Poppies Example: One-way ANOVA with Unequal Variance

In this example, five herbicides are compared in their ability to limit the number of poppy plants in oats. The five herbicide treatments are randomly assigned to twenty plots. The results, in number of poppy plants per 3.75 sqft of oats.

We consider several approaches:

1. Analysis on the original scale shows unequal variance.
2. After square root transformation the diagnostic plots look better.
3. Finally, we use gls() from nlme to allow different variances for the 5 treatments.

```
library(dplyr)
library(emmeans)
Poppies <- read.csv("C:/hess/STAT512/RNotes/ExtraTopics/CH8_Poppies.csv")
str(Poppies)
```

```
## 'data.frame':    20 obs. of  2 variables:
##  $ Trt   : Factor w/ 5 levels "A","B","C","D",..: 1 1 1 1 2 2 2 2 3 3 ...
##  $ Plants: int  438 442 319 380 538 422 377 315 77 61 ...
```

```
SumStats <- summarize(group_by(Poppies, Trt),
                n = n(),
                mean = mean(Plants),
                sd = sd(Plants))
SumStats$scalesd <- (SumStats$sd)/(SumStats$sd[1])
SumStats
```

```
## # A tibble: 5 x 5
##   Trt       n  mean    sd scalesd
##   <fct> <int> <dbl> <dbl>   <dbl>
## 1 A         4 395.   57.9   1.00
## 2 B         4 413.   94.2   1.63
## 3 C         4  86.8  48.0   0.828
## 4 D         4  37.8  33.5   0.579
## 5 E         4  35.2  28.0   0.484
```
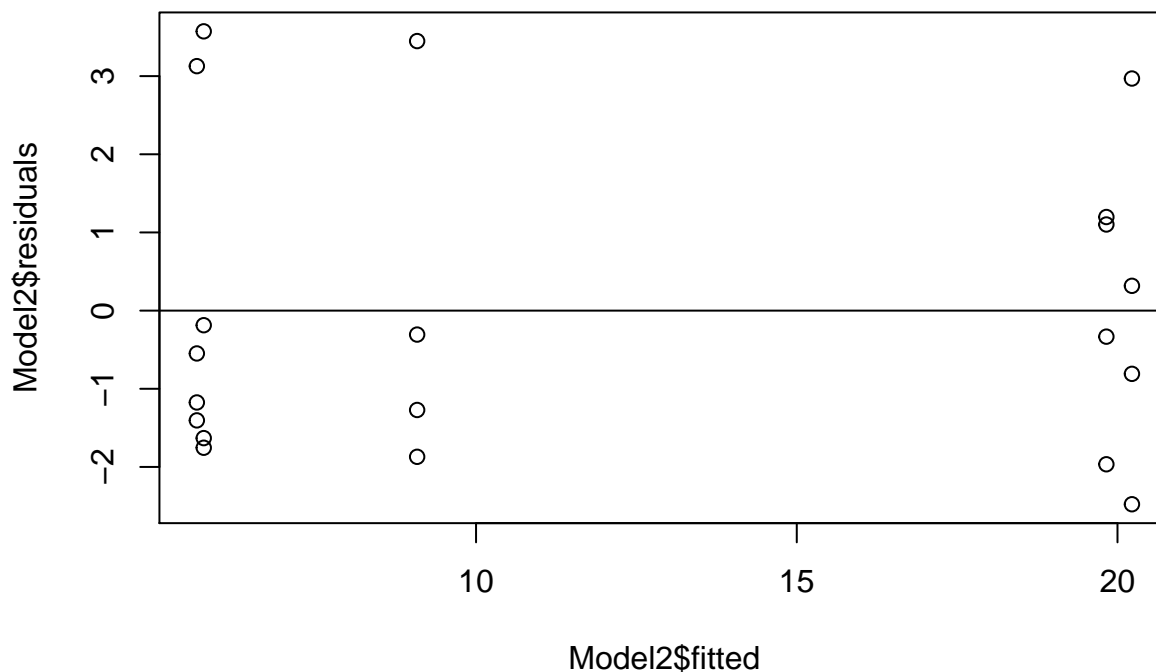
Original Scale

