**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**

Based on all the data available from the CDC based on flu cases and vaccines.

If a state has a high number of vulnerable populations. Then will that state have more deaths than other states.

**Data Overview**

* Population data by geography

Source: US Census Bureau

This data shows population from 2009 to 2017

Population is broken down into 5-year intervals and by male and female

* Influenza deaths by geography, time, age, and gender

Source: CDC

This data shows total influenza deaths

This data is from 2009 to 2017

This data is broken down total number of deaths per month and into 10-year age groups

**Data Limitations**

* The Census data is only taken every 10 years, so this data is changing between data being collected. Not everyone participates in the Census, it gets mailed to everyone but not everyone fills it out. This can cause the data not to completely accurate.
* The Influenza Death data does account for deaths under 10. I adjusted the data to reflect 5 deaths for every area that didn’t disclose the death rate. I chose 5 as it is the average between 0 and 10. It is unknown how well this number reflects the true death rates in these areas.

**Descriptive Analysis**

|  |  |  |
| --- | --- | --- |
|  | Variable 1 | Variable 2 |
| Data Set Name | **Vulnerable Death Rates** | **Non-Vulnerable Death Rates** |
| Sample or Population | Sample | Sample |
| Normal Distribution? | No | No |
| Variance | 0.000001% | 0.000003% |
| Standard Deviation | 0.01194% | 0.01781% |
| Mean | 0.02352% | 0.01753% |
| Outlier Percentage | 0.01% | 1% |
|  |  |  |
| Correlation |  |  |
|  |  |  |
| Variables: | **Vulnerable Death Rates** | **Non-Vulnerable Death Rates** |
| Proposed Relationship: | Strong | Strong |
| Correlated Coefficient: | 0.95 | 0.95 |

In this data we compared deaths between the vulnerable population (1-5years and 65 years and older) verse the non-vulnerable population’s death due to influenza. We can see that the vulnerability population, even though it is a smaller portion of the total population, does have more deaths due to influenza.

**Results and Insight**

* Hypothesis: If a state has a high number of vulnerable populations. Then will that state have more deaths than other states.
* Null Hypothesis: If the state has a high number of the vulnerable group, then the influenza deaths in that state will be lower.
* Alternative Hypothesis: If the vulnerable population is higher in the state, then the influenza deaths will be higher.

I used a One-Tailed Test as we are only interested in one direction (higher or lower). If the vulnerable group deaths due to influenza are higher or lower than the non-vulnerable group.

Since our P-value is 0.000 and our significant level of 0.05. We can determine our null hypothesis was disproven. The states with a higher population of the vulnerable group will have a higher death rate from influenza.

**Remaining Analysis and Next Step**

The next step is to look deeper into what states will need the most staff to accommodate for the higher number of cases and what states need less staffing due to the key variables. This will help us decide what states are more effected by influenza. We will then make visuals of our findings for the stakeholders on Tableau.

**Appendix**

* Project Management Plan
* Business Requirement Questions

|  |  |  |
| --- | --- | --- |
| Data Spread | Influenza | |
|  | Variable 1 | Variable 2 |
| Data Set Name | **Vulnerable Death Rates** | **Non-Vulnerable Death Rates** |
| Sample or Population | Sample | Sample |
| Normal Distribution? | No | No |
| Variance | 0.000001% | 0.000003% |
| Standard Deviation | 0.01194% | 0.01781% |
| Mean | 0.02352% | 0.01753% |
| Outlier Percentage | 0.01% | 1% |
|  |  |  |
| Correlation |  |  |
|  |  |  |
| Variables: | **Vulnerable Death Rates** | **Non-Vulnerable Death Rates** |
| Proposed Relationship: | Strong | Strong |
| Correlated Coefficient: | 0.95 | 0.95 |
| Strength of Correlation: | Strong | Strong |
| Usefulness/interpretation: | useful | useful |
|  |  |  |

