Report

Danya Zhang

2023-04-15

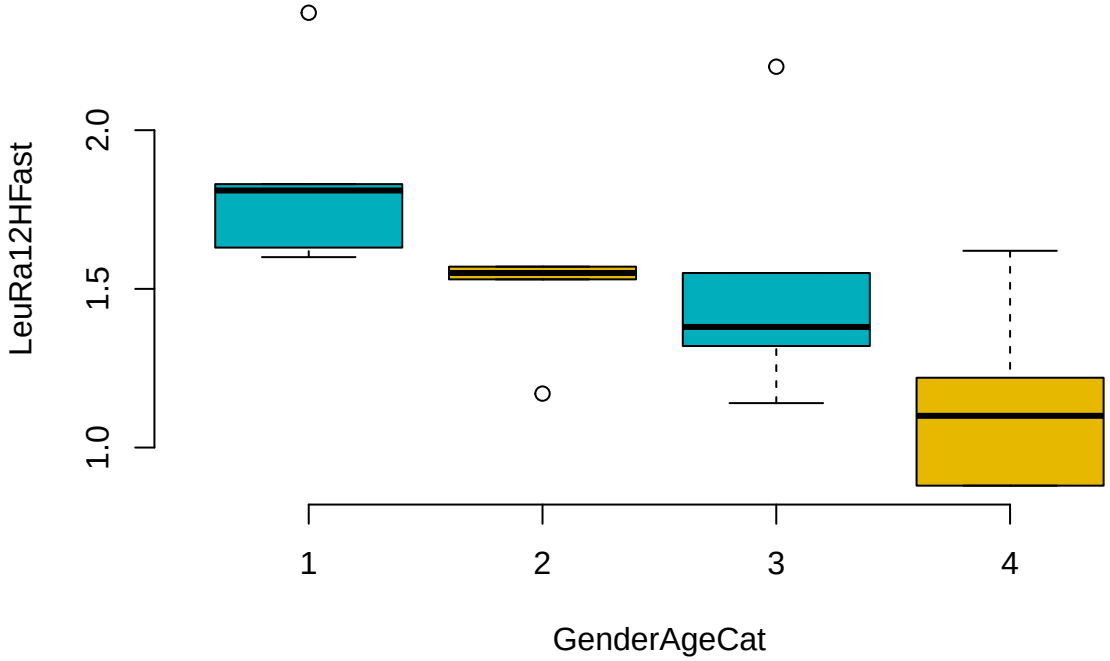
Introduction

The elderly are more prone than the young to periods of involuntary fasting because of disease, dementia, depression and social isolation, but the utilization of their stores of the three major metabolic substrates during short-term periods of fasting has not been systematically examined. This report details some findings based on our client's experiment.

Exploratory Data Analysis (EDA)

