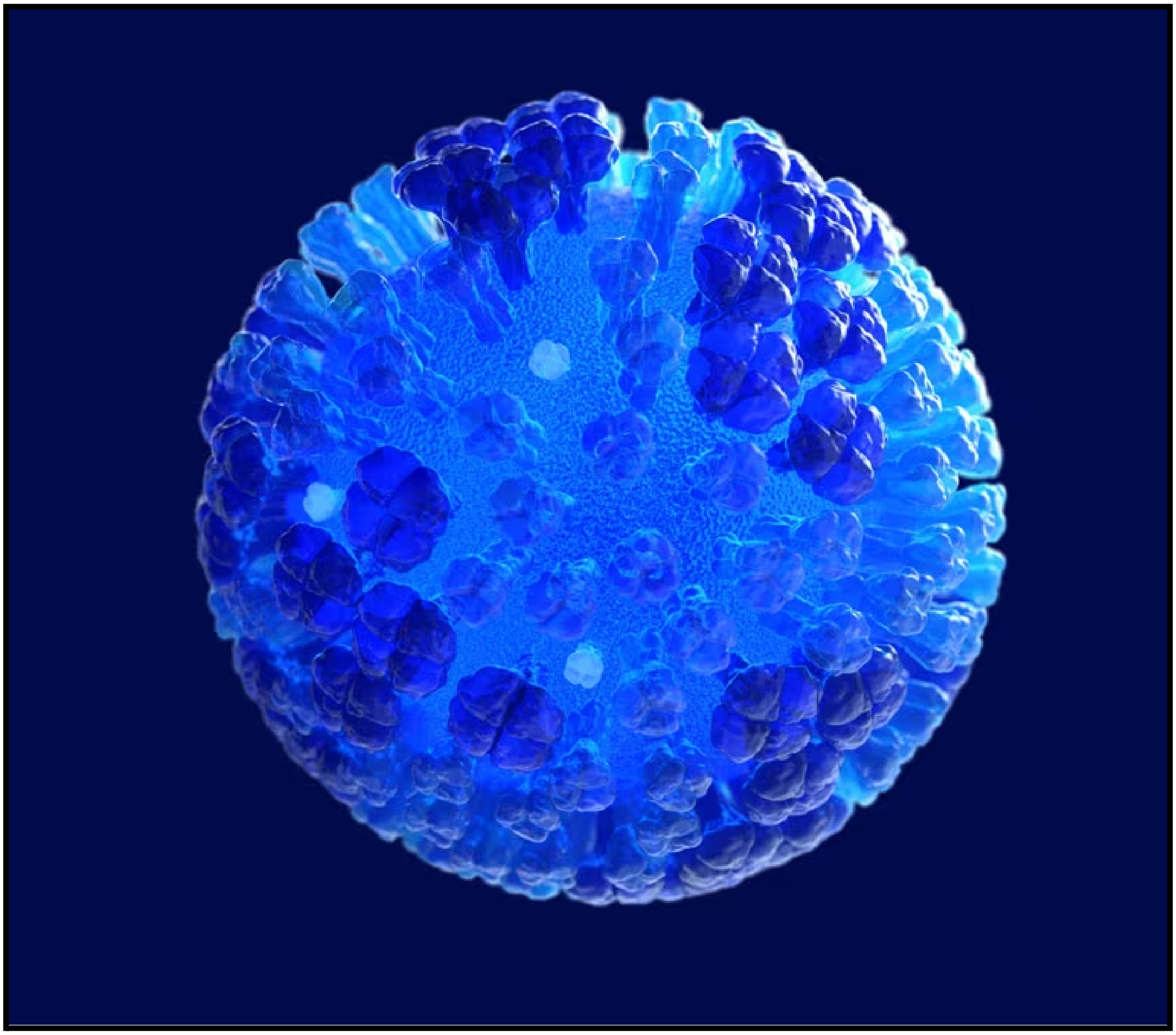
Influenza Season Prep Interim Report



Justin Kim

1.10 Assignment

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

1. If a person is an elderly person above the age of 65, then the person has a higher possibility of dying due to influenza.

