

BodyWeight Linear Mixed Model Analysis

Michelle Ha

06 October 2024

Contents

Data Visualization	1
Model Fitting	3
Model Comparison	5
Assumption Checking	5

```
str(BodyWeight)
```

```
## Classes 'nfnGroupedData', 'nfGroupedData', 'groupedData' and 'data.frame':  176 obs. of  4 variables:
## $ weight: num  240 250 255 260 262 258 266 266 265 272 ...
## $ Time : num  1 8 15 22 29 36 43 44 50 57 ...
## $ Rat : Ord.factor w/ 16 levels "2"<"3"<"4"<"1"<...: 4 4 4 4 4 4 4 4 4 4 ...
## $ Diet : Factor w/ 3 levels "1","2","3": 1 1 1 1 1 1 1 1 1 1 ...
## - attr(*, "outer")=Class 'formula' language ~Diet
## ..$.- attr(*, ".Environment")=<environment: R_GlobalEnv>
## - attr(*, "formula")=Class 'formula' language weight ~ Time | Rat
## ..$.- attr(*, ".Environment")=<environment: R_GlobalEnv>
## - attr(*, "labels")=List of 2
## ..$ x: chr "Time"
## ..$ y: chr "Body weight"
## - attr(*, "units")=List of 2
## ..$ x: chr "(days)"
## ..$ y: chr "(g)"
## - attr(*, "FUN")=function (x)
## ..$.- attr(*, "source")= chr "function(x) max(x, na.rm = TRUE)"
## - attr(*, "order.groups")= logi TRUE
```

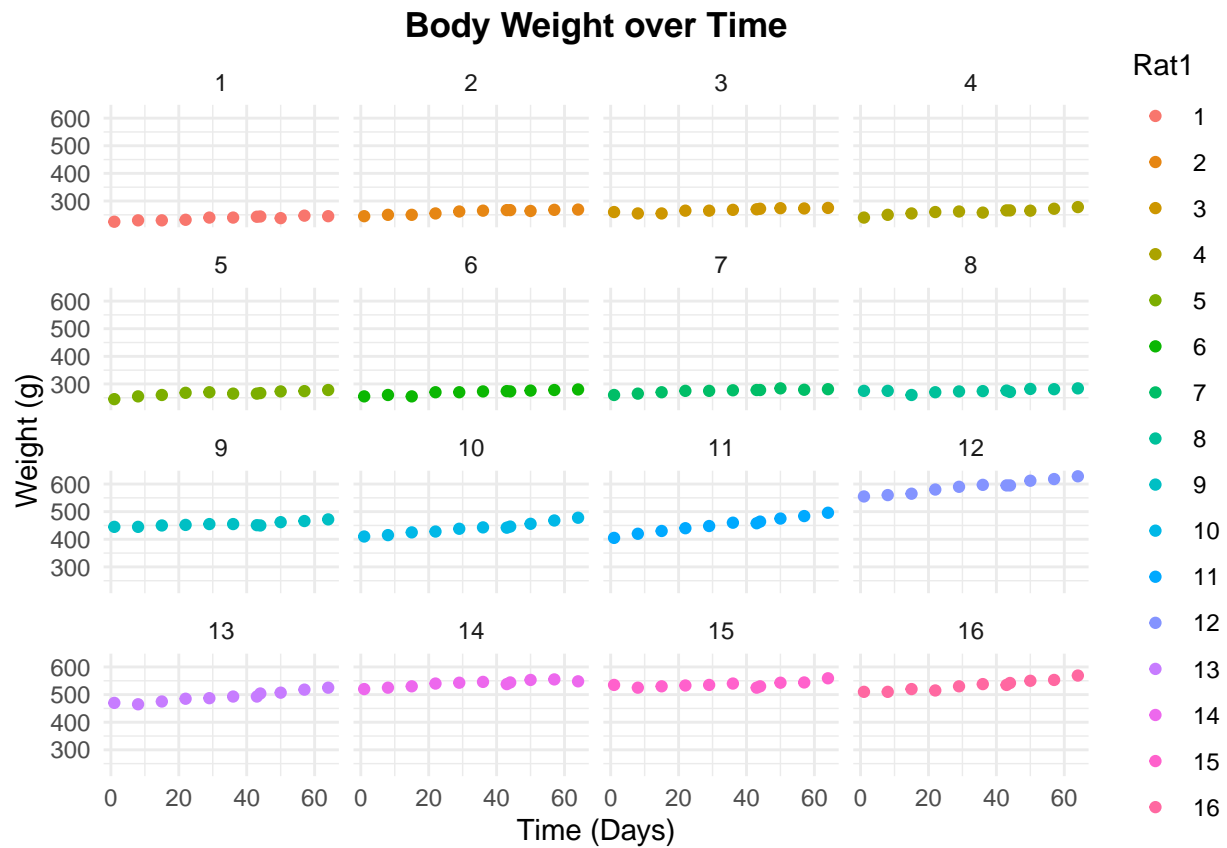
Data Visualization

```
BodyWeight |>
  mutate(Rat1 = factor(as.numeric(Rat))) |>
  ggplot(aes(Time, weight, group = Rat1, color = Rat1)) +
```

```

geom_point() +
facet_wrap(~Rat1) +
theme_minimal() +
labs(
  x = "Time (Days)",
  y = "Weight (g)",
  title = "Body Weight over Time"
) +
theme(
  plot.title = element_text(hjust = 0.5, face = "bold")
)

