Preparing for Influenza Season.

Interim Report

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 population has more people over 65 years old, then it has a higher Influenza death rate."

Data Overview

We used two primary data sources:

- Census Data.

This data shows the population of each state by year, gender, and age group.

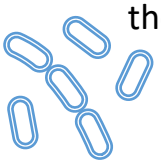
- Influenza Deaths Data.

The data contains monthly death counts for influenza-related deaths in the United States from 2009 to 2017. It has two main categories: state and age.

Data Limitations

- Census Data.

The numbers from this data set are estimates, hence the sum of the numbers from the different age groups may not sum up to the total in the first columns.



This data set does not contain all the years of the census. The records with some years are missing.

- Influenza Deaths Data.

The data contains suppressed values - fewer than ten (one to nine) deaths are suppressed. All numbers of deaths of children under 5 years old, for example, were suppressed. This significantly limits our data source, considering that is one of the important groups of vulnerable populations we need to examine.

Descriptive Analysis Summary

We performed descriptive analysis to determine the mean, standard deviation, and outlier percentage of our data, thereby establishing the quality of the data (Data spread).

Data Source	Influenza Deaths			Census Populations		
Variable	Under 5 years	5-64 years	65+ years	Under 5 years	5-64 years	65+ years
Mean	119.8605664	417.5490196	897.1917211	386302	3994622	806945
Standard Deviation	12.21255316	119.4970322	971.5247427	457038.5712	4577527.202	886876.8546

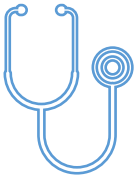
Also, we discovered the relationship between the core variables (correlation):

Variables in Correlation	1. Deaths 5-64 years 2. Population 5-64 years old	1. Deaths 65+ years 2. Population 65+ years old
Correlation Coefficient	0.92	0.94
Strength of Correlation	Strong relationship	Strong relationship

Variables have a strong relationship. We can state that there is obviously a correlation between the age of the person and the death rate. Especially important for our analysis group of people aged over 65 (a vulnerable population) should be analyzed further.

Results and Insights

In order to test our research hypothesis, we used inferential statistics. For this, we transformed our research hypothesis into statistical ones.



Null Hypothesis (H0): People 0-65 years old have a higher or equal Influenza death rate than people over 65 years old.

Alternative Hypothesis (H1): People 65+ years old have a higher Influenza death rate than younger people.

For our inferential statistics we used „Student's t-Test "and performed it in Excel.

t-Test: Two-Sample Assuming Unequal Variances

	Death rate 0-64 years old	*Death rate 65+ years old*
Mean	0.002015227	0.00662632
Variance	4.3776E-06	5.72752E-06
Observations	458	458
Hypothesized Mean Diffe	0	
df	898	
t Stat	-31.0431539	
P(T<=t) one-tail	1.26E-144	
t Critical one-tail	1.646552237	
P(T<=t) two-tail	2.5218E-144	
t Critical two-tail	1.962609216	

The results show that our p-value is the number that is extremely close to zero and

< 0.05 (our significant level), which means we can disprove the null hypothesis.

With a confidence level of 95 percent, our test shows that the population of people over 65 years old is more likely to have complications after the flu and possibly die than younger people.

Remaining Analysis and Next Steps

The next steps will be:

- Spatial Analysis.
- Textual Analysis.
- Creating statistical visualizations to provide insights and key findings using the Tableau Storyboard.
- Final Deliverable (the result will be provided to the stakeholders in the shape of a presentation and will be introduced by the analyst with clarification).

Appendix

Data Sources:

1. Influenza Deaths Data.

Data collection:

- Administrative data.
- Part of the government's vital statistics program.
- Each of the U.S. states and territories is required to record all births, deaths, marriages, and divorces within their jurisdiction. Death records come from death certificates, in which a doctor codes the primary cause of death as "Influenza" or "Pneumonia" (ICD-10 codes J09-J18).

2. Census Data.

Data collection:

- Administrative data.
- Data is collected through a survey.
- Data collection is carried out every year.

Hypotheses Development:

1. Clarifying questions.

- When is influenza season?
- Which states will suffer more from the flu?
- Why does flu season differ by state?
- What can affect the intensity of the flu?
- How does the number of vulnerable populations vary relative to the state?
- How does influenza death differ depending on the state?
- How does the number of flu shots vary by state?

2. Funneling questions.

• Does influenza occur seasonally or throughout the entire year?

• If seasonally, does it start and end at the same time (month) in every state?

• Why will some states have more patients in hospitals in influenza season than others?

• How will I determine the vulnerable population in each state?

• Does the number of flu shots have connection with number of influenza death in each state?

3. Privacy and Ethics.

• Are there relevant data privacy laws to consider?

• Can I use the data of hospitals and clinics using the staffing agency's services?

• Are there privacy laws we need to add here to analyze data from Influenza patients.