

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.

Stakeholder Identification

- Medical agency frontline staff (nurses, physician assistants, and doctors)
- Hospitals and clinics using the staffing agency's services
- Influenza patients
- Staffing agency administrators

Hypothesis

If a state has a larger proportion of vulnerable population, then more deaths from flu will occur.

Data Overview

1. Population data by geography (Source: US Census Bureau)

The Decennial Census is the once-a-decade population and housing count of all 50 U.S. American states. The exercise is conducted by the U.S. Census Bureau. The data is collected by sending a survey to all households. Each household has then to return the filled survey to the U.S. Census Bureau that aggregates, organizes, and publishes the data. Complementary to the census, the U.S. Census Bureau also conducts several ongoing surveys, on different topics such as employment, crime, health, consumer expenditures, and housing.

The data collected by the U.S. Census Bureau covers a broad list of variables that aim to characterize the country's people, places, and economy. It is considered the most complete effort to collect information by one single entity in the USA. It covers topics such as

employment, crime, health, consumer expenditures, and housing. The Census Bureau also conducts regular surveys of economic variables such as manufacturing, retail, service.

These data were relevant to the project because it enabled us to estimate the proportion of vulnerable population per state. The resulting variable used to test the hypothesis was named “Proportion of Vulnerable Population per State”.

2. Influenza deaths by geography, time, age, and gender (Source: CDC)

The Underlying Cause of Death database contains mortality and population counts for all U.S. counties. Data are based on death certificates for U.S. residents. Each death certificate identifies a single underlying cause of death and demographic data.

The information on deaths is desegregated by region (state and county), age group, race, gender, year, and cause-of-death. Data are also available for place of death, month, and weekday of death, and whether an autopsy was performed.

These data were relevant to the project because it enabled us to know the amount of deaths caused by Influenza per State. The resulting variable used to test the hypothesis was named “Average of Flu Deaths per State”.

Data Limitations

A census, being based on a complete count of the population and not an extrapolation of a sample, is a robust method for constructing a representation of reality. The methodologies used by the U.S. Census Bureau for data collection meets the highest quality standards in the world. There are no reasons to believe the data is particularly biased or incorrect.

The CDC data covering influenza deaths is also a complete count of the population and not an extrapolation of a sample. It is therefore a robust method for constructing a representation of reality. The methodologies used by the CDC for data collection meets the highest quality standards in the world. There are no reasons to believe the data is particularly biased or incorrect.

