**Does an individual’s household income affect the amount of different Junk Foods they eat? And how does the total amount of Junk Food eaten affect an individual’s BMI?**

**Data-Management:**

To investigate this question, we first extract the relevant data from the large, raw, .csv files provided into R. To avoid unnecessary processing, we first choose the columns required from each dataset, and preserve the “SEQN” column for data linking. Then, use the merge() function to link the datasets. The table below shows which columns we require from each dataset. I initially identified all ‘Junk Foods’ in the raw data set, then randomly selected the 10 junk foods listed.

|  |  |  |
| --- | --- | --- |
| BMI | “SEQN” – ID for merging | |
| “BMXBMI” – BMI for second study (Numeric) | |
| DIET  Foods as explanatory variables (EVs): | “SEQN” | |
| Fries-“FFQ0047” | Spareribs-“FFQ0081” |
| Pancakes-“FFQ0059” | Fried Fish-“FFQ0094” |
| Mac and Cheese-“FFQ0061” | Cake-“FFQ0114” |
| Peanut Butter-“FFQ0068” | Cookies-“FFQ0115” |
| Hamburgers-“FFQ0075” | Chocolate Candy-“FFQ0120” |
| DEMO | “SEQN” | |
| “INDHHINC” – Household Income grouped into 13 categories, EV in first study | |

Next, we remove erroneous or useless rows of data. The first ‘cleaning’ removes individuals with BMI outside valid range, such that they are considered erroneous or unrepresentative. We set boundaries at 10 and 80 and remove any individuals in the study with BMI outside of this.

The next task is to remove individuals with incomplete data for the relevant categories. Amount of each food eaten is scored on a scale of 1-13 (see appendices) and errors are marked with a score of “77”, “88”, “99” or “.”. To remove individuals with these values they are swapped with “NA” and we use the “na.omit” function to remove the incomplete rows.

Finally, we add one column to the data labelled “jfs” (Junk-Food-Score). This sums the scores from the food variables to indicate the total junk food an individual eats. Doing this produces a column with range 9 to 94, median 36 and an IQR from 30 to 43.

Now we can produce visualisations.

**Data-Visualisation Approach:**

To address the research question, we aim to produce one plot demonstrating the associations between different foods (and jfs) with Household income and another for the association with jfs as the EV and BMI as the dependent variable. The first will show the relationships as a series of fitted linear models. The second will present the raw data – in boxplot form – with 2 adjusted models fitted on top of it: one uses stratification, and the other doesn’t (we will use weights of 1/(s\_i^2) to combine stratification predictions).

To support the first study, we also produce a shinyapp. It can be accessed by running the ‘us\_food\_shinapp.R’ in the ‘visualisation’ directory – see ‘readme.md’. This allows you to view the raw data for each food type along with the fitted model, and for comparisons between any three of the food models. There are also toggles for Confidence Intervals (CIs) and for linear/categorical regression.

**Results:**

A graph of colored lines

Description automatically generated

Plot 1 consolidates the findings from the shinyapp (CIs removed for clarity). It shows the overall general decrease in junk food eaten by individuals with higher household income. The ‘Junk-Food-Score’ against Income model (shown by the dotted-line) gives a p-value of 0.021: reasonable evidence for this trend. Further detail in the shinyapp.

**A graph with a line of black and white squares

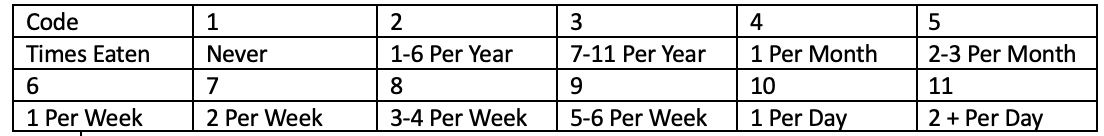
Description automatically generated with medium confidence**Plot 2 investigates BMI’s association with Junk-Food-Score. The results show, perhaps contrary to intuition, an individual with higher JFS had lower BMI. We display fitted values, with CIs, from two models. Model 1 is shown in black and adds Household\_Income as a parameter to manage confounding. Model 2 is red and stratifies over household income. The resultant models are very similar.

The outputs from model 1 (as models are similar we consider the simpler model) give a JFS parameter of -0.477 (-0.547, -0.407) with an associated p-value of <0.001 indicating very strong evidence of a negative association in our data.

To conclude, we have some evidence that an individual with higher household income eat less junk food; and we have strong evidence that an individual who eats more junk food has lower BMI. Further study would be needed to establish association between income and BMI. A possible explanation for an association between higher JFS and lower BMI is high BMI individuals may diet by eating less junk food.

**Appendices: Variable Scales**

* A number on a white background

  Description automatically generatedHousehold Income Categories:
* ****Food Categories: