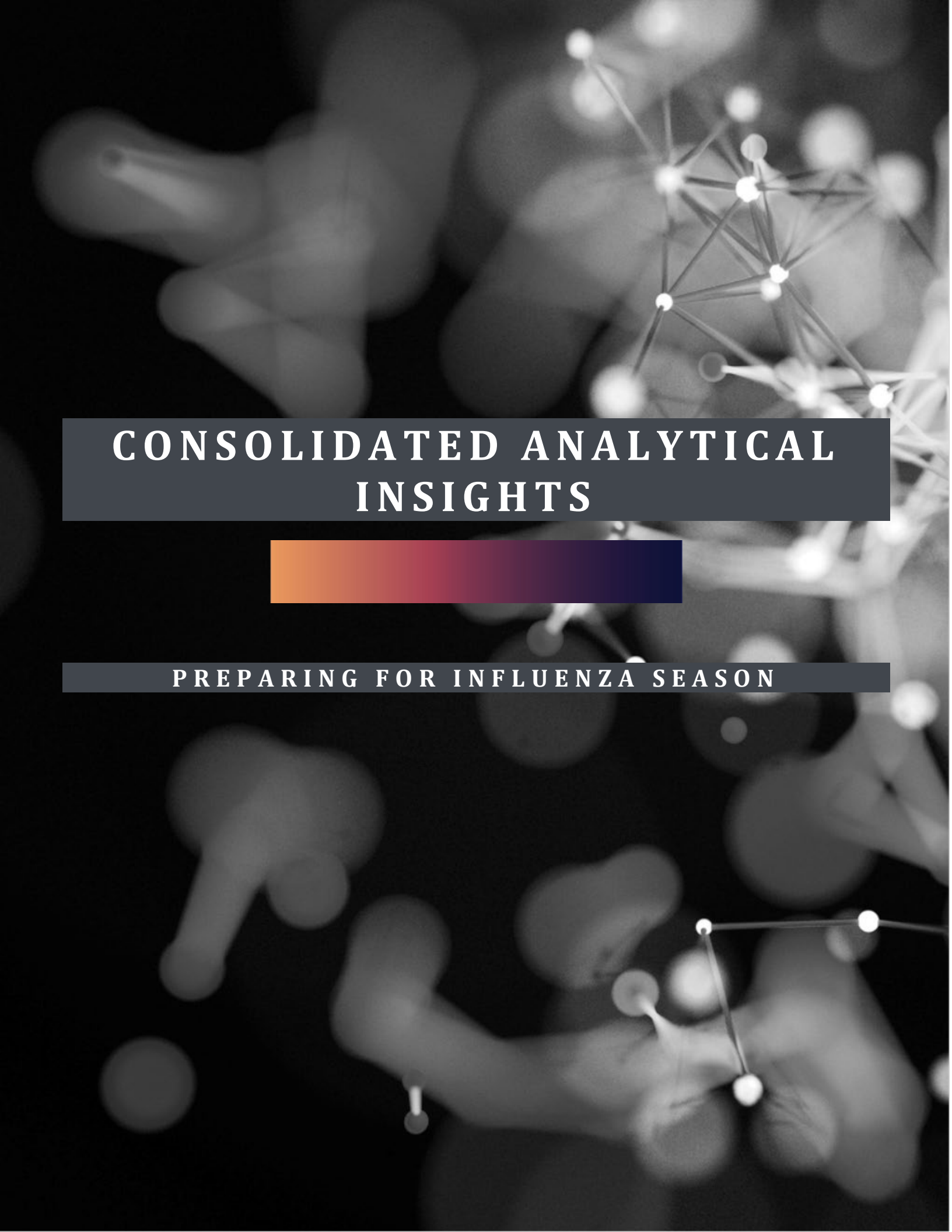

CONSOLIDATED ANALYTICAL INSIGHTS

PREPARING FOR INFLUENZA SEASON

SUMMARY

The goal of this analysis is 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 the influenza season when additional staff is in high demand. The final results will examine trends in influenza and how they can be used to plan for staffing needs across the country proactively.

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 treat these extra patients adequately. 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.

