

Preparing for Influenza Season: Interim Report (Task 1.10)

by Tristan Savella

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.

Objectives

- ❖ 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

- ❖ Older residents are more likely to suffer from influenza-related complications, with the risks of complications increasing the older the age range. Therefore, states with higher populations of older residents (aged 65 and older) require more medical staffing.

Data Overview

Population Data by Geography

- ❖ External data source, owned by the Census Bureau
- ❖ Additional data is retrieved from the operations of other organizations (e.g. IRS)
- ❖ Provides an estimate of population count in each county in the US each year between 2009-2017
- ❖ Population counts are available by gender and five-year age range

Influenza Deaths by Geography, Time, Age and Gender

- ❖ External data source provided by the Center for Disease Control (CDC)
- ❖ Lists the number of influenza related deaths by state, ten-year age range and each year between 2009-2017

Data Limitations

Population Data by Geography

- ❖ The data is based on population estimates and is therefore prone to error
- ❖ Some counties are missing data from certain years
- ❖ Some counties have multiple entries for certain years
- ❖ Listed population of different age ranges may not exactly add up to the total population
- ❖ Time lag with updating of this data as the Census Bureau conducts their population count every ten years
- ❖ Subject to administrative data collection limitations of other organizations on which the Census Bureau relies on for data (e.g. IRS)

Influenza Deaths by Geography

- ❖ To protect the identity of individuals, the death count has been labeled as “suppressed” for certain states/years if the number of deaths was less than 10
- ❖ These suppressed values were counted as “0” during this project’s statistical hypothesis testing

Descriptive Analyses

Comparing Three Core Variables in Different Age Ranges

- ❖ State Population by Year (2009-17)
- ❖ Flu-Related Death Count by State/Year (2009-17)
- ❖ Flu-Related Death Rate by State/Year (2009-17)
- ❖ Three Age Ranges: 85+, 75-84 & 55-64
- ❖ Data from Selected Younger Age Ranges in Appendix

Ages 85+

	State Population by Year	Flu-Related Death Count	Flu-Related Death Rate
Average	110,737	463	0.372%
Standard Deviation	122,903	551	0.00169%

Ages 75-84

	State Population by Year	Flu-Related Death Count	Flu-Related Death Rate
Average	262,808	246	0.0717%
Standard Deviation	284,946	306	0.0487%

Ages 65-74

	State Population by Year	Flu-Related Death Count	Flu-Related Death Rate
Average	455,885	116	0.0163%
Standard Deviation	487,666	168	0.0159%

Relevant Correlations Between Variables

❖ Correlation 1

- The higher the state population, the higher the death count for those aged 85+ will be
- Strong Correlation (Coefficient = 0.932)

❖ Correlation 2

- The higher the populations of 85+ residents, the higher the death rate for that age range
- Moderate Correlation (Coefficient = 0.398)

Results & Insights

Summarization

- ❖ Older residents are at higher risk for death due to influenza-related complications
- ❖ With each ascending age range, both the death count and death rate increase
- ❖ States with higher populations of older residents will require more medical staffing
- ❖ See tables below for details of hypothesis testing
 - Test 1: Death Count of Age Ranges 85+ and 55-64
 - Test 2: Death Count of Age Ranges 75-64 and 45-54

Statistical Hypothesis Test 1: Death Count (2009-17) of Age Ranges 85+ and 55-64

t-Test: Two-Sample Assuming Unequal Variances		
	Deaths 55-64	Deaths 85+
Mean	58.51198257	463.48366
Variance	10337.72638	303916.06
Observations	459	459
Hypothesized Mean Difference	0	
df	489	
t Stat	-15.47714077	
P(T<=t) one-tail	1.451E-44	
t Critical one-tail	1.647975667	
P(T<=t) two-tail	2.902E-44	
t Critical two-tail	1.96482708	

