

# STATS 501 Project Proposal

Brody Erlandson and Danielle Fanizzi

December 11, 2021

## 1 Introduction

Risky behavior is a problem among the high school age group; in particular, violence can be a problem in some areas and not others. Furthermore, the violence in an area can change over time. We want to look further into what factors predict the trend of violence over time. With the Youth Risk Behavior Surveillance Survey (YRBSS), we can look into how violent the youth is in sepcific counties; moreover, with the data about the Supplemental Nutrition Assistance Program (SNAP) and Small Area Income and Poverty Estimates (SAIPE), we can look at how the socioeconomic factors of the district of the school can possibly predict the change of violence over time.

### 1.1 YRBSS Data Description

The CDC data set YRBSS is responses of a survey given to high school students from 1991 to 2019 every 2 years. It ask questions pertaining to violent behavior, drug use, sexual activity, depression and suicide, and general risky behavior. The survey was taken in multiple districts throughout the country.

### 1.2 SNAP Data Description

The SNAP data set is collected every year by the Food and Nutrition Service from the United States Department of Agriculture. It provides the number of

people who participate in SNAP for each state and the price each state spends for each participant during a year.

### 1.3 SAIPE Data Description

The United State Census collects data every year based around poverty and unemployment rates and median household income for each county, which makes up the SAIPE data set.

## 2 Making a continuous Response

We will start by grouping questions based on their topic, like violence or depression and suicide. These groupings will be used as the predictors for our models, starting with violence. Because not all participants answered each question, we impute the values four different ways: (1) using KNN, (2) using the median, (3) using the mean, and (4) using the most frequent. We look to compare the performance of each imputing method by comparing AIC of the models. We created a singular predictor value for each year, given the survey takers change each year. We assign values to answer choices for each question, and determine a weight for each question. For example, the weight of each question is determined by assessing which question signify more violent tendencies and assigning these with higher weighted values. Then we add up all the responses, multiplied by the weights, to get a singular number. Given each county has a different number of survey takers and different violence questions answered, we also take the saturated violence score which is calculated by finding the violence score if every survey taker responded with the most extreme value. We then take the violence score divided by saturated violence score This means the final score is between zero and one, which we multiplied by 1000 to put on a scale of zero to 1000.

## 2.1 Methods of Analysis

We will fit the violence scores with a random intercept and group by county in a mixed model. We will use AIC to compare all combinations of the fixed predictors, so that we can see what socioeconomic factors predict the change of violence over time. We also want to compare fitting methods of our best predictor mixed models: (1) using 'REML' and (2) using 'ML'. We hope to use AIC to compare this also.

Next, we will split the data on sex because we believe sex influences a child's violent tendencies. We will fit a model on the female data and fit a model on the male data using the same response variable and predictors. Finally, we will create synthetic data to fit to our two sex models, and see if the two models create significantly differing results.

## 3 Time Permitting Method

Given time, we would like to perform a similar analysis for other risk factors, such as risk of depression and suicide or use of drugs/alcohol. Since we would be using the same predictors and models, it would be very similar to the above analysis. Additionally, we may also look to split on age/grade similar to how we split on sex.