Preparing for Influenza Season: Interim Report

Table of Contents

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 the flu patients age is +65 then the mortality rate will increase.*
- *if the population in a state is big then the mortality rate will increase.*

DATA OVERVIEW:

- **Population data:**

  Information about U.S census in each state by gender (male, female), and by age (from 0 to 85+) from 2009 -2017.

  Source: US Census Bureau.

  https://coach-courses-us.s3.amazonaws.com/public/courses/data-immersion/A1-A2_Influenza_Project/Census_Population_transformed_202101.csv

- **Influenza deaths by geography:**

  Information about influenza related deaths in U.S, from 2009-2017, categorized by age (from 0 to 85+) and state.

  Source: Centers for Disease Control and Prevention (CDC).

  https://wonder.cdc.gov/ucd-icd10.html

Limitation:

in the **Influenza deaths by geography** dataset, before data are released from CDC to the public, data cells with fewer than 10 case counts are suppressed, due to that we can not know the exact numbers for deaths caused by influenza.

Both datasets contain information from 2009-2017, which means that dataset is not up to date to the current time (2024), in other words it is not reliable.

Descriptive Analysis:

Dataset Name	total population	total deaths	Sum of 65+ deaths	sum of 65+ population
Sample or population?	sample	sample	sample	sample
normal distribution?	right skewed	right skewed	right skewed	right skewed
Variance	11,879,558,054,678	953,084	294,203	786,799,499,984
Standard Deviation	3,446,673	976	542	887,017
mean	3,035,913	889	475	806,989
median	2,075,601	575	320	546,938
upper outlier	9,929,260	2,842	1,560	2,581,023
lower outlier	-3857434	-1063	-609	-967045
count of outlier	26	18	7	29
outlier percentage	5%	3.92%	3.71%	6.30%

Statistical analysis on total population & total influenza deaths.

Correlation results:

Variables	total population & total deaths	65+ population & 65+ deaths
proposed relation	in the bigger population, deaths caused by flu increases.	if age increase then influenza deaths increase
strength of correlation	0.956093627	0.940623644
usefulness / interpretation	strong relationship	strong relationship

Based on the correlation results, there is a strong relationship between total population and total deaths that means that the bigger the population the higher, the death caused by influenza. And the correlation between the population of people age more than 65 and the death caused by influenza for people age more than 65 is also strong that means that the older people dies because of influenza more than young people.

Results and insight:

Statistical hypothesis:

a) The null hypothesis:

I. People age 65 or younger, have a higher mortality rate.
II. Deaths caused by influenza remain the same or increases in small populations.

b) Alternative hypothesis:

I. People older than 65 have a higher mortality rate.
II. Deaths caused by influenza increases in bigger populations.

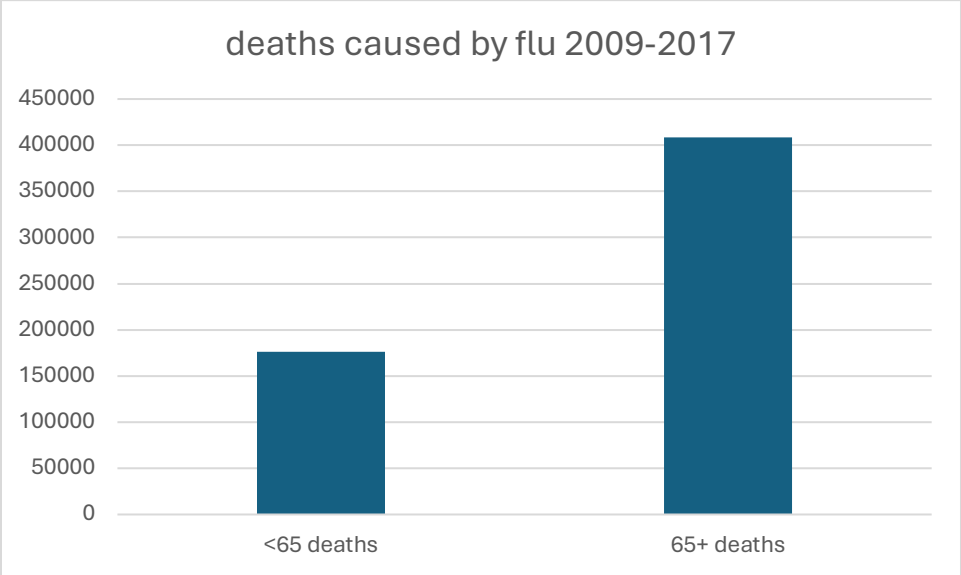
The one-tailed t-test shows that there is 95% confidence level for both variables that were tested (total population of people 65+ and total deaths for 65+, total population and total deaths). In both test, the p-values were lower than the significant level of 0.05, that means we can reject the null hypothesis.

t-Test: Two-Sample Assuming Unequal Variances		
	65+ deaths	*65+*
Mean	889.3006536	806988.939
Variance	953083.9793	7.868E+11
Observations	459	459
Hypothesized Mean Difference	0	
df	458	
t Stat	-19.46985729	
P(T<=t) one-tail	2.92576E-62	
t Critical one-tail	1.648187415	

p-value=2.92576E-62, Alpha=0.05.

t-Test: Two-Sample Assuming Unequal Variances		
	total deaths	*total population*
Mean	1380.60349	3035913.38
Variance	1181920.71	1.188E+13
Observations	459	459
Hypothesized Mean Difference	0	
df	458	
t Stat	-18.862446	
P(T<=t) one-tail	1.9025E-59	
t Critical one-tail	1.64818741	

p-value=1.9025E-59, Alpha=0.05.

