# PLAN FOR INFLUENZA SEASON


Vanessa Núñez Peñas

# 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 severe 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.

# HYPOTHESIS

If flu patients are over 65 years old, influenza complications and mortality increase notably.

If people at high risk are vaccinated, influenza complications are probably less, and the medical staffing needs are not overcrowded.

If patients are contagious by influenza in determined states with specific demographic conditions, the spread of the virus is higher than in others.

# DATA OVERVIEW

## INFLUENZA DEATHS

This data contains monthly death counts for influenza-related deaths in the United States from 2009 to 2017. Counts are broken into two categories: state and age.
**Limitations:** It is manual data entry by doctors who can only list one cause of death. Vulnerable populations who died infected with influenza may be counted as a victim of the virus or death related to their previous condition, being a personal decision of the medical staff to write down one code or another.

## GEOGRAPHY US CENSUS DATA

It contains the census of the United States population from 2009 to 2017. Population is classified by gender and age.
**Limitations:** The numbers are estimates, and the data come primarily from surveys, although they can be partially inaccurate. Furthermore, we can only know how many people live if they are formally registered.

## PATIENT VISITS

Influenza Visits tracks patient visits to a medical provider for influenza by week and state from late 2010 to early 2019.
**Limitations:**
Survey data. It is not a complete count of all influenza visits in the United States; instead, it comes from 3,500 outpatient healthcare providers.

## INFLUENZA LABORATORY TEST

This data contains monthly death counts for influenza-related deaths in the United States from 2009 to 2017. Counts are broken into two categories: state and age.
**Limitations:** Survey data recollected from 100 public health providers and over 300 clinical laboratories through out the United States and its territories.

## SURVEY OF FLU SHOT RATES IN CHILDREN

This data file includes information about children between 1.5 and 3 years and their families recollected in 2017. It shows diverse household information, census regions, and medical data.
**Limitations:** Group of phone surveys collecting data from parents or guardians in all 50 states, the District of Columbia, and some U.S. territories. Cell phone numbers are randomly selected and called to enroll one or more age-eligible children or teens from the household. The demographics are self-reported and may be biased, but the flu shot information is verified with health providers and can be considered accurate.

# DESCRIPTIVE ANALYSIS

| | Variables | Mean | Standard Deviation | Outliers % |
|---|---|---|---|---|
| Hypothesis 1 | Mortality in ages group under 5 years | 0 | 0 | 0% |
| | Mortality in ages group 5-14 years | 0 | 0 | 0% |
| | Mortality in ages group 15-24 years | 0 | 1 | 0% |
| | Mortality in ages group 25-34 years | 1 | 5 | 3% |
| | Mortality in ages group 35-44 years | 3 | 12 | 5% |
| | Mortality in ages group 45-54 years | 17 | 42 | 9% |
| | Mortality in ages group 55-64 years | 59 | 102 | 15% |
| | Mortality in ages group 65-74 years | 116 | 168 | 13% |
| | Mortality in ages group 75-84 years | 246 | 306 | 13% |
| | Mortality in ages group over 85 years | 463 | 551 | 12% |
| Hypothesis 2 | Percentage of positive influenza tests | 14.32% | 0 | 13% |

These results confirm that the mortality rate increases according to age, especially from 65 years, and there is a relevant difference among age. Regarding the second hypothesis, relative to the percentage of positive influenza tests, we can see that there are 14.32% positives on average, with a minimum distance of the data from the mean. However, there are 13% of outliers.

Another interesting descriptive analysis done at this first stage has been the *correlation coefficient*, a statistical measure of the strength of a linear relationship between two variables.The goal was to see the relationship between the number of deaths and the population over 65 years old.The result is a strong correlation (0.9), meaning that States with higher populations over 65 years have higher influenza mortality.

# RESULTS AND INSIGHTS

Considering we have only a sample of data, we need to determine the statistical significance of this subset of the US population to draw conclusions about the entire population.

**Statistical Hypothesis 1:**
- *Null Hypothesis:* people over 65 years do not have a higher mortality rate than people under 65 years.
- *Alternative Hypothesis:* people under 65 years have a lower mortality rate than people over 65 years.
- *t-Test:* the p-value is 0,00000000, although less than the significance level (0,05).
This means we can reject the null hypothesis with a 95 percent confidence level. In other words, there is a significant difference between people under and people over 65 years in terms of mortality rate at a confidence level of 95 percent.

**Statistical Hypothesis 2:**
- *Null Hypothesis:* most populated states have no higher virus spread than others.
- *Alternative Hypothesis:* less populated states have a lower virus spread than others.
- *t-Test:* the p-value is 0,00000000, although less than the significance level (0,05).

