Adjusting Standard ANOVA Methods to Account for Heterogeneous Variances With an Application to Turfgrass Management

Michael Dumelle

November 30, 2020

Section 1

Introduction

Background

- Research statistician at the Environmental Protection Agency
- PhD in Statistics from Oregon State University (2020)
- Met Alec Kowalewski and Clint Mattox through OSU Statistics Consulting Practicum
  - Encourage you to sign up!
  - Long format vs drop-in
  - Faculty are encouraged too - separate process

Background

- Use Analysis of Variance (ANOVA) to study designed experiments
  - Are there *statistically significant differences* among treatment effects?
- One common problem: unequal variance / standard deviation within treatment groups
  - Focus of the talk!

Accessing Slides

- I will interweave R code to illustrate ideas (I will also provide SAS code!)

```
# this is a comment
this_is_some_code <- this_is_a_function(this_is_an_argument)
print(this_is_some_code)
#> [1] "this is output"
```

- Slides and code available on my GitHub here
- Slide numbers in bottom left

Section 2

ANOVA

Experiment Roadmap

1. Formulate a hypothesis
2. Choose an experimental design
3. Choose an analysis method
4. Randomize treatments
5. Collect data
6. Analyze data (focus of the talk)
7. Report results

Experiment Roadmpap

- Often use ANOVA to analyze data
  - Focus on one-way ANOVA with categorical (group) structure
  - $Y_i = \mu + \alpha_i + \epsilon_i$
  - Response = True Mean + Treatment Effect + Random Error
- Several attractive propreties (accurate, precise, p-values reliable) **when assumptions are satisfied**
- One important assumption **constant variance**
  - All ϵ_i have the same variance (standard deviation)
  - standard deviation $= \sqrt{\text{variance}}$
- What about using ANOVA when the data do not have constant variance?

Percent Green Cover Data

Figure 1: Healthy vs. Non-Healthy Turfgrass. Percent green cover is the proportion of healthy turfgrass.

Percent Green Cover Data

- Use simulated data to study analysis methods
  - So helpful because we know the truth!
  - Study several scenarios without having to design an experiment, collect data, etc.

Table 1: Treatment Means, Standard Deviations (StDev), and Replicates

Treatment	Mean	StDev	Replicates
Trt1	50	5.0	8
Trt2	50	2.0	8
Trt3	58	1.0	8
Trt4	60	0.5	8

Percent Green Cover Data

```
set.seed(1130)
data <- create_data(treatments = c("Trt1", "Trt2", "Trt3", "Trt4"),
                    means = c(50, 50, 58, 60),
                    stdevs = c(5, 2, 1, 0.5),
                    replicates = c(8, 8, 8, 8))
head(data, n = 9)
#>   treatments pct_green
#> 1       Trt1  43.98231
#> 2       Trt1  54.94049
#> 3       Trt1  45.64911
#> 4       Trt1  50.33370
#> 5       Trt1  45.03723
#> 6       Trt1  54.45938
#> 7       Trt1  47.62064
#> 8       Trt1  40.34577
#> 9       Trt2  50.07621
```

Visualizing the Data

- Visualization always a good first step – notice the difference in spread!

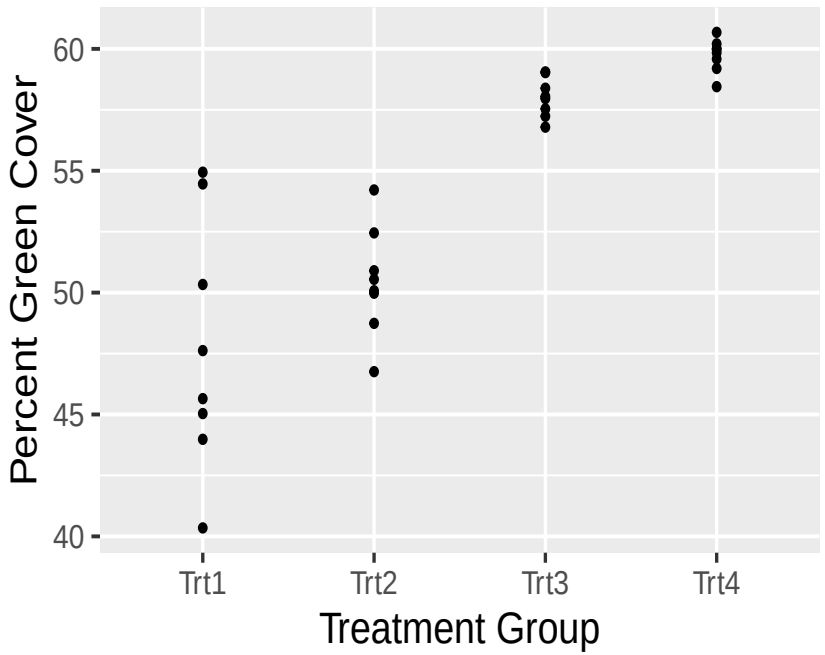


Figure 2: Percent Green Cover by Treatment Group.

ANOVA Code

- Let's perform an ANOVA

```r
# Perform the ANOVA
anova_model <- gls(pct_green ~ treatments, data = data) # nlme package
# Pairwise comparisons among treatments
anova_trtmeans <- emmeans(anova_model, "treatments") # emmeans package
pairs(anova_trtmeans, adjust = "bonferroni") # emmeans package
```

```
# SAS Code
proc mixed data=data;
  class treatments;
  model pct_green = treatments;
  lsmeans treatments / diff adjust=BON;
run;
```

ANOVA Results

Table 2: ANOVA Pairwise Comparison Results. P-values have been adjusted for multiple comparisons using the Bonferroni method.

contrast	estimate	SE	df	t.ratio	p.value
Trt1 - Trt2	-2.660	1.424	28	-1.868	0.434
Trt1 - Trt3	-10.209	1.424	28	-7.166	0.000
Trt1 - Trt4	-11.946	1.424	28	-8.386	0.000
Trt2 - Trt3	-7.548	1.424	28	-5.299	0.000
Trt2 - Trt4	-9.286	1.424	28	-6.519	0.000
Trt3 - Trt4	-1.737	1.424	28	-1.220	1.000

ANOVA Residuals

- Use residuals (normalized) to check assumptions!
  - Regular residuals divided by their standard deviations

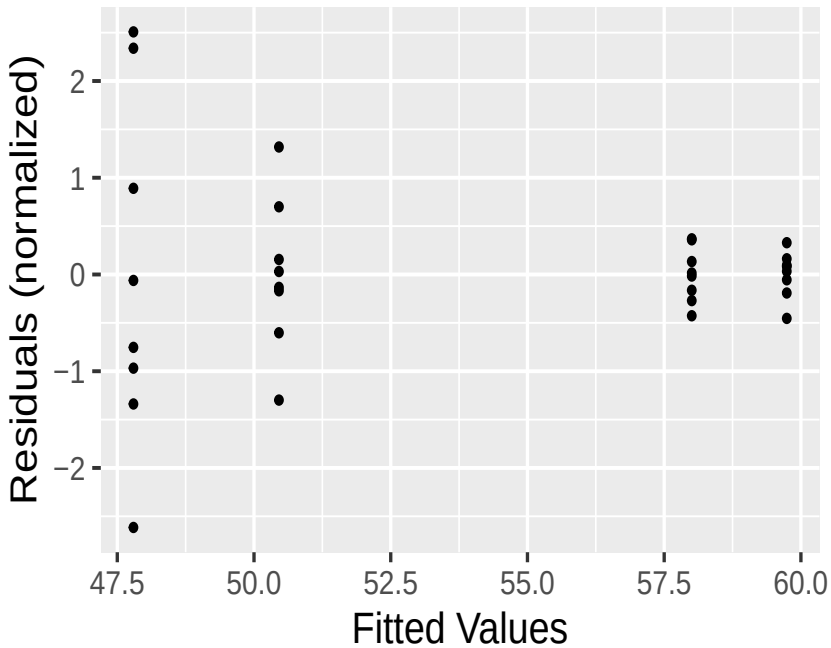
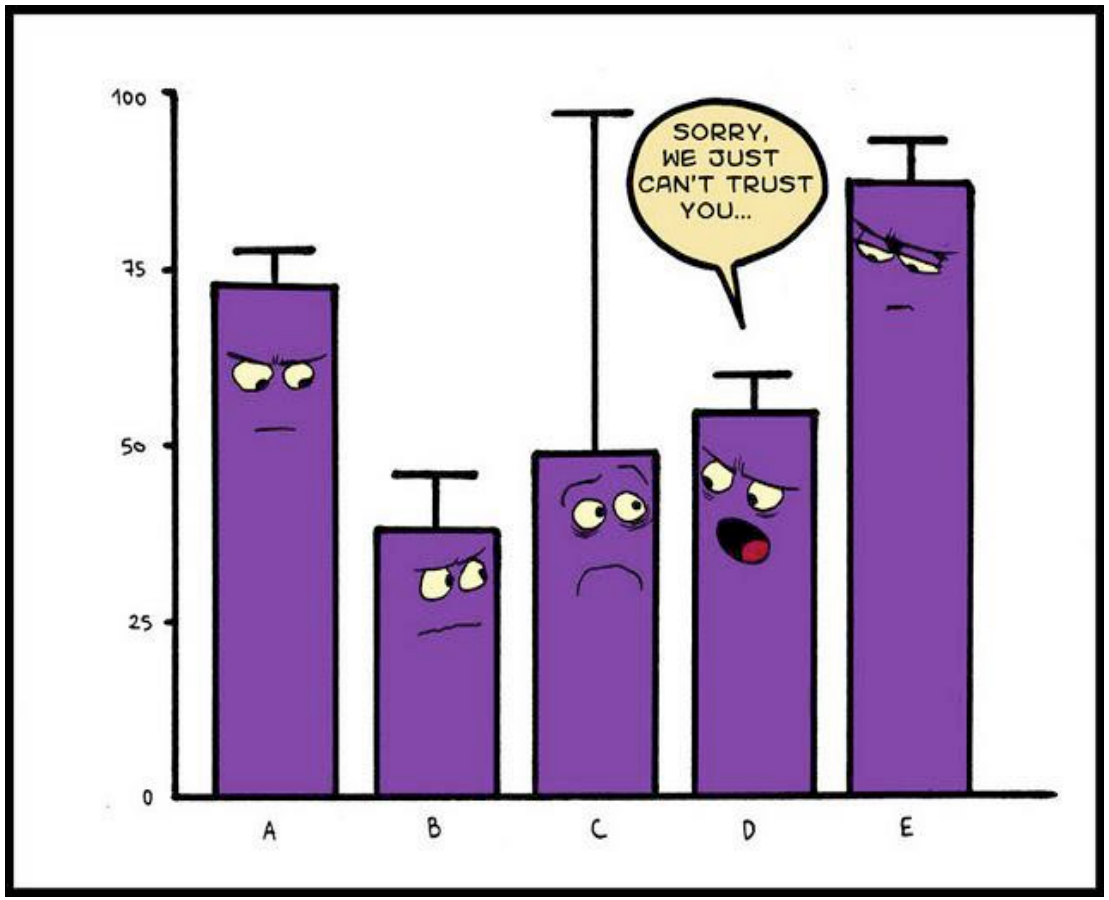