Test 1 Summary

- ❖ *Null Hypothesis* = $\mu_{85+} \text{ Deaths} \leq \mu_{55-64} \text{ Deaths}$
- ❖ *Alternative Hypothesis* = $\mu_{85+} \text{ Deaths} > \mu_{55-64} \text{ Deaths}$
- ❖ *P-Value* = 1.45099958771555E-44
- ❖ *Alpha* = 0.05
- ❖ **Null Hypothesis = Successfully Rejected**

Statistical Hypothesis Test 2: Death Counts (2009-17) of Age Ranges 75-84 and 45-54

t-Test: Two-Sample Assuming Unequal Variances		
	Deaths 45-54	Deaths 75-84
Mean	16.9477124	246.3638344
Variance	1727.31604	93695.60314
Observations	459	459
Hypothesized Mean Difference	0	
df	475	
t Stat	-15.911236	
P(T<=t) one-tail	2.6599E-46	
t Critical one-tail	1.64806787	
P(T<=t) two-tail	5.3198E-46	
t Critical two-tail	1.96497077	

Test 2 Summary

- ❖ Null Hypothesis = $\mu_{75-84} \text{ Deaths} \leq \mu_{45-54} \text{ Deaths}$
- ❖ Alternative Hypothesis = $\mu_{75-84} \text{ Deaths} > \mu_{45-54} \text{ Deaths}$
- ❖ P-Value = 2.65988684937604E-46
- ❖ Alpha = 0.05
- ❖ **Null Hypothesis = Successfully Rejected**

Remaining Analysis & Next Steps

- ❖ Creating Temporal & Statistical Visualizations to Examine the Following:
 - Time Forecasts for Selected Variables
 - Distribution of Selected Variables
 - Correlation Between Relevant Variables
- ❖ Spatial Analysis: Determining Where and When Staff Should Go
- ❖ Final Presentation to Stakeholders

Appendix

Business Requirements

Goal

- ❖ To help a medical staffing agency that provides temporary workers to clinics and hospitals on an as-needed basis. The analysis will help plan for influenza season, a time when additional staff are in high demand. The final results will examine trends in influenza and how they can be used to proactively plan for staffing needs across the country.

Stakeholders

- ❖ Administrators of the Medical Staffing Agency
- ❖ Medical Agency Frontline Staff (Nurses, Physician Assistants, Doctors)
- ❖ Hospitals and Clinics Using the Staffing Agency's Services
- ❖ Influenza Patients

Success Factors

- ❖ A staffing plan that utilizes all available agency staff per state requirements, without necessitating additional resources.
- ❖ Minimal instances of understaffing and overstaffing across states (a state can be considered understaffed if the staff-to-patient ratio is lower than 90% of the required ratio and overstaffed if greater than 110%).

Assumptions

- ❖ Vulnerable populations suffer the most-severe impacts from the flu and are the most likely to end up in the hospital.
- ❖ Flu shots decrease the chance of becoming infected with the flu.

Constraints

- ❖ The staffing agency has a limited number of nurses, physician assistants, and doctors on staff
- ❖ There's no money to hire additional medical personnel.

Requirements

- ❖ Provide information to support a staffing plan, detailing what data can help inform the timing and spatial distribution of medical personnel throughout the United States.
- ❖ Determine whether influenza occurs seasonally or throughout the entire year. If seasonal, does it start and end at the same time (month) in every state?
- ❖ Prioritize states with large vulnerable populations. Consider categorizing each state as low-, medium-, or high-need based on its vulnerable population count.

- ❖ Assess data limitations that may prevent you from conducting your desired analyses.

Glossary

- ❖ *Influenza*: a contagious viral infection, often causing fever and aches.
- ❖ *Vulnerable populations*: patients likely to develop flu complications requiring additional care, as identified by the Centers for Disease Control and Prevention (CDC). These include adults over 65 years, children under 5 years, and pregnant women, as well as individuals with HIV/AIDs, cancer, heart disease, stroke, diabetes, asthma, and children with neurological disorders.

Additional Context

A count of the historical influenza deaths gives an indication of the severity of flu in an area. Deaths can be prevented with flu shots and adequate medical staff. In the United States, each state has a different population composition, meaning that some states will have more vulnerable populations. In this project, you should pay particular attention to influenza deaths, vulnerable populations, and (optionally) flu-shot rates—particularly in vulnerable populations—to determine medical staffing needs.

