Interim Report: Staffing for 2018 Influenza Season

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 state has a high population of vulnerable patients (65+ years old) relative to the entire population of the state, the more likely influenza hospitalizations and mortality will occur for that vulnerable group compared to the rest of the population (64 and under).

Data Overview

Influenza Deaths by geography, time, age, and gender: This data source contains a sample of the number of influenza deaths by state, from the year 2009-2017. It is broken down into total population, female population, male population, and age groups. This data is from the CDC.
Population data by geography: This data source contains a sample of the US population broken down by county and state for the years 2009-2017. The data is further broken down into population by age groups. This data is from the US Census Bureau.

Data Limitations

Influenza Deaths: the data set is a sample of the total deaths and not a showing of the total numbers of influenza deaths. Also, the suppressed deaths (deaths numbering 10 and under) were over 5% of the data. Imputation of a random number (5) was used to replace the suppressed deaths using =randbetween in excel to pick the value. I chose a random number as there was no specific reason one number should be used over another.
Population: The data set is a sample of the total population of the US. The data is entered manually and therefor may contain errors. Additionally, not everyone responds to the US Census.

Descriptive Analysis

The hypothesis is associated to influenza deaths and the population of those aged 65+. A descriptive analysis of these two variables was conducted and is shown below in Table 1.

	Number of Influenza Deaths (65+)	Population Age 65+
Mean	897	806,990
Standard Deviation	972	887,017
1st Lower Deviation	-75	-80,027
1st Higher Deviation	1869	1,694,007
2nd Lower Deviation	-1045	-967,044
2nd Higher Deviation	2840	2,581,024

Table 1: Summary of descriptive analysis of core variables

The mean is the average of a series of numbers, in this case we took the average of the number of influenza deaths of those aged 65+, also calculated for the population aged 65+. On average there were 897 influenza deaths in that age group across the US. The average population for those aged 65+ was 806,990. The standard deviation provides how much the values may vary from the mean. We usually look at 2 standard deviations from the mean and go higher and lower. In Table 1, you can see the standard deviations from the mean for both Influenza Deaths and Population. Since

there is negative numbers and we cannot have negative people or numbers of deaths, those values will be analyzed as "0".
The hypothesis states that the number of influenza deaths for those aged 65+ will rise as the population for that age group does. A positive correlation was found in the analysis of the two variables. This positive correlation lends credence to the hypothesis being correct.

Variables	Influenza Deaths & Population of 65+
Proposed Relation	A person aged 65+ is at higher risk of influenza death. This makes for a positive relationship between two variables.
Correlation Coefficient	0.94
Strength of Correlation	Strong

Table 2: Correlation between number of influenza deaths and people aged 65+

Results and Insights
With varying a strong correlation between influenza deaths and populations aged 65+, a testing of the hypothesis was completed using a T- Test. We measured the influenza death rate by those aged 0 to 64 and those aged 65+ using an alpha (significance) level of 0.05. The results are summarized in Table 3.

Null Hypothesis	The mortality rate of people 65+ years old is lower than or equal to people under 65 years old.						
Alternative Hypothesis	The mortality rate of people 65+ years old is greater than people under 65 years old.						
Dependent Variable	Influenza Death Rate						
Independent Variable	Age Group (65+)						
Two-Tailed or One-Tailed	One-Tailed as we are only focused on influenza deaths for people 65+						
P-Value	1.7027E-171						
Significance Level	The p-value is below 0.05, meaning the null hypothesis is false. With 99%						
	Aged 65+	*Aged 0 to 64*					
Mean	0.0013165	0.000269203					
Variance	2.744E-07	7.60736E-08					
Observations	459	459					
Hypothesized Mean Difference	0						
df	694						
t Stat	37.901722						
P(T<=t) one-tail	1.7E-171						
t Critical one-tail	1.6470522						
P(T<=t) two-tail	3.41E-171						
t Critical two-tail	1.9633881						

Table 3: Hypothesis Testing

Next Steps

- A closer analysis of the states and their populations and influenza deaths
- Create a time forecast
- Create visualizations
- Spatial Analysis
- Textual Analysis
- Data Presentation

Appendix

Data Profile

Influenza Deaths	Total Deaths	<5	5-14	15-24	25-34	35-44	45-54	55-64	1 to 64	65+	NS
min	720	120	60	60	60	60	60	60	480	180	60
max	6812	120	65	66	123	183	346	589	1339	5694	60
mean	1493	120	60	60	60	62	71	103	537	897	60

Table 4: Influenza Deaths by age groups. NS is not stated

Population	Total Pop.	Under 5	5 to 14	15 to 24	25 to 34	35 to 44	45 to 54	55 to 64	1 to 64	65+
min	490148	24255	52027	65014	58373	57037	62572	61509	417526	47816
max	38572021	2705688	5120725	5593680	5762760	5350966	5241677	4543110	33617086	5078704
mean	5973849	386283	786272	838031	805692	793104	844879	712777	5167039	806990

Tables 5: Population by age group

Influenza Deaths Variables	time -variant/-invariant	structured/unstructured	qualitative/quantitative	qualitative: nominal/ordinal quantitative: discrete/continuous
State	Invariant	Structured	Qualitative	ordinal
Year	Variant	Structured	Qualitative	ordinal
Ten-Year Age Groups	Variant	Structured	Quantitative	continuous
Deaths	Variant	Structured	Quantitative	discrete

Table 6: Data Profile for Influenza Deaths

Population Variables	time -variant/-invariant	structured/unstructured	qualitative/quantitative	qualitative: nominal/ordinal quantitative: discrete/continuous
State	Invariant	Unstructured	Qualitative	nominal
Year	Variant	Structured	Quantitative	continuous
Total Pop.	Variant	Structured	Quantitative	discrete
Under 5	Variant	Structured	Quantitative	discrete
5 to 9	Variant	Structured	Quantitative	discrete
10 to 14	Variant	Structured	Quantitative	discrete
15 to 19	Variant	Structured	Quantitative	discrete
20 to 24	Variant	Structured	Quantitative	discrete
25 to 29	Variant	Structured	Quantitative	discrete
30 to 34	Variant	Structured	Quantitative	discrete
35 to 39	Variant	Structured	Quantitative	discrete
40 to 44	Variant	Structured	Quantitative	discrete
45 to 49	Variant	Structured	Quantitative	discrete
50 to 54	Variant	Structured	Quantitative	discrete
55 to 59	Variant	Structured	Quantitative	discrete
60 to 64	Variant	Structured	Quantitative	discrete
0 to 64	Variant	Structured	Quantitative	discrete
65 +	Variant	Structured	Quantitative	discrete

Table 7: Data Profile for Population