Introduction

Zayed Shahjahan

12/8/2020

This report attempts to understand the links between State-level obesity rates and health indicators for individuals aggregated at the State level. The original dataset can be obtained from the following link: https://www.kaggle.com/spittman1248/cdc-data-nutrition-physical-activity-obesity

The main objective is to use multivariate linear regression analysis to investigate the effects of healthy practices, income inequality and geography on obesity rates in the US. The dataset is from the CDC Study on Nutrition, Physical Activity and Obesity.

Obesity is a major health concern in the developed world. This issue is even more pronounced in the United States. Obesity is connected to a host of diseases and the CDC reports that the obesity costs the US economy approximately $149Bn more in added expenditures. The focus will be in examining conventional wisdom which dictates that healthy eating habits and regular exercise is key to reducing a person's chance of becoming obese. However, this belief may not hold at the state-level owing to the fallacy of composition which dictates that what is true for an individual may not necessarily be true for the collective.

The questions of interest are as follows:

"Percent of adults aged 18 years and older who have obesity". This is our main question of interest. In the original dataset, it is stratified on multiple bases. Our stratification category of interest is the income stratification. This consists of five categories: "$15,000 - $24,999", "$25,000 - $34,999", "$35,000 - $49,999", "$50,000 - $74,999", "$75,000 or greater".

"Percent of adults who engage in muscle-strengthening activities on 2 or more days a week". This one of our health indicators. It is stratified the same way as the first question. This will be one of the predictor variables pertaining to healthy practices at the individual level aggregated at the State level. This particular question focuses on strength training.

"Percent of adults who achieve at least 150 minutes a week of moderate-intensity aerobic physical activity or 75 minutes a week of vigorous-intensity aerobic activity (or an equivalent combination)" Same as above but with a cardiovascular health focus.

"Percent of adults who achieve at least 150 minutes a week of moderate-intensity aerobic physical activity or 75 minutes a week of vigorous-intensity aerobic physical activity and engage in muscle-strengthening activities on 2 or more days a week" Composite question containing practices of both of the above questions. Used as an indicator of overall physical activity.

"Percent of adults who report consuming fruit less than one time daily". Question pertaining to fruit consumption. One of two questions that focus on nutrition.

"Percent of adults who report consuming vegetables less than one time daily". Question pertaining to vegetable consumption. Second question that focuses on nutrition.

All of these percentages do not take into account population variation across States and across income brackets. In our analysis survey sample sizes will be used as a proxy for population variation. These will then be converted to relative weights and the percentages for each question will be adjusted by the weights.