Zahid Niloy

CareerFoundry

Assignment: Interim Excel Report

September 16, 2024

Project Overview

Motivation

Every year, the United States experiences an influenza season during which there is a noticeable increase in flu cases. Certain individuals, particularly those from vulnerable groups, are at risk of developing severe complications that may require hospitalization. To handle the surge in patients, hospitals and clinics often need extra staff, which is provided by the medical staffing agency.

Objective

The goal is to determine the optimal timing and number of staff to deploy toeach state.

Scope

This project covers all hospitals across the 50 states in the United States, focusing on preparations for the upcoming influenza season.

Hypothesis

Individuals who are 65 years of age or older face an increased likelihood of experiencing fatal outcomes as a result of influenza infection.

Data Overview

U.S. Census Data

The U.S. Census Bureau, a government agency, gathers detailed information about people across the country. They collect this data every year using different methods like mail, online surveys, and face-to-face interviews. The information includes details about where people live, their age, gender, and total population counts for each area. This data covers the years 2009 to 2017 and breaks down age groups into five-year segments.

This population data is very important for research projects. It helps researchers understand how many people of different ages live in each area, which can be used to study various topics. The total population numbers are also useful for comparing states of different sizes fairly.

The Census Bureau uses advanced techniques to ensure the data is accurate and representative of the entire country. They combine information from multiple sources and use statistical methods to fill in gaps where direct data collection might be difficult.

Influenza Deaths by Geography

The data on influenza-related deaths is provided by the Centres for DiseaseControl and Prevention (CDC). The Underlying Cause of Death database includes mortality and population statistics for all U.S. counties. Mortality dataare derived from death certificates for U.S. residents, each identifying a single underlying cause of death along with demographic details.

The CDC Influenza

Deaths Data Set provides monthly counts of influenza-related deaths in the U.S.from 2009 to 2017, categorized by state and age group.

Integrated Data

I have combined two sets of information into one big dataset. This new dataset has details about flu deaths and population numbers. It shows how many people live in each state and how many died from the flu. The information is sorted by different age groups, states, and years.

This combined dataset makes it easier to see patterns and trends. For example, we can now compare flu deaths in different age groups across states or over time.

Combining datasets like this is a common practice in data analysis. It allows researchers to see connections between different types of information that might not be obvious when looking at each dataset separately

Data Limitations

The information I am using has some issues that might need some speculation:

People make mistakes: Since people collect this data by hand every year, there might be some errors. It's like when you copy numbers from one place to another - sometimes you might write the wrong thing.

Human error is a common problem in data collection. Even with careful procedures, mistakes can happen when entering or transferring data.

Old information: The data we have is from 2009 to 2017. This means we don't have the most recent information, which could be different from what we're looking at.

Data that isn't current can sometimes lead to outdated conclusions. It's important to consider how things might have changed since the last data point.

Missing pieces: Some parts of the flu death data are missing, marked as "Suppressed."

Also, the data might not show all flu deaths because death certificates only list one main cause of death. If the flu was part of why someone died but not the main reason, it might not be counted.

Incomplete data is a challenge in many research fields. Scientists often have to develop methods to work around missing information or estimate what the full picture might look like.

Descriptive Analysis

Variables	**Average**	**Standard Deviation**
Total Population (Ages 65+)	806,529	886,759
Influenza Deaths (Ages 65+)	826	1014

The statistical analysis unveiled a robust association (correlation coefficient of 0.91) between advanced age, particularly those 65 and older, and elevated influenza-related mortality rates. This finding underscores the heightened vulnerability of seniors to influenza complications.

Statistical Testing

- H0: Influenza mortality rates for 65+ age group ≤ rates for under 65
- H1: Influenza mortality rates for 65+ age group > rates for under 65

The extremely low p-value ($p < 0.001$) provides strong evidence to reject the null hypothesis. We can conclude with high confidence that the observed increased influenza mortality among those 65 and older is statistically significant and not attributable to chance.These results align with previous research showing that older adults face a substantially higher risk of severe influenza outcomes. The age-related decline in immune function, coupled with the increased prevalence of chronic conditions, likely contributes to this disparity.

Further Scope for Exploration

1. Create interactive visualizations using Tableau to illustrate age-specific mortality trends over time.
2. Develop geospatial maps to identify regional variations in influenza impact on older populations.
3. Conduct additional analyses to examine potential confounding factors and comorbidities.
4. Prepare a comprehensive report and presentation to effectively convey key insights to public health stakeholders.
5. Consider policy implications, such as targeted vaccination campaigns and enhanced preventive measures for high-risk age groups.

Appendix

1. **Influenza deaths by geography**

   Source: CDC

2. **Population data by geography, time, age, and gender**

   Source: U.S. Census Bureau

I. Statistical Testing:

t-Test: Two-Sample Assuming Unequal Variances

T Test: Two-Sample Assuming Unequal Variances

	0-64 yrs	65+ years
Mean	8.695E-06	0.00091994
Variance	1.045E-10	1.3909E-07
Observations	423	423
Hypothesized Mean Differ	0	
df	422	
t Stat	-50.23404	
P(T<=t) one-tail	1.87E-180	
t Critical one-tail	##########	
P(T<=t) two-tail	3.73E-180	
t Critical two-tail	1.9656013	

II. Statistical Analysis

Hypothesis
Regions where a greater proportion of residents are elderly (65+) are likely to experience more influenza-related fatalities than areas with a smaller percentage of senior citizens. The prevalence of older adults in a state's demographic makeup correlates positively with the number of deaths

Data Spread				
Variable	**Vulnerable Population**	**Total Population**	**Total Death of Vulnerable Population**	**Total Death**
Dataset Name	Integrated Data Set	Integrated Data Set	Integrated Data Set	Integrated Data Set
Sample or Population?	Sample	Sample	Sample	Sample
Normal Distribution?	left-skewed	left-skewed	left-skewed	left-skewed
Variance	786342285401	46336577337755	1028484	1332864
Standard Deviation	886759	6807098	1014	1154
Mean	806529	5975286	826	905
1 standard deviation lower b	-80230	-831812	-188	-249
1 standard deviation upper b	1693289	12782383	1840	2060
Outlier count	49	37	59	59
Outlier Percentage	11%	8%	13%	13%