

# Biomass Example: Three-way ANOVA with Unequal Variance

In this example, there are 2 Types (grass or forb), 3 Herbicide treatments (A, B or C) and 2 Water regimes (Low or High) for a total of 12 treatment combinations. There are three reps per treatment combination.

Based on the summary statistics and residual diagnostic plot, it appears that the variance is different for grass versus forb. We can deal with this two ways: (1) running a 2way analysis separately for grass and forb (2) using `gls()` from `nlme` to allow different variances by Type.

```
library(XLConnect)

## Warning: package 'XLConnect' was built under R version 3.4.4

library(dplyr)
library(emmeans)
excel.file <- file.path("C:/hess/STAT512/RNotes/ExtraTopics/Biomass.xlsx")
InData <- readWorksheetFromFile(excel.file, sheet = "Sheet1")
str(InData)

## 'data.frame': 36 obs. of 4 variables:
## $ Type : chr "Grass" "Grass" "Grass" "Grass" ...
## $ Herb : chr "A" "A" "A" "A" ...
## $ Water : chr "Low" "Low" "Low" "High" ...
## $ Biomass: num 18.8 16.4 17.6 20.5 18.1 ...

SumStats <- summarize(group_by(InData, Type, Herb, Water),
  n = n(),
  mean = mean(Biomass),
  sd = sd(Biomass))

SumStats

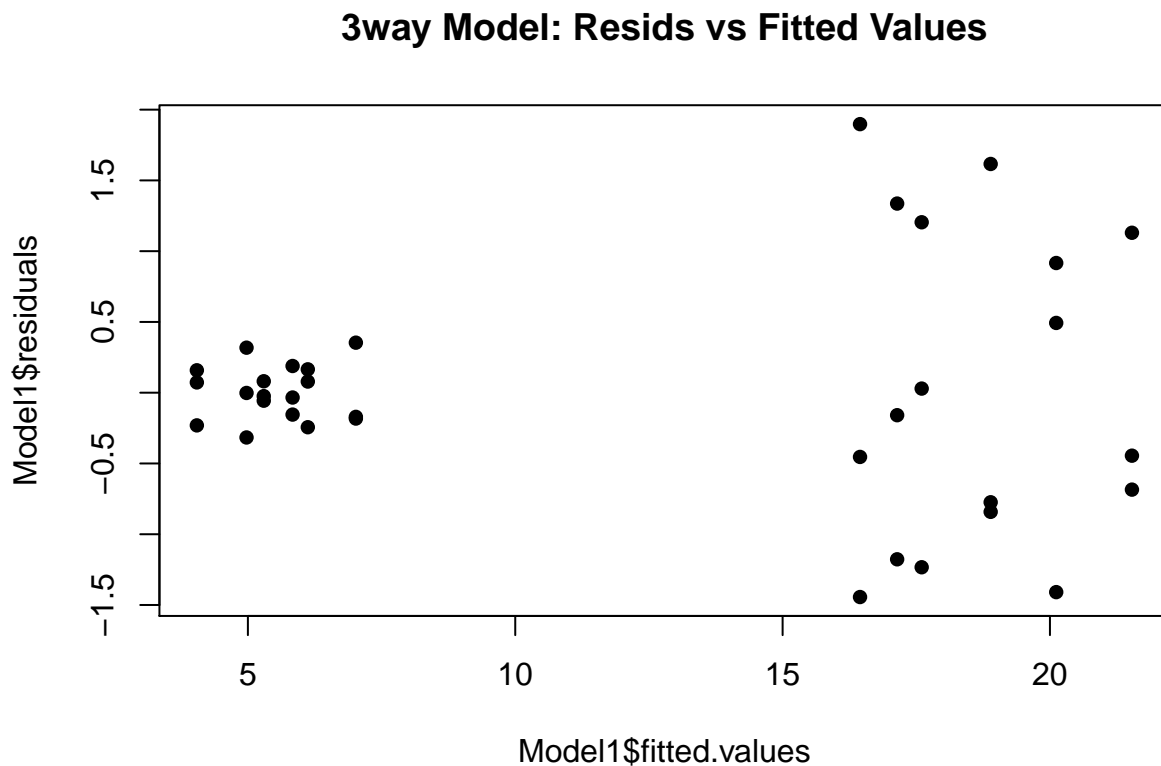
## # A tibble: 12 x 6
## # Groups: Type, Herb [?]
##   Type Herb Water    n mean    sd
##   <chr> <chr> <chr> <int> <dbl> <dbl>
## 1 Forb A High     3  6.12 0.216
## 2 Forb A Low      3  4.97 0.317
## 3 Forb B High     3  5.30 0.0718
## 4 Forb B Low      3  4.05 0.205
## 5 Forb C High     3  7.02 0.306
## 6 Forb C Low      3  5.84 0.174
## 7 Grass A High     3 18.9  1.40
## 8 Grass A Low      3 17.6  1.22
## 9 Grass B High     3 20.1  1.24
##10 Grass B Low      3 17.1  1.26
##11 Grass C High     3 16.4  1.72
##12 Grass C Low      3 21.5  0.986
```

