**Research Question**

Relationship between frequency of eating potato chips and BMI.

**Data Management**

* Outcome: BMXBMI (Body Mass Index)
* Explanatory variable: FFQ0102 (How often did you eat potato chips (including low-fat, fat-free, or low salt)?)

|  |  |
| --- | --- |
| Code | Value Description |
| 1 | never |
| 2 | 1-6 times per year |
| 3 | 7-11 times per year |
| 4 | 1 time per month |
| 5 | 2-3 times per month |
| 6 | 1 time per week |
| 7 | 2 times per week |
| 8 | 3-4 times per week |
| 9 | 5-6 times per week |
| 10 | 1 time per day |
| 11 | 2 or more times per day |
| 88 | Blank |
| 99 | Error |

* Confounder: INDFMPIR (poverty income ratio)
* Data cleaning:
  1. The values I need to use are from DEMO\_D.csv, BMI.csv and FFQRAW\_D.csv. So, I put these three datasets together first.
  2. Then I choose the value I need and create a new dataset.
  3. Since FFQ0102 coded by 88, 99 are blank or error. So, I remove all rows contain 88, 99.
  4. Finally remove all NA rows and output new data file.

**Data Visualization Approach**

Since frequency of eating potato chips are coded as different number. So, I use logistic regression model to visualize data.

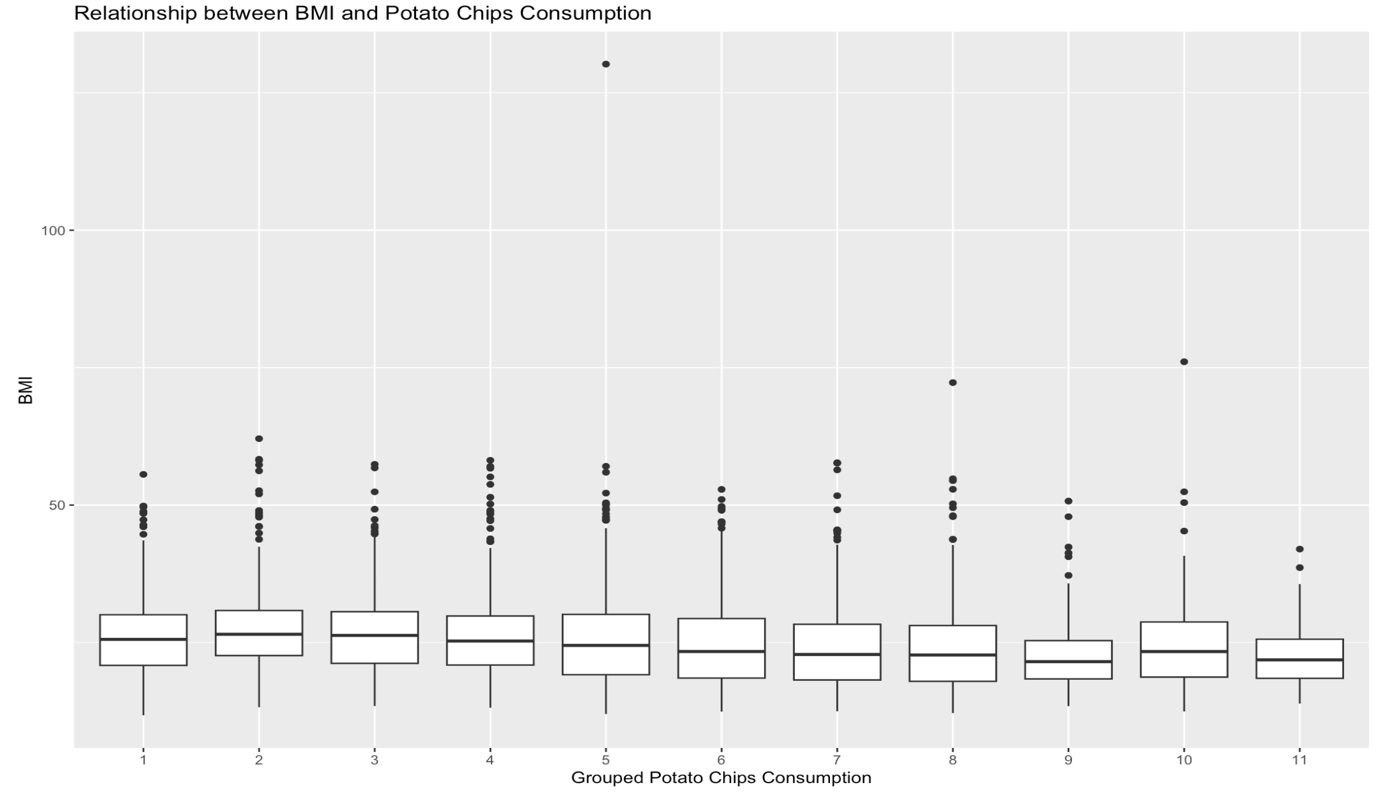
Explanatory variable FFQ0102 is continuous- discrete, so I use a histogram to examine the distribution and see if I need to collapse the data.

Then I use a boxplot to observe the overall distribution of BMI in each group, including whether there is an outlie.

After that, I tried to fit data with linear model, quadratic model, fractional polynomial model, linear spline model and cubic spline model. Put them in the same graph, compare their AIC and find the best fitted.

Finally, I think poverty income ratio might be a confounder for this model, so I add it in and test if there is an interaction between two variables.

**Result**

****

From above graph I notice that the median of each group is similar and in the fifth group there is a person with a BMI as high as 130.21. I found this is a poor person eating potato chips two to three times a month, I thought this was an outlier due to measurement error, so I decided to remove this data.

A graph with colored lines

Description automatically generated

From the predicted relationship plot for BMI and Potato Chips Consumption, I found that the overall change trend of BMI decreased as the frequency of people eating potato chips increased, which was different from the conclusion I originally imagined.

|  |  |  |
| --- | --- | --- |
|  | AIC | BIC |
| Linear | 38519.45 | 38539.35 |
| Quadratic | 38521.44 | 38547.98 |
| Fractional polynomial | 38508.71 | 38535.25 |
| Linear spline | 38509.37 | 38542.54 |
| Cubic spline | 38514.02 | 38553.83 |

Based on AIC and BIC of each model, together with their fitted graph, I think fractional polynomial model is the best fitted one.

Then I consider poverty income ratio as a confounder and fit it with the fractional polynomial of Potato Chips Consumption. I found that its coefficient is 0.182 with a small p value (0.0039), which shows that poverty income ratio has indeed adjusted the model to a certain extent.

My shiny app is a 3D plot about relationship between BMI, potato chip eating frequency group and poverty income ratio. The blue points represent the real data from my sample dataset. Users can select how often they eat potato chips and enter their poverty income ratio to predict their BMI by click the Predict button, all poverty income ratios greater than or equal to 5 are treated as 5. The prediction is represented by a red dot in the image.

Finally, my conclusion is that BMI will decrease as the frequency of eating potato chips increases, and the overall BMI of rich people will be higher than that of poor people. But I think this conclusion is not reliable, because the human diet structure is very complex, and everyone’s eating habits are also different. However, the frequency of eating potato chips can indeed be used as one of the factors that affect BMI in research.

**Reference**

<https://wwwn.cdc.gov/nchs/nhanes/2005-2006/FFQRAW_D.htm#FFQ0102>

<https://wwwn.cdc.gov/nchs/nhanes/2005-2006/DEMO_D.htm>