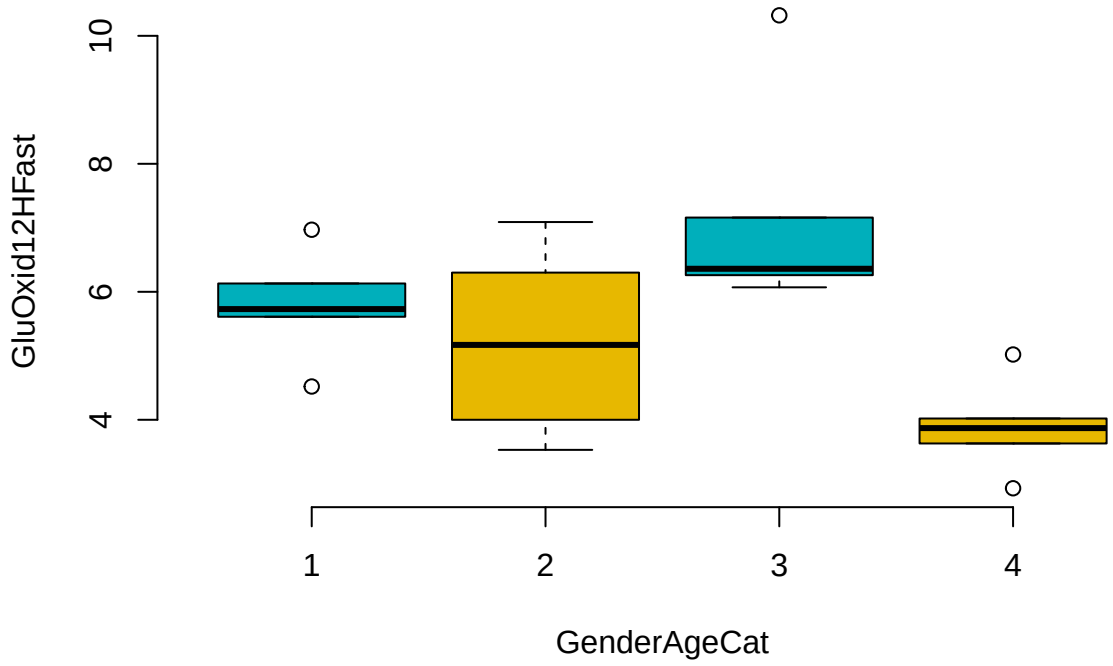
Visualizations

First, we made some visualizations to examine the spread of the various age and gender groups.



The two yellow boxes represent the senior age group. From this visualization, it seems like the `LeuRa12HFast` distributions among the two young people groups (the blue boxes) are similar. The dots represent outliers. However, we will further examine if there is a difference between the means using an ANOVA test. It should be noted that the client is free to change the variable on the y-axis, simply change the `LeuRa12HFast` in the code to the variable of choice, like so...

```
boxplot(GluOxid12HFast ~ GenderAgeCat, data=data_fasting, frame = FALSE,
        col = c("#00AFBB", "#E7B800"), ylab="GluOxid12HFast")
```