```
Model1 <- lm(Plants ~ Trt, data = Poppies)
plot(Model1$residuals ~ Model1$fitted);abline(h=0)
```
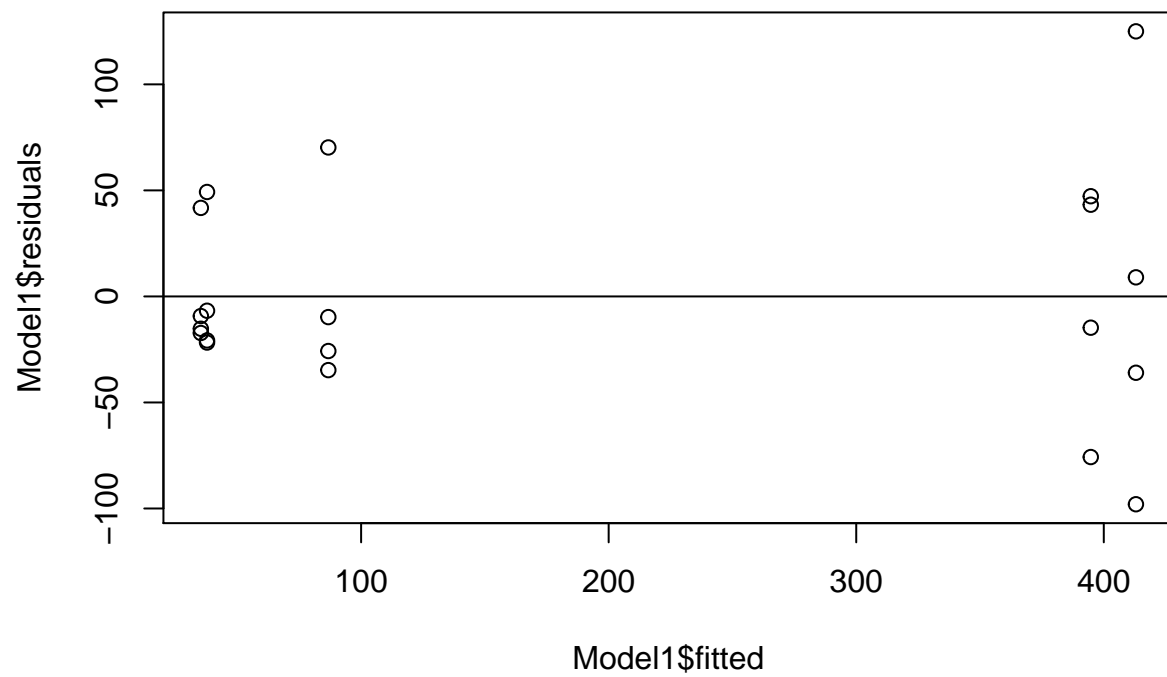
Analysis after Square Root Transformation

```
Poppies$sqrtPlants <- sqrt(Poppies$Plants)
Model2 <- lm(sqrtPlants ~ Trt, data = Poppies)
plot(Model2$residuals ~ Model2$fitted);abline(h=0)
```

```r
emmeans(Model2, pairwise ~ Trt)
```

```
## $emmeans
##  Trt     emmean       SE df  lower.CL  upper.CL
##  A    19.826601 1.090945 15 17.501307 22.151896
##  B    20.225548 1.090945 15 17.900254 22.550843
##  C     9.081570 1.090945 15  6.756276 11.406865
##  D     5.754562 1.090945 15  3.429268  8.079857
##  E     5.647190 1.090945 15  3.321896  7.972485
##
## Confidence level used: 0.95
##
## $contrasts
##  contrast    estimate       SE df t.ratio p.value
##  A - B     -0.3989469 1.542829 15  -0.259  0.9989
##  A - C     10.7450312 1.542829 15   6.964  <.0001
##  A - D     14.0720391 1.542829 15   9.121  <.0001
##  A - E     14.1794112 1.542829 15   9.191  <.0001
##  B - C     11.1439780 1.542829 15   7.223  <.0001
##  B - D     14.4709859 1.542829 15   9.380  <.0001
##  B - E     14.5783581 1.542829 15   9.449  <.0001
##  C - D      3.3270079 1.542829 15   2.156  0.2481
##  C - E      3.4343800 1.542829 15   2.226  0.2229
##  D - E      0.1073721 1.542829 15   0.070  1.0000
##
## P value adjustment: tukey method for comparing a family of 5 estimates
```

