

# 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.

## Hypothesis

---

If a population has more elderly (65+) people, then they will be more at risk of influenza related deaths and require more staff

## Data Overview

---

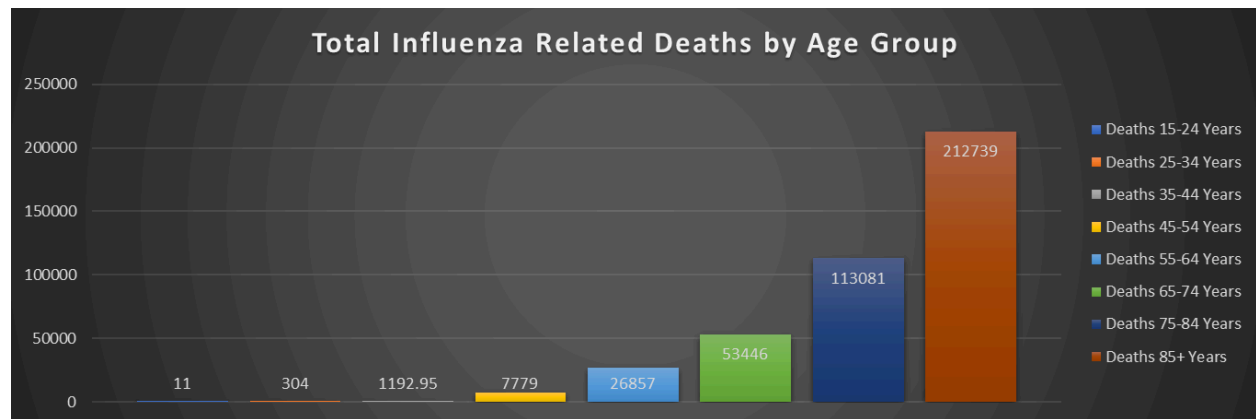
- ❖ Influenza Deaths by Geography
  - **Source:** CDC
  - [https://coach-courses-us.s3.amazonaws.com/public/courses/da\\_program/CDC\\_Influenza\\_Deaths\\_edited.xlsx](https://coach-courses-us.s3.amazonaws.com/public/courses/da_program/CDC_Influenza_Deaths_edited.xlsx)
  - This data set contains the monthly influenza related deaths, for different age groups in every state for the years 2009-2017
  - **Limitations:** Since data is only collected annually there is a potential issue of time-lag within the data. Additionally, for the sake of privacy if certain age groups recorded a small number of deaths within a given month the information was suppressed, so there is some missing data as a result.
- ❖ Population data by geography, time, age, and gender
  - **Source:** US Census Bureau
  - [https://coach-courses-us.s3.amazonaws.com/public/courses/data-immersion/A1-A2\\_Influenza\\_Project/Census\\_Population\\_transformed\\_202101.csv](https://coach-courses-us.s3.amazonaws.com/public/courses/data-immersion/A1-A2_Influenza_Project/Census_Population_transformed_202101.csv)
  - The data contains the total population, gender, and age (measured in ranges of 5 years) for every county for the years 2009-2017.
  - **Limitations:** It is survey data so there is always the chance of human error affecting the results. Additionally as this is annually compiled data from a decennial survey there is an issue of potential time lag in the data.

# Preparing for Influenza Season: Interim Report

## Descriptive Analysis

Variable	Total Population	Population 65+	Total Influenza Deaths	Influenza Deaths 65+
Mean	5,973,702	806,964	905	826
Standard Deviation	6,806,766.773	887,010.1408	1,154	1,014.141877

\*See Appendix **Figure 1** for more detailed analysis



\*\*Shows total deaths for the years 2009-2017

## Correlation Results

Variables	Total Population and Total Death	Vulnerable Population (65+) and Total Deaths
Correlation Coefficient	96%	95%
Usefulness/Interpretation	This correlation shows that as the population increases so does influenza deaths. Meaning states with larger populations will require more staffing.	This does indeed show a strong relationship between the number of vulnerable people in a population and the number of deaths within that population. It is useful because it shows that populations with more vulnerable people do suffer from more influenza related deaths

# Preparing for Influenza Season: Interim Report

## Results and Insights

<b>Independent / Dependent Variable:</b>	Vulnerable (65+) population / Total Deaths
<b>Null Hypothesis:</b>	As the vulnerable population increases the total deaths will stay the same or decrease
<b>Alternative Hypothesis:</b>	As the vulnerable population increases so will total deaths
<b>1 or 2 Tailed:</b>	1 Tailed
<b>Significance Level:</b>	Alpha = .05

### t-Test: Two-Sample Assuming Unequal Variances

	65+	Total Deaths
Mean	806963.6707	905.0522876
Variance	7.87E+11	1332863.727
Observations	459	459
Hypothesized Mean Difference	0	
df	458	
t Stat	19.46901661	
P(T<=t) one-tail	2.95E-62	
t Critical one-tail	1.648187415	
P(T<=t) two-tail	5.90E-62	
t Critical two-tail	1.965157098	

The p-value is 2.95215112412343E-62 since this number is so small it shows that we have more than a 99% percent chance that the increase the total death count is a result of the increase in vulnerable population

**Final Results:** Our findings show a statistically significant positive correlation between the number of elderly vulnerable people in a population and the number of influenza related deaths suffered in that population. With this information alone we will know to allocate more staffing in States with a large elderly population.

\*\*See Appendix **Figure 2** for additional hypothesis testing results

# Preparing for Influenza Season: Interim Report

## Further Analysis & Next Steps

- To be able to better implement our findings we should research which states have a statistically significant larger elderly population compared to other states to be able to identify the states that will need this increased staffing
- Other potential null hypotheses that can be tested to supplement our findings are
  - The average income of a states population has no effect on the rates of influenza related deaths
  - The season has no effect on the number of influenza related deaths within a given state
- Begin construction of data visualization and video presentation of findings

## Appendix

**Figure 1:** Complete Descriptive Analysis

DATA SPREAD				
Dataset Name / Variable	Integrated Data Set: Total Population	Integrated Data Set: Population 65+	Integrated Data Set: Total Deaths	Integrated Data Set: Deaths for Vulnerable Pop
Sample or Population	Sample	Sample	Sample	Sample
Normal Distribution	Right Skewed	Right Skewed	Right Skewed	Right Skewed
Variance	46332073902299	786786989961	1332864	1028484
Standard Deviation	6806766.773	887010.1408	1154	1014.141877
Mean	5973702	806964	905	826
Median	4079507	546938	546	521
Outlier Percentage	5%	6%	4%	4%

## Preparing for Influenza Season: Interim Report

**Figure 2:** Additional Statistical Hypothesis Testing

<b>Independent / Dependent Variable:</b>	Total Population/ Total Deaths
<b>Null Hypothesis:</b>	As the total population increases total deaths will stay the same or decrease
<b>Alternative Hypothesis:</b>	As the total population increases so will total deaths
<b>1 or 2 Tailed:</b>	1 Tailed
<b>Significance Level:</b>	Alpha = .05

t-Test: Two-Sample Assuming Unequal Variances		
	<i>Total Population</i>	<i>Total Deaths</i>
Mean	5973702.057	905.0522876
Variance	4.63E+13	1332863.727
Observations	459	459
Hypothesized Mean Difference	0	
df	458	
t Stat	18.79936678	
P(T<=t) one-tail	3.72E-59	
t Critical one-tail	1.648187415	
P(T<=t) two-tail	7.45E-59	
t Critical two-tail	1.965157098	
The p-value is 3.72304240098572E-59 since it is so close to 0 it means we have a 99% chance that the increase in total deaths is a result of the increase of total population		