2. If a person is a child under the age of 5, then the person has a higher possibility of dying due to influenza.

Data Overview

1. Data Set 1- Influenza deaths by geography, time, age, and gender
    a. Source: CDC
    b. Description - This data displays influenza deaths related to age group that includes the month, year, and state
2. Data Set 2- Population data by geography
    a. Source: US Census Bureau
    b. Description - This data displays influenza deaths related to age group that includes the month, year, and state

Data Limitations

1. Data Set 1- Influenza deaths by geography, time, age, and gender
    a. Source: CDC
    b. Limitations: One limitation about this data set is that it has missing data on

influenza deaths under 9, labeled as “suppressed”. Another limitation is that the data is recorded through death certificate recording, meaning non documented people in the US are not accounted for.

2. Data Set 2- Population data by geography
    a. Source: US Census Bureau
    b. Limitations: The first limitation in this data is that data is collected every 10 years.This creates estimates in between those years, and if there are changes within the data,it would take a long time to see a more accurate result. Another factor that limits the data is that not every resident in the US submitted data or lied about their entries,creating either a bias count or incomplete data for the census.

Descriptive Analysis Summary

Elderly (65+ years) Young Children (5 years or less)	Elderly Influenza Deaths	Elderly Population	Young Children Influenza Deaths	Young Children Population
Mean	904	947,082	108	445,539
Standard Deviation	980	1,110,750	14	545,122

RESULTS

In my analysis to determine the relationship between elderly population and elderly deaths, I calculated a correlation coefficient of 0.855274813. Because the Correlation Coefficient falls between 0.5 and 1, there is evidence that there is a strong relationship between the two variables. This is useful because it helps support our initial hypothesis in my 1.3 assignment that "If a person is an elderly person above the age of 65, then the person has a higher possibility of dying due to influenza." I can now use the data to confidently say that the data can support my hypothesis and I can make further insights and recommendations to send more healthcare staff to states with a higher elderly population".

Statistical Hypothesis and Interpretation

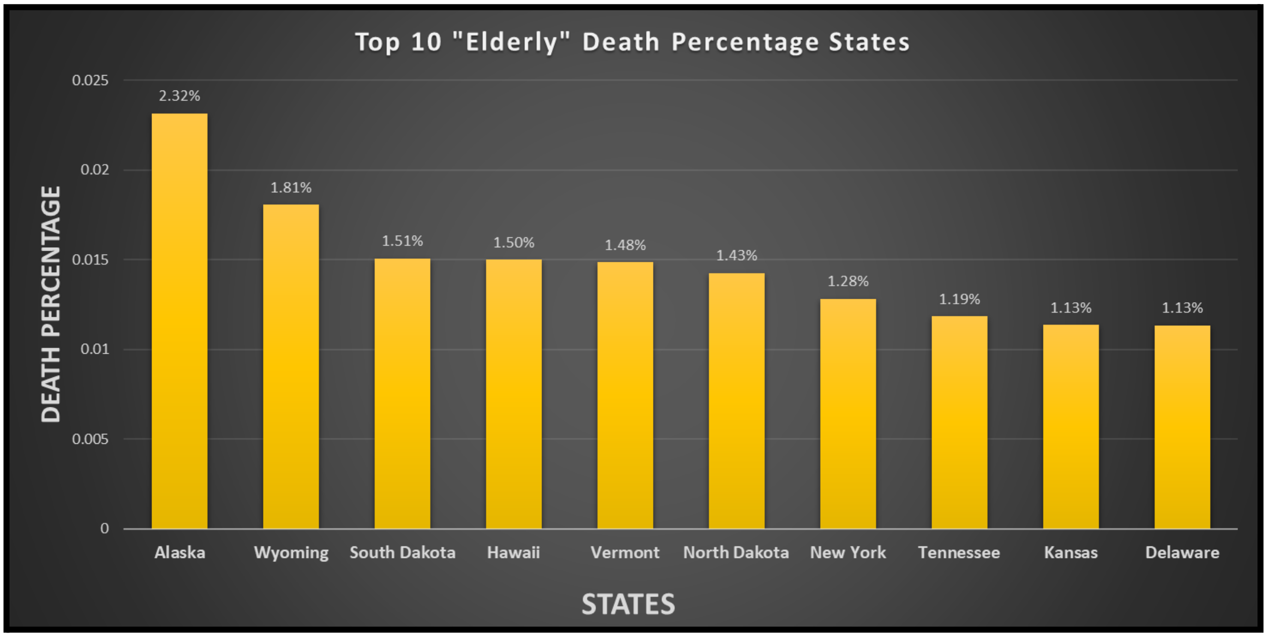
Elderly Population Hypothesis:

