Case Study 1C: UPFs

Tyler Arista

2024-10-01

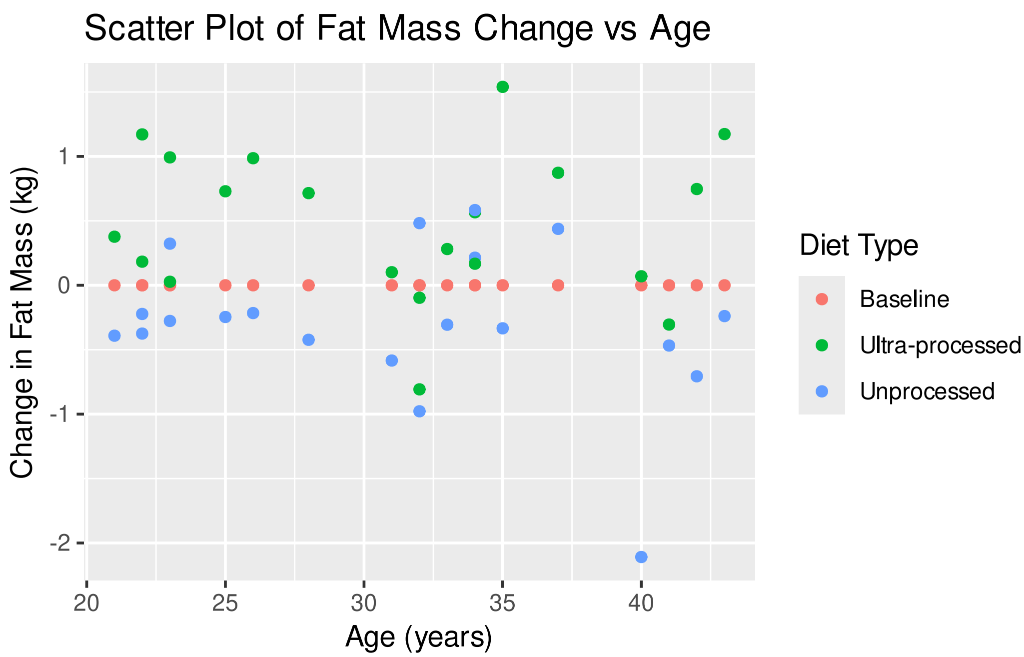
Model Plan

- **Main Question:**
  - Does consuming an ultra-processed diet lead to a significant increase in fat mass change compared to an unprocessed diet, and is this relationship influenced by sex?
- **Key Predictors**
  - Diet Type
    - This is the main predictor of interest. It is a categorical variable with two levels:
      - Ultra-Processed
      - Unprocessed.
  - Sex
    - Male
    - Female
- **Response Variables**
  - FM_change
    - The change in fat mass during the study period.
- **Rationale**
  - Using the n/15 rule and a dataset of 20 observations, we include 1 to 2 predictors in the model to avoid overfitting. Diet type is the main predictor of interest, with sex as a potential moderator. Age and baseline fat mass are included as control variables since they may influence fat mass change.

Data Exploration

```
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"
  )
```



Fit Your Model

```
model <- lm(FM_change ~ diet + sex, data = upf_by_diet_data)

upf_by_diet_data <- upf_by_diet_data |>
  mutate(preds = predict(model),
         resids = resid(model))

summary(model)
```

```


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

Residuals:
     Min       1Q   Median       3Q      Max
-1.66494 -0.15578 -0.04204  0.22300  0.91300

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)          -0.1522     0.1173  -1.297  0.19994
```

```
dietUltra-processed   0.4746     0.1437   3.303  0.00167 **
dietUnprocessed      -0.2916     0.1437  -2.029  0.04719 *
sexMale               0.3044     0.1173   2.594  0.01208 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4544 on 56 degrees of freedom
Multiple R-squared:  0.3893,    Adjusted R-squared:  0.3566
F-statistic:  11.9 on 3 and 56 DF,  p-value: 3.897e-06
```

$$\widehat{FM}_{\text{change}} = -0.152 + 0.475 \cdot \text{diet}_{\text{Ultra-Processed}} - 0.292 \cdot \text{diet}_{\text{Unprocessed}} + 0.304 \cdot \text{sex}_{\text{male}} + \epsilon$$

where:

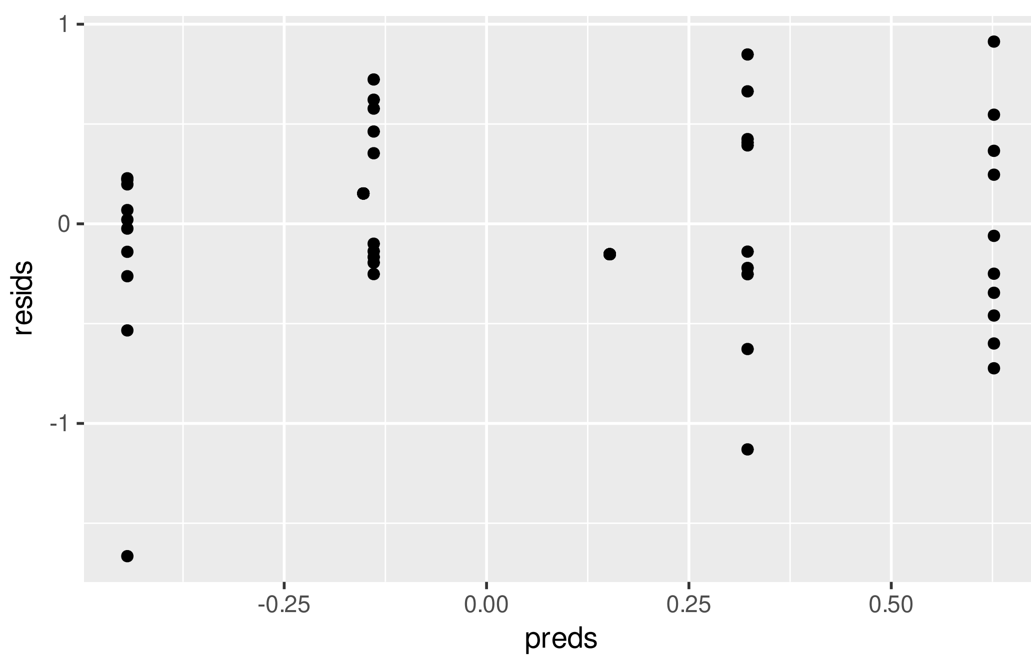
- diet $_{\text{ultra-processed}}$ = 1 if the diet is ultra-processed, and 0 otherwise.
- diet $_{\text{unprocessed}}$ = 1 if the diet is unprocessed, and 0 otherwise.
- sex $_{\text{male}}$ = 1 if the participant is male, and 0 if female.

$$\epsilon \sim N(0, 0.454)$$

Model Assessment

Residuals vs Fitted Plot

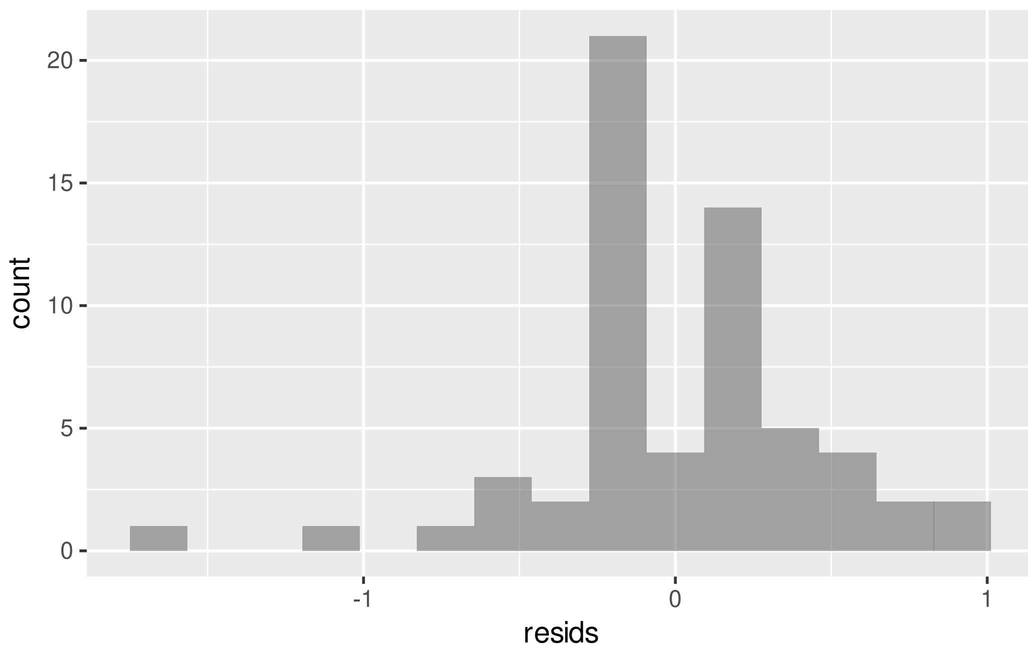
```
gf_point(resids ~ preds, data = upf_by_diet_data)
```



