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.

Research hypothesis

If adults are older than 65 years or children are younger than 5 years, then they are more likely to die from flu.

Data overview

1. Influenza death by geography

This data shows the death counts throughout the United States by states and age groups. The data comes from the Centers for Disease Control and Prevention (CDC) and represents the death counts by month over 9 years from 2009 to 2017.

2. Population data by geography, time, age and gender

This data shows the population throughout the United States by state / county, gender, and five-year age groups. The data source is the US Census Bureau, covering 9 years from 2009 to 2017.

Data limitations

1. Influenza death by geography

Incompleteness:

- Data based on death certificates for U.S. residents and does not contain the data on non-resident aliens. The patients from this category can increase the influx of hospital admissions, but the count of non-residents is not covered by these statistics.
- The statistics representing fewer than ten people (0-9) are suppressed. In this data set 'Suppressed' values account 82% of the death values. The missed values were replaced by random selection from 1 to 9.
- "Not Stated" age group category represents the observations not distributed among any age group. These observations account for 8% and were removed from the data set as they cannot be used for the analysis.

The Influenza Death data set was adjusted to be acceptable for the project analysis.

Inaccuracy: Data contains a single underlying cause of death certified by a domain expert. However, there is risk that the deaths due to complications of flu are not certified as 'caused by flu'.

2. Population data by geography, time, age and gender

Sampling error: The census methods are less biased because the entire population is the subject of investigation. However, if the census data come from the survey of a sample of population, there may be a sampling error because only a part of population was observed.

Incompleteness: If data is collected by questionnaire sent by mail, there is a risk of non-respondents. In such cases, the data imputation and adjustments were used by the state agency.

Inaccuracy: Data collected in the process of personal interview may lead to input errors.

In general, census data is collected directly from all eligible units in the entire target population. It can be considered a sample with a 100 percent sampling rate. The Census data set is considered acceptable for the project analysis.

The count of either population and influenza deaths are based on patterns in historical data for nine years from 2009 to 2017 and can be accepted according to our project context.

Descriptive Analysis

Variable	Average	Standard deviation
Population (under 5 and over 65 years old)	1,082,577	1,204,099
Influenza Mortality (under 5 and over 65 years old)	999	975

We have tested the average and standard deviation for the age groups that we assume to be vulnerable population for developing complications because of flu infection.

Variables are recorded by state and by year.

The results show that the population and influenza mortality data are widely spread; and most of data points are far away from the mean / average value.

Correlation

We have tested the correlation between two variables: Influenza death and Population of vulnerable population group.

Correlation coefficient 0.95 witnesses a strong relationship between the death count and population of age group under 5 and over 65 years. The states with higher numbers of vulnerable people have more deaths due to influenza.

Results and Insights

Null Hypothesis: People 65 years of age and older have the same or lower death rate from influenza than people under 65 years of age.

Alternative Hypothesis: People 65 years and older have a higher death rate from the influenza.

The strong significance level is assumed to be 0.05.

The p-value takes on an extremely low value 5.3559E-196.

This means that there is a significant difference in average mortality rates for two testing groups (population under 65 and adults over 65 years old)

The average mortality rate for adults over 65 is 0.1474% vs 0.0243% for people under 65 years old.

P-value is significantly low. That means that there is a low probability that difference in average mortality rate is due to random chance. Thus, we can reject the null hypothesis.

The influenza mortality percentage (dependent variable) is impacted by age factor (independent variable)

Remaining analysis

We have confirmed that adults over 65 years are more likely to develop flu complications and may need to be admitted to hospital.

Further, we will range states by population groups and categorize the states as having high-, medium- and low-needs in medical assistance for influenza season based on the vulnerable population count. Hospitals in the states with a high level of aging population are more likely to need additional medical staff.

Another point to consider is the seasonal nature of influenza in the U.S. We will need to determine (i) whether influenza occurs seasonally or throughout the entire year; (ii) whether the flu season starts and ends at the same time (month) in each state.

As a result of the analysis, insights will be provided to support the staffing plan, including spatial and timing distribution of medical personnel based on the hospitals needs and staff available with the agency.

Next Steps

Create time forecast and spatial visualization.

Create a Story board presenting the sequence of visualization, findings, insights, and recommendations.

Present the result of analysis to the stakeholders.

