higher percentage of individuals that are in the vulnerable population for influenza (ages under 5 or over 65), then that state will have a higher rate of influenza deaths.

**Data Overview**

“Population Data by Geography” US Census Data: This data set includes the population of each US county and state from 2009 to 2017, broken down by age group and gender.

“Influenza Deaths by Geography, Time, Age, and Gender” CDC Data: This data set includes a count of all deaths from 2009 to 2017 in which influenza was listed as the cause of death broken down by state, time, and age group.

**Data Limitations**

Population Data by Geography US Census Data:

**Time Lag:** There is a time lag in the data as the census is only performed every ten years.

**Collected Manually via Survey:** There is the possibility of human error when data was collected due to the collection process being manual and via survey (users can make mistakes when completing it). There is also potential bias in the results as some individuals may have not responded to the survey.

“Influenza Deaths by Geography, Time, Age, and Gender” CDC Data:

**Comorbidities:** Due to the data set only counting deaths in which influenza was recorded as the main cause of death, it is possible that deaths caused by comorbid conditions of influenza were not included.

**Suppressed Data:** For months in which fewer than 10 deaths were recorded, the true values were suppressed due to privacy concerns.

**Descriptive Analysis**

|  |  |  |
| --- | --- | --- |
|  | **Average** | **Standard Deviation** |
| **Total State Population** | 5972685.85 | 6799110.851 |
| **Deaths by State Population for Vulnerable Population (< 5 and Over 65 Years)** | 388.78 | 290.2946566 |
| **Correlation Coefficient** | 0.959378223 | |
| **Relationship** | There is a strong, positive relationship between these variables; the higher a state’s population is, the higher the number of deaths due to influenza. | |

In the hypothesis testing, a normalized version of the influenza deaths data was used in order to prevent this relationship from causing a falsely significant result to occur.

**Results & Insights**

Null Hypothesis: The influenza death rate of the population over the age of 5 and under the age of 65 (the non-vulnerable population) will be greater than or equal to the influenza death rate of the population under 5 and over 65 (the vulnerable population).

H0 : 𝜇NV ≥ 𝜇V

Alternative Hypothesis: The influenza death rate of the population over the age of 5 and under the age of 65 (the non-vulnerable population) will be less than the influenza death rate of the population under 5 and over 65 (the vulnerable population).

HA : 𝜇NV < 𝜇V

Results:

α = 0.05

p-value = 1.75441E-32

p < α, therefore we reject the null hypothesis

With a 95% confidence level, we reject the null hypothesis that the influenza death rate of the non-vulnerable population (ages over 5 and under 65) is greater than or equal to the influenza death rate of the vulnerable population (ages under 5 and over 65). There is statistically significant evidence that the influenza death rate of the non-vulnerable population is less than that of the vulnerable population.

**Remaining Analysis and Next Steps**

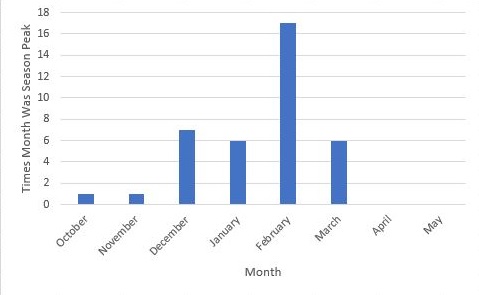
Remaining Analysis: In order to communicate to stakeholders how resources of the staffing agency should be distributed, states could be separated into different risk categories (low, medium, high) based on their percentage of vulnerable populations. Visual analyses will also be performed, including spatial and textual analyses.

Next Steps: Various deliverables (visualizations, written reports, presentations) will be created to communicate results to stakeholders.

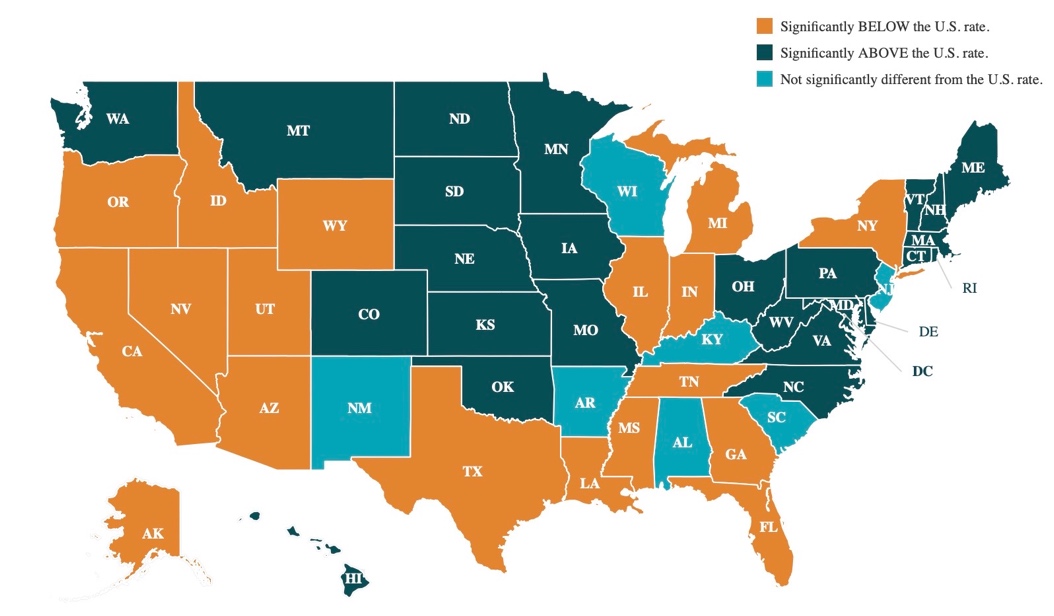
**Appendix**

Hypothesis Development:

* When is flu season?
  + In the United States, flu season is primarily in the fall and winter. Although the virus spreads year round, the peak of positive tests for the virus usually occurs between December and March. In some years, high activity can last until May.
  + Flu Activity Peak Months in the U.S. from the 1982-1983 through 2019-2020 Seasons:



* + Source: <https://www.cdc.gov/flu/about/season/flu-season.htm>
* Which individuals are included in the term “vulnerable population”?
  + Vulnerable population = adults over 65, children under 5, pregnant women and individuals with HIV/AIDS, cancer, heart disease, stroke, diabetes, asthma, and children with neurological disorders
* Which states have the lowest flu-shot rates?
  + Nevada, Louisiana, and Georgia have the lowest flu shot rates for adults



* + Source: <https://www.shadac.org/publications/state-level-flu-vaccination-rates-among-key-population-subgroups-50-state-profiles>

Results & Insights:

|  |  |  |
| --- | --- | --- |
| t-Test: Two-Sample Assuming Unequal Variances | |  |
|  |  |  |
|  | *Non-Vulnerable Population Death Rate* | *Vulnerable Population Death Rate* |
| Mean | 0.001344227 | 0.002676035 |
| Variance | 1.91579E-06 | 3.45152E-06 |
| Observations | 459 | 459 |
| Hypothesized Mean Difference | 0 |  |
| df | 847 |  |
| t Stat | -12.31599942 |  |
| P(T<=t) one-tail | 1.75441E-32 |  |
| t Critical one-tail | 1.646654627 |  |
| P(T<=t) two-tail | 3.50881E-32 |  |
| t Critical two-tail | 1.962768716 |  |