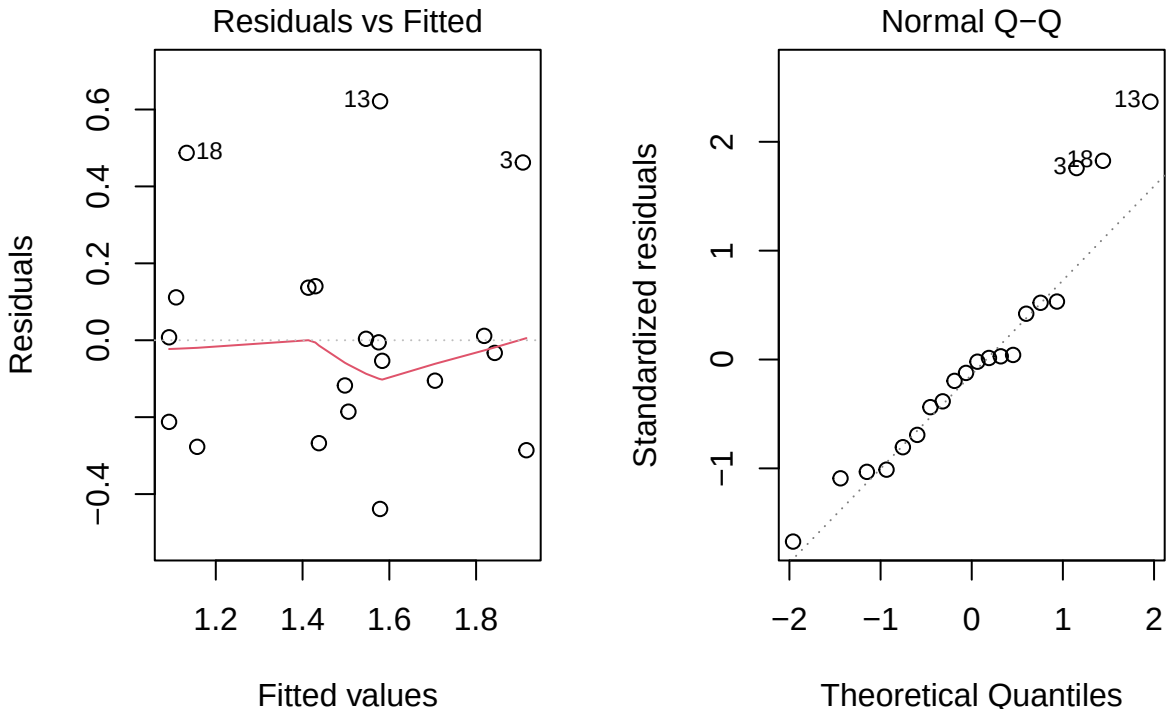


ANOVA

As per the client's request, we have performed a two-way ANOVA test. We are using a **two-way** ANOVA because we would like to consider two different factors on which the means of two groups could be vary: `Age` and `Gender`. In this case, we will use `LeuRa12HFast` and `LeuRa36HFast`.

We'd like to check if the ANOVA test assumptions are valid before viewing the results. The assumptions are as follows: independent observations, equal variances, and normal distributions.

```
par(mfrow = c(1,2))
plot(lm(LeuRa12HFast ~ Age + Gender, data = data_fasting), which=1)
plot(lm(LeuRa12HFast ~ Age + Gender, data = data_fasting), which=2)
```



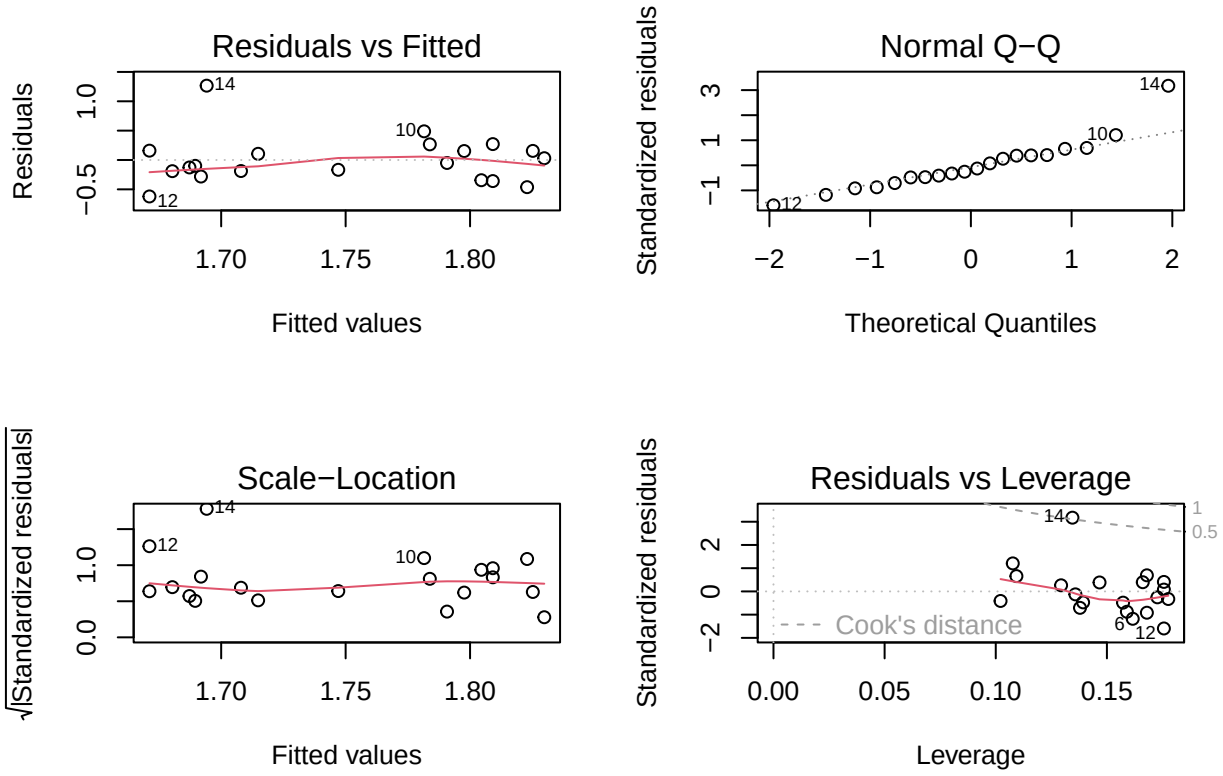
The above plot demonstrates why we may be able to trust the ANOVA results even if they show significance. The panel in the top right shows that the normality assumption is violated. Ideally, you would want the points to lie on the line. However, since there are quite a few points that stray from the line, we can conclude that the normality assumption is violated and therefore may not be able to trust the ANOVA results.

```
LeuRa12HFast_aov <- aov(LeuRa12HFast ~ Age + Gender, data = data_fasting)
summary(LeuRa12HFast_aov)
```

```
##             Df Sum Sq Mean Sq F value  Pr(>F)
## Age          1 0.7861  0.7861   9.417 0.00696 **
## Gender       1 0.5959  0.5959   7.138 0.01610 *
## Residuals   17 1.4191  0.0835
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Based on the p-values, it seems like `Age` and `Gender`, on some level, are significant in telling of `LeuRa12HFast` levels. However, we are hesitant to conclude so based on the QQ-plot above as well as the small sample size.