This means we can reject the null hypothesis with a 95 percent confidence level. In other words, most populated states have a higher virus spread than the less populated at a confidence level of 95 percent.

## Remaining Analysis and Next Steps

There are remaining critical analyses to do.

- Identifying the states needing more medical staff and resources for the flu season. This means states with a high percentage of people over 65 years, the most vulnerable population, and the most populated states, where the spread of the virus will be higher.
- To better identify the most vulnerable states, we need more medical data about other patients who should be considered as vulnerable populations besides age criteria.
- For a more efficient plan during the flu season, it is necessary to identify the peak weeks of the virus. This would facilitate the optimization of the medical staff geographically and timely.
- To verify the hypothesis regarding the influence of vaccination rate, we need additional data beyond the children between 1.5 and 3 years.

# APPENDIX

## Business Requirements Document

<u>Success Factors</u>

The project's success will be based on:

- 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%)

<u>Assumptions:</u>

- 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.

<u>Constraints:</u>

- The staffing agency has a limited number of nurses, physician assistants, and doctors on staff.
- There's no money to hire additional medical personnel.

<u>Requirements</u>

- Provide information to support a staffing plan, detailing what data can help inform medical personnel's timing and spatial distribution throughout the United States.
- Determine whether influenza occurs seasonally or throughout the entire year. If seasonal, does it start and end simultaneously (month) in every state?
- Prioritize states with large vulnerable populations. Consider categorizing each state as low-, medium-, or high-need based on its vulnerable population count.
- Assess data limitations that may prevent you from conducting your desired analyses.

<u>Glossary</u>

*Influenza*: a contagious viral infection, often causing fever and aches.

*Vulnerable populations*: 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 years, children under 5 years, and pregnant women, as well as individuals with HIV/AIDs, cancer, heart disease, stroke, diabetes, asthma, and children with neurological disorders.

## Hypotheses development

<u>Clarifying questions:</u>

- What kind of virus is the influenza?
- Why is this virus so contagious?
- Who is most affected by this virus?
- When is the virus most potent?
- Where have there been historically more problems with resources and staff during influenza season?
- Which treatments do usually need the patients?

<u>Adjoining questions:</u>

- How does the flu relate to other hospital admissions or deaths?
- Are there ways to decrease flu admissions or deaths (such as flu shots)?

<u>Funneling questions:</u>

- *When is the virus most potent?*
    - When is the influenza season?
    - Does it depend on the weather?
    - Are there regional differences?
- *Where have historically been more problems with resources and staff during influenza season?*
    - Which states and hospitals had more problems?
    - What kind of problems did they face? Staff, treatment, hospital capacity, ICU occupancy rate, health equipment…
- *Which treatments do usually need the patients?*
    - What is the rate of hospital ingressions?
    - What capacity are there in the hospitals?
    - Could some patients receive the treatment at home?
    - How many people are already vaccinated?
    - How many vaccines must be shot next season?
    - When is the best moment to display a vaccination campaign?

<u>Elevating questions:</u>

- Can the government (or workplaces, schools, health departments, etc.) do something to help prepare for the flu nationally (hand washing public health campaigns, flu shots, etc.)?
- How does the flu impact the community, for instance, school and work attendance?

<u>Privacy and Ethical questions:</u>

- What does the data privacy law say regarding clinical histories?
- May patients be affected somehow after categorizing them as flu-vulnerable?
- Could a patient refuse treatments or vaccines?

## Hypothesis Testing result

t-Test: Two-Sample Assuming Unequal Variances

| | *Population Under 65* | *Population Over 65* |
|---|---|---|
| Mean | 0,01% | 0,48% |
| Variance | 0,00000000 | 0,00000514 |
| Observations | 459 | 459 |
| Hypothesized Mean Difference | 0 | |
| df | 459 | |
| t Stat | -44,700944 | |
| P(T<=t) one-tail | 0,00000000 | |
| t Critical one-tail | 1,64818014 | |
| P(T<=t) two-tail | 0,00000000 | |
| t Critical two-tail | 1,96514575 | |

| | *Total Population* | *Percentage Positive* |
|---|---|---|
| Mean | 6379160 | 13,64% |
| Variance | 5,04552E+13 | 0,03246136 |
| Observations | 10283 | 10283 |
| Hypothesized Mean Difference | 0 | |
| df | 10282 | |
| t Stat | 91,06900721 | |
| P(T<=t) one-tail | 0,0000000000 | |
| t Critical one-tail | 1,645001838 | |
| P(T<=t) two-tail | 0,0000000000 | |
| t Critical two-tail | 1,960194732 | |