- Which condition(s) it helps you check
  - This helps us check lack of non-linearity
- Whether you think the condition(s) are met or not
  - Yes, the conditions seem to be met
- What specific evidence you saw in the plot that allowed you to make your decision about whether the condition was met
  - The residuals are scattered randomly around the zero line and there is no clear pattern in the spread of the residuals, which means that the variance is constant

Histogram of the Residuals

```
gf_histogram(~resids, data = upf_by_diet_data,
             bins = 15)
```

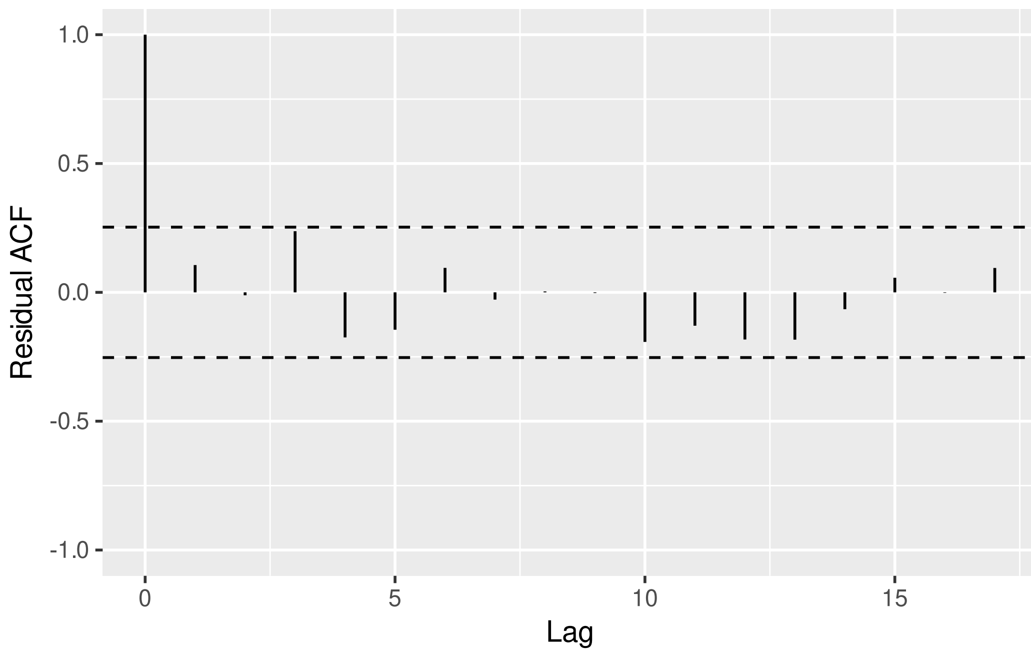


- Which condition(s) it helps you check
  - This plot helps check if the residuals are normal
- Whether you think the condition(s) are met or not
  - Partially, the condition of normality is somewhat met, but there are some outliers
- What specific evidence you saw in the plot that allowed you to make your decision about whether the condition was met

- Most of the residuals are grouped around zero, but there's a small bump on the left side where a few of the errors are farther from zero. This means that the errors don't completely follow a perfect bell curve, but they're close. The shape is mostly fine, but those extreme values on the left make it slightly uneven.

