Preparing For Influenza Season: Interim Report.

Paul Maden.
Last Update: 03/10/2024

Project Overview:

The project aims to prepare for the upcoming influenza season by determining when and how many temporary medical staff to allocate to hospitals across the United States.

- **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.

Hypothesis: If a state has a large *vulnerable population* (like adults over 65 years – see appendix 1), then more influenza deaths will occur.

Data Overview/Limitations:

Aspect	Population Data by Geography (US Census)	Influenza Laboratory Tests and Patient Visits Data
Data Source **(See appendix 2, 3)**	External, US Census Bureau	External, Centres for Disease Control and Prevention (CDC), NNDSS, Influenza Surveillance System
Trustworthiness	Trusted (US Government source, no conflict of interest)	Trusted (US Government source, no conflict of interest)
Collection Method	Administrative data; Decennial Census (every 10 years) and ACS (ongoing surveys)	Administrative data; Weekly reports from healthcare providers, labs, and local health departments
Data Coverage	2009-2017; includes state/county population, age groups, and gender data	Lab Tests: 2010-2015, Patient Visits: 2010-2019
Variables Included	County/State, Year, Total Population, Male/Female, Age in 5-year increments	Patient Visits: Year, Week, State/Region, Providers, Patients Lab Tests: Year, Week, Specimens, Influenza Type
Limitations	Time lag; Risk of manual errors; Timeliness issues (data updated every 10 years with ongoing surveys)	Time lag; Inconsistencies in columns for both datasets, making the lab data unreliable*
Relevance to Project	Relevant: Provides data on population demographics and vulnerable groups (e.g., 65+), useful for analysing influenza-associated deaths	Relevant: Patient visit data aligns with population data for comparison analysis Not Relevant: Lab test data is unreliable and period too short

*** Limitations such as the time lag in US Census data and inconsistencies in lab reports may affect the reliability of the results by introducing outdated or incomplete information, potentially leading to less accurate predictions about influenza deaths and staffing needs.**

Descriptive Analysis:

Core Variables: Means and Standard Deviations (See appendix 4).

Variable	Mean	Standard Deviation
Influenza Deaths (Adults > 65 years)	826	1,014
Total Population (Adults > 65 years)	806,990	887,017
Total Population (All ages)	5,974,029	6,807,710
Total Influenza Deaths (All ages)	905	1,154

(NB. Minimal outliers (data points that significantly differ from the other values in a dataset) - See Appendix 5).

Correlation Results:

Two key correlation analyses were conducted to test the hypothesis (see appendix 6):

1. **Population Aged 65+ vs. Total Influenza Deaths:**
   - **Correlation Coefficient:** 0.94 (Strong positive correlation) or 94% confidence level.
   - **Interpretation:** As the population of individuals aged 65 and older increases, the number of influenza deaths also increases. This infers that states with larger elderly populations are more vulnerable to higher influenza-related mortality.
2. **Overall Population vs. Total Influenza Deaths:**
   - **Correlation Coefficient:** 0.95 (Strong positive correlation) or 95% confidence level.
   - **Interpretation:** Larger states (in terms of population size) experience more influenza deaths, regardless of age structure. This infers that it is important to allocate more healthcare staff to states with larger total populations.

Results & Insights:

To test this analysis, I used this statistical hypothesis and interpretation:

Null Hypothesis	***The size of the vulnerable population (like the number of adults over 65 years) has no effect on the number of influenza deaths.***
	i.e. There is no relationship between the size of the vulnerable population and the number of influenza deaths.
Alternative Hypothesis	***The size of the vulnerable population (such as the number of adults over 65 years) influences the number of influenza deaths.***
	i.e. There is a relationship between the size of the vulnerable population and the number of influenza deaths.

Results	***Reject the Null Hypothesis:*** The p-value helps us understand how likely it is that the results we are seeing happened just by chance. In this case, the p-value is very small, meaning it is highly unlikely that the link between larger elderly populations and more influenza deaths is due to random chance. Because the p-value is much smaller than our threshold 0.05 (see Appendix 7), we have strong evidence to believe that states with more elderly people really do experience more flu-related deaths, not just as a coincidence. This is why we can confidently reject the idea that there is no connection (the null hypothesis).
	This infers that the likelihood of observing the results (or more extreme ones) under the null hypothesis is extremely low.

Final Analysis	Since the p-value is far smaller than the alpha, we reject the null hypothesis and conclude that the size of the vulnerable population (like adults over 65 years) ***does*** have a statistically significant effect on the number of influenza deaths.	
	The result is highly statistically significant, suggesting a strong relationship between the size of the vulnerable population and the number of influenza deaths.	
	In other words, the data strongly supports the idea that as the number of vulnerable individuals increases, so does the number of deaths from influenza.	

Conclusion: The descriptive analysis shows strong positive correlations between both the elderly population and total population sizes with influenza deaths. This data provides actionable insights for healthcare resource allocation, especially in states with larger elderly or overall populations, e.g. the data infers those states with a large elderly population, such as Florida and New York, are more susceptible to influenza-related deaths.

