Data Immersion
Exercise 1.10
Jordan Novelli

Influenza Data: Interim Report

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

Research Hypothesis:

- If hospitals/clinics are staffed appropriately in states that are hit the hardest by the flu, then more patients will be treated sooner and recover quicker.

Data Overview:

- CDC Influenza Death Data Set: Documentation of deaths caused by Influenza across the United States categorized by ages.
- Census Population Transformed Data Set: Population sizes of different counties within the United States categorized by Total Population, Sex, and Age in ten year groups.
- CDC Influenza Visits Data Set: Number of Medical Providers and Total Patients ratios by State.

Data Limitations:

- The Death data set is limited by the number of deaths. Any documentation that are 9 – 20 deaths per that state are not documented and will be labeled as suppressed.
- The Population Consensus is limited to year 2009-2017. Furthermore, according to the United States Census Bureau the population statistic come from decennial censuses which means it's counted every 10 years.
- The Influenza Visits data set is limited to the years 2010-2019 however, these statistics are reported each week according to Fluview Interactive.

Data Immersion
Exercise 1.10
Jordan Novelli

<u>Influenza Data: Interim Report</u>
<u>Descriptive Analysis:</u>

- These stats are derived from the CDC Influenza Death data set Using the reported deaths of 85+ years.
- Hypothesis: If people 85 years + are dying from Influenza, then flu vaccines should be provided more readily for them prior to their most potent flu season (based on state).

Variance (sample = VAR.S()):	1381	One standard deviation is -3 to 71
Standard Deviation (sample = STDEV.S()):	37	Two standard deviation is -40 to 108
Average (=average()):	34	Outliers are -77 to 145

- Use the CDC Influenza Visits consensus data set.
- Hypothesis: If hospitals/clinics are staffed appropriately in states that are hit the hardest by the flu, then more patients will be treated sooner and recover quicker.

Variance (sample = VAR.S()):	939	One standard deviation is 3 to 63
Standard Deviation (sample = STDEV.S()):	31	Two standard deviation is -27 to 93
Average (=average()):	33	Outliers are -57 to 123

- The correlation results of the CDC Influenza Death data set:

CORRELATION

Variable	Death/Population
Proposed Relationship	Age to population ratio
Correlation Coefficient	1
Strength of Correlation	Strong Relationship
Usefulness/Interpretation	Useful

<u>Summary:</u>

- The statistical Hypothesis: If hospitals/clinics are staffed appropriately in states that are hit the hardest by the flu, then more patients will be treated sooner and recover quicker, still needs some further analysis. We can indicate that Influenza causes many deaths for people over 85+ years.

Influenza Data: Interim Report

Additional Analysis: Next steps are to determine the top states with the highest vulnerability and the amount of medical staff provided in the past and how to provide additional help. We will categorize states via a low, medium and high need based on their vulnerability and indicate if there is a season in which Influenza hits the hardest.

There will be visualizations created to help see states affect by influenza and their influenza season as well as the amount of deaths that occurred by year.

Appendix:

Among one of the final tasks there were other hypotheses created and tested to help indicate whether specific age groups where affect more heavily by the Influenza virus.

Research Hypothesis	If hospitals/clinics are staffed appropriately in states that are hit the hardest by the flu, then more patients will be treated sooner and recover quicker.

Independent	Increasing staff in states hit the hardest by the flu.
Dependent	The amount of ill individuals by state.
Null Hypothesis	More children (<15 years) die from influenza than the elderly (75 years +).
Alternative Hypotheis	More elderly die from influenza (75 years +)
Tailed Test	Two-Tailed Test
Alpha =	0.05

t-Test: Two-Sample Assuming Unequal Variances

	NUM. OF PROVIDERS	TOTAL PATIENTS
Mean	33.37770594	14481.55134
Variance	939.1721139	385085458.5
Observations	24437	24437
Hypothesized Mean Difference	0	
df	24436	
t Stat	-115.0953578	
P(T<=t) one-tail	0	
t Critical one-tail	1.644915987	
P(T<=t) two-tail	0	
t Critical two-tail	1.96006107	

Interpretation of Results

a. The relevent p-value for the one-tailed test is 0.

b. Our significance level for alpha was 0.05 or (95%). The P-value is zero therefore we have successfully rejected our null hypothesis.

Data Immersion
Exercise 1.10
Jordan Novelli

Influenza Data: Interim Report

To summarize, my hypothesis that more children (<15years) die from influenza than the elderly (75year+) is rejected per the T-Test and the p-value obtained. This indicates that there the alternative hypothesis is correct.

Further steps I could take would be:

1. Determine the states that have the largest amount of 75 years + deaths
2. Determine the states that have the least amount of 75 years + deaths
3. Determine if these deaths happened in states with lower providers stationed there