```
par(mfrow = c(2,2))
plot(lm(LeuRa36HFast ~ Age + Gender, data = data_fasting))
```




```
LeuRa36HFast_aov <- aov(LeuRa36HFast ~ Age + Gender, data = data_fasting)
summary(LeuRa36HFast_aov)
```

```
##             Df Sum Sq Mean Sq F value Pr(>F)
## Age          1 0.0668 0.06680   0.363  0.555
## Gender       1 0.0009 0.00095   0.005  0.944
## Residuals   17 3.1249 0.18382
```

The same cannot be said for `LeuRa36HFast`, due to the larger p-values. In other words, `Age` and `Gender` are not significant in telling `LeuRa36HFast`.

Model Fitting

Data Cleaning

Continuing forward, in order to build a statistical model, we will need to reorganize and clean the data. To summarize, what we've done is basically separated the 36H measurements and the 12H measurements into two data frames and stacked the two, so that we may create an indicator variable for `Ind36H` which is a binary, 1 if the trial was a 36H fast and 0 if the trial was a 12H fast. We are doing this to combine the power of paired observations per patient, at the cost of assuming the correlation within patients comes about as a change in intercept. Also, it should be noted that there were some spelling errors within the variables so those were changed simply for consistency.

Initial Model Fitting

Now, we begin to our multilevel models. The reason behind using a multilevel model is so that we can have a different model for each subject, to account for genetic differences.

```
library(lme4)
```

```
## Loading required package: Matrix
```

```r
#multilevel model, varying intercepts, constant slopes
m1 <- lmer(LeuRaHFast ~ Ind36H + Age + (1 | Name), data=new_data_fasting)
summary(m1)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: LeuRaHFast ~ Ind36H + Age + (1 | Name)
##    Data: new_data_fasting
##
## REML criterion at convergence: 52.5
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -1.87132 -0.69248 -0.00259  0.50441  2.84858
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Name     (Intercept) 0.009484 0.09738
##  Residual             0.147434 0.38397
## Number of obs: 40, groups:  Name, 20
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept)  1.634776   0.155205  10.533
## Ind36H       0.254500   0.121422   2.096
## Age         -0.002804   0.002575  -1.089
##
## Correlation of Fixed Effects:
##        (Intr) Ind36H
## Ind36H -0.391
## Age    -0.821  0.000
```

```r
coef(m1)
```

```
## $Name
##        (Intercept) Ind36H          Age
## ACAUSI    1.604189 0.2545 -0.002803561
## ACOTE     1.660723 0.2545 -0.002803561
## CC        1.690227 0.2545 -0.002803561
## CF        1.659738 0.2545 -0.002803561
## CG        1.669363 0.2545 -0.002803561
## CL        1.632665 0.2545 -0.002803561
## CW        1.661500 0.2545 -0.002803561
## DZ        1.651861 0.2545 -0.002803561
## GG        1.621810 0.2545 -0.002803561
## GS        1.627556 0.2545 -0.002803561
## GW        1.669677 0.2545 -0.002803561
## HA        1.654161 0.2545 -0.002803561
## HK        1.614219 0.2545 -0.002803561
## JH        1.666013 0.2545 -0.002803561
## LF        1.598150 0.2545 -0.002803561
## NB        1.605622 0.2545 -0.002803561
## NT        1.610959 0.2545 -0.002803561
## SC        1.603049 0.2545 -0.002803561
## SW        1.629548 0.2545 -0.002803561
## TS        1.564498 0.2545 -0.002803561
```

