STATS 501 Project Proposal

Brody Erlandson and Danielle Fanizzi

October 20, 2021

1 Introduction

Suicide has become an increasing problem over the years, especially in the high school age group. We want to look further into what factors may be predictors for someone at risk of attempting suicide. With the Youth Risk Behavior Surveillance Survey (YRBSS), we can take a look at what types of behaviors may indicate this behavior; moreover with the Food Environment Atlas (FEA), we can look at the socioeconomic factors of the district the school is in. Further, we can also look at how these risk factors change overtime with the socioeconomic change in a district.

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 FEA Data Description

The FEA data set is collected every 2 years for every state and county by the Economic Research Service from the United States Department of Agriculture.

It provides food environment factors based around each population's socioeconomic including food insecurity for varying demographics, school nutrition assistance programs, average family income, and poverty rate.

2 Methods of Analysis

In our initial analysis we would like compare two models: a logistic model and a hierarchical logistic model. Both would have the response of attempted suicide (1) vs considered but not attempted suicide (0)[1]. In the hierarchical model, we may split on factors like: school district[2], sex, and sexual orientation. We would like to see if splitting on these factors make a difference in our result, for example: we believe that males and females will have different predictors for suicide.

3 Time Permitting Method

Given time, we would also like to look into the multiple response longitudinal model where the risk factors are our response and the districts socioeconomic factors as our predictors. We hope to see if any risk factors change overtime with the districts socioeconomic factors. If we have even more time, we would like to try to use a Bayesian approach on our best performing model; however, we don't expect to get to this.

[1] if this turns out to be too nuanced, then we will switch it to suicidal (1) vs not suicidal (0)

[2] or the districts socioeconomic factors