**Research Question**

Relationship between frequency of eating potato chips and BMI status.

**Data Management**

* Outcome: BMI status (bmi)

I coded BMXBMI >= 25 as obesity (bmi = 1) and BMXBMI < 25 as not obesity (bmi = 0)

* 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 |

* 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 calculate the observed BMI odds of the outcomes across my categories of frequency of eating chips.

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**

**A graph with black dots and lines

Description automatically generated**

From above graph I found the odds ratio for BMI decrease as frequency of eating chips increase. This suggests that people who eat less potato chips may have higher rates of obesity which is very wired and inconsistent with common sense.

|  |  |  |
| --- | --- | --- |
|  | AIC | BIC |
| Linear | 7646.76 | 7660.03 |
| Quadratic | 7648.43 | 7668.31 |
| Fractional polynomial | 7646.76 | 7660.03 |
| Linear spline | 7644.71 | 7671.25 |
| Cubic spline | 7646.16 | 7679.34 |

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

A graph showing a number of chips

Description automatically generated

Then I consider poverty income ratio as a confounder, and I found that the effect of frequency of eating chips has reduced slightly following adjustment for poverty income ratio.

"PIR<1" = 0, "1<=PIR<2" = 1, "2<=PIR<3" = 2, "3<=PIR<4" = 3, "4<=PIR<5" = 4, "PIR>=5" = 5

**Reference**

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

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