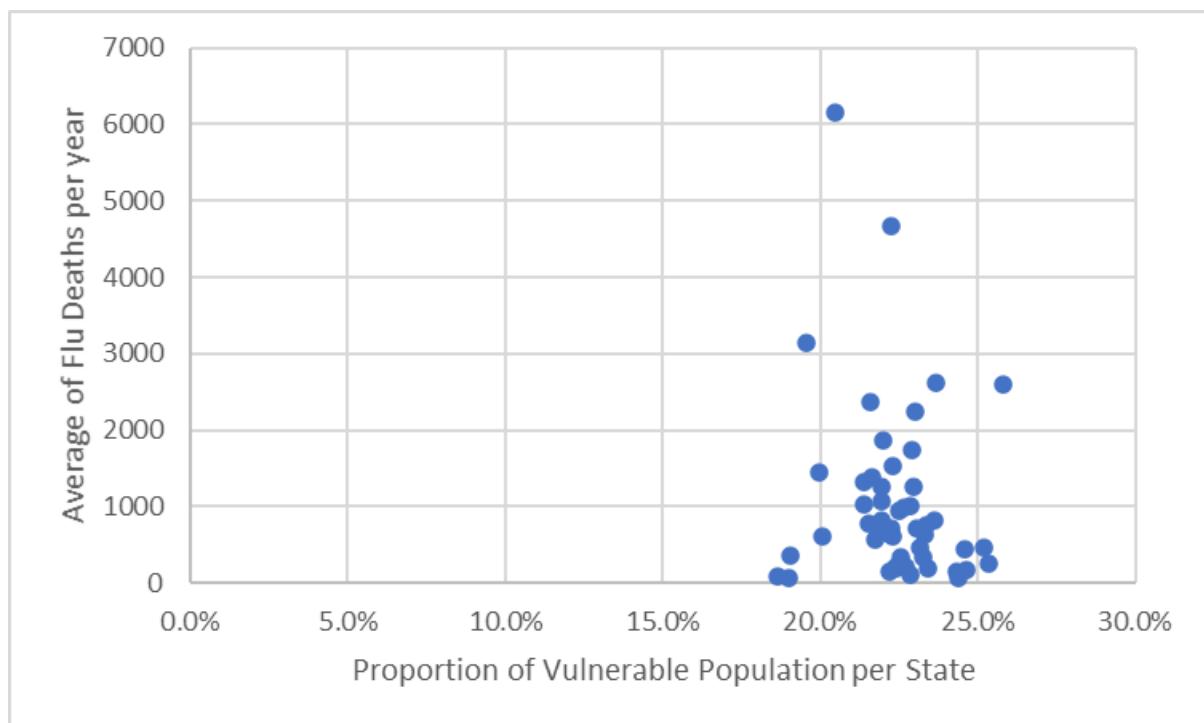
Summary of Descriptive Analysis

The following table presents a summary of the descriptive analysis:

| | Proportion of Vulnerable Population per State | Average of Flu Deaths per State |
|----------------------|---|--|
| Original data set | Population data by geography (Source: US Census Bureau) | Influenza deaths by geography, time, age, and gender (Source: CDC) |
| Sample or population | Population | Population |
| Normal distribution | Yes | Yes |
| Mean | 22% | 1056 |
| Variance | 0.0002 | 1319646.94 |
| Standard deviation | 0.0154 | 1148.76 |
| Outlier percentage | 7.8% | 3.9% |

To test the hypothesis, we proceeded by calculating the strength of the correlation between the variables. The following table and graph present a summary of the results:

| Correlation | |
|-------------------------|--|
| Variables: | "Proportion of Vulnerable Population" and "Average of Flu Deaths per year" |
| Preposed relationship: | If a state has a larger proportion of vulnerable population, then more deaths from flu will occur. |
| Correlation Coefficient | -0.16 |
| Stranght of Correlation | weak |

