Preparing for Influenza Season

Data Analysis – Interim Report

Table of Contents

1. 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.

2. Hypotheses

a. If a state has a high number of populations, it will have more flu death.
b. If you are over 65, then you are likely to have flu death.
c. If state has more people who got vaccinated, then it will have the lower flu deaths.
d. If the temperature goes down, more flu deaths will occur.
e. Influenza virus can spread. If a location has a high travel traffic, it will likely to have more flu patients.
f. People in rural areas have relatively lower education levels in comparison to the people in urban areas.

3. Data Overview

Following data sets were used for the analysis:

Name	Source (Owner)	Trustworthy Level	Collection Method	Overview
Influenza deaths by geography, time, age, and gender	External (CDC)	Trustworthy (government)	• Administrative data collected as part of the National Vital Statistics Cooperative Program	Monthly death counts for influenza-related deaths in the United States from 2009 to 2017. Counts are broken into two categories: • State • Age
Population data by geography	External (US Census Bureau)	Trustworthy (government)	• Combination of administrative and surveys – while some census data are collected directly from the respondents via census surveys, the primary additional data source are federal, state, and local governments and commercial entities according to US Census Bureau • Cannot conclude whether the data was collected manually or automatically; however, it is likely that both methods were used (manual collection of surveys and automatic feeds from government and commercial entities) • Data coverage – 2009 to 2017	Total census population data with the following breakdowns • Year (2009-2017) • County, State • Age buckets (increments of 5 years) • Gender
Influenza Vaccination Coverage for All Ages 6 months and up	External (CDC)	Trustworthy (government)	• Survey of sample population • Manual collection/consolidation	• Vaccination % by various geography level from the period of 2018 to 2022
Historical monthly average temperatures by state	External (NOAA National Centers for Environmental information)	Trustworthy (government)	• Administrative • As part of the government agency's operations • Appears to be that the data was collected/consolidated automatically by/from local weather stations	• Historical average monthly temperatures from January 2009 to December 2017 by state

4. Data Limitations

Following are the limitations related to the data sets that were used:

Data	Limitations
Influenza deaths by geography, time, age, and gender	• **Accuracy**: Deaths on a death certificate only list 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
Population data by geography	• **Time lag**: Data appears to be collected yearly, while US Census mentions the collection is done through census surveys and through government and commercial entities, the timing of the data collection is not defined and it is highly likely that the data is not updates on a real-time basis • **Completeness**: As part of the collection method is using the manual household level surveys, the data may not be complete if candidates did not elect to respond
Influenza Vaccination Coverage for All Ages 6 months and up	• **Accuracy**: Given that the collection method was using the manual sample surveys, the accuracy is questionable (e.g., candidates who did not get vaccinated responding that they were vaccinated) • **Completeness**: Given that the collection method was using the manual sample surveys, the data may not be complete if candidates did not elect to respond
Historical monthly average temperatures by state	• **None noted**: historical weather data collected from weather stations and appears to be consolidated in an automated fashion

5. Descriptive Analysis:

Category	Data Sets							
	Temperature X Flu Death		Flu Death X Vaccination Rate		Flu Death X Census Population		Flu Death X Age	
	Variable 1	Variable 2	Variable 1	Variable 2	Variable 1	Variable 2	Variable 1	Variable 2
Variable Name	Temperature (in F)	# Flu Death	FluDeathRate (%)	VaxRate (%)	FluDeaths	Population	Age (artificial figure assigned based on age bucket)	# Flu Death
Variable Value Arrangement/ Grouping Method:	Average temperature by month (using data for all states)	Total Influenza Deaths by each month (using data for all states)	For each year and state, total flu deaths/total population	For each year and state, # of samples who reported vaccinated (previously calculated based on %) / # of total sample population	Total influenza deaths for each state for the entire scope period (2009-2017)	Total census population for each state for the entire scope period (2009-2017)	The data was already grouped and re-arranged by age buckets previously. Artificial figures based on age bucket (e.g., 25-34 years old = 29.5) assigned to analyze the trend between age and flu deaths	Total census population for each age bucket and each state for the entire scope period (2009-2017)
Average	53.05	3823.09	0.01%	33.91%	7,988.82	53,321,843.57	44.80	88.76
Standard Deviation	15.43	1221.05	<0.0001	0.05	10,300.13	60,665,968.57	28.26	252.40
Correlation Coefficient	-0.75		-0.04		0.96		0.45	
Relationship usefulness	The temperature had a strong inverse correlation (-0.75) with the number of influenza deaths, meaning that colder the temperature more influenza deaths were reported.		The vaccination rate had almost no correlation (-0.04) with the influenza death rate, meaning that the people's vaccination did not impact the number of influenza deaths occurred for the period/state.		The number of influenza deaths had a very strong correlation (0.96) with the number of populations that a state has, meaning that the higher the state's overall population figures, there will be more people dying from influenza		This analysis was conducted as a step to test one of my hypotheses where the people over 65 will have higher chances of dying from flu. People's age had a moderate correlation (0.45) to the number of flu deaths, meaning that older people were more likely to die from influenza.	