Figure 3: Fitted Values vs Normalized residuals Using ANOVA for Percent Green Cover

- That variance does NOT look constant!

Can We Trust Our Analysis?



Section 3

Warning Signs?

Warning Signs?

- Were there any warning signs?
  - YES
- What were they?

1. Graphics
2. Ratio of largest variance and smallest variance
3. Statistical hypothesis tests for constant variance

Graphics

- It it looks off, it probabily is!
- Similar to the spread we saw in data (should also visualize data!)

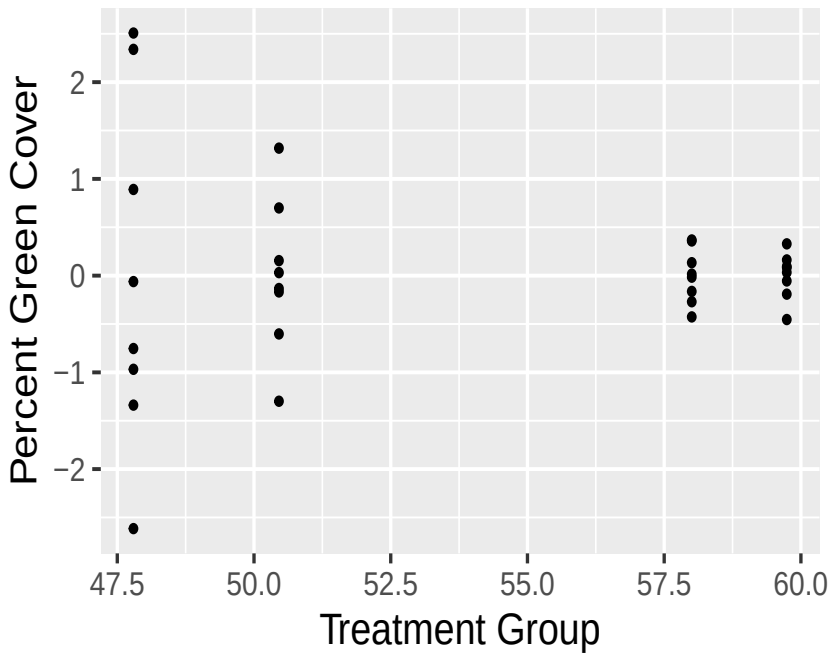


Figure 4: Fitted Values vs Normalized Residuals Using ANOVA for Percent Green Cover