To meet the needs of the clients' datasets from the US Census (2010-2017), the CDC US Influenza Deaths (2010-2017), and the CDC Testing/Visits (2010-2017) are being analyzed.

INTRODUCTION

HYPOTHESIS

The Null Hypothesis is that with decreased visits of under less than or equal to 14%, there will be an increase in deaths in general. The ancillary effect would be an increase in the deaths of vulnerable populations throughout the United States.

Null

H_0 = DR(14% +) >= DR(14% -)

Alt

H_1: DR(14% +) < DR(14% -)

Note: DR = Death Rate

METHOD

After cleaning the datasets, it was decided that using the years 2010-2017 would allow for more advanced analysis, such as a one-tailed T-test. T-testing provides the standard deviation from each population group, the amount of data values from each group, and the mean difference between the values of the data sets, which are critical for statistical analysis.

SUMMARY OF RESULTS

The P-value was <α=.05, so the null hypothesis at this point is rejected. A secondary test of deaths versus the number of providers yielded more promising results. Further exploration of these two variables at this point is recommended.

Figure 1

DATA OVERVIEW

CLIENT PRIORITIES

- 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%)

STAKEHOLDERS

- Medical agency frontline staff (nurses, physician assistants, and doctors)
- Hospitals and clinics using the staffing agency's services
- Influenza patients
- Staffing agency administrators

OVERVIEW

The null hypothesis was created to meet the client's goal of lessening the impact of the seasonal flu on vulnerable populations throughout the US and increasing the number of flu shots given during the pre-season to decrease the infection rate during the flu season. Vulnerable populations, as defined in by the CDC, are 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, children under five, pregnant women, individuals with HIV/AIDs, cancer, heart disease, stroke, diabetes, asthma, and children with neurological disorders.

Influenza deaths by geography, time, age, and gender (Dataset #1)

Observation
The data is grouped by age groups. The data set only contains information on death caused by influenza. There is no identifying information that would reveal the identities of the person nor their family. The only identifier is the state where the deaths occurred, month and year of death, and age group.

Variables
There are several variables listed in data terms. The main ones for this analysis will be State, Year, Month, and Ten-Year Group Code.

Describing & Summarizing Data
According to the CDC website, it receives its data from 50 states and 3000+ local jurisdictions and territories. Additionally, data comes to CDC in a variety of formats. Some data are sent via excel spreadsheets, fax machines, or even by phone or hand. The data covers from 2009 to 2017, and there is the anomaly of 20133, which can be attributed to a clerical error (i.e., human error). There is data for all 50 states, and US territories such as Puerto Rico or Guam are excluded.
Limitations: It is up to each city, county, and state to decide what information is collected, as well as how and when it can be shared with CDC. These decisions can vary widely, leading to big differences in the data CDC receives.

Counts of influenza laboratory test results by state (survey) (Dataset #2)

Observation
The data is grouped by state groups. The data set only contains information on test results for influenza by strain. There is no identifying information that would reveal the identities of the person nor their family. The only identifiers are the state where the tests were taken, the week, and the amount of test given.

Variables
There are several variables listed in data terms. The main ones to be used for this analysis will be State, Year, and number of tests given.

Describing & Summarizing Data
Influenza Visits tracks patient visits to a medical provider for influenza. It counts the number of visits, providers, and total patients seen by week and state from late 2010 to early 2019. This report comes from 3,500 outpatient healthcare providers.
Limitations: It is up to each city, county, and state to decide what information is collected, as well as how and when it can be shared with CDC. These decisions can vary widely, leading to significant differences in the data CDC receives.

Survey of flu shot rates in children (Dataset #3)

Observation
The data is grouped by state, providers, age and number shots given. The data set only contains information for influenza shots. There is no identifying information that would reveal the identities of the person nor their family. The only identifiers are the state where the shots were taken given and a general timeframe.

Variables
There are several variables listed in data terms. The main ones to be used for this analysis will be State, Year, and number of shots given. At some point, the number of providers might be used.
Data Types

Describing & Summarizing Data
The National Immunization Surveys (NIS) provide survey data collected through telephone interviews with parents in all states and U.S. territories. The University of Chicago runs the surveys on behalf of the Centers for Disease Control (CDC).

