Preparing for Influenza Season: Interim Report

1. 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.

2. Research Hypothesis

The influenza death rate for patients aged 65 and older is higher than the death rate for patients under 65.

3. Data Overview

Population Data by Geography US Census Data

- ➔ The data, sourced from the U.S. Census Bureau is administrative data collected annually through mailed questionnaires, online responses, and in-person interviews. This multi-modal approach helps maximize response rates and accuracy, although there may be some nonresponse bias that the Bureau mitigates with follow-up and statistical adjustments. The dataset includes population counts by county, state, sex, and age groups, with the file name suggesting it is from early 2021 ("2021 01").

Influenza Deaths

- ➔ This data is collected as part of the National Vital Statistics Cooperative Program, where U.S. states and territories record vital events, including deaths. Deaths attributed to influenza or pneumonia (ICD-10 codes J09-J18) are documented on death certificates. The dataset includes monthly influenza-related death counts from 2009 to 2017, categorized by state and age.

4. Data Limitations

Population Data by Geography US Census Data

➔ Data accuracy can be compromised by manual entry errors and delays in data processing, resulting in outdated information. Non-response from certain groups and privacy concerns, particularly among immigrants, can create gaps in the dataset. Moreover, social and geographical challenges, such as difficulty reaching remote areas and varying trust in government surveys, can affect participation rates.

Influenza Deaths

➔ The CDC's influenza deaths dataset is limited by underreporting and delays in data collection, which can lead to inaccuracies. Estimation models used by the CDC may not fully capture all cases or account for demographic and regional variability. While this system provides a comprehensive count, it only records one primary cause of death, which may lead to underreporting in cases involving multiple contributing factors.

5. Descriptive Analysis

	Total population for Ages 65+	**Influenza Deaths for Ages 65+**
Population	171457303	379266
Average	373545	826
Standard Deviation	407235	1014

Correlation results at 0,95 indicated that States with a larger population over 65 years old will most likely have higher death rates caused by influenza. Recommended to send more medical staff to those areas (larger populations of people 65+).

6. Results and insights

- Null Hypothesis: The influenza death rate for patients aged 65 and older is less than or equal to the death rate for patients under 65.
- Alternative Hypothesis: The influenza death rate for patients aged 65 and older is higher than the death rate for patients under 65.
- Conclusion: The analysis resulted in a p-value close to 0, which is below the significance level of 0.05. This allows us to reject the null hypothesis, indicating that the observed higher mortality rate among older individuals (aged 65+) due to influenza is *unlikely* to be due to random chance. **Therefore, the difference in mortality rates is *statistically significant*.**

7. Recommendations

Given the significant results, I suggest investigating the vaccination status of the 65+ age group. Understanding the underlying factors that contribute to this group's vulnerability can help focus efforts on prevention rather than just treatment. Additionally, it might be beneficial to examine different age groups, such as children under 5 years old.

8. Remaining Analyses and Next Steps

Remaining Analyses:

- Investigate states with significant populations in the 65+ age group, focusing on vaccination statuses.
- Determine the current and projected patient-doctor ratios to address under- and overstaffing in healthcare institutions.
- Identify the times when influenza has the most impact on patients.

Next Steps:

- Develop a dashboard featuring heatmaps (states with most influenza, and deaths), line charts (influenza timelines), and scatter plots (vaccination vs. death concentration)
- Prepare a final report including findings, conclusions, and visualizations.
- Present findings to stakeholders.

9. Appendix

Data Spread:	Total population for Ages 65+	Influenza Deaths for Ages 65+
Dataset Name	Integrated Dataset	Integrated Dataset
Sample or Population?	Sample	Sample
Normal Distribution?	right skewed, mean higher than median	right skewed, mean higher than median
Mean	373545	826
Median	251652	521
Variance	165839997457	1028484
Standard Deviation	407235	1014
1 Standard Deviation		
Lower	-33689	-188
Upper	780780	1840
2 Standard Deviations		

Lower	-440924	-1202
Upper	1188014	2855
Count of Outliers	31	18
Outlier Percentage	7%	4%

Correlation:	
Variables	Population (Ages 65+), Influenza Deaths (All ages)
Proposed Relationship	As the population of people aged 65 and older increases, the number of influenza-related deaths rises.
Correlation Coefficient	0.95
Strength of Correlation	Strong relationship
Usefulness/Interpretation	Correlation results at 0,95 indicated that States with a larger population over 65 years old will most likely have higher death rates caused by influenza. Recommended to send more medical staff to those areas (larger populations of people 65+)

t-Test: Two-Sample Assuming Unequal Variances		
	0-64	***65+***
Mean	0.00000771	0.00181486
Variance	0.00000000	0.00000088
Observations	459	459
Hypothesized Mean Difference	0.00	
df	458	
t Stat	-41.34	
P(T<=t) one-tail	0.00	
t Critical one-tail	1.65	
P(T<=t) two-tail	0.0000000	
t Critical two-tail	1.97	