# Return to Original Scale, but allow 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 Satterthwiate df but does not work with nlme.

```
library(nlme)
```

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

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

```
Model3 <- gls(Plants ~ Trt, weights=varIdent(form = ~1|Trt), data = Poppies)
summary(Model3)
```

```
## Generalized least squares fit by REML
##   Model: Plants ~ Trt
##   Data: Poppies
##        AIC      BIC    logLik
##   185.4238 192.5043 -82.71188
##
## Variance function:
##  Structure: Different standard deviations per stratum
##  Formula: ~1 | Trt
##  Parameter estimates:
##         A         B         C         D         E
## 1.0000000 1.6263645 0.8282781 0.5792271 0.4842493
##
## Coefficients:
##               Value Std.Error    t-value p-value
## (Intercept)  394.75  28.95218  13.634550  0.0000
## TrtB          18.25  55.27564   0.330164  0.7458
## TrtC        -308.00  37.59377  -8.192847  0.0000
## TrtD        -357.00  33.45830 -10.669997  0.0000
## TrtE        -359.50  32.16817 -11.175642  0.0000
##
##  Correlation:
##      (Intr) TrtB   TrtC   TrtD
## TrtB -0.524
## TrtC -0.770  0.403
## TrtD -0.865  0.453  0.666
## TrtE -0.900  0.471  0.693  0.779
##
## Standardized residuals:
##        Min         Q1        Med         Q3        Max
## -1.3081915 -0.6160591 -0.2923072  0.7641910  1.4889364
##
## Residual standard error: 57.90437
## Degrees of freedom: 20 total; 15 residual
```

```
plot(Model3)
```
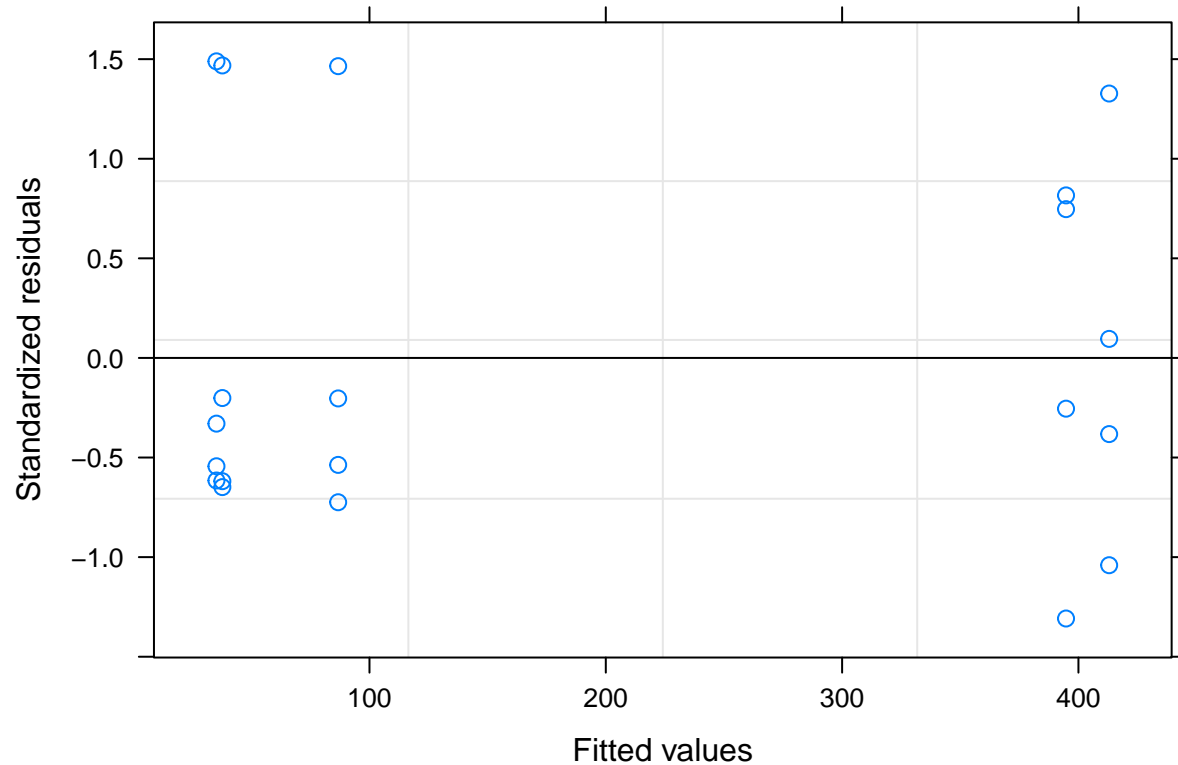
```r
emmeans(Model3, pairwise ~ Trt)
```

```
## $emmeans
##  Trt emmean       SE df   lower.CL   upper.CL
##  A   394.75 28.95218 15 333.039880 456.46012
##  B   413.00 47.08680 15 312.636853 513.36315
##  C    86.75 23.98046 15  35.636859 137.86314
##  D    37.75 16.76989 15   2.005828  73.49417
##  E    35.25 14.02007 15   5.366918  65.13308
##
## Confidence level used: 0.95
##
## $contrasts
##  contrast estimate       SE df t.ratio p.value
##  A - B      -18.25 55.27564 15  -0.330  0.9971
##  A - C      308.00 37.59377 15   8.193  <.0001
##  A - D      357.00 33.45830 15  10.670  <.0001
##  A - E      359.50 32.16817 15  11.176  <.0001
##  B - C      326.25 52.84155 15   6.174  0.0001
##  B - D      375.25 49.98396 15   7.507  <.0001
##  B - E      377.75 49.12972 15   7.689  <.0001
##  C - D       49.00 29.26246 15   1.675  0.4769
##  C - E       51.50 27.77814 15   1.854  0.3813
##  D - E        2.50 21.85845 15   0.114  1.0000
##
## P value adjustment: tukey method for comparing a family of 5 estimates
```