Hypothesis Development

Key Questions

- ❖ Which age range(s) suffer(s) the most from flu-related complications?
 - Which states have the highest population of this age range/these age ranges?
- ❖ Do states with higher populations have higher flu-related death counts? Death rates?
 - Does the death count/death rate vary by age range?
- ❖ When does flu season usually begin and end?

Project Overview

Stakeholder Communication

- ❖ Kickoff Meetings/Calls
 - Initial Meeting: Medical Frontline Agency and Frontline Staff
 - Initial Video Call: All Hospitals and Clinics
- ❖ Email: Medical Staffing Agency, Frontline Staff, and Hospitals/Clinics
- ❖ Video Call/Meeting: Presentations at Various Stages of Completion (See Project Deliverables)
- ❖ Available Weekly Calls (for those who wish for more frequent updates)
- ❖ For Urgent Matters:
 - Email All Stakeholders (Agency Admin, Frontline Staff, Patients, Hospitals & Clinics)
 - If Necessary: Phone Call or Meetings with Involved Stakeholders

Scheduling & Milestones

- ❖ Step 1: Kickoff Meetings: Discussing Business Requirements and Asking Questions
- ❖ Step 2: Project Design: Turning Business Requirements into Questions and Form Hypothesis
- ❖ Step 3: Gathering Data: Sourcing, Cleaning, Ensuring Data Quality and Integrating Datasets
- ❖ Step 4: Statistical Analyses: Identify Key Correlations Between Variables
- ❖ Step 5: Statistical Hypothesis Testing: Interpret Results of the Testing
- ❖ Step 6: Consolidation of Analytical Insights: Communicate Insights Via Interim Report
- ❖ Step 7: Create Visualization Checklist: Decide What Visualizations are Needed
- ❖ Step 8: Statistical Visualizations: Create the Visuals to Display Correlation Between Variables
- ❖ Step 9: Spatial Analysis: Plan Where Staff Should Go
- ❖ Step 10: Storytelling: Create the Narrative to Communicate Insights to Go With Visuals
- ❖ Step 11: Final Presentation to Stakeholders

Project Deliverables

- ❖ Kickoff Meetings: Turning Business Requirements Into Research Questions (Step 1)
- ❖ Email: Communication of Initial Hypotheses and Project Design (Step 2)
- ❖ Interim Report: Communication of Consolidated Analytical Insights (Step 6)
- ❖ Final Presentation (Step 11)

Descriptive Analyses

Comparing Three Core Variables in the Different Age Ranges (continued)

- ❖ State Population by Year (2009-17)
- ❖ Flu-Related Death Count by State/Year (2009-17)
- ❖ Flu-Related Death Rate by State/Year (2009-17)
- ❖ Age Ranges 55-64, 45-54 and 35-54

Ages 55-64

	State Population by Year	Flu-Related Death Count	Flu-Related Death Rate
Average	731,001	59	0.0044%
Standard Deviation	777,161	102	0.0053%

Ages 45-54

	State Population by Year	Flu-Related Death Count	Flu-Related Death Rate
Average	863,696	17	0.00015%
Standard Deviation	951,968	42	0.00084%

Ages 35-44

	State Population by Year	Flu-Related Death Count	Flu-Related Death Rate
Average	821,275	2.6	0.00011%
Standard Deviation	966,976	11.6	0.000367%

Original Data Sets

CDC: "Influenza deaths by geography, time, age, and gender"

https://coach-courses-us.s3.amazonaws.com/public/courses/da_program/CDC_Influenza_Deaths_edited.xlsx

Census Bureau: "Population data by geography"

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

Cleaned & Integrated Data Set (CDC and Census with Statistical Hypothesis Testing)

https://docs.google.com/spreadsheets/d/1O_JFEZ39TguHsBPawF7u4W_w_FyuePCT4KJTMKI9xQ/edit?usp=sharing