```
##
## attr(,"class")
## [1] "coef.mer"
```

Above is a varying intercepts, constant slopes multilevel model. The output from `coef(m1)` shows the model for each subject. For example, the model for the first subject, `ACAUSI`, is written as follows:

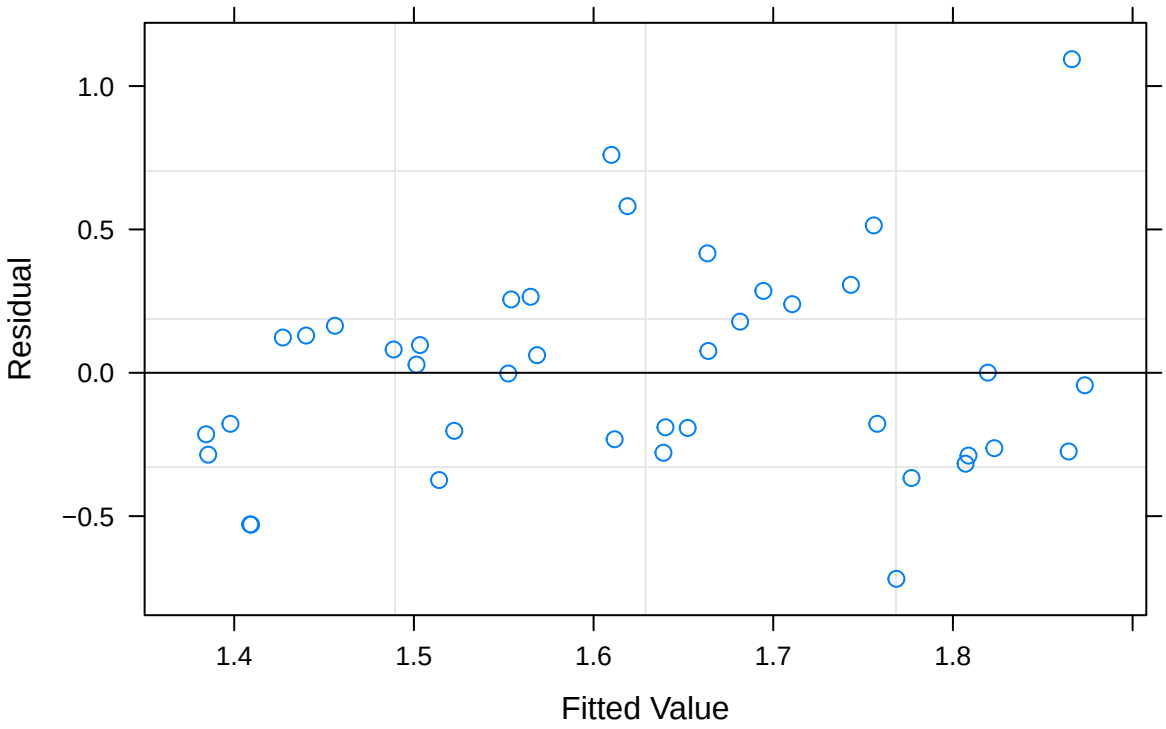
$$\widehat{ACAUSI} = 1.604 - 0.2545Ind36H + -0.00280Age$$

Again, the outcome here can be changed to your choosing, as so...