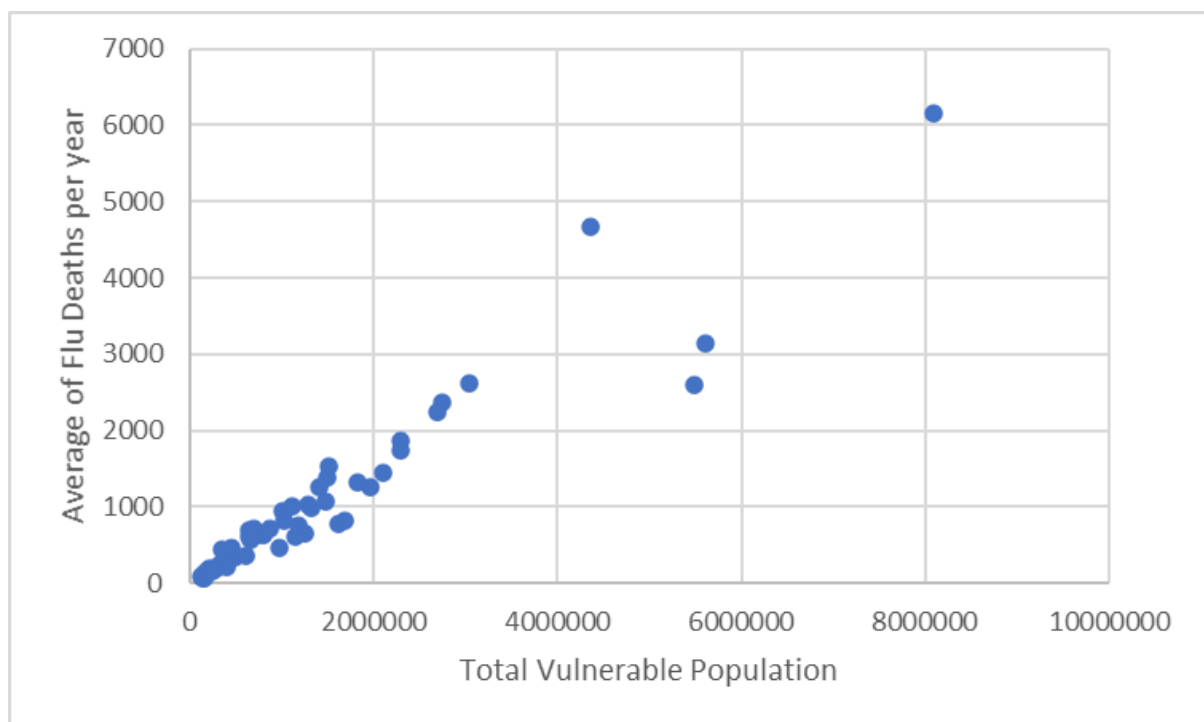


As it can be seen, the data does not support the hypothesis. The strength of the correlation is weak (-0.16) and the visual representation shown no visible pattern suggesting a correlation between the variables.

When analysing why the data did not support the hypothesis, we see that, although the number of people considered vulnerable in each State varies significantly from State to State, the proportion of vulnerable population between States does not vary significantly between States. States that have more vulnerable population also tend to have more non-vulnerable population. This explains why the variance and standard deviation of the proportion of vulnerable population per State is relatively small, which, in turns, explains the absence of a statistical correlation between the variables considered.

If we were to test a different hypothesis relating the absolute number of people considered vulnerable in each State, instead of the proportion of vulnerable population in each State, then the correlation would be evident, like it is shown by the following table and graph.

| Correlation | |
|-------------------------|--|
| Variables: | "Total Vulnerable population" and "Average of Flu Deaths per year" |
| Preposed relationship: | If a state has more vulnerable population, then more deaths from flu will occur. |
| Correlation Coefficient | 0.95 |
| Strenght of Correlation | Strong |



Summary of results and insights

We can say with confidence that if a State has more vulnerable population it will have more deaths due to influenza (flu).

This conclusion has a direct impact for the objective of the project: to determine when to send staff, and how many, to each state. More medical staff will be needed in States that have more people considered vulnerable.

We cannot however affirm that if a State has a higher proportion of vulnerable population, then it will have more deaths due to influenza. This was our original hypothesis and it was rejected.

It is important to note that no inferential statistics tests were performed to test the hypotheses. Inferential statistics uses a random sample of data taken from a population to describe and make inferences about the population. In this case however, we are working with population data and not with sample data. Therefore, it is not adequate to use inferential statistics.

Remaining Analysis and Next Steps

The objective of the project is to determine when to send staff, and how many, to each state.

So far, we have uncovered a relevant insight, that more staff will be needed in States that have more people considered vulnerable.

Although this is helpful, it is far from sufficient. More hypotheses must be tested to meet the project's objective to a satisfactory degree.

Our recommendation is to proceed with the testing of new set of hypotheses with the potential of bringing us closer to the objective of the project.

Some of the research questions currently being discussed include:

- To what extent does the flu season occurs in different moments throughout the year in different States?
- To what extent does the number of hospitalized patients due to the influenza in the previous years can inform us about the staff needs in the future?
- To what extent does the population density of a State determines the number of patients suffering from the flu?

The next step is to perform additional planning and meetings to determine which research questions have priority in the next round of testing.

Appendix

1. The business requirements document