**Preparing for Influenza Season: Interim Report**

* Project overview
* Hypothesis
* Data overview
* Data limitations
* Descriptive analysis
* Results and insight
* Remaining analysis and next steps
* Appendix

**PROJECT OVERVIEW**

* **Goal:**

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 influenza season, a time when additional staff are in high demand. The final results will examine trends in influenza and how they can be used to proactively plan for staffing needs across the country.

* **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 all vulnerable population peoples injected with flu shots before flu season, then patients admitting for influenza will be decreased.
* If high priority vulnerable populations (which includes 65+ age people) has more staffs, then patient’s death rate will be reduced.
* If minimising instances of understaffing and overstaffing across states, then staff shortage during flu season will be reduced.
* If patient is older than 65+ years (Vulnerable population) then they are more death from influenza.

**DATA OVERVIEW**

|  |  |  |
| --- | --- | --- |
|  | **Data Source** | **Data Content** |
| **US Census Population Data** | This medical staff agency getting data from US Census Bureau and it provided from united state government so it is trustworthy source. This is External source of data. The data is relevant because from this data contains dataset for our wish list, we can tell which state has more states with more vulnerable populations. These are the population that is most vulnerable to the flu, which tells us staffing needs during the influenza season. | The data is about the total population on a yearly basis from 2009 to 2017. This has information’s about county, year, population, age and gender. It can be say as “Population data by geography”. |
| **Influenza Laboratory Test Data** | This data is from CDC which collects, compiles, and analyses information on influenza activity weekly in the United States. This is external source of data. This source includes both Public health and clinical Laboratories data. This Data is relevant as this provide data for wish list of relationship between age and influenza. Also this shows timing of flu activity seasons. | This data has information of both public and clinical Laboratory results of influenza illness, hospitalization, deaths, Vaccination status, age group and patient’s visits from 2010 to early 2019 from all states. |

**DATA LIMITATIONS**

|  |  |
| --- | --- |
|  | **Data Limitations** |
| **US Census Population Data** | As Data collected in **Time lag** of 1 year and have prone to human errors and this is official data provided by government so there is no bias in data. |
| **Influenza Laboratory Test Data** | This data’s collected from both Public and Clinical Laboratories, these data’s entered manually so leads to **Inaccuracy, duplicate** and bias in data. These data’s collected over weekly so there time lag occurs. |

**DESCRIPTIVE ANALYSIS**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Variable** | **Total Death** | **Total Death of Vulnerable Population** | **Total Population** | **Vulnerable Population** |
| **Dataset Name** | Integrated Data Set | Integrated Data Set | Integrated Data Set | Integrated Data Set |
| **Sample or Population** | Sample | Sample | Sample | Sample |
| **Normal Distribution** | Left Skew | Left Skew | Left Skew | Left Skew |
| **Variance** | 1322659 | 1021568 | 46025346980846 | 9279871018291 |
| **Standard Deviation** | 1150 | 1011 | 3046288 | 6784198 |
| **Mean** | 888 | 810 | 3974369 | 6059734 |
| **Outlier Percentage calculated by calculated Empirical Rule** | 9% | 8% | 5% | 7% |

|  |  |  |
| --- | --- | --- |
| **Correlation** | | |
| **Variable** | Total Vulnerable Population & Total Death of Vulnerable Population | Total Population & Total Death |
| **Proposed Relationship** | To test the relationship between total vulnerable population and total death of the vulnerable population due to Influenza. | To test the relationship between total population and total death of population due to Influenza. |
| **Correlation Coefficient** | 0.999825511 | 0.998375226 |
| **Strength of Correlation** | Strong Correlation | Strong Correlation |
| **Interpretation** | This has strong correlation, as vulnerable population increases death in vulnerable area also increase and also in reverse. | This has strong correlation, as Total population increases then death also increase and also in reverse. |
|

**RESULTS AND INSIGHT**

|  |  |  |
| --- | --- | --- |
|  | *0 to 65 years* | *65+ years* |
| Mean | 1.09E-05 | 0.000251 |
| Variance | 4.755E-10 | 9.68E-08 |
| Observations | 469 | 469 |
| Hypothesized Mean Difference | 0 |  |
| df | 473 |  |
| t Stat | -16.67478 |  |
| P(T<=t) one-tail | 9.699E-50 |  |
| t Critical one-tail | 1.6480815 |  |
| P(T<=t) two-tail | 1.94E-49 |  |
| t Critical two-tail | 1.964992 |  |
| This two sample of 0 to 65 years and 65+ years have huge observations of 469 total numbers. P value is 9.69871802608052E-50(equals to 0), this is less than alpha value (0.05).So we can neglect null hypothesis here and we can add Alternate Hypothesis of **High vulnerable population affect more than Low priority vulnerable populations**.  **REMAINING ANALYSIS AND NEXT STEP** | | | | |
| Furthermore, steps to understand the underlying causes of this influenza death rate difference can identify risk factors in individuals. Knowing the risk factors can allow us to predict the vulnerable patients, flu shots and staff hospitals accordingly. | | | | |

**APPENDIX**

**Reference:**

https://www.cdc.gov/flu/vaccines-work/vaccineeffect.htm

<https://www.cdc.gov/flu/weekly/overview.htm>

https://www.cdc.gov/vaccines/imz-managers/nis/about.html