



Preparing for Influenza Season: Interim Report

Project Overview:

During Influenza season in the United States there are more than usual number of people suffer from the flu. Some people in vulnerable conditions develop more serious conditions and end up getting admitted in the hospital. During these times the hospitals need more staff to meet with high need. The medical staffing agency provides temporary staffing to these hospitals.

Objective:

Determining how many staff members to send and when to send to each state.

Scope:

The staffing agency covers all the 50 states of the United States, and the plan is for the upcoming influenza season.

Hypothesis:

- If a state has a high number of vulnerable population (such as children under 59 months, elderly (65+), then there will be more cases of Influenza related deaths.

Data Overview:

Data	Population by geography
Source	US Census Bureau
Content	Population by State from the year 2009 to 2017
Limitations	Since the data is collected manually, there is a chance for errors, typos. The data could be biased as not everyone takes the surveys.

Data	Influenza deaths by age, state, year
Source	CDC
Content	Influenza deaths by age group, states and by year from 2009 to 2017
Limitations	This is survey type of data and is not complete counts of all influenza visits or lab tests in the U.S. Since the data is collected through surveys, it could have human errors, bias and time lag. This is an external data.

Descriptive Analysis:

Table1: Data Spread

	Vulnerable Deaths (below 5)	Vulnerable deaths (65+)	Vulnerable population (below 5)	Vulnerable population 65+
Variance	132	942310.2674	2.08327E+11	7.85085E+11
Standard Deviation	11	970.7266698	456428.7315	887017.1926
Mean	120	896.8366013	386282.6406	806988.9386
Outlier Percentage	5%	4%	4%	13%

Table 2: Correlation

Variables	Deaths and 65+ vulnerable population	Deaths of non-vulnerable population (between 15 to 24)
Proposed Relationship	Vulnerable pop. deaths should be high as flu deaths are high.	Non vulnerable pop. Deaths should not be as high as flu deaths
Correlation Coefficient	0.94	0.16
Strength of Correlation	Relationship is strong as it is between 0.5 to 1.0	Very weak relationship as the value is 0
Interpretation	It supports the hypothesis that more vulnerable population deaths are related.	There is very weak relationship between non vulnerable population deaths
Direction	Strong Positive association	Weak positive

The above table shows the comparison between deaths of vulnerable population who are 65+ and deaths of non-vulnerable population of ages between 15 and 24. We can see that there is a strong positive relation between vulnerable population and deaths. Meaning, increase in the number of vulnerable population, leads to increase in the number of Influenza related deaths.

Results and Insights: Statistical Hypothesis Testing

Hypothesis:	If a state has more vulnerable population (ages 65+) then there will be more flu related deaths in that state
-------------	---

Dependent Variable:	Number of vulnerable population (ages 65+)
Independent Variable:	Flu deaths
Null Hypothesis:	Death rate in vulnerable population (ages 65+) is less than or equal to deaths in non-vulnerable population
Alternate Hypothesis:	Death Rate in vulnerable population (ages 65+) is greater than deaths in non-vulnerable population
Test type:	This is a one tailed test as the test will be done to see if the deaths in vulnerable population (ages 65+) is higher
Alpha:	0.05
P-value:	2.07E-178
Significance level:	Since the p-value is less than alpha, we can reject the null hypothesis. Thus, we can say that the deaths in the States with vulnerable people (ages 65+) are higher than in the states with less vulnerable people.

	<i>Flu mortality rate for 65+</i>	<i>Flu mortality rate for < 65</i>
Mean	0.001317078	0.000217792
Variance	2.80137E-07	4.84072E-08
Observations	459	459
Hypothesized Mean Difference	0	
df	612	
t Stat	41.08844817	
P(T<=t) one-tail	2.07E-178	
t Critical one-tail	1.647347242	
P(T<=t) two-tail	4.1444E-178	
t Critical two-tail	1.963847792	

The main objective of this t-test is to determine the number of staff to be sent to the states depending on the number of Influenza related deaths. The test was done to see if increase in vulnerable population (ages 65+) leads to increase in Influenza related deaths. The result of the one-tailed (since we are checking if the number of populations increase leads to increase in death rate) t-test, $P = 2.07E-178$ which is much less than the significance value of 5%, we can reject the null hypothesis and confirm that death rate in vulnerable population is greater than death rate in non-vulnerable population.

Remaining analysis and next steps:

Additional Analysis:

- For the vulnerable population deaths of ages less than 5, the death numbers are suppressed (meaning assigned randomly between 1 and 10). A more accurate number is required for accurate analysis.
- The vaccination record of both vulnerable and non-vulnerable population would be very useful in determining if the vaccination has correlation with the number of deaths. If there is a negative correlation, then measures can be taken to vaccinate more people to protect from Influenza deaths.
- Historical data on the number of staff in each area, their burn out rate and no of days they had to take off due to sickness can help determine how to prepare staff for the Influenza season and allocate resources by taking into consideration the staff burn out rate and absenteeism due to sickness.

Visualization:

- A special visualization using Tableau needs to be created depicting the number of vulnerable deaths per State.

Appendix

Questions asked to gather requirements.

1. Business Requirements from the project brief to analyze past trends across states and demographics to determine how best to distribute medical staff.

2. Clarifying Questions:

1. What states have the highest rate of flu patients?
2. Which hospitals have the highest rate of flu patients?
3. When is the flu season?
4. What is the staff-to-patient ratio during non-flu and flu season?
5. Which hospitals have below the recommended level of staff-to-patient ratio?

3. Funneling Questions for Clarifying Questions:

- What states have the highest rate of flu patients?
 1. Are these states with the highest rate of flu patients regularly and consistently have the highest number of patients during all seasons?
- Which hospitals have the highest rate of flu patients?
 1. What is the wait time of the patient to be seen by a staff during flu season in these hospitals?

2. What is the mortality rate of flu patients in these hospitals?
3. What is the staff-to-patient ratio in these hospitals below the required/recommended level?

- When is the flu season?
 1. Is the flu season start and end during the same months every year?
 2. Is it during the same time in every state?
- What is the staff-to-patient ratio during non-flu and flu season?
 1. Is there a huge difference in the ratio?
 2. What is the wait time of the patient to be seen by a staff during flu and non-flu season?
- Which hospitals have below the recommended level of staff-to-patient ratio?
 1. How many of the seasonal staff are coming back for the next season?
 2. What are the average working hours of these staff?
 3. Is there a system to measure the burnout level and overall job satisfaction among the staff?

4. Privacy and ethics questions

1. Are there precautions taken to ensure that data that is being collected and used complies with the HIPAA (Health Insurance Portability and Accountability Act)?
2. Is the data being collected secure and not hackable?
3. Is the data about the demographics of patients namely race, economic condition collected in ethical manner?
4. If the patient is a minor, did the parents/guardian give consent to use the patient data?