Misty Ann Stone
April 2023

Exercise 2.10: Influenza Presentation
Project Limitations and Metrics: Additional Technical Information

YouTube Link: https://youtu.be/4DTQf1lsMN0

Tableau Link: https://public.tableau.com/views/Flu_Final/FluPrep?:language=en-US&publish=yes&:display_count=n&:origin=viz_share_link

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.

Data Overview

- Influenza Deaths by State, Month and Year, and Age, is sourced from the CDC. It includes monthly death counts for influenza-related deaths in the United States from 2009-2017, broken into two categories: state and age category. This dataset has several missing data points, but the sample size is more than sufficient for the analysis.[1]
- Population data by geography is sourced from the US Census Bureau. This data set includes yearly population totals by county 2009-2017. Population in recorded in two categories: male and female, as well as by age group. This data set required quite a lot of cleaning.[2]

---

[1] The CDC may suppress many data points for several reasons, particularly for reasons of privacy. For example, we are unable to compete analysis on children, or indeed any population group under the age of 35, due to missing data. Several states had more missing data than others. Particularly, Vermont, Wyoming, Delaware, North Dakota, Montana, South Dakota, Idaho, Rhode Island, New Hampshire, Maine, and New Mexico have many missing observations. For these states, I included the provided data in the principal statistical tests as we do have important data *for that particular age group*. Nonetheless, care was taken in the "Conclusions and Recommendations" phase of the analysis. For example, the Vermont and South Dakota data was completely missing the younger (non-vulnerable) group, so making recommendations based exclusively on mortality rates would send relatively too many staff members to those states because their stated mortality rates (as a state) are much higher than in reality. Recommendations were made solely based on population age groups for this reason.

[2] The population data set required a lot of cleaning. Ideally, we need population measures for each state for each year 2009-2017. This data set includes population measures for each *county* for each year 2009-2017, however, there are inconsistencies. [1] The data set included multiple measures (not duplicates) for the same dates and counties. To address this, I averaged each population of the measures for each county and date. [2] There are also many states who have different counties included for different years. Therefore, there is significant variation in the population measures for each year, but the variation is based on the number of counties. To address this, I calculated the *average* population measures for each county within each state for each year. I then multiplied those (average by county) population measures by the *average* number of counties available for that state for each year.

Data Limitations

This data is from a government source, so we can assume that it is trustworthy and unbiased.

- Influenza Deaths Data Set: The number of data points for age groups under age 35 is too small to do a meaningful analysis, so we delete those observations from the analysis. This data set includes information based on death certificates for U.S. residents, which identify a single underlying cause of death, and so many miss some deaths in which influenza is not the primary cause.
- Population Data Set: This data set included many instances of multiple observations of a single county within a year as well as many instances of missing county observations across years. See footnote 1 for details as to how I addressed these problems. Interim Census data is based on estimates, which may introduce potential for error.

Descriptive Statistics

Population Data, per month, per state

	TOTAL	MALE	FEMALE	AGES 35-64	AGES 65-85+
MEAN	6,285,777	3,091,153	3,194,624	2,445,349	843,106
STANDARD DEVIATION	6,870,984	3,392,839	3,478,725	2,689,274	896,619

Deaths, per 1000 people, per month

	ENTIRE SAMPLE: AGES 35 THROUGH 85+	YOUNG SAMPLE: AGES 35-64	ELDER SAMPLE: AGES 65 AND OLDER
MEAN	0.1842	0.0124	0.2186
STANDARD DEVIATION	0.2095	0.2186	0.2134

- *Correlations:* The correlation between number of deaths (each age group, per month, per year) and total population is 0.50. So, as population increases, deaths increase moderately strongly. Interestingly, the correlation between the number of deaths (per 1000 people, for each month, for each state, and the total population is -0.2137. Therefore, the higher the population of a state, the lower the death rate is, on average. This is a weak relationship, and could have a variety of causes – perhaps the healthcare is better in more populated states? The correlation coefficient between deaths in a state

and the % of people over 65 is 0.206, which is weak. This is not surprising due to the large variance between the states.

- *Note:* See Appendix for several more interesting descriptive statistics.

Hypothesis

Vulnerable people (>65yrs) are more likely to die of influenza than non-vulnerable (<65yrs).

Results and Insights

I conducted a Two sample t-test, assuming unequal variables with alpha 0.05, comparing average number deaths (per 1000 people) of the younger population (ages 35-64) with the average number of deaths (per 1000 people) of the elder population (ages 65-85+). The results indicate that there ***is a significant difference*** between these averages. We should anticipate regions with older populations to need more medical staff. (Note: See appendix for detailed t-test results.)

Appendix

Our data includes the number of deaths per Number of observations, average, and standard deviation of the number of deaths per 1000 people, for each month, for each state:

AGE GROUPS	NUMBER OF OBSERVATIONS	AVERAGE OF DEATHS PER 1000 PEOPLE, FOR EACH MONTH, FOR EACH STATE	STDDEV OF DEATHS PER 1000 PEOPLE, FOR EACH MONTH, FOR EACH STATE
15-24	1	0.0021	#DIV/0!
25-34	21	0.0042	0.0017
35-44	81	0.0062	0.0044
45-54	510	0.0078	0.0058
55-64	1418	0.0144	0.0079
65-74	2307	0.0334	0.0141
75-84	3435	0.0996	0.0416
85+	4309	0.4125	0.1937

Interesting ***Correlations*** between number of deaths (per 1000 people) for each year, for each state and the population of each age group for each month, for each state:

AGE GROUP:	DEATHS PER 1000 PEOPLE, FOR EACH MONTH, FOR EACH STATE
70 TO 74	-0.30
75 TO 79	-0.30
80 TO 84	-0.29
85 YEARS AND OVER	-0.29

T-Test Details

t-Test: Two-Sample Assuming Unequal Variances		
	Ages 35-64	*Ages 65-85+*
Mean	0.012402177	0.218551848
Variance	6.32119E-05	0.045554968
Observations	2009	10051
Hypothesized Mean Difference	0	
df	10188	
t Stat	-96.49774241	
P(T<=t) one-tail	0	
t Critical one-tail	1.645003206	
P(T<=t) two-tail	0	
t Critical two-tail	1.960196861	

SCRIPT

Hello, my name is Misty Ann Stone and I am a data analyst working on this project: Preparing for Influenza Season.

Here, I will highlight my key findings and present a set of recommendations for additional medical staff by state.

By way of foreshadowing, I find that flu season is the winter months: December, January, February, and March. People over the age of 65 are particularly vulnerable to the flu and medical staff should be sent to states with a higher proportion of their population over the age of 65.

Influenza is a contagious respiratory illness caused by a flu virus. Approximately 8% of the population of the United States get the flu every year. The flu can cause mild to severe illness, and we will be focusing on those people who develop serious complications and end up in the hospital. Hospitals and clinics need additional staff to treat these additional patients.

The bar graph on this slide shows the number of people in the US who die of flu each year 2009 to 2017. Flu also affects each state differently. The choropleth map simply shows the states with higher or lower mortality from the year 2013. The states with higher populations, predictably, have a higher mortality rate.

Flu season peaks in the winter months: December, January, February, and March. Flu mortality peaks in January every year, in every region of the US, as you can see by this line graph. The bar chart below shows flu related mortality by month for all of the years considered. The winter months have significantly higher mortality. 13% of flu related deaths take place in January.

Which population group is most likely to need medical care during flu season? People over age 65 are more likely to develop serious complications and require augmented medical care.

The Histogram is a graphic representation of the mortality rates for each age group considered. The darker color bars represent older population groups. Notice that the higher mortality bins on the right side of the chart are dominated by these darker bars. The left side of the chart shows the lowest mortality rates, and this side contains nearly all of the younger age groups. This histogram shows that older population groups have a higher mortality rate.

The US Center for Disease Control (CDC) maintains that people 65 years and older are at higher risk of developing serious flu-related complications. Our analysis is consistent with this conclusion. The footnote as well as the accompanying pdf file contains the details of this statistical test.

Nonetheless, consider the Scatterplot. This contains one point for each state with their average deaths (per 1000 people) on the vertical axis and the percentage of their population over the age of 65 on the horizontal axis. The trend line shows that the older population states tend to have a higher mortality rate.

In conclusion, my analysis indicates that the medical staffing agency should plan to send additional staff during the months of December, January, February, and March. More staff should be deployed to states with higher percentage of people over the age of 65. In particular, Florida, Maine, and West Virginia have the relatively highest percentage of people in the vulnerable group. States such as Utah and Texas will need fewer staff as they have the lowest percentage of people in the vulnerable group.

At the bottom of this slide, I have included a map and a chart with specific staffing recommended percentages.

Thank you, and I'm looking forward to following up on this analysis with this agency in the future. I recommend continuing to monitor the influenza death rates and surveying the staff about other possible factor affecting hospital and clinic staffing needs during flu season.