### 3Way Model

```
options(contrasts = c("contr.sum", "contr.poly"))
Modell1 <- lm(Biomass ~ Type*Herb*Water, data = InData)
emmeans(Modell1, ~ Herb*Water*Type)
```

```
## Herb Water Type      emmean      SE df lower.CL upper.CL
## A      High  Forb      6.120484  0.5480887 24  4.989285  7.251684
## B      High  Forb      5.296773  0.5480887 24  4.165574  6.427973
## C      High  Forb      7.019339  0.5480887 24  5.888139  8.150538
## A      Low   Forb      4.974372  0.5480887 24  3.843172  6.105571
## B      Low   Forb      4.045132  0.5480887 24  2.913933  5.176332
## C      Low   Forb      5.835487  0.5480887 24  4.704287  6.966687
## A      High  Grass     18.891215  0.5480887 24  17.760016  20.022415
## B      High  Grass     20.117282  0.5480887 24  18.986083  21.248482
## C      High  Grass     16.448119  0.5480887 24  15.316919  17.579318
## A      Low   Grass     17.601195  0.5480887 24  16.469996  18.732395
## B      Low   Grass     17.142260  0.5480887 24  16.011060  18.273459
## C      Low   Grass     21.533570  0.5480887 24  20.402371  22.664770
##
## Confidence level used: 0.95
```

```
plot(Model1$residuals ~ Model1$fitted.values, main = "3way Model: Resids vs Fitted Values", pch = 16)
```