```



The plot displays the body weight (in grams) of the rats over 64 days at interval of 7 days.

```

# Calculate mean weights for each Time and Diet combination
mean_weights <- BodyWeight |>
  summarize(mean_weight = mean(weight), .by = c("Time", "Diet"))

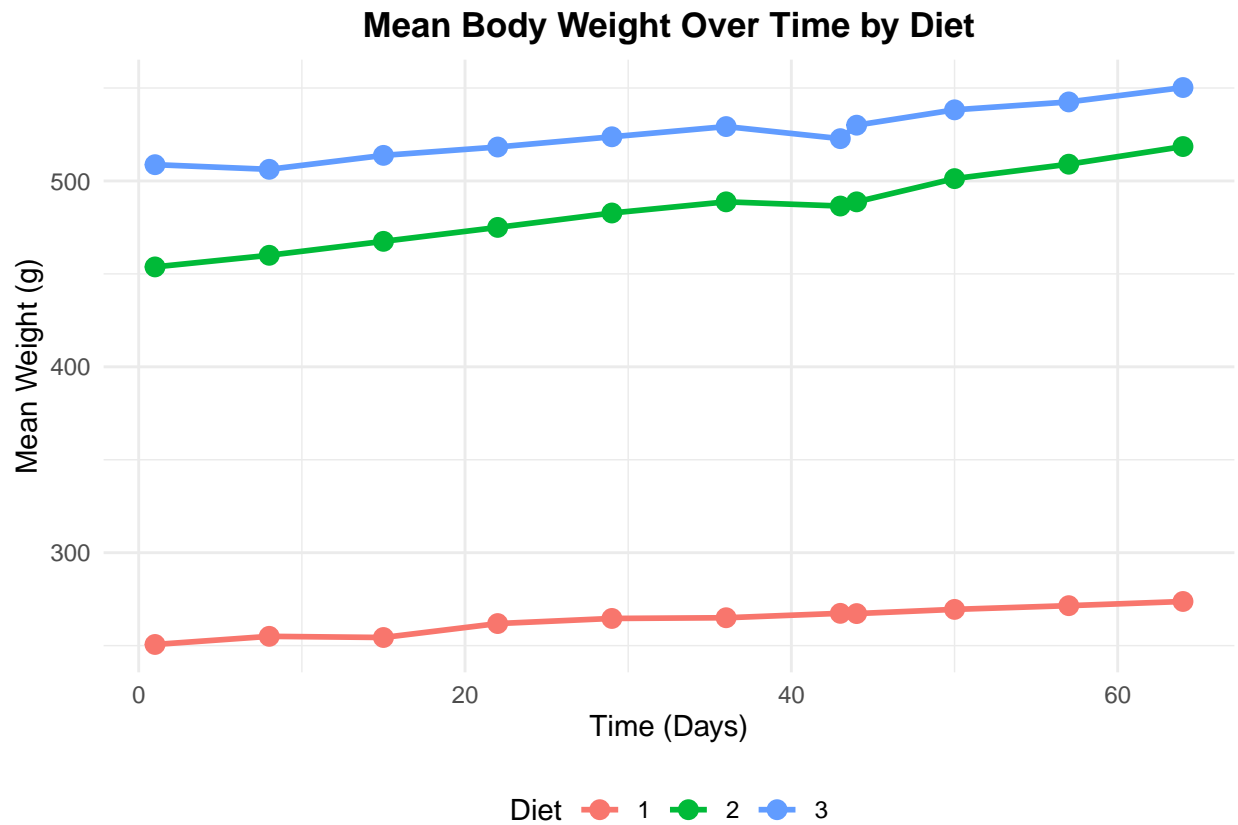
mean_weights |> ggplot(aes(x = Time, y = mean_weight, color = Diet, group = Diet)) +
  geom_line(size = 1) +
  geom_point(size = 3) +
  theme_minimal() +
  labs(
    title = "Mean Body Weight Over Time by Diet",
    x = "Time (Days)",
    y = "Mean Weight (g)",

```

```

    color = "Diet"
  ) +
  theme(
    legend.position = "bottom",
    plot.title = element_text(hjust = 0.5, face = "bold")
  )

```



The figure showing the mean weight of each type of diet over time displays an increasing trend in the body weight of the rats as the days progress.