Appendix

Project overview related information.

The project's success will be based on:

- A staffing plan that utilizes all available agency staff per state requirements, without necessitating additional resources.
- Minimal instances of understaffing and overstaffing across states (a state can be considered understaffed if the staff-to-patient ratio is lower than 90% of the required ratio and overstaffed if greater than 110%.)

Assumptions:

- Vulnerable populations suffer the most-severe impacts from the flu and are the most likely to end up in the hospital.
- Flu shots decrease the chance of becoming infected with the flu.

Constraints:

- The staffing agency has a limited number of nurses, physician assistants, and doctors on staff.
- There's no money to hire additional medical personnel.

Questions to ask to develop and test the project hypothesis

	Clarifying questions for better understanding of the project issues
1.	Does influenza occur seasonally or throughout the entire year? How can we measure the start and the end of influenza season?
2.	What percentage of people with influenza are hospitalized?
3.	Which states are considered as low-, medium- and high-needed in staff due to higher hospitalization level during flu season? How can we measure the low-, medium- and high-needed states?
4.	What is the number and the percentage of vulnerable people by each category in each state?
5.	What is the trend of flu vaccination over time and by state?
6.	What is the trend of influenza death rate over time and by states?
	Adjoining questions to look at other issues related to influenza
1.	Does the flu impact other hospital admissions or deaths level due to other diseases?
2.	Are there any way of decreasing flu admissions or deaths (such as flu shots)?
	Funneling questions to look deeper into the project issues
1.	Which hospitals did meet the staff-to-patient ratio (overall and by categories**,** nurses, physician assistants, and doctors? Which hospitals did not meet the ration?
1.1	How did it affect the death ration?
2.	What is the trend of flu vaccination over time and by state?
2.1.	What is the vaccination rate of vulnerable population over time and by state?
2.2.	Is there any correlation between the number of vaccinated patients and hospitalizations?
3.	What is the trend of influenza death rate over time and by states?
3.1.	What is the rate of influenza death among vaccinated people? hospitalized patients?
3.2.	Is there any correlation between death rate and the number of vaccinated people?
3.3.	Is there any correlation between the death rate and number of vulnerable populations?
	Elevating questions to look at how the project may affects other issues
1.	Can the government (or workplaces, schools, health departments, etc.) do something to help prepare for the flu nationally (hand washing public health campaigns, flu shots, etc.)?
2.	How does the flu impact the community, for instance, school and work attendance?
	Privacy and ethical questions
1.	Are there privacy laws related to collecting and analyzing data?
2.	Do we need to ask the patients permission to use their data?

3.	Are there any laws related to medical staff hiring which might concern this project (e.g. the requirements to keep the gender parity) ?

Data Source

Influenza Mortality: CDC (Centers for Disease Control and Prevention, U.S. Government Agency)

Population Data: U.S. Census Bureau, Government Agency.

Statistical Analysis

T Test: Two-Sample Assuming Unequal Variances

	Influenza Death Percentage, age group under 65	Influenza Death Percentage, age group over 65
Mean	0.000243	0.001474
Variance	0.0000000606	0.0000002971
Observations	459	459
Hypothesized Mean Difference	0	
df	637	
t Stat	-44.08977782	
P(T<=t) one-tail	5.3559E-196	
t Critical one-tail	1.647249173	
P(T<=t) two-tail	1.0712E-195	
t Critical two-tail	1.96369499	

	Influenza Death Percentage, age group under 5 years	Influenza Death Percentage, age group over 5 years
Mean	0.000811119	0.000361103
Variance	7.79776E-07	5.10448E-08
Observations	459	459
Hypothesized Mean Difference	0	
df	517	
t Stat	10.5774421	
P(T<=t) one-tail	4.18987E-24	
t Critical one-tail	1.647806216	
P(T<=t) two-tail	8.37974E-24	
t Critical two-tail	1.96456301	

Two tests were done separately to verify the research hypothesis that the flu can lead to serious complications for 2 age categories: adults over 65 years and children under 5 years.

The analysis for age group 65+ is presented above in the report.

The result of testing the age group under 5 years shows that the average death rate is higher in comparison with other population. Although the difference in mean values for this age group is not as significant as for adults over 65 years, nevertheless, the low p-value suggests that this age group (children under 5 years) vulnerable to influenza as well.