Medical Staffing Plan for ILI 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 the patient is 65 of age or older, then the risk of mortality is higher because they are especially vulnerable to Influenza-like-Illnesses.

Data Overview

CDC Influenza Deaths (Source: CDC)
This data shows the monthly number of influenza-related deaths per age group from 2009 to 2017 for all U.S. states.

U.S. Census Bureau Population Data (Source: U.S. Census Bureau)
This data shows the population data for all U.S. states from 2009 to 2017 broken down into total population and the population per age group.

Data Limitations

CDC Influenza Deaths
The data is administrative data collected as part of the National Vital Statistics Cooperative Program. Each of the U.S. states and territories is required to record all births, deaths, marriages, and divorces within their jurisdiction. Death records come from death certificates, in which a doctor codes the primary cause of death as “Influenza” or “Pneumonia” (ICD-10 codes J09-J18).
Because this data is part of the government's vital statistics program we can assume a complete and accurate count of deaths.

As the deaths listed in this data don't include any information about preexisting conditions that could be a critical factor for the event of death, it 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.
We can only relate the number of deaths accounted in this data to their geographic location (U.S. state) of occurrence as well as the age group the dead person falls into.
Another limitation is that the data only contains information from 2009 to 2017 and is not up-to-date. This could result in inaccuracies as current events, that could influence the number of deaths (e.g. COVID) are not part of this data.

U.S. Census Bureau Population
The data is administrative data collected to be used for federal programs, policies, and decision-making as well as to produce timely, relevant statistics about the population and the economy of the United States. The U.S.Census Bureau collects administrative data at other U.S. agencies and combines it with survey and census data. It is therefore partly collected manually and automatically.

As the data in this set is collected anonymously, it is unlikely to be biased but the manually collected information is always vulnerable to human error.
The data set only contains information from the years 2009 to 2017. As such the data is not up-to-date and does not display any change in population (births, deaths, migration) after 2017. Furthermore, the numbers in the data set are estimates. This could result in further inaccuracies.

During the data cleaning and integration process of both data sets, the age groups were redefined as follows:
Under 5 years, 5-14 years, 15-24 years, 25-34 years, 35-44 years, 45-54 years, 55-64 years, 65-74 years, 75-84 years, 85+ years

Later on, age groups were summarized for statistical purposes as follows:
0-65 years (sum of Under 5 years, 5-14 years, 15-24 years, 25-34 years, 35-44 years, 45-54 years)
65+ years (sum of 65-74 years, 75-84 years, 85+ years)

Descriptive Analysis

The comparison between the mortality rates of the different age groups shows, people over 65 (65+ age group) have on average a much higher mortality rate than people under 65 years old.

	Mortality Rate patients 0-65 years	Mortality Rate patients 65+ years
Mean/Average	0,000684%	0,05219%

The table below shows measures of central tendency for the sum of all ILI deaths in the age group of 65+ and the population sum of the same age group.

	Sum of ILI Deaths 65+	**Sum of pop 65 years and over**
Data Set	Integrated Data Set	Integrated Data Set
Sample/Population	Sample	Sample
Normal Distribution	normal distribution (sample size over 30)	normal distribution (sample size over 30)
Variance	1028484	399931255021565000
Standard Deviation	1014	632401182
Mean	826	518127324
Count of Records	459	459
95% Data Range	-1201 to 2854	-746675040 to 1782929687
Outlier count	18	25
Outlier Percentage	3,92%	5,45%

To measure the strength of their relationship, we compared the **sum of ILI deaths of patients age 65+** to the **Sum of population age 65 and above** which resulted in a correlation coefficient of 0,8, which is a strong relationship. As the population of people 65 of age or older increases, the number of ILI deaths in the same age group rises as well.

The table below shows measures of central tendency for the mortality rate of all vulnerable age groups (**% all Age Groups**) and the mortality rate of the total population (**% of Total population**).

	% all Age Groups	**% of Total Population**
Data Set	Integrated Data Set	Integrated Data Set
Sample/Population	Sample	Sample
Normal Distribution	normal distribution (sample size over 30)	normal distribution (sample size over 30)
Variance	0,0003%	0,0000%
Standard Deviation	0,1656%	0,0066%
Mean	0,0529%	0,0124%
Count of Records	459	459
95% Data Range	-0,0027% to 0,0038%	0% to 0,0002%
Outlier count	34	3
Outlier Percentage	7,41%	0,65%

Comparing the **mortality rate of vulnerable age groups** with the **mortality rate of the total U.S. population** resulted in a correlation coefficient of 0,2. Which is a weak relationship.
The weak relationship between the two variables shows that they change somewhat independently from each other. A change in the age group mortality does not mean a significant change in the mortality rate of the total population but a change in values will still be accounted for in the overall mortality rate.

Results and Insights

On the basis of the descriptive analysis, statistical hypotheses were created:

Null Hypothesis: The mortality rate of people that are 65 years or older **is equal to or less than** the mortality rate of people under 65 years of age.
Alternative Hypothesis: The mortality rate of people that are 65 years or older is **higher than** the mortality rate of people under 65 years of age.

To disprove the null hypothesis we compared the two mortality rates in a one-tailed t-test. The results showed that the p-value (2,23E-11) is lower than the significance level (alpha=0,05). Which concluded that the probability of the null hypothesis being significant or happening due to chance is less than 1%. The data provided enough evidence to reject the null hypothesis and to accept the alternative hypothesis as statistically significant.
In conclusion, we can say with 95% confidence, that the two groups are significantly different and that the mortality rate of people 65 years of age and older is significantly higher than the mortality rate of people below 65 of age.

Remaining Analysis & Next Steps

Based on the descriptive analysis and the statistical test (hypotheses test) we determined that people age 65 and above have the highest mortality rate. To counter the potential number of ILI deaths, we need to find out more about the population distribution of that age group in the different U.S. states and adjust medical staffing plans proportionally (while taking the time factor of influenza season into account). Other populations especially vulnerable to ILI (e.g. children) should also be part of the medical staffing planning. Their population distribution needs to be determined as well.

Another deciding factor for the distribution of medical staff could be the geographical location of the individual U.S. state, as research shows that climate and the level of urbanization can influence the transmission potential for influenza.

Other factors such as vaccination status and rate, accessibility to medical facilities, as well as any ILI prevention measures that were made could be crucial to getting more insights, but data for those factors is currently not available.

The final deliverable for the project will include a video presentation and a Tableau storyboard with temporal and statistical visualizations, spatial and textual analysis, as well as any key results and recommendations.

Appendix

Project Overview

Stakeholder Identification

- Medical agency frontline staff (nurses, physician assistants, and doctors)
- Hospitals and clinics using the staffing agency's services
- Influenza patients
- **Staffing agency administrators**

Project Goal

The project's success will be based on a staffing plan that utilizes all available agency staff per state requirements, without necessitating additional resources. At the same time, 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%) should be ensured.

To reach the goal we have to:

- Provide information that supports the 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.

Hypothesis

During the initial stages of this project, we started creating a series of key questions important for the data analysis and the subsequent creation of the medical staffing plan.
Excerpt:

1. Which state was historically most affected by influenza?
2. Which age group has the highest mortality rate?
3. Which state has the most citizen in vulnerable populations?
4. Which states had the highest/lowest mortality rates?

We later used these questions to formulate our research hypothesis: **If the patient is elderly, then the risk of mortality is higher.**

Additional Sources:

https://www.science.org/doi/10.1126/science.aat6030

Urbanization and humidity shape the intensity of influenza epidemics in U.S. cities

By Benjamin D. Dalziel, Stephen Kissler, Julia R. Gog, Cecile Viboud, Ottar N. Bjørnstad, Jessica E. Metcalf and Bryan T. Grenfell

https://www.cdc.gov/flu/index.htm

CDC FluView

Data Overview

Additional Data Profiles

CDC Influenza Deaths Data

Variables	Description	time -variant/-invariant	structured/unstructured	qualitative/quantitative	quantitative: discrete/continuous
State	US State in full name	time-invariant	structured	qualitative	nominal
State Code	State Code from 1-56	time-invariant	structured	qualitative	ordinal
Year	Year from 2009 - 2017	time-invariant	structured	qualitative	ordinal
Month	Abbriviation Month with Year	time-invariant	structured	qualitative	ordinal
Month Code	Year and Month in Code (Year, Month)	time-invariant	structured	qualitative	ordinal
Ten-Year Age Groups	One of 11 age groups or "Not Stated"	time-invariant	structured	qualitative	ordinal
Ten-Year Age Groups Code	One of 11 age groups in code or "NS" for Not Stated	time-invariant	structured	qualitative	ordinal
Deaths	Deaths in numbers or "suppressed"	time-variant	structured	quantitative	discrete

U.S. Census Bureau Population Data

Variables	Description	time -variant/-invariant	structured/unstructured	qualitative/quantitative	qualitative: nominal/ordinal quantitative: discrete/continuous
County	County Name	time-invariant	structured	qualitative	nominal
State	US State full Name	time-invariant	structured	qualitative	nominal
Year	Year from 2009-2017	time-invariant	structured	qualitative	ordinal
Total population	Population Numbers as whole numbers	time-variant	structured	quantitative	discrete
Male Total population		time-variant	structured	quantitative	discrete
Female Total population		time-variant	structured	quantitative	discrete
Under 5 years		time-variant	structured	quantitative	discrete
5 to 9 years		time-variant	structured	quantitative	discrete
10 to 14 years		time-variant	structured	quantitative	discrete
15 to 19 years		time-variant	structured	quantitative	discrete
20 to 24 years		time-variant	structured	quantitative	discrete
25 to 29 years		time-variant	structured	quantitative	discrete
30 to 34 years		time-variant	structured	quantitative	discrete
35 to 39 years		time-variant	structured	quantitative	discrete
40 to 44 years		time-variant	structured	quantitative	discrete
45 to 49 years		time-variant	structured	quantitative	discrete
50 to 54 years		time-variant	structured	quantitative	discrete
55 to 59 years		time-variant	structured	quantitative	discrete
60 to 64 years		time-variant	structured	quantitative	discrete
65 to 69 years		time-variant	structured	quantitative	discrete
70 to 74 years		time-variant	structured	quantitative	discrete
75 to 79 years		time-variant	structured	quantitative	discrete
80 to 84 years		time-variant	structured	quantitative	discrete
85 years and over		time-variant	structured	quantitative	discrete

Results and Insights

Null Hypothesis Testing

t-Test: Two-Sample Assuming Unequal Variances		
	Mortality Rate of 0-65 years	*Mortality Rate of 65+ years*
Mean	6,83855E-06	0,000521897
Variance	8,86698E-10	2,67101E-06
Observations	459	459
Hypothesized Mean Difference	0	
df	458	
t Stat	-6,750765261	
P(T<=t) one-tail	2,23E-11	
t Critical one-tail	1,648187415	
P(T<=t) two-tail	4,47E-11	
t Critical two-tail	1,965157098	

The test above shows the results of the null hypothesis testing (one-tailed, two-sample t-test).