Aymen Touihri

04.12.2023

Preparing for Influenza Season: Interim Report

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

- "If patients' age increases, then the Influenza death rate increases."

Refer to the Research Hypothesis Development in the Appendix section for more information about the key questions drawn to form the research hypothesis.

3. Data overview

Influenza Death by Geography

- The dataset contains information about Influenza mortality by geography and age group from 2009 to 2017.

U.S Census Population Data

- The dataset contains information about population counts by state, gender, and age group from 2009 to 2017.

**Data profiling and quality assessment for the datasets used in this analysis project can be found in the GitHub project repository.*

4. Data limitations

Influenza Deaths by Geography

- **Missing/Incomplete data**: Death counts for certain age groups (less than 9) are not present in the dataset. This limitation prevents us from analyzing the Influenza mortality for some of the vulnerable population groups. For example: Children under 5 years.

U.S Census Population Data

- **Inaccuracy**: As stated in the project overview document, the numbers in the dataset for the various age groups are estimates and may not represent the actual population counts.

5. Descriptive analysis

The first step in our statistical analysis proposes that a relationship exists between two key variables — **Death Rate** and **Population Count** for the age group 65+ years, which ties back to our research hypothesis.

Variable	Average	Standard Deviation
Influenza death rate (65+ years)	810	1,011
Population count (65+ years)	801,596	879,279

The analysis shows a strong correlation between the population count and influenza mortality rate for the age group 65+ years with a correlation coefficient of **0.8**.

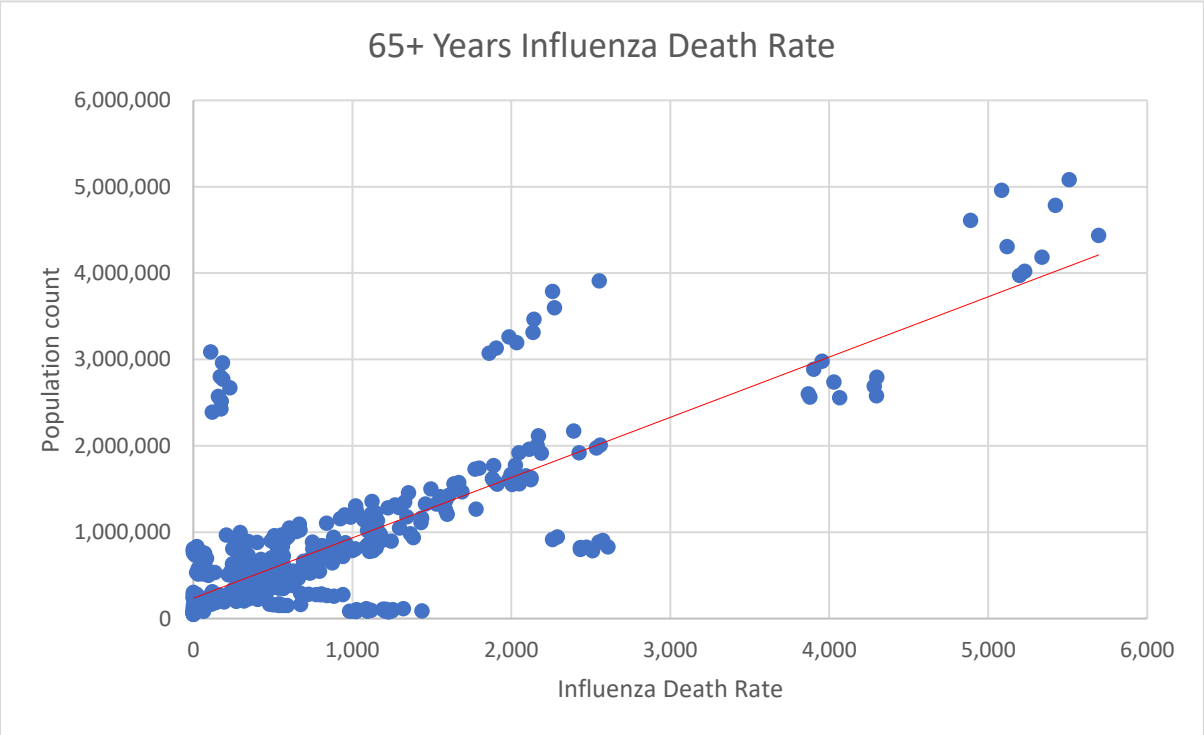

Figure 1: Scatter Plot that shows the relationship between the population count and Influenza mortality for the age group of 65+ years.

Based on the obtained result, we can conclude that the Influenza mortality rate increases as the population count of the 65+ years age group grows higher.

6. Results and insight

The next step in our analysis was to look at the Influenza mortality rate for the two variables: *Less than 65 years* and *65+ years* age groups, and test for a statistical significance difference between the two age groups.

The following Null and Alternative Hypothesis were proposed:

- **Null Hypothesis**: "The Influenza death rate is the same for both age groups."
- **Alternative Hypothesis**: "The Influenza death rate is higher for the 65+ years age group."

Our statistical test shows that, at a confidence level of 95% (0.05 significance level), there is a significant difference in the average Influenza death rate between the two age groups. The statistical test allows us thus to reject the null hypothesis and conclude that the Influenza mortality does increase with patients' age.

Conclusion: The test results obtained from this statistical hypothesis, as well as the strong correlation found between age and Influenza death rate can be used to prioritize medical staffing for regions with the highest concentration of 65+ years age group in the upcoming Flu season.

Refer to the Statistical Test in the Appendix section for more information about the test results.

7. Remaining analysis and next steps

- Examine the regions with the highest concentration of the vulnerable populations to determine medical staff allocation for the next Flu season.
- Create Data Visualizations incl.:
    - Composition & Comparison charts
    - Temporal Visualization & Forecasting
    - Statistical Visualization

- Spatial & Textual Visualization
- Storytelling & Data Presentation

- Present Findings to Stakeholders & Project Wrapping

8. Appendix

Research Hypothesis Development

The project requirements document identifies the Vulnerable Populations as follows:

- Adults over 65+ years
- Children under 5 years
- Pregnant Woman
- Patients with chronic diseases (HIV, cancer, heart disease, diabetes, etc.)

The datasets provide information about the population count and Influenza mortality rate by geography and age group. We can use this information to form and answer the following questions, and draw our research hypothesis:

- How is age affecting Influenza mortality rate?
- Is there a correlation between the two variables? How strong/weak is it?
- Can we make inferences on the rest of the Vulnerable Population groups? i.e., Children under 5 years, pregnant women, and patients with chronic diseases.

Statistical Test

T-Test: Two-Sample Assuming Unequal Variances

	< 65 years death rate	> 65 years death rate
Mean	78.76470588	826.2875817
Variance	22903.91395	1028483.747
Observations	459	459
Hypothesized Mean Difference	0	
df	478	
t Stat	-15.61886217	
P(T<=t) one-tail	**4.96E-45**	
t Critical one-tail	1.648047593	
P(T<=t) two-tail	9.91009E-45	
t Critical two-tail	1.964939189	

The t-test performed on the Influenza mortality for the two age groups; less than 65 years, and greater than 65 years yielded a p-value result of **4.96E-45**, which is way less than the significance level of 0.05.