6. Results and insight

Research Hypothesis	If a state has a high number of populations, it will have more flu death	If you are over 65, then you are likely to have flu death	If state has more people who got vaccinated, then it will have the lower flu deaths	If the temperature goes down, more flu deaths will occur
Null Hypothesis	The number of flu deaths of the states with populations above the mean is equal to the number of flu deaths of the states with the populations below the mean	The number of flu deaths of the people who are 65 years or older is equal to the number of flu deaths of the people who are under 65	The flu death rate of the people who are vaccinated is equal to the flu death rate of the people who are not vaccinated	The number of flu deaths under the colder temperature is equal to the number of flu deaths under the warmer temperature
Alternative Hypothesis	The number of flu deaths of the states with populations above the mean does not equal to the number of flu deaths of the states with the populations below the mean	The number of flu deaths of the people who are 65 years or older does not equal to the number of flu deaths of the people who are under 65	The flu death rate of the people who are vaccinated does not equal to the flu death rate of the people who are not vaccinated	The number of flu deaths under the colder temperature does not equal to the number of flu deaths under the warmer temperature
Summary	At an alpha of 0.05, it was noted that there is a significant difference in the independent variable between the states with populations above the mean and the states with the populations below the mean (p-value is smaller than the alpha value)	At an alpha of 0.05, it was noted that there is a significant difference in the independent variable between the people who are 65 years or older and the people who are under 65 (p-value is smaller than the alpha value)	At an alpha of 0.05, it was noted that there is not a significant difference in the independent variable between the people who are vaccinated and the people who aren't vaccinated (p-value is greater than the alpha value)	At an alpha of 0.05, it was noted that there is a significant difference in the independent variable between the lower temperature and higher temperature (p-value is smaller than the alpha value)

7. Remaining Analysis and Next Steps

a. Additional analysis of the influenza deaths for vulnerable groups
   - As for the interim, only the age 65 or over was focused compared to the rest of the population for the flu death ratio comparison. Need to expand this upon the other vulnerable groups (e.g., population under five years)

b. Data Visualization including a dashboard creation, including:
   - Spatial Visualization
   - Temporal Visualization

c. Hold meetings with project stakeholders to discuss the following:
   - The total number of staffs that agency has and their availabilities and the default staff to patient ratio
   - Variable Prioritization – based on the hypotheses tested, there are more than one variable that may impact the number of influenza deaths; different variables need to be prioritized differently based on the stakeholders' inputs to facilitate deployment.

     E.g., State A has a population of 1 and the flu death rate of 100% (1 total flu death) while State B has a population of 1 million and the flu death rate of 1% (10,000 total flu deaths). The agency has a total of 100 staff to deploy and a staff to patient ratio is 1:10. Given this situation, what would the agency focus on? To cover the greatest number of flu patients (all-in to the big state) or to have the effective flu patient care across different states (smaller state will have 100% coverage, while the big state may have only a certain level of coverage).

d. Create a formula based on a weighted average from the variable prioritization and other inputs collected from the stakeholders

e. Draft conclusions and recommendations based on analysis/calculation

f. Final presentation to the stakeholders

8. Appendix

a. Success Factors
The project’s success will be based on:
- A staffing plan that utilizes all available agency staff per state requirements, without necessitating additional resources
- Minimal instances of understaffing and overstaffing across states (a state can be considered understaffed if the staff-to-patient ratio is lower than 90% of the required ratio and overstaffed if greater than 110%)

b. Project Requirements
- 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.
- Determine whether influenza occurs seasonally or throughout the entire year. If seasonal, does it start and end at the same time (month) in every state?
- Prioritize states with large vulnerable populations. Consider categorizing each state as low-, medium-, or high-need based on its vulnerable population count.
- Assess data limitations that may prevent you from conducting your desired analyses.

c. Initial Research Hypotheses that were not tested
The following hypotheses that were drafted at the project initiation were not tested due to lack of reliable data availability:
- *Influenza virus can spread. If a location has a high travel traffic, it will likely to have more flu patients.*
- *People in rural areas have relatively lower education levels in comparison to the people in urban areas.*