

# STAT 3494W Proposal

Jessica Zambuto

Department of Statistics, University of Connecticut

October 7, 2022

## 1 Introduction

## 2 Specific Aims

## 3 Data Description

The data I will be using for this research project comes from the Center for Disease Control and Prevention Website called, "Influenza Vaccination Coverage for All Ages (6+ Months)". The dataset was collected using the National Immunization Survey-Flu and Behavioral Risk Factor Surveillance System at the national, regional, and state levels by age and race. The dataset is updated monthly and includes the name of the influenza vaccine distributed, the state, the year, the age group, the percent estimate of those vaccinated, the 95 percent confidence interval for the estimate, and the sample size. I am interested in analyzing the data at a race, age, and year level. This dataset contains 169,293 observations. I will also be using the CDC's, "Vaccine Coverage Among Health Care Personnel" dataset, which contains the same variables as the first dataset mentioned except for age and is only based on healthcare workers. The data was collected using the influenza vaccination performance measurement from the National Healthcare Safety Network. I am interested in analyzing this dataset on a yearly level. This dataset contains 1456 observations.

## 4 Research Design and Methods

To begin my research, I will first compile a literature review to have a better understanding of existing results. I plan on only analyzing the, "Influenza Vaccination Coverage for All Ages" dataset on a year, race, and age basis and the, "Vaccine Coverage Among Health Care Personnel" dataset on a year basis. However, these variables of interest are subject to change if I find confounding variables or others that are associated with receiving the influenza vaccination based on my literature review. I will analyze the two datasets by using the two sample z test for proportions. I will first compare the proportions of the general population who received the influenza vaccine prior to the COVID-19 pandemic defined as 2018-2019 and during the pandemic defined as 2020-2021. I will then compare the proportions of the

general population who recieved the influenza vaccine during the pandemic (2020-2021) to those who did in 2022. I will repeat these steps for the, "Vaccine Coverage Among Health Care Personnel" dataset. The steps above will be repeated for the "Influenza Vaccination Coverage for All Ages" dataset, but this time not only comparing before, during, and after the pandemic, but different races, so the z test will be conducted two times for each race (white, black, asian, hispanic, and other). The Z test will again be repeated for before, during, and after the pandemic, but for different age groups defined by 6 months-17 years, 18-49 years, 50-64 years, and greater than 65 years of age.

## **5 Discussion**

## **6 Conclusion**

## **References**