Influenza Project: Interim Report

**Project Overview**

* **Motivation**: The United States experiences a seasonal increase in influenza cases, which leads to a higher demand for hospital resources, particularly for vulnerable populations. Temporary staffing is essential to support hospitals in managing this influx.
* **Objective**: Determine the optimal timing and number of staff required in each state to prepare for the influenza season.
* **Scope**: The project covers hospitals across all 50 states in the United States, focusing on planning for the upcoming influenza season.

**Research Hypothesis**

If flu shot coverage is higher in vulnerable populations, then influenza mortality rates will be lower in those populations.

**Data Overview**

1. **US Census Data:** This dataset, sourced from the U.S. Census Bureau, provides demographic information, including population counts and age distribution by state. It helps identify the number of vulnerable individuals in each state who may be more affected by influenza.
2. **Influenza Deaths Data:** Sourced from the Centers for Disease Control and Prevention (CDC), this dataset contains historical data on influenza-related deaths, broken down by state, age, and time period. This information is essential for assessing the severity of influenza across regions.
3. **NIS Flu Shot Survey:** This survey dataset, also provided by the CDC, includes survey data on flu shot coverage, broken down by age groups across different states. This data is key for understanding vaccination rates in various demographics, which may correlate with influenza outcomes.
4. **Influenza Lab Tests and Patient Visits:** Sourced from CDC’s FluView, This dataset records counts of laboratory-confirmed flu cases and patient visits, categorized by state. It supports trend analysis on influenza outbreaks, helping to predict staffing needs.

**Data Limitations**

1. **US Census Data**: Since this data provides population estimates, there may be minor differences, especially in age distributions across different groups. Additionally, since census data isn’t updated all the time, it’s possible that some recent changes in demographics aren't completely shown.
2. **Influenza Deaths Data**: This dataset relies on reporting practices that may vary between states, which may introduce inconsistencies. Also, influenza deaths are sometimes underreported if influenza is not identified as the primary cause, resulting in incomplete counts.
3. **NIS Flu Shot Survey**: Survey data is based on samples and may not accurately represent the entire population, especially for smaller or underrepresented demographic groups. This limitation could lead to potential biases in flu shot coverage estimates.
4. **Influenza Lab Tests and Patient Visits**: Testing rates for influenza can vary state by state and are influenced by the availability of healthcare resources. Meaning that some states may report higher case counts simply because testing is more accessible. Additionally, lab-confirmed cases only account for a portion of actual influenza cases, as everyone with flu-like symptoms gets tested.

**Descriptive Analysis Summary**

1. **Core Variables**:
   * **Flu Shot Rate**:
     + **Mean**: 556.10
     + **Standard Deviation**: 376.00
   * **Influenza Mortality Rate**:
     + **Mean**: 0.81
     + **Standard Deviation**: 2.96
2. **Correlation Result**:
   * A weak negative correlation was calculated between the **Flu Shot Rate** and the **Influenza Mortality Rate** with a correlation coefficient of **-0.09**. This weak correlation suggests only a minimal association between higher flu shot rates and lower mortality rates, aligning with the hypothesis but not providing strong evidence.

**Results and Insights**

1. **Statistical Hypothesis:**
   * **Null Hypothesis (H₀):** There is no significant difference in influenza mortality rates between groups with high flu shot rates and those with low flu shot rates.
   * **Alternative Hypothesis (H₁):** Groups with high flu shot rates have significantly lower influenza mortality rates compared to groups with low flu shot rates.
2. **Hypothesis Test Results:**
   * A two-sample t-test was conducted to compare mortality rates between high and low flu shot rate groups. The test produced a **t-statistic of 0.166** and a **p-value of 0.867** (two-tailed), indicating no statistically significant difference between the two groups.
3. **Interpretation:**
   * The results fail to reject the null hypothesis, suggesting that there is no statistically significant evidence that higher flu shot rates are associated with lower influenza mortality rates. Although there is a weak negative correlation between flu shot rates and mortality rates, it is not strong enough to demonstrate a meaningful impact on mortality.
   * These findings imply that while flu vaccinations may have benefits, other factors may play a more significant role in reducing mortality rates.

**Remaining Analysis and Next Steps**

1. **Remaining Analyses:**
   * **Visual Analyses:** To provide stakeholders with a clearer understanding of the data, visualizations will be created to illustrate key findings. These include:
     + A bar chart comparing flu shot rates across states to highlight variability.
     + A scatter plot showing the relationship between flu shot rates and influenza mortality rates to visually represent the correlation.
     + A line or bar chart comparing influenza mortality rates among groups with different levels of influenza vaccination coverage highlighting the differences in outcomes.
     + Maps to show regional differences in flu shot rates and influenza mortality, offering spatial insights.
2. **Next Steps:**
   * **Final Report and Presentation:** After completing the visualizations, a comprehensive final report will be prepared to integrate all analyses and findings. This report will be presented to stakeholders in a format that includes both narrative explanations and visual elements to enhance understanding.
   * **Stakeholder Review and Feedback:** Following the presentation, feedback from stakeholders will be gathered to refine any remaining deliverables. Additional analyses may be conducted based on stakeholder questions or suggestions.

**Appendix: Additional Resources for Stakeholders**

1. **Project Brief**

* **Document:** "A1-A2 Data Immersion Project Brief"
* **Overview:** This document gives a clear view of the project’s purpose, outlining why we’re focused on preparing for flu season with adequate hospital staffing. It also includes the project’s main goals and the specific stakeholders involved.

2. **Data Sources**

* **US Census Data:** Population data by state and age, helping us identify which groups might be more vulnerable to influenza impacts.
* **Influenza Mortality Data:** Historical data from the CDC on flu-related deaths, broken down by region and age. This helps assess severity across different areas.
* **NIS Flu Shot Survey:** Vaccination data showing flu shot coverage rates by age group across states, helping us understand vaccine distribution.
* **Lab Tests and Patient Visits Data:** CDC data from FluView on lab-confirmed flu cases and patient visits, giving a closer look at flu trends by region.

3. **Statistical Analysis Methodology**

* **Document:** "1.9 Hypothesis Testing - Yarisel Velacanto"
* **Overview:** This document walks through the hypothesis testing process used in the project. It includes the statistical tests, like t-tests, that we used to see if higher flu shot rates might impact mortality rates.

4. **Glossary of Key Terms**

* **Influenza:** A viral infection that can lead to hospitalizations, especially for vulnerable groups.
* **Flu Shot Rate:** The proportion of the population vaccinated for influenza—an important variable in our analysis.
* **Statistical Testing:** Methods like t-tests that help us see whether flu vaccinations correlate with lower mortality rates or not.

5. **Methodology Notes**

* **Data Cleaning and Integration:** A quick summary of the steps taken to clean and integrate the data, like removing duplicates and standardizing values.
* **Hypothesis Development:** How we set up our research questions and structured our hypothesis tests to explore the impact of flu shot rates on mortality.