Model Fitting

```

# get weight at baseline
BodyWeight <- BodyWeight |>
  group_by(Rat) |>
  arrange(weight, .by_group = TRUE) |>
  mutate(initial_weight = first(weight)) |>
  ungroup()

# fit LMM with interaction effect
model11 <- lmer(weight ~ Time * Diet + initial_weight + (1 | Rat), BodyWeight)
summary(model11)

```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: weight ~ Time * Diet + initial_weight + (1 | Rat)
## Data: BodyWeight
##
## REML criterion at convergence: 1198.1
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.7956 -0.4758 -0.0059  0.4842  3.2162
##
## Random effects:
## Groups Name Variance Std.Dev.
## Rat (Intercept) 75.84 8.708
## Residual 41.02 6.404
## Number of obs: 176, groups: Rat, 16
##
## Fixed effects:
## Estimate Std. Error df t value Pr(>|t|)
## (Intercept) 13.07809 16.89727 12.11731 0.774 0.454
## Time 0.35964 0.03509 157.00001 10.248 < 2e-16 ***
## Diet2 2.95591 14.90964 12.46144 0.198 0.846
## Diet3 5.08494 18.10691 12.31007 0.281 0.784
## initial_weight 0.96151 0.06673 11.99998 14.408 6.15e-09 ***
## Time:Diet2 0.60584 0.06079 157.00001 9.967 < 2e-16 ***
## Time:Diet3 0.29834 0.06079 157.00001 4.908 2.28e-06 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
## (Intr) Time Diet2 Diet3 intl_w Tm:Dt2
## Time -0.070
## Diet2 0.857 0.079
## Diet3 0.891 0.065 0.913
## initil_wght -0.980 0.000 -0.920 -0.947
## Time:Diet2 0.040 -0.577 -0.137 -0.038 0.000
## Time:Diet3 0.040 -0.577 -0.046 -0.113 0.000 0.333
```

```
anova(model1)
```

```
## Type III Analysis of Variance Table with Satterthwaite's method
## Sum Sq Mean Sq NumDF DenDF F value Pr(>F)
## Time 26193.0 26193.0 1 157.000 638.6116 < 2.2e-16 ***
## Diet 4.1 2.0 2 13.675 0.0497 0.9517
## initial_weight 8514.3 8514.3 1 12.000 207.5865 6.155e-09 ***
## Time:Diet 4190.4 2095.2 2 157.000 51.0833 < 2.2e-16 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The interaction term `Time*Diet` has $p\text{-value} < 0.05$ therefore it is reasonable to say the type of die significantly affect the rats body weight and the effect is consistent over time. The initial weight is also a strong indicator for the body weight.

Model Comparison

```
# fit LMM with no interaction effect
model2 <- lmer(weight ~ Time + Diet + initial_weight + (1 | Rat), BodyWeight)
```

We will fit a simpler model without the interaction term to see if it could potentially performs as well as model1. The simpler model still has Rat as random effect and initial_weight as fixed effect, except Time and Diet are now the fixed effects.

```
# compare models
anova(model1, model2)
```