Null: The elderly age group (65+) death by influenza is equal to or less than deaths in adults (25-64).

Alternative: The elderly age group (65+) death by influenza is more than adults (25-64).

Significance Level: 0.05 or 5%

Results: Seeing that the p-value is 0.000000000000000000000000000000000000003585, it is safe to assume with 95% confidence that we can reject the null hypothesis and assume the alternative is correct. The elderly population is dying of influenza much higher than adults.



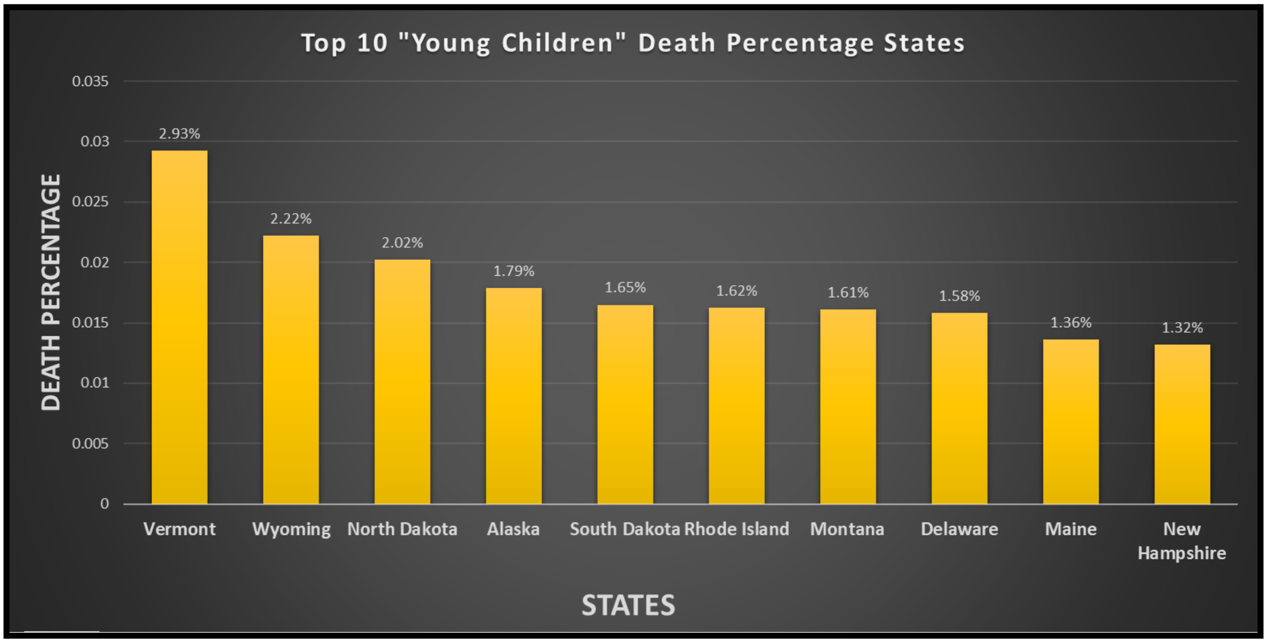
Young Children Population Hypothesis:

Null: The young children (5 years or less) death by influenza is equal to or less than deaths in adults (25-64).

