Preparing & Analyzing Data
1.10: Consolidating Analytical Insights

Motivation:

In the United States, the influenza season is characterized by extremely high flu cases, when people more than normal, get infected with the virus. Some individuals, particularly those from high-risk groups, experience major health difficulties and are assigned to medical institutions. In order to cover high demands for treatments, hospitals and clinics will require additional personnel to appropriately handle these additional patients. This temporary workforce is provided by a medical staffing service, therefore, it is required to understand the amount of this personnel to be prepared when sick cases will reach its pick.

Objectives:

Determine the amount of staff for each US state.

Scope of Work:

The agencies which provide temporary workforce operate in hospitals in all 50 US states, and the effort will prepare for the forthcoming influenza season which is normally taking place from November until March (may slightly vary from state to state) with its pick in December-January each year.

Research Hypothesis:

Null Hypothesis: *If the state has more inhabitants who belong to vulnerable groups (children under 5 years old and elderly people 65+), then mortality rates will be higher.*

Alternative Hypothesis: *If the state has more inhabitants who belong to vulnerable groups (children under 5 years old and elderly people 65+), then there will be no influence on mortality rates.*

The assumption is that the amount of the population which is considered a risk - group has an influence on the mortality rate as influenza spreads from person to person via airborne throughout the contacts and these segments are the most sensitive due to health reasons. The group “under 5 years old” and “Seniors 65+“were chosen as one of the most vulnerable segments.

Data Overview

US Census Data:

The data set contains the populations of every US state, broken into 5-year age groups (which have been consolidated to focus on the two primary groups being researched), with data from 2009 - 2017. Additional US Census data detailing both urbanization percentage and population density for every US state in 2010 has been integrated as well.[1]

CDC Influenza Deaths Data:

This data set includes influenza deaths reported by every US state, broken into 5 and 10-year age groups (which have been consolidated to focus on the two primary groups being researched), with monthly data from 2009 - 2017. In conjunction with the US Census Data, these deaths have been calculated as a percentage of the total population within their corresponding age groups.[2]

Data Limitations

The fact that the information given by two data sets is no more current, as it shows the timeline from 2009 until 2017, hence, the tendency can be observed in this period, which limits all data sets. Since then, the covid-19 pandemic has had a significant influence on both yearly flu outbreaks and the demographic situation of the US population, in all age groups.

US Census Data:

Nearly 15% of all demonstrated data elements, or a significant number of observations for certain states of different years, were missing from this dataset. Although it may be expected that the data provided, coming from the US government, which is a trustworthy source, values were interpolated for every missing case using the original data. With the significant exception of the effects of the COVID-19 pandemic situation, it may also be assumed that overall population trends have remained stable throughout the years since statistics have been unavailable and estimated deaths were compensated through immigrant acceptance.

CDC Influenza Deaths Data:

For the states that record flu fatalities, this data from the US government may be regarded as accurate because it was obtained from the indicated cause of death reflected in The Death certificates. However, neither Alaska nor Washington, DC have reported any flu-related fatal cases in this data set. Additionally, data was withheld due to privacy concerns for any month in which there were less than 10 fatalities for a particular age group in a given state. Therefore, any state that claims no influenza

[1] https://www.census.gov/
[2] https://www.cdc.gov/

deaths, may not genuinely be so, but rather represent a very tiny number. Although Washington, DC, and Alaska may not have any documented deaths, this is perhaps the case. The same types of restrictions were applied to this data collection.

Descriptive Analysis

The bigger population group, then more death occurred, more death occurred in the very young group and very old group - Risk groups. The correlation is supposed to be strong – 0,95

The statistical analysis shows that there is a high correlation between influenza death rates and vulnerable groups (children under the age of five). Because of this correlation, states with a larger number of vulnerable people are more likely to have higher rates of influenza-related mortality. As a result, they will require more medical personnel throughout the influenza season.

Correlation	**Variable 1**	**Variable 2**
Variables	Census Population Data	Influenza Deaths
Correlation coefficient	**0,95**	
Strength of Correlation	The correlation is extremely strong	
Data Spread	**Variable 1**	**Variable 2**
Dataset Name	Census Population Data	Influenza Deaths
Sample or Population	Sample	Sample
Normal Distribution?	Yes	Yes
Variance	46330288139601	1185438
Standard Deviation	6806636	1089
Mean	5973696	1379
Outlier Percentage	88,02%	100,00%
Standard Deviation	13613271	2178
Standard Deviation Upper Limit	12780332	2468
Standard Deviation Lower Limit	832940	-290
Count > 2 SD	404	459

Outliers are within two standard deviations of the mean in 100% of influenza death and 88,02% of population count variables. Outliers include residents of New York, California, Texas, Illinois and Florida. These states have much more inhabitants. As a result, they will not be deleted from the dataset.

Results and Insights

The analysis has revealed that a group younger than 5 years old and those people who are over the age of 65 die from flu more frequently (Null Hypophysis) than the other group 5 to 64 years old (Alternative Hypophysis).

A (two-tail) hypothesis was tested to determine if influenza mortality in the age groups of “under 5 and 65+ years old” is equal to or different from those in the age group of 5-64 years old (null hypothesis).
As a result, P value shows 8, 08274E-61 (see Appendix for details. This means that the vulnerable segment will face the death situation more likely than the other non-vulnerable group.

Therefore, it is recommended to reinforce the medical institutions with more labor in the regions with higher population density and where these vulnerable groups prevail. Also, the states with lower vaccination rates would require more medical staff for the patients.

Remaining Analysis and Next Steps

Three important factors—vaccination status, population density, and urbanization rate—have been successfully isolated from the consideration whereas these factors play a significant role in identifying further needs and probabilities. Initial attempts to weight those variables and correlate them with flu mortality have yielded encouraging results, but additional work may be done to fine-tune the weighting method so that it better reflects the distribution of flu deaths. Which states will require more medical personnel throughout this year's flu season will depend on the relative intensity of these elements and the size of their population.
The other variables should be taken into consideration e.g. the level of medical services and medical services availability of this group; The climate aspects (e.g. the flue cases in the warmer climate happened not so frequently which leads to lower mortality rates), the social behavioral features (if people have enough will to cure when there are symptoms, do they have medical insurance).

Appendix

Project Goals:

To organize a viable strategy that will assist staffing agencies to restrain the spread of infection, protection of the population, and involvement a sufficient amount of personnel.

Audience and Stakeholders:

Medical personnel (doctors, nurses)

Health Care Institution Management and Administration

Patients (potential, current)

Staffing agencies

Data Wishlist:

- Flu mortality rates in each state divided by age groups.
- Flu vaccination rates in each state divided by age groups.
- Flu hospitalization rates.
- The data regarding days spent in Hospital during flu season by each age group.
- The number of hospitals in each state and their capacity.
- Information regarding Healthcare staff.

t-Test: Two-Sample Assuming Unequal Variances

	Vulnerable Population	***Death Rates of Vulnerable Population***
Mean	1195222,45	997,2925764
Variance	1,76622E+12	954055,0696
Observations	458	458
Hypothesized Mean Difference	0	
df	457	
t Stat	19,23074328	
P(T<=t) one-tail	4,04137E-61	
t Critical one-tail	1,648194724	
P(T<=t) two-tail	8,08274E-61	
t Critical two-tail	1,965168491	

Sources:

1. Influenza deaths by geography, time, age, and gender Source: CDC
2. Population data by geography Source: US Census Bureau
3. Counts of influenza laboratory test results by state (survey) Source: CDC (Fluview)