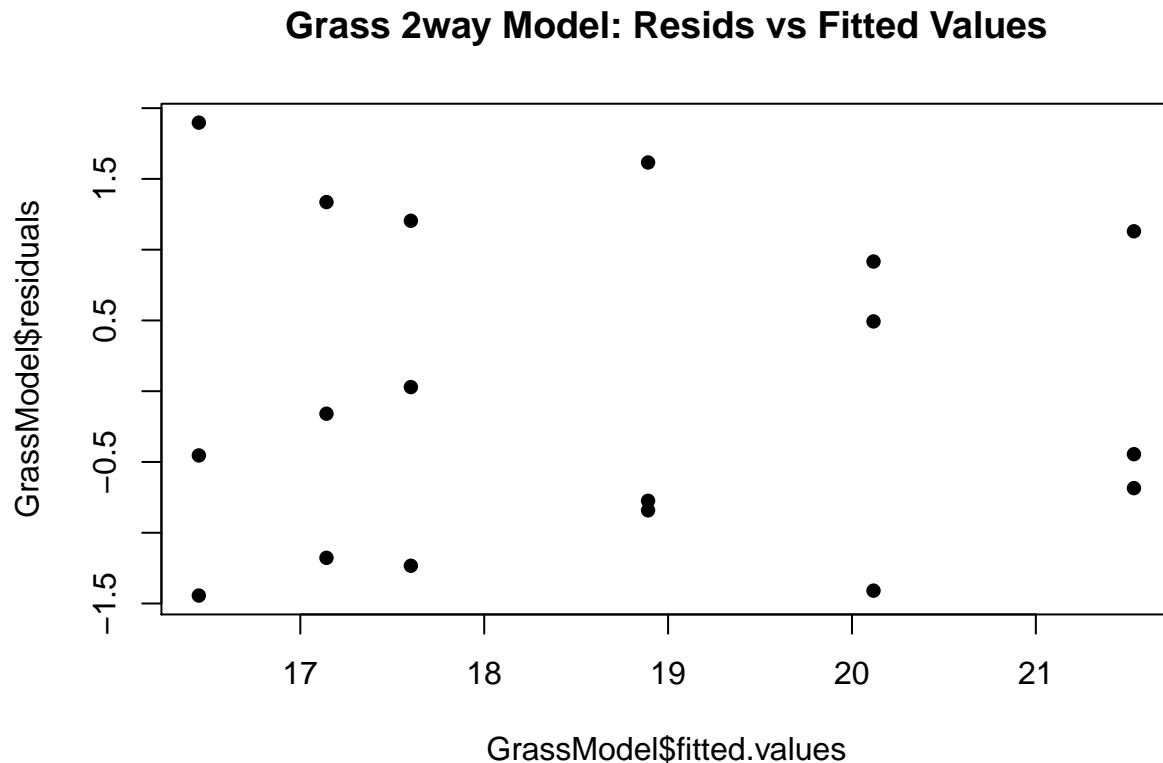


## Grass: 2way Model

```
GrassModel <- lm(Biomass ~ Herb*Water, data = InData[InData$Type == "Grass",])
emmeans(GrassModel, ~ Herb*Water)
```

```
## Herb Water    emmean      SE df lower.CL upper.CL
## A   High  18.89122 0.7636266 12 17.22742 20.55501
## B   High  20.11728 0.7636266 12 18.45348 21.78108
## C   High  16.44812 0.7636266 12 14.78432 18.11192
## A   Low   17.60120 0.7636266 12 15.93740 19.26499
## B   Low   17.14226 0.7636266 12 15.47846 18.80606
## C   Low   21.53357 0.7636266 12 19.86977 23.19737
##
## Confidence level used: 0.95
```

```
plot(GrassModel$residuals ~ GrassModel$fitted.values,
     main = "Grass 2way Model: Resids vs Fitted Values", pch = 16)
```



## Forb: 2way Model

```
ForbModel <- lm(Biomass ~ Herb*Water, data = InData[InData$Type == "Forb",])
emmeans(ForbModel, ~ Herb*Water)
```