Alternative: The number of young children (5 years or less) dying by influenza is more than adults (25-64).

Significance Level: 0.05 or 5%

Results: Seeing that the p-value is 0.000…0002897, it is safe to assume with 95% confidence that we can strongly reject the null hypothesis and assume the alternative hypothesis is true, meaning that young children are dying more to influenza than adults.



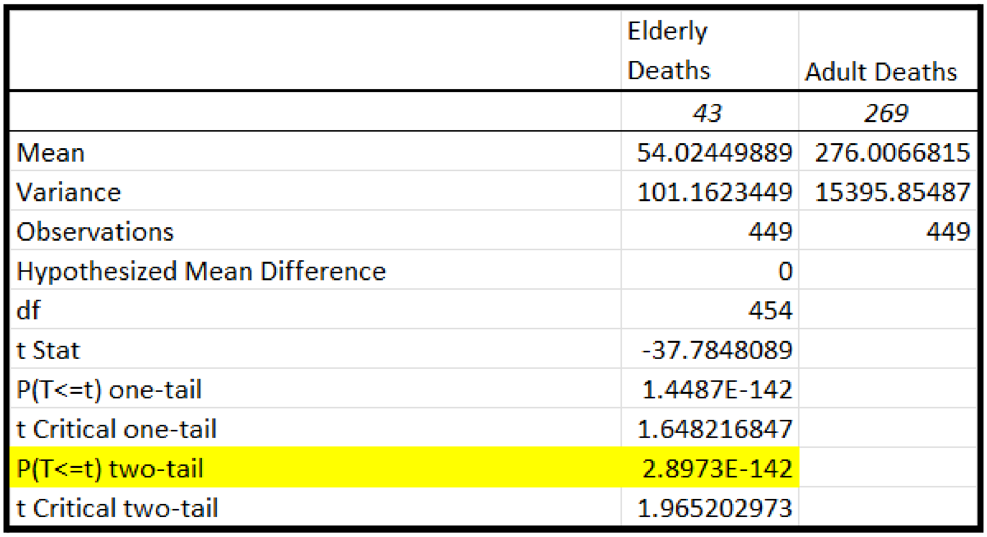
Remaining Analysis and Next Steps:

1. Analyze the correlation between young children population and influenza death rate
2. Using statistical hypothesis results to find states that have a high young children and elderly population
3. Connect data to a visualization program such as Tableau to create visuals and graphs to display findings
4. Create storytelling narrative to present to stakeholders

Appendix

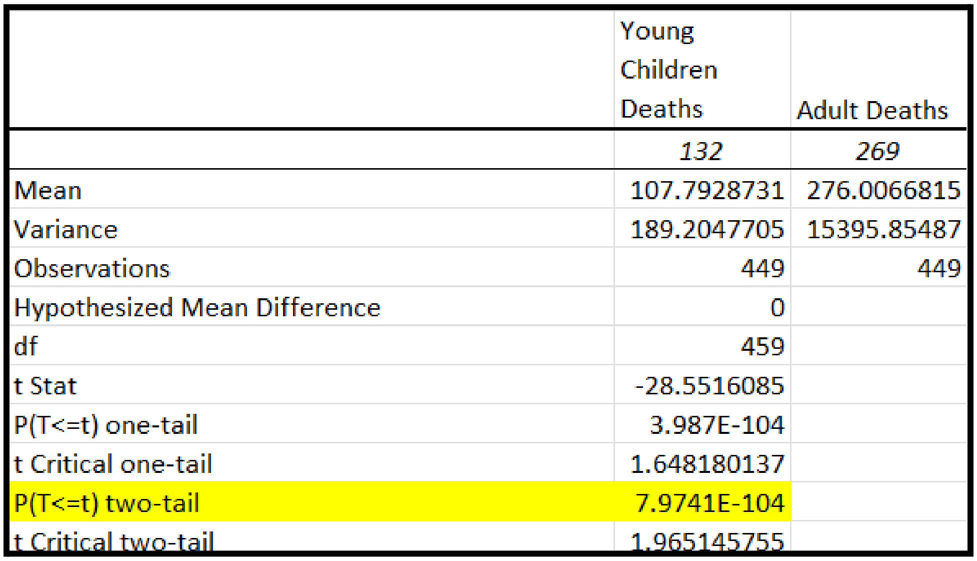
Detailed Results

Elderly Population Hypothesis:



	Elderly Deaths	Adult Deaths
	43	*269*
Mean	54.02449889	276.0066815
Variance	101.1623449	15395.85487
Observations	449	449
Hypothesized Mean Difference	0	
df	454	
t Stat	-37.7848089	
P(T<=t) one-tail	1.4487E-142	
t Critical one-tail	1.648216847	
P(T<=t) two-tail	2.8973E-142	
t Critical two-tail	1.965202973	

Young Children Hypothesis:



	Young Children Deaths	Adult Deaths
	132	*269*
Mean	107.7928731	276.0066815
Variance	189.2047705	15395.85487
Observations	449	449
Hypothesized Mean Difference	0	
df	459	
t Stat	-28.5516085	
P(T<=t) one-tail	3.987E-104	
t Critical one-tail	1.648180137	
P(T<=t) two-tail	7.9741E-104	
t Critical two-tail	1.965145755	

Additional Methodology Details

Elderly and Young Children Hypothesis:

1. State the null hypothesis and alternative hypothesis for each age group
2. Establish the significance level
3. Decide if the statistical test was a one or two tailed test
4. Use the following formula to find the t-statistic value:
    a. $t = (\bar{x} - \mu) / (s / \sqrt{n})$
        i. "t" is the test statistic.
        ii. "$\bar{x}$" is the sample mean.
        iii. "μ" is the hypothesized mean.
        iv. "s" is the standard deviation of the sample.
        v. "n" is the size of the sample.
5. Use the t distribution table to find the p-value
6. Determine if the p-value is strong enough to reject the null hypothesis and accept the alternative hypothesis

Assumptions and Limitations

1. For the CDC influenza data set, deaths counted under 9 were recorded as “suppressed”
    a. In order to solve this limitation, I used a random number generator to choose a random number from 1-9 to replace all of the suppressed entries.
2. For the Census Population data set, each count was recorded as a separate county within each set
    a. In order to solve this limitation, I combined each county into each respective state in order to integrate with the CDC Influenza data set.

Hypothesis Development:

Clarifying questions:

1. Who is most vulnerable to influenza in the United States?

2. How long is a typical influenza season?

3. Which type of weather conditions (rainy, sunny, cloudy, etc.) contributes to catching the flu?

Funneling questions:

Who is most vulnerable to influenza in the United States?

1. Are males more vulnerable to females? And vice versa?

2. Since influenza is spread through droplets when people cough, sneeze, or talk,does population density play a part in influenza spreading?

3. Which demographic of the population is admitted to hospitals due to flu symptoms.

How long is a typical influenza season?

1. Does the influenza season differ based on state?

2. Can the overall health of a state's demographic have an impact on influenza season length?

3. What percentage of a hospital's patients admitted with influenza is considered to be still in an "influenza season?"

Which type of weather conditions (rainy, sunny, cloudy, etc.) contributes to catching the flu?

1. Does weather play a big or small impact on catching the flu?

2. Is there a correlation between length of flu symptoms and weather?

Privacy and Ethics Questions:

1. Is it lawful to obtain medical reports and data on mortality in each state?

2. Since young kids are known to have a weaker immune system, is it ethical to obtain information on young kids that have been admitted to the hospital for influenza?