```
tmp <- lmer(GluOxidHFast ~ Ind36H + Age + (1 | Name), data=new_data_fasting)
coef(tmp)
```

```
## $Name
##        (Intercept) Ind36H        Age
## ACAUSI    6.435879 -1.507 -0.0224508
## ACOTE     6.502895 -1.507 -0.0224508
## CC        7.854759 -1.507 -0.0224508
## CF        6.951196 -1.507 -0.0224508
## CG        6.289607 -1.507 -0.0224508
## CL        6.097385 -1.507 -0.0224508
## CW        6.398036 -1.507 -0.0224508
## DZ        6.011011 -1.507 -0.0224508
## GG        6.556297 -1.507 -0.0224508
## GS        6.439759 -1.507 -0.0224508
## GW        6.519630 -1.507 -0.0224508
## HA        6.939176 -1.507 -0.0224508
## HK        6.733723 -1.507 -0.0224508
## JH        6.840445 -1.507 -0.0224508
## LF        6.641492 -1.507 -0.0224508
## NB        6.404997 -1.507 -0.0224508
## NT        7.027528 -1.507 -0.0224508
## SC        6.949902 -1.507 -0.0224508
## SW        6.452997 -1.507 -0.0224508
## TS        6.869579 -1.507 -0.0224508
##
## attr(,"class")
## [1] "coef.mer"
```

How well does `m1` model perform? To check this, we look at a residual plot.



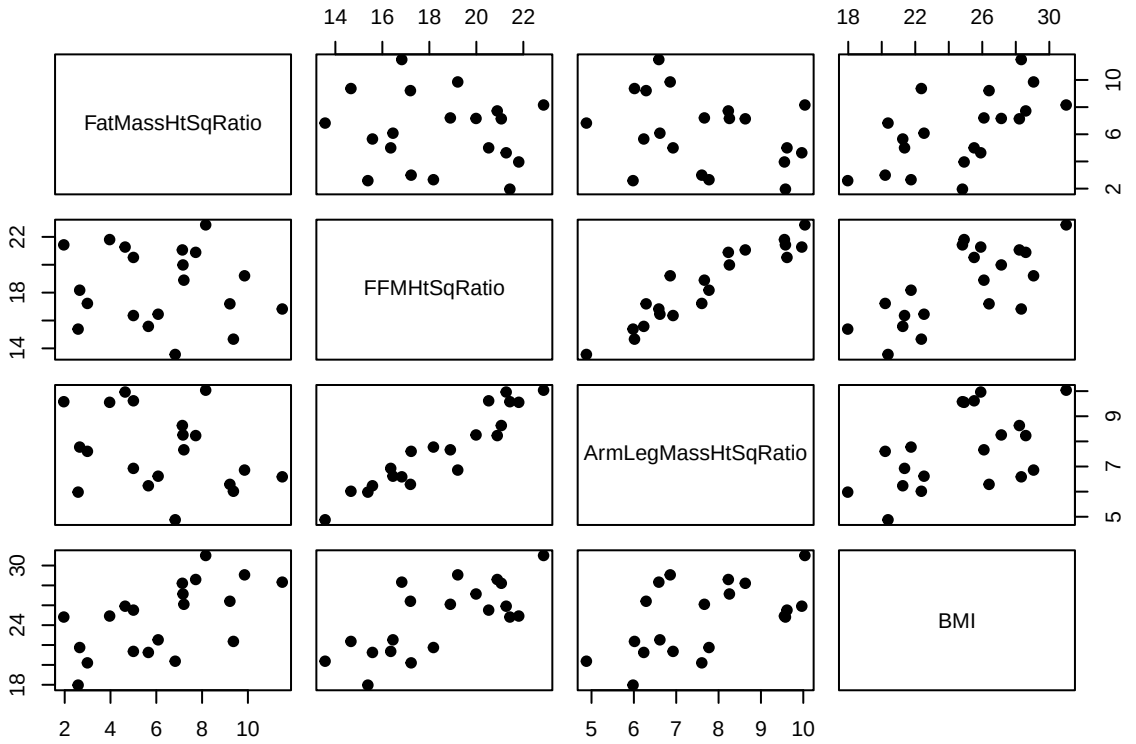
The plot shows that `m1` is a decent fit because of the "randomness" in its pattern. It is roughly evenly distributed along the line with no discerning pattern. Moving forward, we will be working with this model. First, let's check the model performance.