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |  |  |
|  | Variables Comparison | | | | |  |  |  |  |  |
|  |  |  |  |  |  |
|  | **Data Set 1: Influenza Deaths** | | | | **Data Set 2: Census** | | |  |  |  |
|  | **Variables** |  |  | **Examples** |  | **Variables** |  |  |  |  |
|  | Death |  |  | 5 |  |  |  |  |  |  |
|  | Ten-Year Age Groups Code | | | 1 |  |  |  |  |  |  |
|  | Ten-Year Age Groups | |  | 1-4 years |  |  |  |  |  |  |
|  | Month code |  |  | 2009/04 |  |  |  |  |  |  |
|  | Month |  |  | Apr., 2009 |  | Country |  |  | Autauga County |  |
|  | Year |  |  | 2009 |  | State |  |  | Alabama |  |
|  | State Code |  |  | 1 |  |  |  |  |  |  |
|  | State |  |  | Alabama |  | Year |  |  | 2009 |  |
|  |  |  |  |  |  | Total population | |  | 49584 |  |
|  |  |  |  |  |  | Male Total population | | | 24057 |  |
|  |  |  |  |  |  | Female Total population | | | 25527 |  |
|  |  |  |  |  |  | Under 5 years | |  | 3421 |  |
|  |  |  |  |  |  | 5 to 9 years | |  | 3570 |  |
|  |  |  |  |  |  | 10 to 14 years | |  | 4413 |  |
|  |  |  |  |  |  | 15 to 19 years | |  | 4016 |  |
|  |  |  |  |  |  | 20 to 24 years | |  | 2727 |  |
|  |  |  |  |  |  | 25 to 29 years | |  | 2876 |  |
|  |  |  |  |  |  | 30 to 34 years | |  | 3025 |  |
|  |  |  |  |  |  | 35 to 39 years | |  | 3520 |  |
|  |  |  |  |  |  | 40 to 44 years | |  | 4363 |  |
|  |  |  |  |  |  | 45 to 49 years | |  | 3967 |  |
|  |  |  |  |  |  | 50 to 54 years | |  | 3223 |  |
|  |  |  |  |  |  | 55 to 59 years | |  | 2628 |  |
|  |  |  |  |  |  | 60 to 64 years | |  | 2430 |  |
|  |  |  |  |  |  | 65 to 69 years | |  | 1983 |  |
|  |  |  |  |  |  | 70 to 74 years | |  | 1438 |  |
|  |  |  |  |  |  | 75 to 79 years | |  | 893 |  |
|  |  |  |  |  |  | 80 to 84 years | |  | 645 |  |
|  |  |  |  |  |  | 85 years and over | |  | 496 |  |

|  |  |  |
| --- | --- | --- |
| **Question** |  | **Answer** |
| **1. Identify the dependent and independent variables in your research hypothesis.** |  | **Dependent: Deaths due to Influenze Independent: Portion of people who are in the vunlerable group.** |
| **2. Research Hypothesis** |  | **If a state has a higher population of the vulnerable group, then there will be more infuenza deaths in that state.** |
| **Null Hypothesis** |  | **If the state has a high number of the vulnerable group then the influenza deaths in that state will be lower.** |
| **Alternative Hypothesis** |  | **If the vulnerable population is higher in the state, then the influenza deaths will be higher.** |
| **Test Type** |  | **We will use a One-Tailed Test as we are only interested in one direction (higher or lower). If the vulnerable group deaths due to influenza are higher or lower than the non-vulnerable group.** |
| **4. Find the relevant p-value for your one- or two-tailed test.** |  | **Alpha 0.05** |
| **Refer to the significance level to assess whether the two groups are significantly different.** |  | **0.000000000000000000000000000000** |
| **5. Summarize the results of your test** |  | **Since our P-value is 0.000 and our significant level of 0.05. We can determine our null hypothesis was disproven. The states with a higher population of the vulnerable group will have a higher death rate from influenza.** |
| **6. List some steps for how you might proceed with your analysis based on your results.** |  | **Since we have confirmed the vulnerable group does have more deaths than the non-vulnerable group, we can now determine whats states have the larges vulnerable group to adequately send staff to.** |