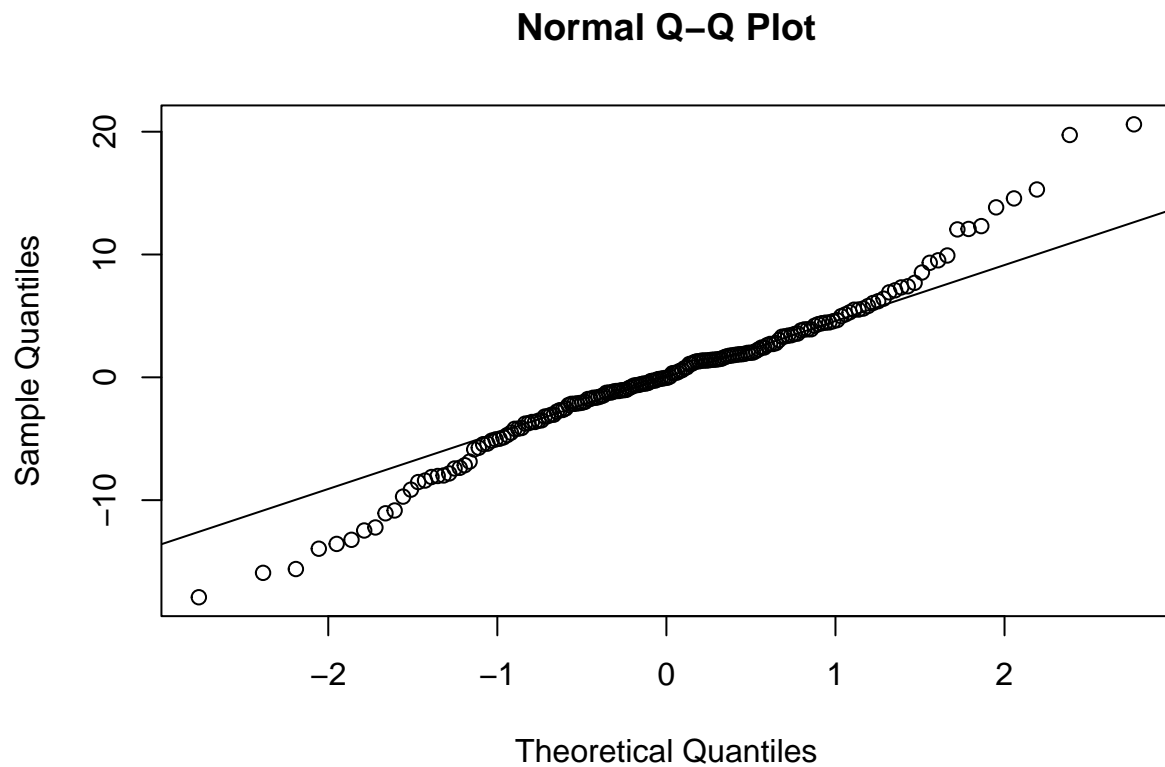
```
## Data: BodyWeight
## Models:
## model2: weight ~ Time + Diet + initial_weight + (1 | Rat)
## model1: weight ~ Time * Diet + initial_weight + (1 | Rat)
##      npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## model2     7 1288.7 1310.8 -637.33   1274.7
## model1     9 1212.5 1241.0 -597.23   1194.5 80.196  2 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

P-value of < 0.05 and lower AIC and BIC indicates that the model1 with the interaction term provides a significantly better fit relative to the simpler model2.

Assumption Checking

```
# Normality of residuals
# Extract residuals
residuals <- resid(model1)

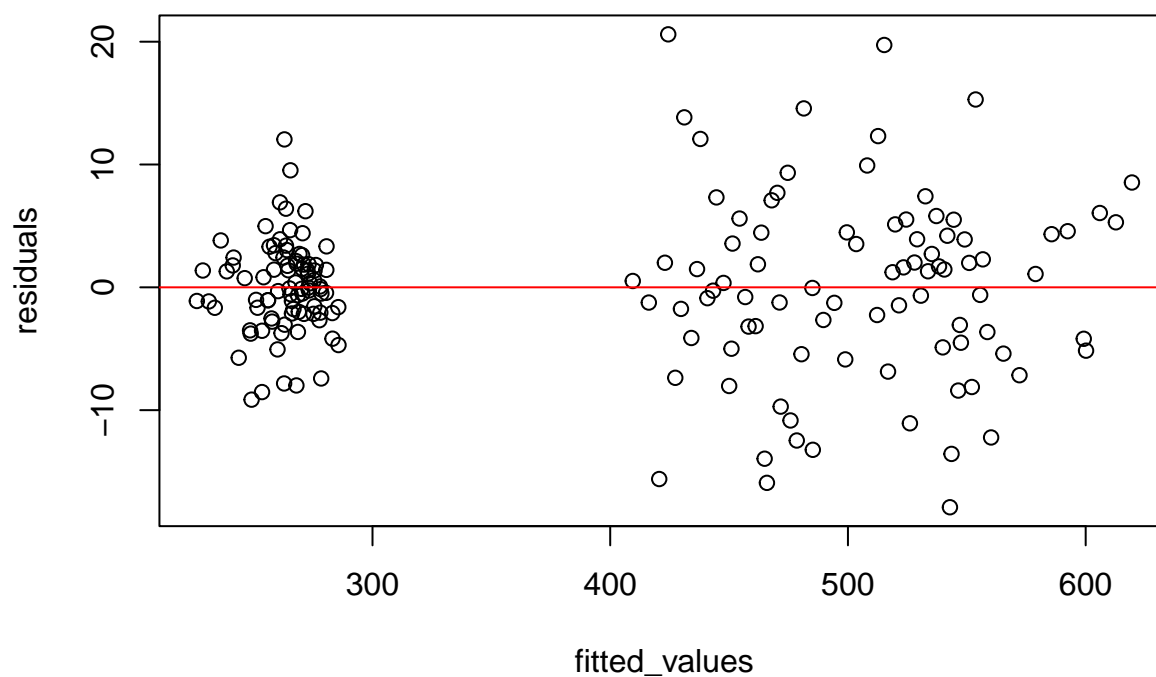
# Q-Q plot
qqnorm(residuals)
qqline(residuals)
```



The majority of the points fall close to the diagonal line, with some departures at the tails. The overall shape suggests mild non-normality.