Variance Ratios

- Rule of thumb: ANOVA problems when variance ratios larger than 1.5 to 9 (seen cutoff suggestions within this range)
  - Standard deviation range of 1.22 to 3

```
(trt_stdevs <- data %>%
  group_by(treatments) %>%
  summarize(grp_stdev = sd(pct_green)))
#> # A tibble: 4 x 2
#>   treatments grp_stdev
#>   <fct>           <dbl>
#> 1 Trt1            5.13
#> 2 Trt2            2.25
#> 3 Trt3            0.809
#> 4 Trt4            0.678
(stdev_ratio <- max(trt_stdevs$grp_stdev) /
    min(trt_stdevs$grp_stdev))
#> [1] 7.563561
```

Statistical Tests for Constant Varince

- Hypothesis test constant variance assumption is questioned
  - Low p-value $\rightarrow$ evidence the variances are NOT equal
  - Levene's test, Brown-Forsythe test are two examples – there are many others
  - Come with their own assumptions

```
leveneTest(pct_green ~ treatments,
           center = "mean", data = data) # car package
#> Levene's Test for Homogeneity of Variance (center = "mean")
#>       Df F value    Pr(>F)
#> group  3  9.0112 0.0002445 ***
#>       28
#> ---
#> Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

What Now?

- We know constant variance assumption is invalid – what now?
- We could transform the response variable
  - Hope the transformed data has constant variance
- This approach can be very useful!
- But there are some drawbacks!

What Now?

1. Generally require relationship between mean and variance to be successful
   - Example: Log transformations successful when mean increases $\rightarrow$ variance increases
2. Analysis on transformed scale – NOT original scale
   - Statistically significant difference on transformed scale does not necessarily imply a statistically significant difference on the original scale

Transformations

- Most common is the $\log_e(Y)$ transformation – lets hope this works!

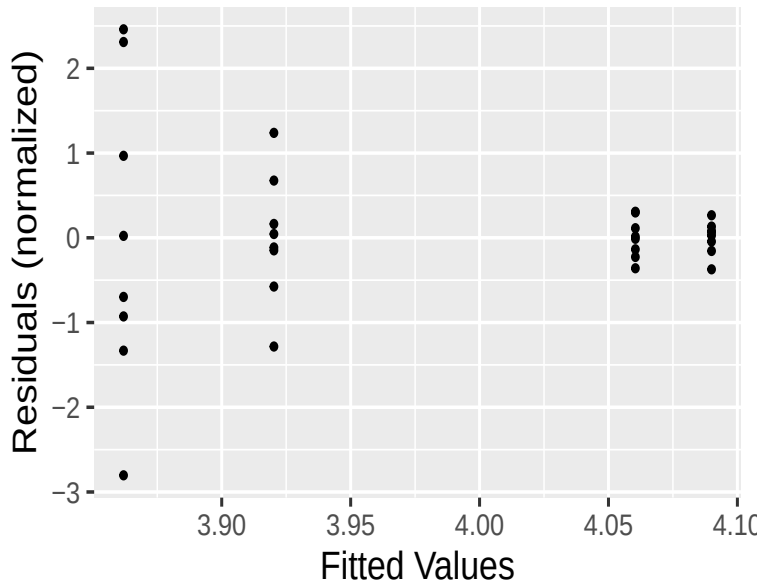


Figure 5: Fitted Values vs Normalized Residuals Using ANOVA for Log_e Percent Green Cover

- Square root, cube root, reciprocal transformations don't work either – we need another approach!

Section 4

GVANOVA

What is GV-ANOVA?

- Can use Generalized Variance ANOVA (GV-ANOVA) to directly model variances within groups
  - Separate variance for each treatment level
  - No mean / variance relationship required
  - Analysis on original scale
  - More variance parameters require estimation
- Goal of this talk is to introduce an alternative approach to transformations
  - Important to be aware of both – transformations are still a useful tool in the toolbox!

GVANOVA Code

- Let's perform an GVANOVA