```
library(performance)
model_performance(m1) #checking R2
```

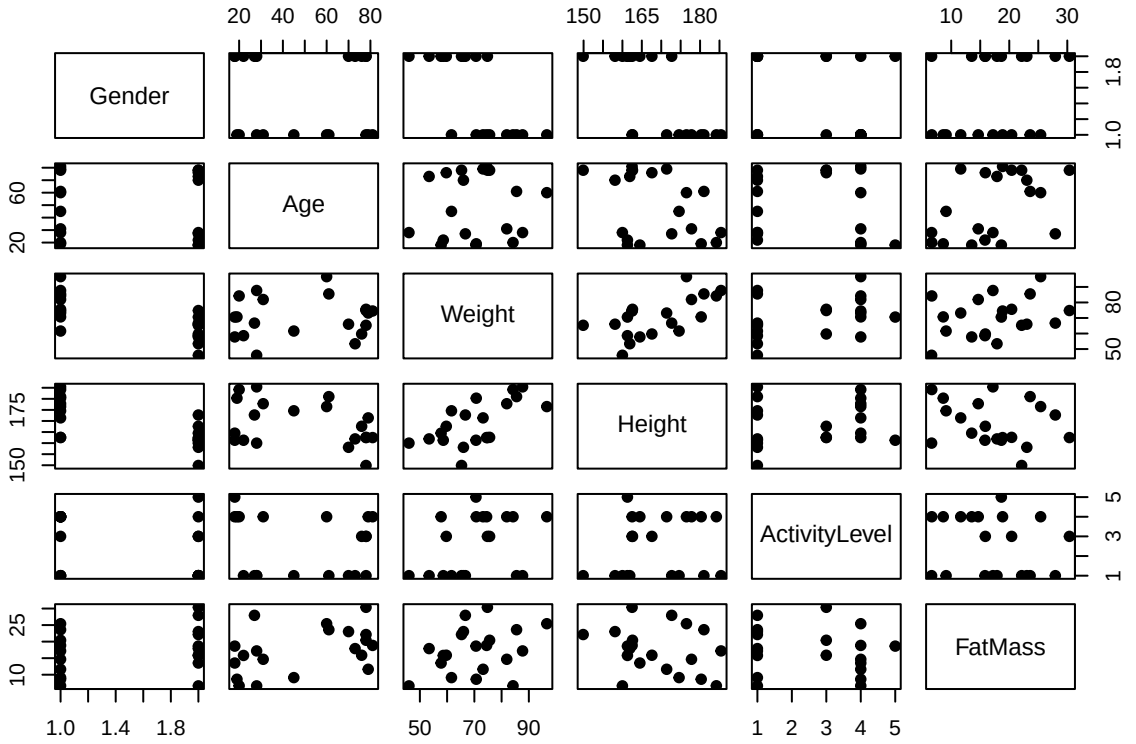
```
## # Indices of model performance
##
## AIC    |   AICc |    BIC | R2 (cond.) | R2 (marg.) |   ICC |  RMSE | Sigma
## ------------------------------------------------------------------------
## 62.470 | 64.235 | 70.915 |      0.174 |      0.121 | 0.060 | 0.359 | 0.384
```

The R2 (marg.) is quite low. The R2 (marg.) considers only the variance of the fixed effects, without the random effects.

Additionally, we'd like to see if there are any interactions between other variables. To do this, let's look at a scatterplot matrix. For our first plot, let's look at a few specific variables: `FatMassHtSqRatio`, `FFMHtSqRatio`, `ArmLegMassHtRatio`, and `BMI`.



According to the plot, there does seem to be an interaction between `FFMHtSqRatio` and `ArmLegMassHtRatio`. The plot at the cross section between these two variables are seemingly in a linear pattern. Also, there may be a weak linear interaction between `FFMHtSqRatio` and `BMI`.



The plot at the cross section between weight and height seems to be the only interaction, as there is a slight discernible pattern, though the correlation is slight. To account for this, let's use `BMI`.

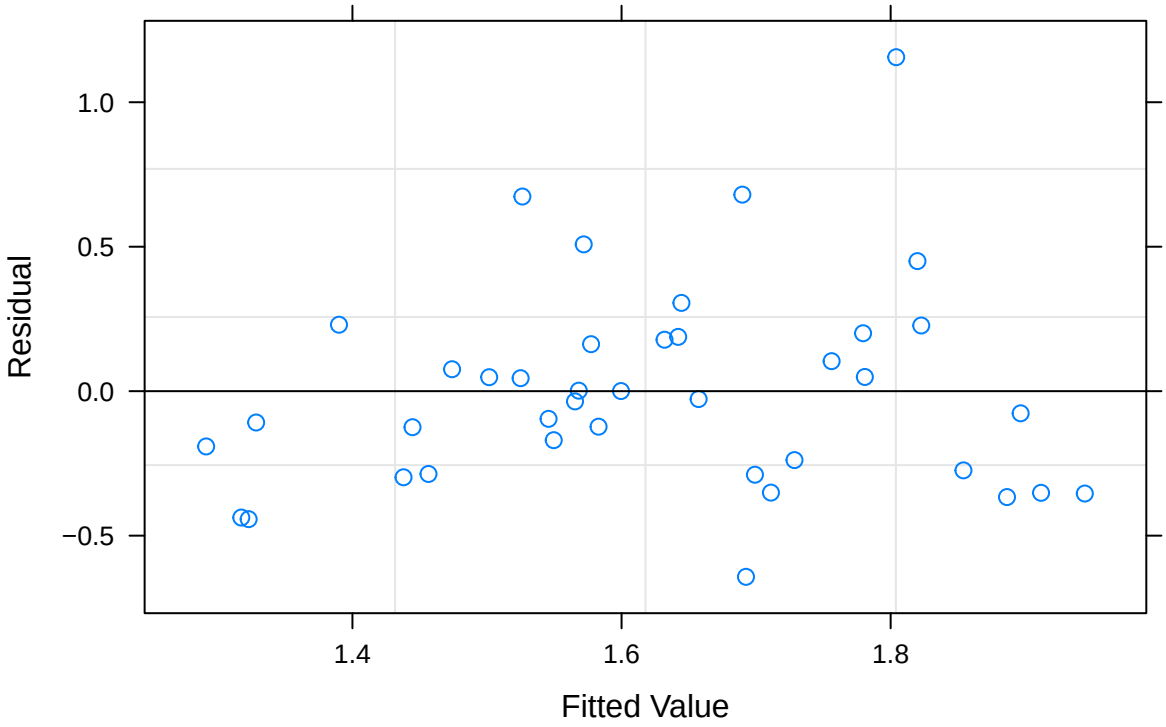
```
m1_alt <- lmer(LeuRaHFast ~ Gender + Age + BMI + Ind36H + (1 | Name), data=new_data_fasting)
model_performance(m1_alt)
```