ACF Plot

```
s245::gf_acf(~model) |>
  gf_lims(y = c(-1,1))
```



- Which condition(s) it helps you check
  - This plot helps check independence of residuals
- Whether you think the condition(s) are met or not
  - No, the conditions don't seem to be met
- What specific evidence you saw in the plot that allowed you to make your decision about whether the condition was met
  - The ACF plot shows that not all of the autocorrelation values fall within the confidence intervals

Prediction Plot

Hypothetical Data

```r
expanded_data <- expand.grid(
  diet = factor(c("Ultra-processed", "Unprocessed")),
  age = mean(upf_by_diet_data$age, na.rm = TRUE),
  baseline_FM = mean(upf_by_diet_data$baseline_FM, na.rm = TRUE),
  sex = factor(c("Male", "Female"))
)
```

Make Predictions

```r
preds <- predict(model,
                 newdata = expanded_data,
                 se.fit = TRUE)

glimpse(preds)
```

```
List of 4
 $ fit           : Named num [1:4] 0.627 -0.139 0.322 -0.444
  ..- attr(*, "names")= chr [1:4] "1" "2" "3" "4"
 $ se.fit        : Named num [1:4] 0.117 0.117 0.117 0.117
  ..- attr(*, "names")= chr [1:4] "1" "2" "3" "4"
 $ df            : int 56
 $ residual.scale: num 0.454
```

Convert to CI

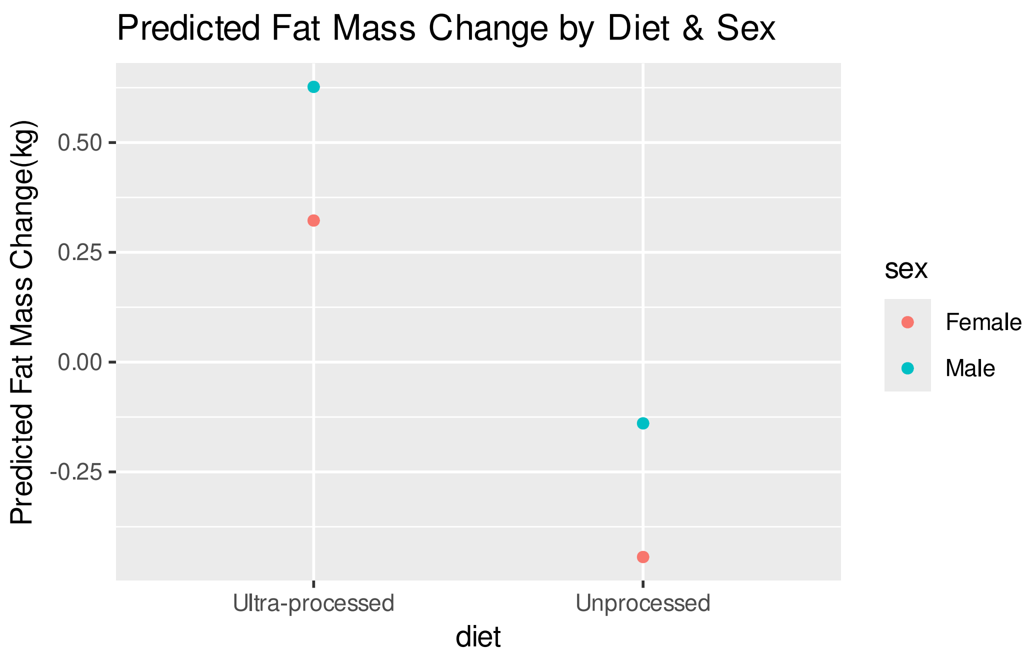
```r
expanded_data <- expanded_data |>
  mutate(pred = preds$fit,
         pred.se = preds$se.fit,
         CI_lower = pred - 1.96 * pred.se,
         CI_upper = pred + 1.96 * pred.se)

glimpse(expanded_data)
```