```
# Homogeneity of Variance  
# Fitted values  
fitted_values <- fitted(model1)  
  
# Residuals vs Fitted plot  
plot(fitted_values, residuals)  
abline(h = 0, col = "red")  
title("Residuals vs Fitted Values")
```

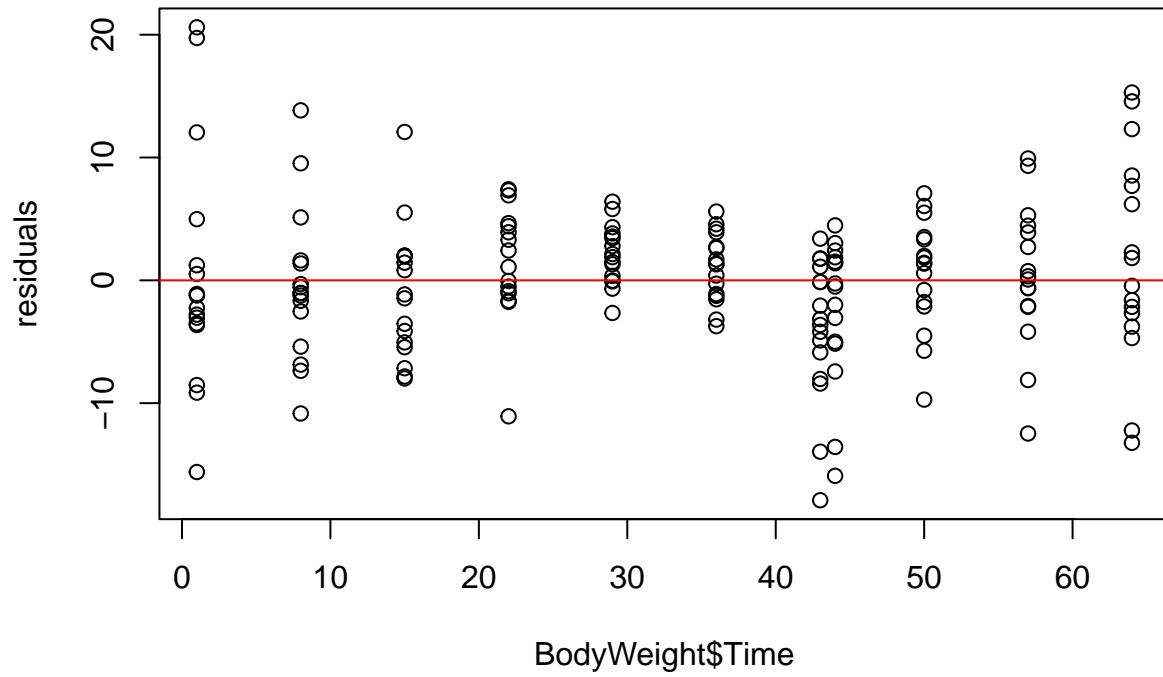
Residuals vs Fitted Values



The spread of residuals appears to vary across the range of fitted values. There is noticeably more spread (larger residuals) for fitted values around 500-600 compared to those around 300-400. This suggests some heteroscedasticity.

```
# Residuals over time  
plot(BodyWeight$Time, residuals)  
abline(h = 0, col = "red")  
title("Residuals over Time")
```

Residuals over Time



The residuals appear to be fairly randomly scattered around the zero line across time. There seem to be a few potential outliers, which might warrant further investigation. Overall, the assumption of independence of residuals is generally met.