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.

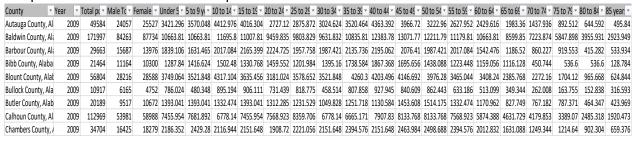
Hypothesis:

"If any state has more vulnerable population (including people over 65 and under 5), then there are high chances of influenza outbreak."

Data Overview:

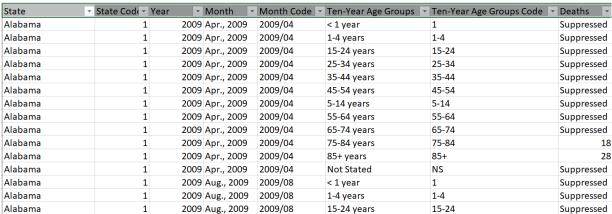
Population Data by geography:

The data is an external source that comes from the U.S Census Bureau. The data contains total population, female/male population, age group by state and county from 2009 to 2017. Snapshot of dataset attached for quick reference:



Influenza deaths by geography, time, age, and gender:

The data is an external source that comes from The Center for Disease Control (CDC). The data for influenza visits tracks patient visits to a medical provider for influenza. It tracks the number of visits, number of providers, and total patients seen from 2010 to 2019 from 3,500 outpatient healthcare providers. Snapshot of dataset attached for quick reference:



Data Limitations:

Population Data by geography:

The survey data is entered manually and is likely to contain unnecessary noise due to errors and omission.

➤ Influenza deaths by geography, time, age, and gender:

In Influenza Deaths data is that the death certificate of patients lists only one cause of death. This could create some discrepancies within vulnerable populations, such as those with AIDs—while the cause of death may be related to AIDs, their decline in health may have been initiated by influenza. However, the data is government administrative data and can be considered trustworthy. Another key point that makes data bias is that count of deaths less than 10 deaths are marked as suppressed.

Descriptive Analysis:

In order to test our project hypothesis relevant datasets were collected, cleaned and integrated together to perform statistical analysis and determine whether those relationships are meaningful—or if they occur by chance. The snapshot of statistical analysis is as under:

Data Spread						
Data Set	Deaths over and	Population over	Deaths less	Population less		
Name	above 65 years	and above 65 years	than 5 Years	than 5 Years		
Mean*	565	268417	55	385316		
Sample or						
Population	Sample	Sample	Sample	Sample		
Normal						
Distribution	Normal	Normal	Normal	Normal		
Variance	149113.0793	1.25354E+11	107.3600887	2.03141E+11		
Standard						
Deviation	386.1516273	354053.6567	10.36147136	450712.0894		
2*Standard						
deviation	772	708107	21	901424		
Upper	1338	976525	76	1286740		
Lower**	-207	-439690	34	-516108		
Total						
Records	459	459	459	459		
Outliers	34	61	8	19		
Outlier						
Percentage	0.07	0.13	0.02	0.04		

Note:

- a) *Mean is calculated on total of deaths (i.e., for the period 2009- 2017) and total population (i.e., for the period 2010- 2019).
- b) **Lower range values are shown in negative since Mean is lower than SQRT of Standard Deviation.

Correlation Coefficient:

Variables	Deaths and Population >65 vears	Deaths and Population <5 vears
Proposed Relationship	Since population >65 is vulnerable to death from influenza, the correlation between these two variables should be high	Since population <5 is vulnerable to death from influenza, the correlation between these two variables should be high
Correlation Coefficient	0.93	0.01
Strength of Correlation	Strong	Very Low
Usefulness/Interpretation	This is a helpful statistic as it supports the hypothesis. It shows a very strong correlation between a state's population over 65 and its number of influenza deaths for people over 65. In other words, it supports the high rate of influenza mortality for this age group	This may appear to be a surprising result, as children under 5 are considered a vulnerable population. However, in the original dataset, random values were inputted for this population, so this statistic isn't truly representative.

Result and Insights:

At an alpha of 0.05, or confidence level of 95 percent, we can reject our null hypothesis and state that the influenza death rate for individuals in age group of 65 years and above is greater than the death rate for individuals in age group of less than 65 years. The snapshot of Statistical testing is attached in the table below:

Statistical Testing				
Hypothesis	If any state has more vulnerable population (including people over 65 and			
to test:	under 5), then there are high chances of influenza outbreak.			
Independent				
Variable:	Proportion of Vulnerable individuals (above 65 and under 5)			
Dependent				
Variable:	Influenza Death Rate			
Null	The death rate for individuals 65+ years of age is less than or equal to the			
Hypothesis:	death rate for individuals less than 65 years of age.			
Alternative	The death rate for individuals 65+ years of age is greater than the death rate			
Hypothesis:	for individuals less than 65 years of age.			
	One-tailed test because we are only interested if the sample mean is higher or lower than the population mean - only interested in one direction not both			
T-test type	simultaneously.			
Significance	Simultanously.			
Level	Alpha = 0.05			
P-Value	0.000000000000096			
Significance				
Level				
Assessment	This p-value is significantly less than 0.05. Therefore, we can reject our null			
:	hypothesis.			

t-Test: Two-Sample Assuming Unequal Variances				
	Total Deaths >=65	Total Deaths <65		
Mean	849.8779956	528.8082789		
Variance	690632.6095	74047.36054		
Observations	459	459		
Hypothesized Mean Difference	0	-		
df	555	-		
t Stat	7.866215327	-		
P(T<=t) one-tail	9.60061E-15	-		
t Critical one-tail	1.647603773	-		
P(T<=t) two-tail	1.92012E-14	-		
t Critical two-tail	1.964247525	-		

Remaining Analysis and Next Steps:

Based on statistical analysis insights, the next step of our analysis would be to dig in deeper and extract valuable insights from our dataset and help stakeholders prepare their staffing plan for the upcoming Flu season. The spatial aspect plays most important role in Project's primary goal i.e., to ascertain which regions are more likely to have vulnerable population (including people over 65 and under 5) and in our previous exercise we focused mainly on formulating and statistically analysis our hypothesis that vulnerable patients are at high risk of Flu.

However, we have not addressed the following geographically yet. Now, we should look at the geographical representation of vulnerable populations and analyze medical facilities available in these regions to plan for medical staff and prepare accordingly for the upcoming Flu Season.

With the help of Data Visualizations, we can effectively compare density of Vulnerable patients across each state with total number of deaths due to Flu and draw insights from it. To do so we will be working on powerful data visualization and analytics tool, **Tableau**. Adding a visualization would make the comparison much more readable, and it would be easier to see the states with a large contrast between vulnerable populations and available staff.

As a final step, we will be presenting final results extracted through combination of statistical analysis and visualization outcome in Powerpoint Presentation to our stakeholders to enable them to make informed decisions and combat upcoming pandemic.

Appendix:

CDC Influenza Deaths Data

https://coach-courses-

us.s3.amazonaws.com/public/courses/da program/CDC Influenza Deaths edited.xlsx

Census Population Data

https://coach-courses-us.s3.amazonaws.com/public/courses/data-immersion/A1-A2_Influenza_Project/Census_Population_transformed_202101.csv