

PREPARING FOR THE
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.

Research Hypothesis:

If a patient is over 65 years old or older, they are more likely to die from influenza.

Data Overview

Population data by geography (US Census)

- The data contains the annual population of each county in each state in the US by gender and age from 2009 to 2017.

Influenza deaths by geography, time, age, and gender (CDC)

- The data contains the number of deaths in each age group (<1 years old to 85+ years) from influenza like illness (ILI) in each state from 2009 to 2017.

Data Limitations

Population data by geography (US Census)

- The likelihood of the data being biased is rare since the data is meant to be informative. However, the data is not up to date since it is collected every 10 years and the last data recorded was in 2017. There is a two-year time lag between the full detail of births and deaths data. As a result of the data being manually collected, the data can contain errors. Besides birth and death, emigration and immigration also affect population, which is not reflected in the data. There is also a discrepancy between the metric for male and female population and for the grouped ages.

Influenza deaths by geography, time, age, and gender (CDC)

- The deaths column in the data set contains many “Suppressed” values due to privacy concerns for individuals. Therefore, exact numbers for that variable is unknown.

Descriptive Analysis

During the descriptive analysis it could be confirmed that a strong correlation exists between the total population and the total death caused by influence. The same can be said about the vulnerable population (population 65 years and older) and the death count in this population. In the following table you can see the mean and standard deviation for the vulnerable population, the total population, and the corresponding deaths due to Influenza.

	Vulnerable Population (65 years and older)	Total Population	Death Vulnerable Population	Total Death
Mean	838790,24	6166644,64	826,29	905,05
Standard Deviation	887222,56	6806912,83	1014,14	1154,50

Results and Insights:

The hypothesis testing confirms that vulnerable populations have more influenza death cases than non-vulnerable populations. Due to the significance level being much higher than the p-value, the null hypothesis can be rejected, and the alternative hypothesis can be accepted. With 95% confidence level, the two groups are significantly different.

Research hypothesis	If a patient is over 65 years old or older, they are more likely to die from influenza
Null Hypothesis	Mortality of a patient with age of 65 + years is less than or equal of a patient with an age lower than 65 years
Alternative Hypothesis	Mortality of a patient with age of 65 + years is greater than of a patient with an age lower than 65 years
p-value (one-tail)	5,03716754155316E-45

Remaining Analysis and Next Steps:

- Based on the analysis, it is evident that the vulnerable populations have more influenza death cases than the non-vulnerable populations. Therefore, it is imperative to allocate more medical staff to those states with higher proportions of vulnerable populations during the influenza season to reduce the mortality rate for those vulnerable populations.
- The final presentation will include visualizations, conclusions, and recommendations

Appendix

Results & Insights: t-Test: Two-Sample Assuming Unequal Variances

t-Test: Two-Sample Assuming Unequal Variances	0-65 years(Non-Vulnerable population)	65+ years(Vulnerable population)
Mean	78,76470588	826,2396514
Variance	22903,91395	1028562,069
Observations	459	459
Hypothesized Mean Difference	0	
df	478	
t Stat	-15,61727902	
P(T<=t) one-tail	5,04E-45	
t Critical one-tail	1,648047653	
P(T<=t) two-tail	1,00743E-44	
t Critical two-tail	1,964939272	

