Modern Applied Data Analysis Project

Aidan Troha - College of Public Health, University of Georgia

5/7/23

# 1. Abstract

# 2. Introduction

## 2.1 Background

## 2.2 Objectives

# 3. Methods

## 3.1 Data Acquisition

## 3.2 Analysis Steps

# 4. Results

## 4.1 Exploratory Analysis

|  |
| --- |
| Figure 1. Number of deaths from COVID-19 by county in Georgia by the projected population for 2020. Trend line represents the estimates association between projected population and number of deaths attributed to COVID-19; grey area represents the 95% confidence interval for the trend line. |

|  |
| --- |
| Figure 2. Distribution of deaths attibuted to COVID-19 by county in Georgia separated by party majority in 2020 (>50%) for each county. |

|  |
| --- |
| Figure 3. Number of deaths from COVID-19 by county in Georgia by hospital bed capacity in 2020. Trend line represents the estimates association between hospital bed capacity and number of deaths attributed to COVID-19; grey area represents the 95% confidence interval for the trend line. |

|  |
| --- |
| Figure 4. Number of deaths from COVID-19 by county in Georgia by log per capita income in 2020. Trend line represents the estimated association between log per capita income and number of deaths attributed to COVID-19; grey area represents the 95% confidence interval for the trend line. |

|  |
| --- |
| Figure 5. (Left) Number of deaths from COVID-19 by county in Gerogia by high school graduation rate in 2020. Trend line represents the estimated association between graduation rate and number of deaths attributed to COVID-19; grey area represents the 95% confidence interval for the trend line. (Right) Distribution of deaths attributed to COVID-19 in the State of Georgia organized by county graduation status (‘Improved’ or ‘Not Improved’). |

## 4.2 Regression Modeling

## 4.3 Model Performance

|  |
| --- |
| Figure 6. (Left) Main effects model predictions compared to observed number of deaths by county. (Right) Saturated model predictions compared to observed values by county. Dashed line represents a 1-to-1 interaction between observed and expected values. |

|  |
| --- |
| Figure 7. (Left) Main effects model residuals compared to the observed number of deaths by county. (Right) Saturated model residuals compared to the observed number of deaths by county. Red line indicates a residual of 0. Blue dashed lines represent 2 RMSE values away from 0. |

### 4.3.1 Predictors

|  |
| --- |
| Figure 8. (Top Left) Association between hospital bed capacity and the number of COVID-19 related deaths by county in Georgia. (Top Right) Association between log per capita income and the number of COVID-19 related deaths by county in Georgia. (Bottom) Association between high school graduation rate and the number of COVID-19 related deaths by county in Georgia. Trend lines represent predictions from the model with grey areas representing 95% confidence interval values. |

# 5. Discussion

## 5.1 Strengths & Limitations

# 6. Conclusion

# 7. References