Infant mortality in the United States

Prayag Gordy

4/13/2021

Introduction

Infant mortality is a widespread issue worldwide. Countless sums of money are poured annually into research and programs to prevent the death of children younger than one year. High rates of infant mortality, which the United States Centers for Disease Control (CDC) defines as the number of infant deaths per 1,000 live births, are common in South Asia and Africa.

Targeted interventions and general improvements in access to and quality of healthcare has decreased the infant mortality rate in the United States, but disparities persist. People of color and of lower income are at higher risk of losing a child before their first birthday. In this project, I will examine infant mortality in 255 counties in the United States.

Data

Data	Source	Link	Collection method
Infant mortality data	CDC WONDER	https://wonder.cdc.gov/controller/datarequest/D159	Download
Demographic data	American Community Survey	https://www.census.gov/programs-surveys/acs	`tidycensus` package

Infant mortality

Courtesy of the CDC, these infant mortality data from 2017-18 provide the number of infant deaths and live births at the county level ($n = 255$). There are more than 3,000 counties in the United States, but these data are limited to 255 because the CDC suppresses data from counties with fewer than nine infant deaths.

Demographics

I will use race, education, income, housing, healthcare coverage, and/or other indicators collected through the 2018 five-year American Community Survey as predictors of infant

mortality rate. These data were collected from 2013-2018, which thereby includes the 2017-18 collection period of the CDC's infant mortality data.

Methods

The primary form of analysis will be a multiple linear regression with infant mortality rate as an outcome and the aforementioned demographic data as predictors. As long as I do not include too many predictors, the sample size of $n = 255$ should be enough to withstand any deviations from the MLR assumptions. This regression will lead to any conclusions about strong predictors of county-level infant mortality, which can thereafter inform policy and funding decisions. Of course, categorical variables like race will be coded as dummy variables.

I may also conduct an F-test to examine the importance of additional demographic variables to this regression versus a simple linear regression with only healthcare coverage as a predictor. This could shed light, for instance, on the impact of race on similarly-uninsured groups.