Staffing Plan 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 state has more population in vulnerable age group (People aged 65 and above), then that state will have more influenza-related deaths

Data Overview

* Influenza deaths by geography, time, age, and gender
* Source: CDC
* The data contains monthly death counts for influenza-related deaths in the United States from 2009-2017. Counts are divided into two categories: state and age.
* Population data by geography
* Source: US Census Bureau
* The data has population by county of each state, grouping the data by gender and range of age.

Data Limitations

Both data sets are sourced from reliable government agencies and are considered trustworthy. However, it is important to keep some limitations in mind that are listed below:

The Influenza deaths data was collected by death certificates. Death certificates only list one cause of death which could be prone to inconsistencies for patients with preexisting conditions such as cancer or heart disease that may have contributed to their deaths.

The census data is collected once every 10 years by the US Census Bureau. This fact means that we are relying on figures that are estimates rather than hard data. The census data is also collected manually which leaves it prone to error.

Descriptive Analysis

Data Spread

Table 1

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Vulnerable-Aged Population | Non-Vulnerable Aged Population | Vulnerable-Aged Deaths | Non-Vulnerable-Aged Deaths |
| Variance | 7.93325E+11 | 3.55312E+13 | 942950.61 | 14178.36 |
| Standard Deviation | 891659.6 | 5960806 | 971.0564398 | 119.0729096 |
| Mean | 826266.8 | 5262450 | 896.3594771 | 536.5 |
| Outlier Percentage | 12.6% | 19.6% | 12.6% | 9.4% |

Correlation

Table 2

|  |  |
| --- | --- |
| Variables | Deaths of Vulnerable-Aged Population |
| Proposed Relationship | A strong relationship should be present between the vulnerable aged population and the number of influenza-related deaths. |
| Correlation Coefficient | .89 |
| Strength of Correlation | Due to the correlation coefficient being so close to 1.0, there is a strong relationship |
| Interpretation | The strong relationship supports the hypothesis that a vulnerable-aged population will have more influenza-related deaths |
| Direction | Strong positive correlation |

The Table above shows the relationship between the vulnerable-aged population (adults 65 years and older) in the United States and the deaths of this population. We can see a strong positive relationship between the vulnerable population and the number of influenza-related deaths. This informs us that if a state has a high population of vulnerable-aged people, there will be a bigger number of influenza-related deaths.

Results and Insights

Statistical Hypothesis Testing

Table 3

|  |  |
| --- | --- |
| Hypothesis | If a state has more population in vulnerable age group (People aged 65 and above), then that state will have more influenza-related deaths. |
| Dependent Variable | Number of influenza-related deaths |
| Independent Variable | Age group |
| Null Hypothesis | Influenza related deaths in the age group of 65 years and older is less than or equal to the death rate for those under 65 years old |
| Alternate Hypothesis | Influenza related deaths in the age group of 65 years and older is greater than the death rate for those under 65 years old. |
| Test Type | This is a one-tailed test type and will inform us if deaths in vulnerable aged populations are higher than in non-vulnerable aged populations. |
| Alpha | 0.05 |
| P-Value | 2.93E-08 (0.0000000292614994790416) |
| Significance Level | Because the P-Value is less than the alpha, we can reject the null hypothesis. It can then be stated that influenza-related deaths in vulnerable-aged populations are higher than in non-vulnerable aged residents. |

Table 4

|  |  |  |
| --- | --- | --- |
|  | Normalized Death rates of Vulnerable Aged Populations | Normalized Death Rates of Non-Vulnerable-Aged populations |
| mean | 0.000647262 | 0.000257145 |
| variance | 1.27684E-07 | 6.64416E-08 |
| Observations | 459 | 459 |
| Hypothesis Mean Difference | 0 | N/A |
| df | 833 | N/A |
| t Stat | 18.97 | N/A |
| P(T<=t) one-tail | 2.82222E-67 | N/A |
| t Critical one-tail | 1.65 | N/A |
| P(T<=t) two-tail | 5.64444E-67 | N/A |
| t Critical two-tail | 1.963 | N/A |

\*N/A = Data is not required for this testing

Table 4 shows the results of the T-test. The purpose of this T-test is to determine the number of medical staff to be sent to each state for the upcoming flu season. The results of this one-tailed test show P=2.82222E-67 which is lower than the significance value of 0.05 or 5%. Because the P value is lower than 0.05, we can reject the null hypothesis and confirm the influenza-related deaths in vulnerable populations are higher than those in non-vulnerable populations.

Remaining Analysis and Next Steps

Additional Analysis

* For Vulnerable population deaths under the age of 5, the death numbers are suppressed so it would be helpful to have a more accurate number for a more precise analysis.
* It would be helpful to have data on the vaccination rates between the vulnerable populations and the non-vulnerable populations. If the data shows that vaccinated individuals have a decreased mortality rate against influenza, we could promote efforts to get more people vaccinated which would decrease the number of medical staff needed to be sent.
* Focus on states with a higher vulnerable-aged population and providing them with additional medical staff than states with low levels of vulnerable-aged populations. We should figure out which states contain a higher number of vulnerable-aged residents by determining the average percentage of vulnerable-aged residents per state and then seeing which states fall above this average.

Visualization

A special visualization using Tableau needs to be created showing the number of vulnerable-aged deaths for each state.

Appendix

Clarifying Questions

1. Business Requirements from the project brief to analyze past trends from each state to determine the best way to distribute medical staff.
2. Clarifying Questions:
3. Which state has the highest number of influenza patients?
4. What percentage of patients who catch the flu in each state end up in the hospital?
5. What months does each state have its flu season?
6. What is each state’s vaccination rate?
7. How quickly does the flu spread within each state’s population?

Funneling Questions from “how quickly does the flu spread within each state’s population?”

Which gender has the highest flu cases in each state?

Do male or female patients have a higher mortality rate?

Which strain of flu spreads the most in each state?

Privacy and Ethics Questions

For patients, do we need to get a signed consent before collecting their data?

Are there any privacy laws such as HIPAA that we need to adhere to?

Which parties have access to the data?

How is the data stored?