Using Automated Web Scraping to Document Variation in Sodium Content of Common School Meal Entrees

Aaron Brown, Kimball Jardine, Joe Price, Nathan Schill, Courtney Webb – BYU
Joanne Guthrie - USDA Economic Research Service

Background

Many school districts consider meeting the National School Lunch Program (NSLP) sodium standards challenging. USDA national survey data indicate that in School Year 2014-15, 28% of lunch menus did not meet the standard. Entrees contributed the largest share of sodium, making their sodium content especially important.

Methods

Raw data is web-scraped from School Nutrition and Fitness, MealViewer, Nutrislice, and SchoolCafe websites. The data were merged with the Per Pupil Current Spending Total (PPCSTOT) from the 2020 US Census Public Elementary-Secondary Education Finance Data and the 2013 Rural-Urban Continuum Codes (RUCC) from the USDA into a file for further statistical analysis and modeling. RUCC codes range from 1 to 9 indicating increasing rurality. The arranged dataset is then filtered for specific lunch menu items, and analysis is narrowed down to a thorough review of two popular entrees prepared at public schools: Pizza and Nuggets. The school districts we studied were organized into groups—state and regional levels were the focus (regions are determined by the U.S. Census Bureau). The variability between these groups was tested. Regression models and other methods were developed to determine predictive relationships between PPCSTOT and RUCC data with Sodium content.

Results

Results for the Pizza Entrée include statistically significant variability between states groups and regional groups, with the South Region seeing the highest mean Sodium level at 816.454 mg—the second highest is the Northeast at 789.026 mg. The Nuggets Entrée also includes significant variability for these groups. Regression models show a negative relationship between PPCSTOT level and Sodium content. In general, school districts in counties of higher RUCC classifications experience greater levels of Sodium content for both entrees—for Pizza, RUCC classifications 7 - 9 see mean levels between 860.214 - 893.429 mg, and 1-3 see between 767.256 - 802.357 mg. These findings provide more information on the health quality of school menu items by school district, state, and region. PPCSTOT and RUCC models can also be used to predict sodium content by group.

Data Collection

The web scraper would attempt to gather monthly school menus from one elementary, middle, and high school (or junior high or K-8) in a given school district. Red dots on the left represent locations of scraped school districts, with the percentage of districts covered on the right. For April 2023, 1000 school districts were scraped.

This research is supported by a Cooperative Agreement with the U.S. Department of Agriculture, Economic Research Service. The findings and conclusions in this paper are those of the authors and should not be construed to represent any official USDA or U.S. Government determination or policy.