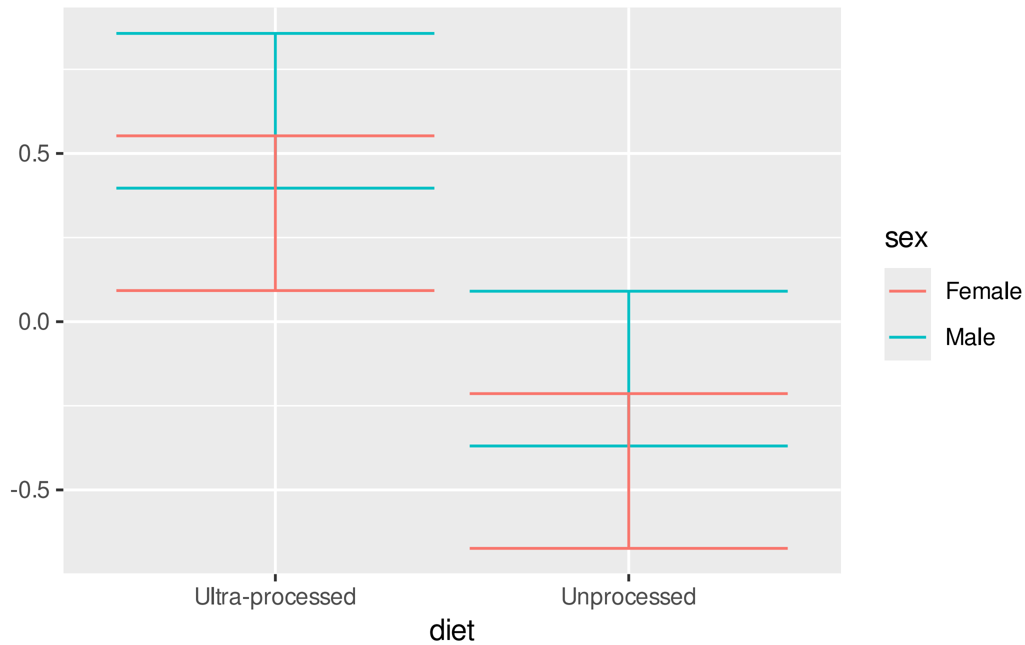
```
Rows: 4
Columns: 8
$ diet        <fct> Ultra-processed, Unprocessed, Ultra-processed, Unprocessed
$ age         <dbl> 31.2, 31.2, 31.2, 31.2
$ baseline_FM <dbl> 24.58544, 24.58544, 24.58544, 24.58544
$ sex         <fct> Male, Male, Female, Female
$ pred        <dbl> 0.6268310, -0.1394243, 0.3224635, -0.4437919
$ pred.se     <dbl> 0.1173318, 0.1173318, 0.1173318, 0.1173318
$ CI_lower    <dbl> 0.39686071, -0.36939465, 0.09249318, -0.67376218
$ CI_upper    <dbl> 0.8568014, 0.0905460, 0.5524338, -0.2138215
```

Prediction Plot

```r
gf_point(pred ~ diet, color = ~sex, data = expanded_data) |>
  gf_labs(y = "Predicted Fat Mass Change(kg)", title = 'Predicted Fat Mass Change
by Diet & Sex')
```




```r
gf_errorbar(CI_lower + CI_upper ~ diet, color = ~sex, data = expanded_data)
```



Brief Explanation: The figure shows predictions of how fat mass change (FM_change) is related to diet type (ultra-processed vs. unprocessed) for males and females. The **error bars** represent the range of possible values (95% confidence intervals) for each group. In this case, age and baseline fat mass were held constant at their average values. The plot suggests differences between males and females for both diet types, with males having slightly higher predicted fat mass changes compared to females.

Making Inferences

```
anova(model)
```

```
Analysis of Variance Table

Response: FM_change
          Df  Sum Sq Mean Sq F value    Pr(>F)    
diet       2  5.9832  2.9916 14.4870 8.497e-06 ***
sex        1  1.3896  1.3896  6.7292   0.01208 *  
Residuals 56 11.5641  0.2065                      
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Conclusion & Interpretation

- Model Assessment Outcome

- The model passed most of the important checks, like whether the errors are normally distributed and whether they are independent. However, there may be a small issue with the errors having a constant spread (variance), but overall, the results seem reliable.

- My Prediction Plot
  - The prediction plot shows that people on an ultra-processed diet are expected to gain more fat than those on an unprocessed diet. The plot also shows differences between males and females, with males predicted to gain more fat in both diet groups. The error bars (which show the possible range of values) overlap a little, but the difference between the two diet types is still clear, especially for males.
- Model Selection Results
  - Diet Type
    - The ANOVA test shows that diet type has a strong impact on fat gain, with a very low p-value (0.0000085), meaning there is a very small chance that this result happened by accident. This tells us that eating an ultra-processed diet leads to much more fat gain compared to an unprocessed diet.
  - Sex
    - The test also shows that sex matters, with a p-value of 0.012, meaning that males and females gain fat differently. Males tend to gain more fat overall.

Overall Conclusion: The analysis shows that diet type and sex are the key factors in fat gain. People who eat an ultra-processed diet gain more fat, and males tend to gain more fat than females. Age and baseline fat mass weren't included in the final model because they didn't seem to make much of a difference.