Preparing for Influenza Season: Interim Report

As we prepare for the upcoming flu season, we look to the past to provide valuable insights to determine how to best support hospitals and clinics across the country by providing additional medical staff to communities with more vulnerable populations.

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 you are of retirement age or older (65+), then you are more likely to be hospitalized or die of influenza.
- If you live in a more densely populated area, then you are increasing your risk of contracting influenza.

Data Overview

CDC Influenza Death Count Data Set

> This data gives the influenza death count by US territories, 10-year age groupings, and year from 2008 - 2017.

US Census Bureau

> This data gives the population count per county in the US territories. It also breaks the population down into gender and 5-year age groupings from 2008 - 2017.

Data Limitations

CDC Influenza Death Count Data:

> The data is administrative data collected as part of the National Vital Statistics Cooperative Program. The death code is recorded by doctors and is limited to one code. This could mean that the death could have been influenced by an underlying health issue that we are unaware of. A limitation that my data is expressing is the death count is not broken down by county so I am only able to perform death rates per state, not county or city.
>
> It is also important to note that the CDC data suppresses data with a death count lower than 9 which may not be giving accurate death rates for states and age ranges with fewer deaths.

US Census Data

> This is administrative data collected by the US Consensus Bureau. Data is collected every 10 years across the nation on multitude of topics about the residence of the US. As noted in the project brief, this data contains a lot of estimates; the census data is not entirely precise. Additionally, most of the data is reliant on self-reporting or interview with a census representative, so there is the possibility that people will lie about their personal information or even not respond at all.

I also do not have vaccination rates on anyone over the age of 18 so I am unable to measure the effects of a vaccine on an older, more vulnerable population.

Descriptive Analysis

Death Rates:

Data Spread		
	0-64 yrs % death	65+ yrs % death
Variance	0.0000000000946	0.0000001770257
Standard Deviation	0.0000097260366	0.0004207442761

The spread of the death rates for those of 65+yrs of age is greater than those of ages 0-64. And on average the data for 65+yrs of age is .00042 units from the mean of .1%.

Avg 0-64 yrs	Avg 65+ yrs	Multiple
0.0015%	0.1000%	66.81192781

The average death rate for ages 0-64 is 0.0015% and the average death rate for those 65 years and older is 0.1%

Results and Insights

One clear result that my analysis produced is that the vulnerable population of people over the age of 65 are at a higher risk of dying from influenza.

Null Hypothesis: The death rate is the same for 65+yrs of age and 0-64 yrs of age.

Alternative Hypothesis: Age 65+ years will have a higher death rate than ages 0-64 yrs.

Results: Since the P value is 0.000 and we used a .05 significance level, this means that we can reject the null hypotheses. There is approximately 100% confidence death rates by age group was not due to random chance. *If fact, with further analysis, we found that people over the age of 65 are 66.8 times more likely to die of influenza than those younger than 64 (see above descriptive analysis).*

Additionally, in my second research hypothesis, I wanted to see how population my have an impact on influenza death rates to provide insights on where to send medical staff.

Null: Hypothesis: The death rate for states with more than 4 million is the same as the death rate for states with less than 4 million people.

Alternative Hypothesis: The death rate for states with more than 4 million is the greater than the death rate for states with less than 4 million people.

Result: Since the P value is 0.000 and we used a .05 significance level, this means that we can reject the null hypotheses. There is approximately 100% confidence death rates per larger populations was not due to random chance.

Note: *While my statistical test proved that I was able to reject my null hypothesis, I think it is important to note the correlation between these variables.*

Correlation		
Variables	Total Deaths and Total Population	Death Rate and Total Population
Correlation Coefficient	0.955991297	0.397366952
Strenght of Correlation	Strong Positive	Weak Positive
Useful/Interpretation	While the bigger the population, the more people who have died from influenza however there is not a strong correlation that the bigger the population the more likely (aka higher percentage) there is to death from influenza.	

Remaining Analysis and Next Steps

Remaining analysis includes but is not limited to identifying states that have the highest populations of vulnerable people, identifying if influenza season is at the same time across the country, and to ultimately give recommendations for the staffing agency to where to send medical staff. Next steps will be to do further analysis and to create a composition, statistical, spatial, and temporal visualizations as well as to make a final presentation of the results for stakeholders.

Appendix

Business Requirements Document (BRD)

Data Sources:

1. Influenza deaths by geography, time, age, and gender

   Source: CDC Data Set

2. Population data by geography

   Source: US Census Bureau Data Set

   NOTE: keep in mind that these numbers are estimates, hence the sum of the numbers from the different age groups may not sum up to the total in the first columns, but that's totally okay.

3. Counts of influenza laboratory test results by state (survey)

   Source: CDC (Fluview)

   Influenza Visits Data Set

   Lab Tests Data Set

Glossary:

Influenza: a contagious viral infection, often causing fever and aches.

Vulnerable populations: patients likely to develop flu complications requiring additional care, as identified by the Centers for Disease Control and Prevention (CDC). These include adults over 65 Page 2 years, children under 5 years, and pregnant women, as well as individuals with HIV/AIDs, cancer, heart disease, stroke, diabetes, asthma, and children with neurological disorders.

Assumptions:

I assumed the 50 US states that the staffing agency is providing staffing to does not include US territories so I did not include the records on District of British Columbia and Puerto Rico.

Previous Analysis:

1.3 Designing a Data Research Project (Initial steps on Project)

1.6 Data Qualities and Measures (Steps in cleaning process)

1.9 Statistical Hypothesis Testing (Clean and Integrated Data)