```
# Perform the GVANOVA
gvanova_mod <- gls(pct_green ~ treatments,
                   weights = varIdent(form = ~ 1 | treatments),
                   data = data) # nlme package
# Pairwise comparisons among treatments
gvanova_trtmeans <- emmeans(gvanova_mod, "treatments") # emmeans package
pairs(gvanova_trtmeans, adjust = "bonferroni") #emmeans package
```

```
# SAS Code
proc mixed data=data;
  class treatments;
  model pct_green = treatments / ddfm=SAT; # this is different
  repeated / group = treatments; # this is different
  lsmeans treatments / diff adjust=BON;
run;
```

GVANOVA Analysis

Table 3: GVANOVA Pairwise Comparison Output. P-values have been adjusted for multiple comparisons using the Bonferroni method.

contrast	estimate	SE	df	t.ratio	p.value
Trt1 - Trt2	-2.660	1.980	9.589	-1.344	1.000
Trt1 - Trt3	-10.209	1.836	7.349	-5.560	0.004
Trt1 - Trt4	-11.946	1.829	7.246	-6.530	0.002
Trt2 - Trt3	-7.548	0.844	8.784	-8.947	0.000
Trt2 - Trt4	-9.286	0.829	8.263	-11.199	0.000
Trt3 - Trt4	-1.737	0.373	13.586	-4.655	0.002

ANOVA vs GVANOVA

Table 4: Standard Errors and P-values of ANOVA (*.a) and GVANOVA (*.gva)

contrast	SE.a	SE.gva	p.a	p.gva	true.diff
Trt1 - Trt2	1.424	1.980	0.434	1.000	0
Trt1 - Trt3	1.424	1.836	0.000	0.004	-8
Trt1 - Trt4	1.424	1.829	0.000	0.002	-10
Trt2 - Trt3	1.424	0.844	0.000	0.000	-8
Trt2 - Trt4	1.424	0.829	0.000	0.000	-10
Trt3 - Trt4	1.424	0.373	1.000	0.002	-2

- More uncertainty refleted in `Trt1 - Trt2`
- Less uncertainty reflected in `Trt3 - Trt4`

GVANOVA Residuals

- Use residuals (normalized) to check assumptions!
- Even spread yields evidence the GVANOVA assumptions are satisfied

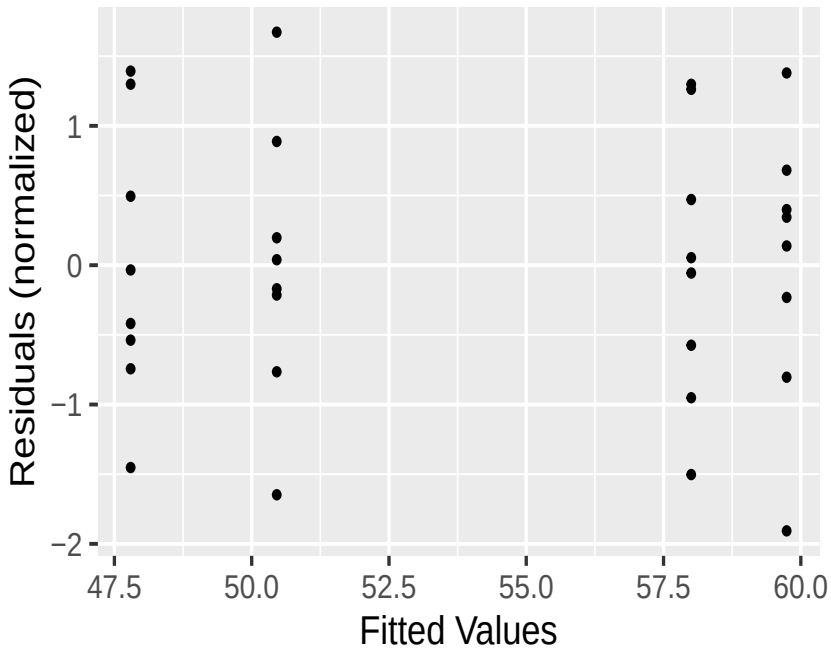


Figure 6: Fitted Values vs Normalized Residuals Using GVANOVA for Percent Green Cover

Section 5

Takeaways

Takeaways

- ANOVA is the best tool we have when assumptions are satisfied
- Constant variance assumption should not be overlooked
  - Remember the warning signs!
- Two approaches: transformations and GVANOVA
- When true variance is not constant, using an analysis approach accomodating this will generally yield **a more accurate representation of the truth**