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.

Hypothesis

- People over the age of 65 are considered part of the vulnerable population who are at a higher risk of hospitalization due to influenza. Increased hospital staff leads to better quality of care and lower mortality rates. If hospitals in geographical areas with a higher average of this age group are provided with more staff, then mortality rates will decrease.

Data Overview

- Influenza Deaths by Geography, Time, Age, and Gender
    - This data source contains monthly death counts for influenza-related deaths in the United States from 2009 to 2017. Counts are broken into two categories: state and age

- Population Data by Geography
    - The data contains information about county & state population from 2009 to 2017 in terms of total population in relation to gender, and age broken down into 4 year increments.

Data Limitations

- Influenza Deaths by Geography, Time, Age, and Gender
    - A limitation of this data set is that death certificates can only list one cause of death. This can create discrepancies within vulnerable populations, while the cause of death could be related to another illness, the decline in health may have been initiated by influenza.

- Population Data by Geography
    - A limitation of this data is that it's only collected on an annual basis which can lead to a lag in timeliness. Also, depending on the survey method, there could be manual entry which could lead to errors.

Descriptive Analysis

Variable	Average	Standard Deviation
Total Population	5,924,386	6,743,019
65+ Population	801,596	878,339
Total Influenza Deaths	1479	1076
65+ Influenza Deaths	944	967

- The 65+ vulnerable age population has a strong correlation (.99) when looking at total deaths across all age groups, meaning that this age group has the highest mortality rate.

Results and Insights

- **Null Hypothesis:** Hospitals in geographical areas that have an average age of 65+ are less likely to have higher mortality rate

- **Alternative Hypothesis:** Hospitals in geographical areas that have an average age of 65+ are more likely to have a higher mortality rate.

- A two- tail sample t-test was done at an alpha of 0.05 or 95% confidence level. We found a significant difference in influenza deaths between the vulnerable populations and non-vulnerable population age groups.

Remaining Analysis and Next Steps

- Due to the high significance and confidence level in the 65+ age group in relation to influenza death rates, it is recommended to continue the analysis. There will be further visualizations delivered through Tableau Storyboard like bar charts, and how the variables are distributed, correlated, time forecasted, as well as a heat map to visually understand peak season locations.

- Next steps include a final presentation that includes the above visuals. Together, this will help outline a plan for the upcoming influenza season that the medical staffing agency can utilize to adequately staff hospitals and clinics across all 50 states. Based on the finding that the age group 65+ are more at risk for hospitalizations, one of the main focuses will be proactively addressing those geographical areas with additional staffing.

Appendix

Project Overview

1. Provide information to support a staffing plan, detailing what data can help inform the timing and spatial distribution of medical personnel throughout the United States.

2. Determine whether influenza occurs seasonally or throughout the entire year. If it is seasonal, does it start and end at the same time (month) in every state?

3. Prioritize states with large vulnerable populations. Consider categorizing each state as low-, medium-, or high-need based on its vulnerable population count.

4. Assess data limitations that may prevent you from conducting your desired analyses.

Hypothesis Development Section

- Who is most affected by influenza?
  - Is there an age group that has a higher rate of hospitalizations?
  - Is there a state with a higher concentration of vulnerable populations?
  - Is there an age group with a lower rate of vaccinations?
  - Are there states more affected by the flu than others?

- Which states have the highest and lowest vaccination rates?
  - Where are these states located?
  - Is there an age group that makes up the highest and lowest rates of vaccination?
  - Is the vaccination location accessible to everyone?
  - What other limitations prevent residents from getting the vaccine?

- Which state has the most residents in vulnerable populations?
  - How is the vulnerable population defined?
  - Where are these states located?
  - Do the state's governing agencies inform the population about flu season and potential risk of not receiving the vaccination?
  - Do states with higher vulnerable residents have higher rates of hospitalizations?

- When does flu season start?
  - Is the flu season the same in every state?
  - Does the flu season last the same amount of time in every state?
  - Does the peak season occur at the same time in every state?
  - Does the location of the state depend on how long the flu season will last?