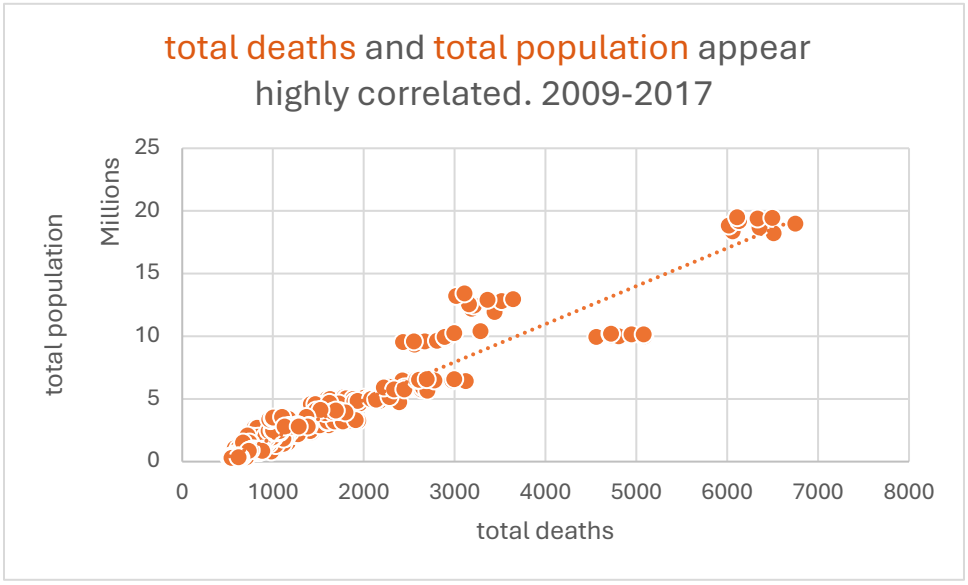


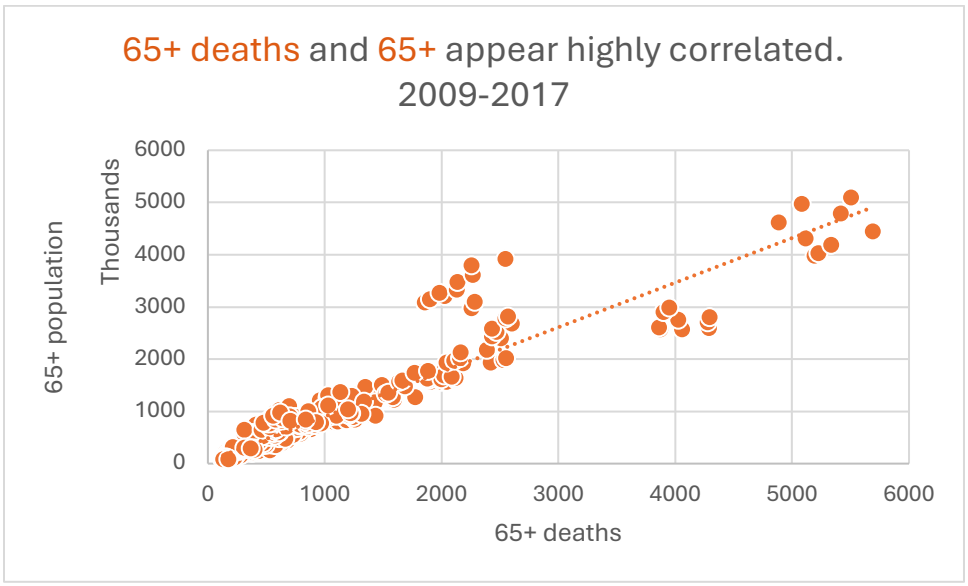
Remaining analysis and next steps:

After analyzing the data sets, we know now that the deaths caused by influenza are impacted by population numbers and people age.

We should dig deeper now to find out which states has the higher populations, and in which states have higher numbers of people older than 65. This way we can determine where to send stuff and how many in each state.

Appendix:


total deaths and total population appear highly correlated. 2009-2017
total population
Millions
25
20
15
10
5
0
0
1000
2000
3000
4000
5000
6000
7000
8000
total deaths


65+ deaths and 65+ appear highly correlated. 2009-2017
65+ population
Thousands
6000
5000
4000
3000
2000
1000
0
0
1000
2000
3000
4000
5000
6000
65+ deaths