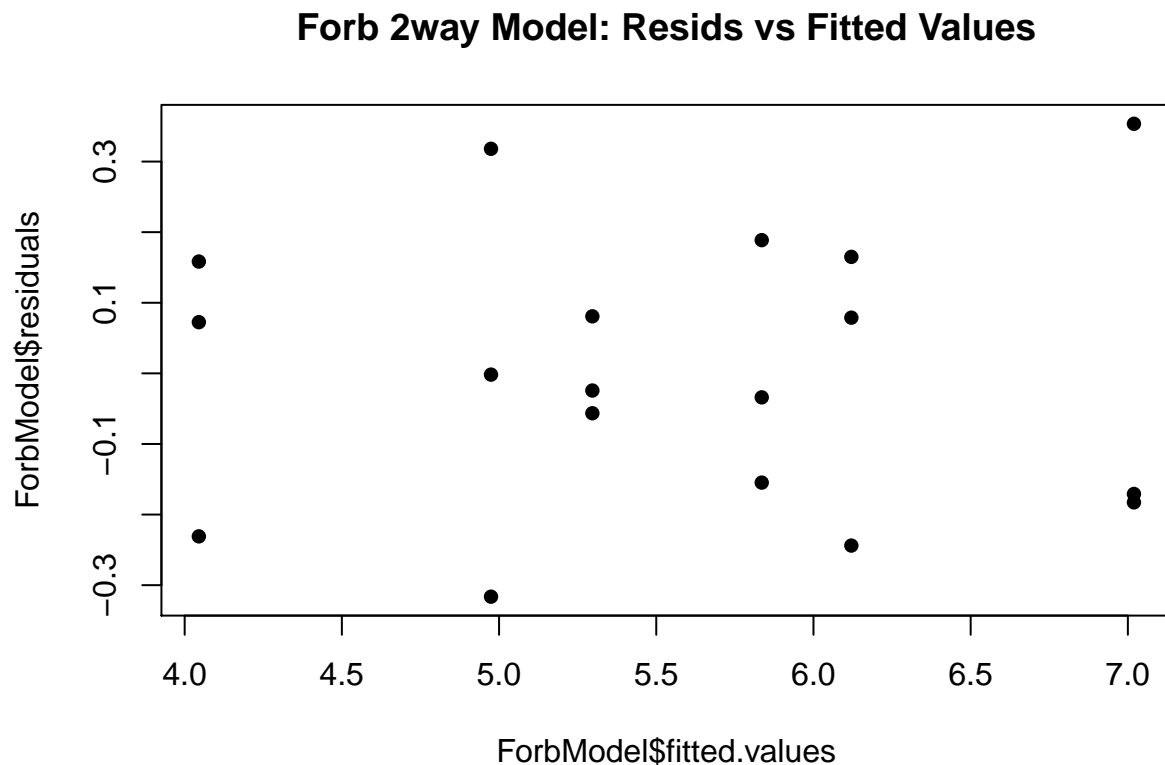
```
## Herb Water    emmean      SE df lower.CL upper.CL
```

```
## A    High  6.120484 0.1329548 12 5.830801 6.410168
## B    High  5.296773 0.1329548 12 5.007090 5.586457
## C    High  7.019339 0.1329548 12 6.729655 7.309022
## A    Low   4.974372 0.1329548 12 4.684688 5.264055
## B    Low   4.045132 0.1329548 12 3.755448 4.334816
## C    Low   5.835487 0.1329548 12 5.545803 6.125171
```

```
##
```

```
## Confidence level used: 0.95
```

```
plot(ForbModel$residuals ~ ForbModel$fitted.values,
     main = "Forb 2way Model: Resids vs Fitted Values", pch = 16)
```



## 3way Analysis allowing Unequal variance

Important Note: Most people recommend using Satterthwaite (or Kenward-Roger) df for this scenario. This option is available in SAS and other programs, but I do not know of a way to do this in R. The lmerTest package uses Satterthwaite df but does not work with nlme.

```
library(nlme)
```

```
##
```

```
## Attaching package: 'nlme'
```

```
## The following object is masked from 'package:dplyr':
```

```
##
```

```
## collapse
```

```
Model4 <- gls(Biomass ~ Type*Herb*Water, weights=varIdent(form = ~1|Type), data = InData)
emmeans(Model4, ~ Herb*Water*Type)
```

```
## Herb Water Type      emmean      SE df lower.CL upper.CL
## A   High  Forb    6.120484 0.1329548 24  5.846079  6.394890
## B   High  Forb    5.296773 0.1329548 24  5.022368  5.571179
## C   High  Forb    7.019339 0.1329548 24  6.744933  7.293744
## A   Low   Forb    4.974372 0.1329548 24  4.699966  5.248777
## B   Low   Forb    4.045132 0.1329548 24  3.770727  4.319537
## C   Low   Forb    5.835487 0.1329548 24  5.561082  6.109892
## A   High  Grass   18.891215 0.7636265 24  17.315168  20.467263
## B   High  Grass   20.117282 0.7636265 24  18.541234  21.693330
## C   High  Grass   16.448119 0.7636265 24  14.872071  18.024166
## A   Low   Grass   17.601195 0.7636265 24  16.025148  19.177243
## B   Low   Grass   17.142260 0.7636265 24  15.566212  18.718308
## C   Low   Grass   21.533570 0.7636265 24  19.957523  23.109618
##
## Confidence level used: 0.95
```

```
plot(Model4)
```