Who is most affected by influenza?	Pregnant women at any stage Children between 6mo and 5yr Elderly Individuals > 65yr Individuals w/ chronic medical conditions Health-Care workers
Which state has the highest vaccination rate? Do these states have educational programs allowing for more accessibility to get vaccinated?	More Research Needed from Data Set
Which state has the most residents in vulnerable populations?	More Research Needed from Data Set
When does flu season start?	Depends on the state and climate, but primarily between December - March. Peak season is commonly in February and can continue into March
What time of year are influenza shots offered?	CDC recommends to get the shot in early fall but ideally by the end of October before the season begins
What time of year has the most hospitalizations for influenza?	More Research Needed from Data Set

Data Overview- Profiles of Data

1. Influenza Deaths Profiles

Counts Expected from the variables	
State	1296
State Code	1296
Year	7344
Month	612
Month Code	612
Ten-Year Age Groups	5508
Ten-Year Age Groups Code	5508
Death	Total is 66096

	State Code	Year	Death
Minimum	1	2009	10
Maximum	56	2017	512
Mean	29	2013	34

Data Types				
Variables	**time -variant/ invariant**	**structured/ unstructured**	**qualitative/ quantitative**	**qualitative: nominal/ordinal quantitative: discrete/continuous**
State	Invariate	Structured	Qualitative	Nominal
State Code	Invariate	Structured	Qualitative	Ordinal
Year	Invariate	Structured	Qualitative	Ordinal
Month	Invariate	Structured	Qualitative	Ordinal
Month Code	Invariate	Structured	Qualitative	Ordinal
Ten-Year Age Groups	Invariate	Structured	Qualitative	Ordinal
Ten-Year Age Groups Code	Invariate	Structured	Qualitative	Ordinal
Death	Variant	Structured	Quantitative	Discrete

2. US Census Data Profiles

Variables	Maximum	Minimum	Mean
Year	2017	2009	2013
Total Population	10,105,722	41	97,842
Male Total Population	4,979,641	23	48,122
Female Total Population	5,126,081	15	49,721
Under 5 Years	733,897	0	6,309
5 to 9 Years	655,400	0	6,378
10 to 14 Years	724,112	0	6,496
15 to 19 Years	753,657	0	6,804
20 to 24 Years	777,988	0	6,909
25 to 29 Years	831,276	0	6,714
30 to 34 Years	762,619	0	6,436
35 to 39 years	753,468	0	6,345

40 to 44 Years	733,897	1	6,614
45 to 49 Years	704,718	0	6,901
50 to 54 Years	683,887	0	6,922
55 to 59 Years	628,513	3	6,293
60 to 64 Years	535,357	0	5,416
65 to 69 Years	415,243	0	4,189
70 to 74 Years	295,420	0	3,129
75 to 79 Years	215,181	0	2,394
80 to 84 Years	161,647	0	1,821
85 Years and Over	177,493	0	1,774

Data Types				
Variables	**time -variant/ invariant**	**structured/ unstructured**	**qualitative/ quantitative**	**qualitative: nominal/ordinal quantitative: discrete/continuous**
Country	Invariant	Structured	Qualitative	Nominal
Year	Invariant	Structured	Qualitative	Ordinal
Total Population	Variant	Structured	Quantitative	Discrete
Male Total Population	Variant	Structured	Quantitative	Discrete
Female Total Population	Variant	Structured	Quantitative	Discrete
Under 5 Years	Variant	Structured	Quantitative	Discrete
5 to 9 Years	Variant	Structured	Quantitative	Discrete
10 to 14 Years	Variant	Structured	Quantitative	Discrete
15 to 19 Years	Variant	Structured	Quantitative	Discrete
20 to 24 Years	Variant	Structured	Quantitative	Discrete
25 to 29 Years	Variant	Structured	Quantitative	Discrete
30 to 34 Years	Variant	Structured	Quantitative	Discrete
35 to 39 years	Variant	Structured	Quantitative	Discrete
40 to 44 Years	Variant	Structured	Quantitative	Discrete

45 to 49 Years	Variant	Structured	Quantitative	Discrete
50 to 54 Years	Variant	Structured	Quantitative	Discrete
55 to 59 Years	Variant	Structured	Quantitative	Discrete
60 to 64 Years	Variant	Structured	Quantitative	Discrete
65 to 69 Years	Variant	Structured	Quantitative	Discrete
70 to 74 Years	Variant	Structured	Quantitative	Discrete
75 to 79 Years	Variant	Structured	Quantitative	Discrete
80 to 84 Years	Variant	Structured	Quantitative	Discrete
85 Years and Over	Variant	Structured	Quantitative	Discrete

Results and Insights

t-Test: Two-Sample Assuming Unequal Variances		
	<65 Years	***65+ Group***
Mean	0.002228582	0.006549856
Variance	5.41115E-06	6.19515E-06
Observations	468	468
Hypothesized Mean Difference	0	
df	930	
t Stat	-27.44022658	
P(T<=t) one-tail	3.2371E-122	
t Critical one-tail	1.646493732	
P(T<=t) two-tail	6.4743E-122	
t Critical two-tail	1.96251808	

- α = 0.05 or 95% confidence level
- P-Value = 6.4743E-122
- P < α = can reject the null hypothesis