Preparing For Influenza Season - Interim report

Sruthy Sreekanth

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.

Research Hypothesis:

- If a person is 65 years or older, they have a higher risk of dying from influenza.

Data Overview:

- **Influenza Death Data set**: This data is provided by CDC. It contains monthly death counts for influenza-related deaths in the United States from 2009 to 2017, state, state code, deaths, and age groups.
- **US Census Data Set**: This data is provided by the US Census Bureau**.** Populations per county and state are included in the data for year between 2009 and 2017.The whole population is categorized into gender and age groups.

Data Limitations:

- **Influenza Death Data set: As** survey data, **it doesn't** represent an accurate count of all laboratory tests for influenza in the US. Since this data is manually gathered, it can be subject to biases and typing errors. The data set appears to have insufficient information**.** Example: State- Puerto Rico data missing in Influenza dataset. However, the data won't be accessible until recent years if we're aiming for the year 2023.
- **US Census Data Set**: Since this data is manually gathered, it can be subject to biases and typing errors. Additionally, these numbers are estimations, therefore the whole population may not equal the total of the numbers from the various age categories. However, the data won't be accessible until recent years if we're aiming for the year 2023.

Descriptive Analysis:

Below table shows the descriptive study for two variables for Vulnerable populations- **Population >=65 Years** and **Deaths>=65 year**

Data Spread for Vulnerable population		
	Population >=65 Years	**Deaths>=65 year**
Dataset Name	US Census Data Set	Influenza Death Data set
Variance	128529399318	151233
Standard Deviation	358510	389
Mean	268996	299
Count of Outliers	61	65
Outlier percentage	4.43%	4.72%

There is a strong correlation between these two variables, as seen in the table below. Therefore, the research's hypothesis seems to be valid. Here the data points that are more than two standard deviations from the mean are considered as outliers.

Variables	**Death & population >=65 years**
Proposed relationship	As people of 65 years or older are more likely to die from influenza, there will be a high correlation between these two variables.
Correlation coefficient	0.94
Strength of Correlation	Strong Correlation

Results and Insights:

STATISTICAL HYPOTHESIS AND TESTING	
Research Hypothesis	If a person is 65 years or older, they have a higher risk of dying from influenza.
Dependent Variable	Influenza death rate (<65 year and >=65 year)
Independent variable	Age groups (<65 year and > =65 years)
Null Hypothesis:	The Influenza death rate of people of 65 years or older is less than or equal to the death rate of people under 65 years
Alternative Hypothesis	The Influenza death rate of people of 65 years or older is greater than the death rate of people under 65 years
Two-tailed or one-tailed test?	One-tailed Test. Because we are testing that the influenza death rate of people of 65 year or older is less than **or** equal to death rate of people under 65 years. We are not testing whether it is greater. We are only interested in one direction
Alpha	0.05
p-value	1.7204E-171
Significance level	In this case, the P-value is much lower than the significance level(alpha). So, our null hypothesis can be rejected. We can confirm with 95% confidence that the risk of influenza-related death is higher in people of 65 years or older.

Next Steps	It has been confirmed that people of 65 years or older have a significantly increased chance of dying from influenza. The states with the highest percentage of residents of 65 or older and the highest death rate can now be examined. This will help us decide where to send front-line staffs. The resources that states already have should also be taken into consideration.

Remaining Analysis and Next Steps:

- Create visual charts, so that stakeholders can understand the analysis done for the vulnerable population and the Influenza death rate.
- Using a pie, bar, or column chart, create visualizations that show the distribution of a variable and the correlation between variables like vulnerable population and death rate.
- Before the final meeting, complete the report and share it to the stakeholders for review.
- Prepare a final presentation for your stakeholders.

Appendix:

Hypothesis Development:

Clarifying and funnelling questions:

- How successful were flu shots in each state in the past?
  - Do flu shots prevent people of all ages from the flu? How effective are flu shots for children, seniors, and other people in each state?
  - Does getting the flu shot help persons with all medical conditions? How effectively do flu shots work for both healthy and vulnerable people in each state?
  - What correlation existed between state-by-state flu shot rates and flu severity rates? Is there any negative correlation?
- How do we measure the flu's severity?
  - Does it depend on the rate of death?
  - What is the vulnerability population's death rate in each state with and without a flu shot?
  - What is the healthy population death rate in each state, both with and without a flu shot?
- Which states experience understaffing the most during flu season?
  - Are the understaffing states have the most flu deaths?
  - Does understaffing only occur during the flu season?

- Which states are overstaffed during flu season? Is the death rate lower than states with understaffing?

Questions concerning Privacy and Ethics:

- Do the people who provided the information have their consent for the data to be shared?
- Is the data entirely accessible to everyone, or are there some restrictions?
- What steps must be taken to collect more information if necessary?

Data Profile:

Influenza Death Data set:

		Data Types			
Variables	Description	time -variant/-invariant	structured/ unstructured	qualitative/quantitative	qualitative: nominal/ordinal quantitative: discrete/continuous
State	States where influenza deaths occurred	Time -invariant	Structured	Qualitative	Nominal
State Code	Codes for 51 States	Time -invariant	Structured	Qualitative	Nominal
Year	Influenza death data from year 2009 to 2017	Time -invariant	Structured	Quantitative	Discrete
Month	Month data from Year 2009 to 2017	Time -invariant	Structured	Quantitative	Discrete
Month Code	Month codes from Year 2009 to 2017	Time -invariant	Structured	Qualitative	Ordinal
Ten-Year Age Groups	Age of patient categorized in various intervals	Time -invariant	Structured	Qualitative	Ordinal

Ten-Year Age Groups Code	Age group codes	Time -invariant	Structured	Qualitative	Ordinal
Deaths	Count of deaths by influenza	Time-variant	Structured	Quantitative	Discrete

US Census Data Set:

		Data Types			
Variables	Description	time - variant/-invariant	structured/unstructured	qualitative/quantitative	qualitative: nominal/ordinal quantitative: discrete/continuous
County	Name of the county	Time - invariant	Structured	Qualitative	Nominal
Year	Year from 2009 to 2017	Time - invariant	Structured	Quantitative	Discrete
Total Population	Population of the county	Time-Variant	Structured	Quantitative	Discrete
Male Total Population	Male population of the county	Time-Variant	Structured	Quantitative	Discrete
Female Total Population	Female population of the county	Time-Variant	Structured	Quantitative	Discrete
5-Year Age Groups	Age categorized in various intervals	Time-Variant	Structured	Quantitative	Discrete
State	State where the county is located	Time - invariant	Structured	Qualitative	Nominal