Remaining Analysis & Next Steps:

Next Steps Recommendations:

- Double-check that the sample size is sufficient to support the strength of the conclusions.
- Determine how strong the relationship is between the size of the vulnerable population and influenza deaths. Even though the p-value is significant, effect size will help understand the practical significance of the results.
- Clearly state any limitations (i.e. mitigated in future analysis or decisions (e.g., using more real-time healthcare data or data from other influenza seasons data constraints), data ethics)
- Start the process to assess actionable recommendations for the staffing agency, the report could suggest the following: Breakdown High-Risk States, Flexible Staffing Models, Early Deployment, Cross-Training Staff – all in preparation for final deliverable or presentation.
- Analyse how the relationship between vulnerable populations and influenza deaths differs across various states, i.e.
    - **Scatter plots** showing the correlation between:
        - Elderly population size vs. influenza deaths across various states
        - Total population vs. influenza deaths across various states
    - **Bar charts** displaying (e.g. example appendix 8):
        - Total influenza deaths across different states
        - Age distribution in states and corresponding influenza death rates

Appendix:

1. **Vulnerable population definition:** (**Table - PMC (nih.gov)**)
2. **Original source data:** Influenza Data CDC & U.S. Census Bureau (Influenza Data Sets)
3. **Integrated Data & Source data**: 1.9_Statistical_Hypothesis_Testing.xlsx
4. **Core Variables:** Means and Standard Deviations.

Variable	Influenza Deaths adults > 65 years	Total Population adults > 65 years	Total Population	Total Influenza Deaths
Dataset Name	Integrated Data	Integrated Data	Integrated Data	Integrated Data
Sample or Population?	Sample	Sample	Sample	Sample
Mean	826	806,990	5,974,029	905
Median	521	546,937	4,080,705	546
Normal distribution?	**positively skewed** or **right-skewed.**	**positively skewed** or **right-skewed.**	**positively skewed** or **right-skewed.**	**positively skewed** or **right-skewed.**
Variance	1,028,484	786,799,847,781	46,344,914,268,576	1,332,864
Standard Deviation	1014	887,017	6,807,710	1,154
1 Standard Deviation - Lower	-188	-80,028	-833,681	-249
1 Standard Deviation - Upper	1840	1,694,007	12,781,739	2,060
2 Standard Deviation - Lower	-1202	-967,045	-7,641,391	-1,404
2 Standard Deviation - Upper	2855	2,581,024	19,589,449	3,214
Number Outliers (*Filtered columns 'greater than upper s')	18	29	22	19
Outlier Percentage	3.9%	6.3%	4.8%	4.1%

5. **Outliers:** Influenza Deaths (Adults > 65 years): 3.9% outliers, Total Population (Adults > 65 years): 6.3% outliers, Total Population (All ages): 4.8% outliers, Total Influenza Deaths (All ages): 4.1% outliers.
(NB. In a statistical table, outliers are data points that significantly differ from the other values in a dataset. They are unusually high or low compared to most of the data and can distort analysis or misrepresent trends. Outliers may occur due to variability in the data, measurement errors, or other anomalies, and identifying them is crucial to ensure accurate interpretation of results)

6. **Correlation:**

Variables	Population aged 65 / Total Influenza deaths (all ages)	Overall Population / Influenza Deaths
Proposed Relationships	As the population of individuals aged 65 and older grows, the number of influenza deaths increases.	The more populous states experience more influenza deaths, independent of the age structure.
Correlation Coefficient	0.94	0.95
Strength of Correlation	Strong or close relationship	Strong or close relationship
Usefulness / Interpretation	This strong positive correlation suggests that states with a larger population of individuals aged 65 and older are likely to experience a higher number of total influenza deaths. Therefore, it would be beneficial for the medical staffing agency to allocate more staff to states with larger elderly populations.	This strong positive correlation indicates that states with larger populations are likely to experience more influenza deaths. To optimize resources, the medical staffing agency should allocate more staff to these states. Further analysis is needed to determine which age groups contribute most to the deaths.

7. **P-Value:** (NB. Children < 5 years old are typically considered vulnerable to severe complications and death from influenza. I.e. data used: 5–64-year variable).

T Test: Two-Sample Assuming Unequal Variances		
	5-64 Total %	65+ Total%
Mean	0.00%	0.09%
Variance	1.27306E-10	3.09948E-07
Observations	459	459
Hypothesized Mean Difference*	0	
df	458	
t Stat	-32.99564219	
P(T<=t) one-tail	2.04E-123	
t Critical one-tail	1.648187354	

(*Hypothesized Mean Difference = “expected difference.”)
(NB. One tailed-tested used as I expect that more vulnerable people lead to more deaths, hence I'm testing whether the number of influenza deaths increases with a larger vulnerable population (e.g., "more," "greater than," or "less than"). A two-tailed test would be used if I were simply testing whether the size of the vulnerable population affects influenza deaths in either direction (i.e., either increases or decreases deaths), but I’m not doing that here.)

8. **Influenza Deaths Bar chart:** by Age Category: