INTERIM REPORT

Ivonne Aspilcueta

2024

PROJECT OVERVIEW:

Motivation: Hospitals and clinics need additional staff to adequately treat patients, particularly those in vulnerable populations, who develop serious complications and end up in the hospital.

Objective: To Determine the time and quantity of temporary medical staff, 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 vulnerable population is 65+ then the mortality rate will be higher.
- If the number of vulnerable population 65+ is higher in the most populated States, then more staff is needed.

DATA OVERVIEW:

Influenza deaths by State

The data sourced is external since is taken from the Centers for Disease Control and Prevention (CDC). This dataset contains 66,096 records and the variables are: State, State Code, Year, Month, Month Code, Ten-year age Group, Ten-Year age group code and Deaths. The data is also reliable and trustworthy coming from a trusted government entity.

US Census Population Data by Geography

The data is external since is taking from the U.S. Census Bureau, a principal agency of the U.S. Federal Statistical System (A government agency). Data is trustworthy since they are dedicated and are the leading provider of quality data about America's people and economy.

This dataset contains variables like: County, Year, Total population, Gender, and Age. It is a detailed breakdown to make it easy to analyze the US population.

Integrated Data

A new dataset was generated by merging the Influenza Deaths data with the Census Population Data. For the Influenza Death I combined key State/Year and changed value "Suppressed" = 0. For the Census Population Data, I combined State/Year in a new Column, aggregated the Age groups into 10 years Age Group and removed Puerto Rico, since is not included as one of the states for this project.

DATA LIMITATIONS:

Influenza deaths by State

This Influenza Deaths Data contains data from 2009 to 2017. This is the limitation of the project for not having the most updated data to prepare us for 2024.

Also, Death counts and death rates are "Suppressed" when the data meets the criteria for confidentiality constraints, according to the CDC about death "Suppressed".

Death rates are not calculated specifically for the "Not Stated" groups because there are no corresponding population denominator data for these groups. "Deaths of persons with Age "Not Stated" are included in "All" counts and rates, but are not distributed among age groups, so are not included in age-specific counts, age-specific rates or in any age-adjusted rates. Wonder.cdc.gov. https://wonder.cdc.gov/wonder/help/ucd.html

US Census Population Data by Geography

The dataset is from the years 2009 to 2017, It is not updated which will not reflect a recent demographic population or trends. This outdated data can affect the accuracy of current population estimations and projections.

Also, can have a margin of error since they collect the data by providing surveys; depending on the survey, they have many options for participating online, by mail, in-person, or over the phone or by fax.

DESCRIPTIVE ANALYSIS:

Data Spread		
Variable	**Vulnerable Population 65+**	**Vulnerable Population Death 65+**
Dataset Name	Integrated Data Set	Integrated Data Set
Sample or Population?	Sample	Sample
Normal Distribution?	Left-Skewed	Left-Skewed
Variance	796788779818	1028484
Standard Deviation	892630	1014
Mean	829430	826
Outlier lower bound	-955831	-1202
Outlier upper bound	2614690	2855
Outlier count	30	18
Outlier Percentage	7%	4%

Correlation		
Variable	Total Vulnerable Population & Total Death of Vulnerable Population	Total Population 65+ & Total Death % 65+
Proposed Relationship	Possible Correlation (Death Increases as vulnerable population increases)	Possible Correlation (Death % increases as vulnerable population increases)
Correlation Coefficient	0.94	0.40
Strength of Correlation	Strong relationship	Weak relationship
Usefulness / Interpretation	This interpretation is useful because let us know that the number of deaths due to the influenza increases as the vulnerable population increases, with that said, they can focus in the states with the most vulnerable population to send more staff than the states with the lowest vulnerable population.	There is a little correlation with the total vulnerable population and the total of Death rate % 65+. That means if we based our conclusions in rates, it would not be as useful as we based our conclusions in numbers, since the rate of death does not increase as the same rate of the population, an important decision may be skipped like the need of more staff in places where the vulnerable population are higher.

RESULTS AND INSIGHT:

Research hypothesis	Vulnerable people 65+ have the higher Rate Deaths for influenza	
Dependent variant	Influenza death rate	
Independent variant	Age Group	
Null hypothesis	The Death rate for influenza on patients 65+ is less than or equal to patients under 65 years old.	
Alternative hypothesis	Death rate for influenza on patients 65+ is greater than patients under 65 yrs.	

One-tailed or two-tailed test	Death by influenza rate of population 65+ is greater than patients under 65 years old.	
alpha	0.05	
p-value (one-tail)	0.0000	The p-value (one-tail) is 1.0838E-180, which is translated as zero. The p-value is lower than the significant level of 0.05, so we can conclude that there is a difference between the mean death rates of 65+ age and under 65 age groups.

t-Test: Two-Sample Assuming Unequal Variances		
	0-64 yrs Total Population rate	*Total Population 65 + rate*
Mean	0.00077%	0.08143%
Variance	9.41616E-11	1.76619E-07
Observations	459	459
Hypothesized Mean Difference	0	
df	458	
t Stat	-41.1072068	
P(T<=t) one-tail	4.30E-156	
t Critical one-tail	1.648187415	
P(T<=t) two-tail	8.6017E-156	
t Critical two-tail	1.965157098	

Summary of findings		Since the mean for 65+ years group is higher (0.0814291797435701%) than the mean for under 65 years old group (0.00077471943227727%), we can disregard the null hypothesis and accept the alternative hypothesis.

Research hypothesis	The most populated states have the higher vulnerable people	
Dependent variant	Vulnerable Population	
Independent variant	Total Population	
Null hypothesis	The number of vulnerable people is less than or equal in highly populated.	
Alternative hypothesis	The number of vulnerable people is greater in highly populated states.	
One-tailed or two-tailed test	It's a one-tailed test, as I'm trying to prove the difference in one direction.	
alpha	0.05	
p-value (one-tail)	0.1904	The p-value (one-tail) is 0.190397783614776. The p-value is higher than the significant level of 0.05, so we can conclude that there is a difference between the states most populated and higher vulnerable 65+ population.

Highly populated states ≥ Q3	
IQR	46,782,692
MIN	5,465,761
Outlier low	-55,791,751
Q1	14,382,287
Median	54,973,128
Q3	**61,164,979**
Outlier high	131,339,017
MAX	339,054,878

t-Test: Two-Sample Assuming Unequal Variances		
	High Populated States	*Low Populated States*
Mean	0.135991694	0.141125153
Variance	0.00035932	0.000163199
Observations	13	34
Hypothesized Mean Difference	0	
df	16	
t Stat	-0.901300879	
P(T<=t) one-tail	0.190397784	
t Critical one-tail	1.745883676	
P(T<=t) two-tail	0.380795567	
t Critical two-tail	2.119905299	

Summary of findings		The t-Test did not reject the Null hypothesis, meaning that States with higher population does not affect the quantity of the vulnerable population 65+ in other words, there is not enough evidence that the higher of the population in some States does not necessarily means the higher of vulnerable population that reside in such States.

NEXT STEPS:

1. Send results and determine next steps in collaboration with the stakeholders.
2. Further Analysis with other/ new hypothesis.
3. Focus Vaccination Campaigns for the group age 65+ by prioritizing that campaign in the age group.
4. Increase the staff number in those states with the higher rates of mortality by influenza.
5. Provide educational material to increase knowledge to change the perceptions towards influenza to increase healthy and preventive behavioral practices, especially in the most vulnerable population.