Case Study 1A: UPFs

Tyler Arista

2024-09-18

Planning a Model

Identify your response & predictor(s)

- Predictors
  - diet(whether the person was eating “Unprocessed” or “Ultra-processed” food during this period
  - age(in years)
- Response
  - FM_change(change in kilograms of fat mass relative to baseline)

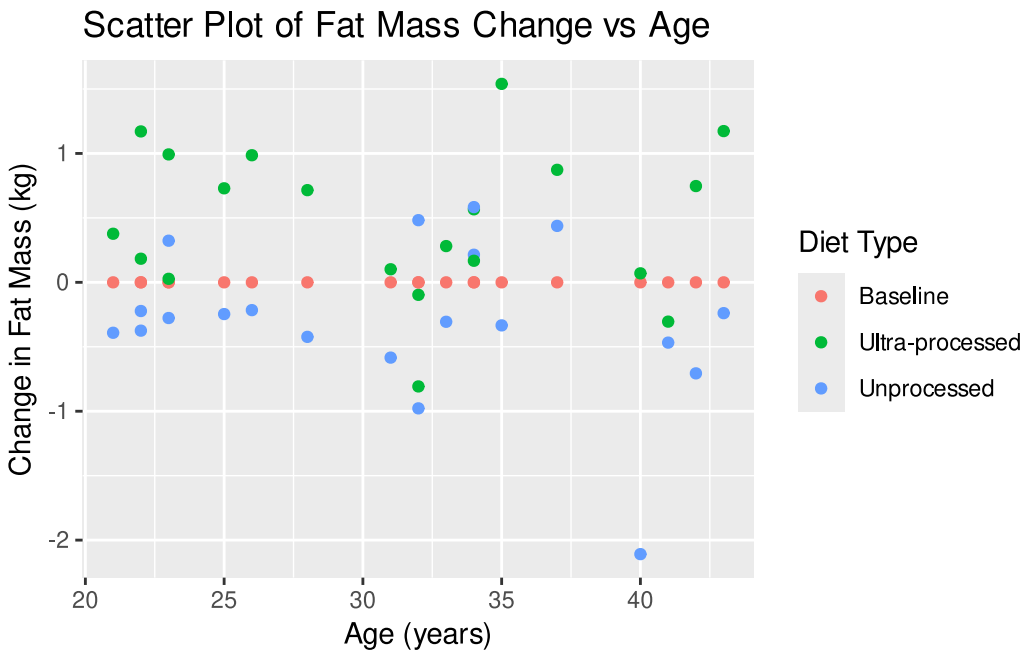
Provide a rationale for your choices based on the size of the dataset & what you know about the data & the research question

- Size of dataset
  - Since the dataset has 20 subjects, we can use two predictors without overfitting, which follows the guidelines that limit the number of predictors to roughly one for every 15 data points.
- Knowledge about data
  - I chose diet type and age as our predictors because I feel like they can have a big impact on the amount of in fat mass either gained/lost. This can help us understand how different diets & age groups affect body fat over time. This fits the research question, which focuses on the effects of ultra-processed and unprocessed diets.
- Research Question
  - Does consuming an ultra-processed diet lead to a greater increase in fat mass compared to an unprocessed diet, and are these effects more pronounced in older individuals?

Graphics & Data Exploration

Create at least one visualization(graph) to explore the relationships between your response & one or more predictors

```r
gf_point(FM_change ~ age, color = ~diet, data = upf_by_diet_data) %>%
  gf_labs(
    title = "Scatter Plot of Fat Mass Change vs Age",
    x = "Age (years)",
    y = "Change in Fat Mass (kg)",
    color = "Diet Type"
  )
```



I decided to use a scatter plot to display the relationship between the change in fat mass and the age of the subjects. I used different colors to separate those on an unprocessed diet from those on an ultra-processed diet, making it easy to compare the two groups. I also added clear labels and a legend to explain which color represents each diet.

One thing I learned from this graph was that people on the ultra-processed diet generally gained more fat mass than those on the unprocessed diet. Additionally, it appears that older individuals on the ultra-processed diet experienced a greater increase in fat mass compared to younger subjects.

Model Fitting

Fit the linear regression model I have planned in R

```r
model <- lm(FM_change ~ diet + age, data = upf_by_diet_data)
summary(model)
```

```
Call:
lm(formula = FM_change ~ diet + age, data = upf_by_diet_data)

Residuals:
     Min       1Q   Median       3Q      Max
-1.75095 -0.16022 -0.00101  0.12008  1.09376

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)          0.234629   0.294838   0.796  0.42952
dietUltra-processed  0.474647   0.151110   3.141  0.00269 **
dietUnprocessed     -0.291608   0.151110  -1.930  0.05871 .
age                 -0.007520   0.008808  -0.854  0.39684
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4779 on 56 degrees of freedom
Multiple R-squared:  0.3247,    Adjusted R-squared:  0.2886
F-statistic: 8.977 on 3 and 56 DF,  p-value: 5.991e-05
```

$$\widehat{FM}_{change} = 0.235 + 0.475 \cdot \text{diet}_{\text{Ultra-Processed}} - 0.292 \cdot \text{diet}_{\text{Unrocessed}} - 0.008 \cdot \text{age} + \epsilon$$

$$where\, \epsilon \sim N(0, 0.478)$$