PREPARING FOR INFLUENZA SEASON: Interim Report

1. 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.

2. HYPOTHESIS

If a state's population has a large number of persons aged 65+, mortality rates will be greater, necessitating more medical help to treat flu sufferers.

3. DATA OVERVIEW

Population Data by Geography

From 2009 to 2017, this data covers the number of persons in each US county, split down by five-year age increments. Every ten years, the US Census Bureau collects this government data in the form of a survey.

Influenza Deaths by Geography, Time, Age, and Gender

From 2009 through 2017, the data includes monthly mortality numbers for influenza-related fatalities in the United States. This is part of the government's vital statistics program generated by the US Centers for Disease Control and Prevention (CDC).

Counts of Influenza Laboratory Test Results by State (Survey)

From 2010 to 2015, the Influenza Lab data monitors specimens tested and % positive rates. The data is broken down by state and week of the year. The CDC gathered data from 3,500 hospitals and clinics around the country to create this sample data set.

4. DATA LIMITATIONS

The high percentage of suppressed data in the Influenza Deaths data set is the most significant source of mistakes in these data sets (over 80 percent of all records). This information was withheld due to privacy concerns, and it reflects any observation with less than ten fatalities. For each record with suppressed data, we imputed " 0 deaths." This will influence the data, particularly in states with lower populations.

Furthermore, the Influenza Lab data set excludes information from Florida and information from 2009, 2016-17. Because Florida has a large vulnerable population (age 65+), its absence must be noted when considering states with a larger elderly population.

5. DESCRIPTIVE ANALYSIS

	Population <65 yrs.	Population 65+ yrs.
Type	Population	Population
Standard Deviation	5,938,164	886,051
Mean	5,167,039	806,990
Outlier %	3.92%	13.07%

Correlation

The death rate from influenza for persons 65+ is highly correlated with the population age. I will postulate that the reason for the significant correlation is that for a high population, the vulnerability is higher even when normalized with the total population. This gives us an interesting perspective, maybe this is so due to negligence or overburdens on public health.

Age Group	<5	5-14	15-24	25-34	35-44	45-54	55-64	65+
Correlation Coefficient	0	.22	.21	.42	.63	.87	.94	.94
Strength of Correlation	No Relationship	Weak	Weak	Moderate	Strong	Strong	Strong	Strong

6. SUMMARY & INSIGHTS

Null Hypothesis

> Influenza-related fatalities in adults 65+ are lower than or equivalent to those among people 64 and younger.

Alternative Hypothesis

> Influenza-related fatalities in adults 65+ are higher than those among people 64 and younger.

A one-tailed t-test evaluating the significance of this connection yielded a p-value (4.955E-45) demonstrating a 99% confidence that the flu death rate is greater in adults 65+ than in people younger than 65 from 2009 to 2017.

Higher death rates indicate higher levels of care and poor outcomes when it comes to treating individuals in clinics and hospitals. More physicians will be needed in areas with a larger number of persons aged 65+ to treat them when they come in with flu symptoms.

7. REMAINING ANALYSIS & NEXT STEPS

Following the rejection of the null hypothesis, a more detailed examination of individual states is required. Trends might also be discovered by visualizing the data as a time series from the available years. These patterns may reveal a shifting positive rate, a growing or diminishing population, and so on.

The analysis is progressing according to plan, and this report will be followed by a meeting with all parties. We'll go through the findings thus far and answer any questions you have about our research during the conference. This will also be a time for us to ask questions so that we can make sure we're on the right track.

8. APPENDIX

Project Brief

Data Profile – Population

Data Types

Variables	time -variant/-invariant	structured/unstructured	qualitative/quantitative	qualitative: nominal/ordinal quantitative: discrete/continuous
County	time-invariant	structured	qualitative	nominal
Year	time-invariant	structured	quantitative	discrete
Total population	time -variant	structured	quantitative	discrete
Male total population	time -variant	structured	quantitative	discrete
Female total population	time -variant	structured	quantitative	discrete
Age Groups (multi)	time -variant	structured	quantitative	discrete

Data Accuracy

	Min	Max	Ave
Year	2009	2017	2013
Total Population	41	10105722	97842
Under 5 years	0	733897	6309
5 to 9 years	0	665400	6378
10 to 14 years	0	724112	6496
15 to 19 years	0	753657	6804
20 to 24 years	0	777988	6909
25 to 29 years	0	831276	6714
30 to 34 years	0	762619	6436
35 to 39 years	0	753468	6345
40 to 44 years	1	733897	6614
45 to 49 years	0	704718	6901
50 to 54 years	0	683887	6922
55 to 59 years	3	628513	6293
60 to 64 years	0	535357	5416
65 to 69 years	0	415243	4189
70 to 74 years	0	295420	3129
75 to 79 years	0	215181	2394
80 to 84 years	0	161647	1821
85 years and over	0	177493	1774

Data Cleaning

Separated all County's from States

Data Profile – Influenza Deaths

Data Types

Variables	time -variant/-invariant	structured/unstructured	qualitative/quantitative	qualitative: nominal/ordinal quantitative: discrete/continuous
State	time -invariant	structured	qualitative	nominal
State Code	time -invariant	structured	quantitative	discrete
Year	time -invariant	structured	quantitative	discrete
Month	time -invariant	structured	qualitative	ordinal
Month Code	time -invariant	structured	qualitative	ordinal
Ten-Year Age Group	time -invariant	structured	qualitative	ordinal
Ten-Year Age Group	time -invariant	structured	qualitative	ordinal
Deaths	time -variant	structured	quantitative	discrete

Data Accuracy

	State Code	*Year*	*Deaths*
minimum	1	2009	10
maximum	56	2017	512
mean	29	2013	34

Data Cleaning

Replaced all abbreviation to the full name

Replaced #N/A to District of Columbia and code

Changed 20133 to 2013

Replaced all “Suppressed” data with 0 deaths.