```
## # Indices of model performance
##
## AIC    |   AICc |    BIC | R2 (cond.) | R2 (marg.) |   ICC |  RMSE | Sigma
## ------------------------------------------------------------------------------
## 72.588 | 76.088 | 84.410 |      0.212 |      0.161 | 0.062 | 0.350 | 0.384
```

```
summary(m1_alt)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: LeuRaHFast ~ Gender + Age + BMI + Ind36H + (1 | Name)
##    Data: new_data_fasting
##
## REML criterion at convergence: 58.6
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -1.67324 -0.72196 -0.08164  0.47011  3.01035
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Name     (Intercept) 0.009672 0.09835
##  Residual             0.147434 0.38397
## Number of obs: 40, groups:  Name, 20
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept)  1.998266   0.593512   3.367
## Gender      -0.186569   0.135919  -1.373
## Age         -0.002697   0.002863  -0.942
## BMI         -0.003600   0.021708  -0.166
## Ind36H       0.254500   0.121422   2.096
##
## Correlation of Fixed Effects:
##        (Intr) Gender Age    BMI
## Gender -0.598
## Age     0.192 -0.111
## BMI    -0.906  0.311 -0.434
## Ind36H -0.102  0.000  0.000  0.000
```

```
plot(m1_alt, ylab="Residual", xlab="Fitted Value", main="m1_alt Resid vs fitted")
```



We use `BMI` in place of `Weight` and `Height` because we are only concerned about the ratio between the two, which is what `BMI` represents. `Age` is added as per the client's request.