The data comes from a random sampling of parents. The demographics are self-reported, but the flu shot information is verified with health providers and can be considered accurate.
The data contains flu shot data for children 6 months to 17 years. It's categorized by geographic state and contains family demographics, including poverty level, race, and parent marital status.

Limitations: It is up to each city, county, and state to decide what information is collected, as well as how and when it can be shared with CDC. These decisions can vary widely, leading to big differences in the data CDC receives.

LIMITATIONS

Before any analysis could be done, a general overview of the datasets had to be completed. What was found was that the datasets are grouped by age groups. The data set only contains information on death caused by influenza. There is no identifying information that would reveal the person's identity or family. The only identifier is the state where the deaths occurred, month and year of death, and age group.

There are several variables listed in data terms. The main ones used for the analysis will be State, Year, Month, and Ten-Year Group Code. There are limitations based on the fact that it is up to each city, county, and state to decide what information is collected, as well as how and when it can be shared with CDC. These decisions can vary widely, leading to big differences in the data CDC receives.

Based on the project brief, there are specific 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.

And 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.

ANALYSIS & INSIGHTS

To complete the analysis, three datasets were used: Counts of influenza laboratory test results by state, Influenza deaths by geography, time, age, and gender, and a Survey of flu shot rates in children.

The null hypothesis was refined to '*The fewer visits, the higher the death rate, below or equal to 14%*'. With the alternative null being: 'Higher *visits, the lower the death rate. Above or equal to 14%*'.

A one-tailed T-test was conducted on the integrated datasets for 2010-2017 across all 50 US states. The reason behind a two-tailed test versus a one-tail test is that in a two-tailed test, regardless of the direction of the relationship, I hypothesize I would be testing for the possibility of the relationship in both directions.

For deaths with visits of more than 14% of the population, the p-value was 4.20823E-87 with a t Critical two-tailed value of 1.97. There is a significant difference from the normal population. Meaning for states with a higher vulnerable population, it would be better to have a higher ratio of flu shots than less.

APPENDIX

I would recommend dropping Florida from further analysis because the state does not provide any information that would directly address the questions being posed by the client. I would further explore the number of providers based on the age demographics of the state and/or vulnerable population.

Data Spread

	Variable 1:	Variable2:	Variable 3:
Data Set Name	US Population by State (2010-2017)	US Influenza Deaths by State (2010-2017)	US Visits Influenza by State (2010-2017)
Sample or Population?	Population of the US based on US census data.	Population of the US based on US census data.	Population of the US based on US census data.
Normal Distribution?	0.19	0.00025	0.50
Variance	4.65974E+13	1330719	9.21979E+11
Standard Deviation	6834613	1154	961376
Mean	5983818	908	910
Outlier Percentage	8%	6%	9%

Correlation

Variables:	US population, two age groups (0-64 & >65)	US deaths from influenza, two age groups (0-64 & >65)	US medical test for influenza, two age groups (0-64 & >65)
Proposed Relationship(s):	There are more than 0-64 in the US population than>65-year-old	There will be higher deaths among the >65 age group than 0-64	There will be a lower death rate if there are higher visits for testing
Correlation Coefficient	1.0/1.0	1.0	0.5
Strength of Correlation	Strong	Strong	Moderate
Usefulness /Interpretation	This data is useful for testing other variables such as death rate or testing.	This information provides a starting point to analyze further where the death rate is highest and start to formulate a why.	This information is inconclusive as it does not indicate clearly if there is a correlation between testing and death from influenza.

t-Test: Two-Sample Assuming Unequal Variances

	DR >14%	*Visits <14%*
Mean	0.000113155	0.060646988
Variance	5.29789E-09	0.001007256
Observations	152	256

Hypothesized Mean Difference	0
df	255
t Stat	-30.5173156
P(T<=t) one-tail	2.10411E-87
t Critical one-tail	1.650851092
P(T<=t) two-tail	4.20823E-87
t Critical two-tail	1.96931057