Preparing for Influenza 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.

Research Hypothesis

If a state has a large population of children under 5 years old and people over 65 years, the death rate from flu is higher for these age groups. People of these age groups fall into the category of vulnerable population, so they are more likely to get flu complications which will require medical assistance.

Data Overview

Dataset	Population Data by Geography	Influenza Deaths
Source	U.S. Census Bureau	CDC
Data Content	Includes yearly population counts for the U.S. counties from 2009 to 2017. Counts are divided by gender and 5-year age groups, with the first group being under 5 years and the last group being over 85 years.	Includes monthly flu death counts for each state from 2009 to 2017. Counts are divided by 10-year age groups.
Data Limitation	Census conducted once per 10 years, so yearly data shows estimated population, that is calculated based on population base, births, deaths and migration. Collected both automatically and manually (e.g., the migration surveys), because of this might contain errors due to a human factor.	Death counts of 9 or fewer people are not allowed to be published, therefore are suppressed (81.7% of the records). Data is obtained from death certificates and inputted manually, thus might contain errors.

Descriptive Analysis

- **Data Spread**

Variable	Population under 5 years	Death rate under 5 years
Dataset	Census Population Data	Influenza Mortality Data
Standard Deviation	456,429	0.10
Mean	386,283	0.09
*Outlier % (Mean+/-2*StandDev)*	3.92%	6.54%

- **Correlation**

Variables	Population under 5 years	Death rate under 5 years
Proposed Relationship	A higher number of children under 5 years might be related to higher death counts from flu for this age group, due to the lack of medical staff to take care of this vulnerable group.	
Correlation Coefficient	-0.51	
Strength of Correlation	Strong, negative	
Interpretation	When population of children under 5 years is increasing, flu-related death rate is decreasing. This might be related to the vaccination status of a child, assuming that flu vaccine decreases the likelihood of flu complications and death.	

Statistical Hypothesis Testing

Null Hypothesis	Flu death rates for vulnerable population (under 5 and over 65 years) are smaller or equal to the death rates for the rest of the population (5-64 years).
Alternative Hypothesis	Flu death rates for vulnerable population (under 5 and over 65 years) are higher than the death rates for the rest of the population.
Test type	One-tailed, since we want to know if the death rates for one age group are smaller or equal to the death rates for another group, so we are focused only on one direction.
Significance level	0.05
p-value	1.92E-124
Results	Because p-value is less than alpha, we can reject the null hypothesis. At a confidence level of 95%, we can state that the death rates from flu for vulnerable population are higher than for people from 5 to 64 years. Based on this, we can determine the states with the highest number of vulnerable population in order to allocate more medical staff there for the upcoming flu season.

Remaining Analysis

- Create composition and comparison charts
- Design temporal visualization and forecasting in Tableau
- Make statistical visualizations to show variables distribution and correlation
- Create spatial visualization
- Design final presentation in a form of Tableau Storyboard
- Record a video presentation

Appendix

1. Refer to Project Brief for project requirements, success factors, original data sets, as well as other details.
2. Refer to Project Management Plan for project milestones, deliverables and initial hypothesis.
3. Refer to Data Profiling & Integrity and Data Quality Measures to review the data cleaning process.
4. Refer to Data Spread & Correlation for descriptive analysis.
5. Refer to Statistical Hypothesis Testing for t-Test details.