Modeling Unequal Variance

Outline:

1. Unequal variance in an ANOVA setting
2. Weighted Least Squares Regression

Examples:

1. Poppies: One way ANOVA with unequal variance
2. Biomass: Three way example
3. Weighted Least Squares

Introduction

We have already discussed using transformations to satisfy the assumption of equal variance.

In this group of notes, we will talk about additional model options that allow for unequal variances.

1. Unequal Variances in an ANOVA setting

We can use the gls() function from the nlme package with the varIdent option to allow the variance to be estimated separately by group.

Warning: 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.

Poppies Example: One Way ANOVA with Unequal Variance We return to the poppy data that we used in STAT511. There are 5 treatments with 4 reps per treatment.

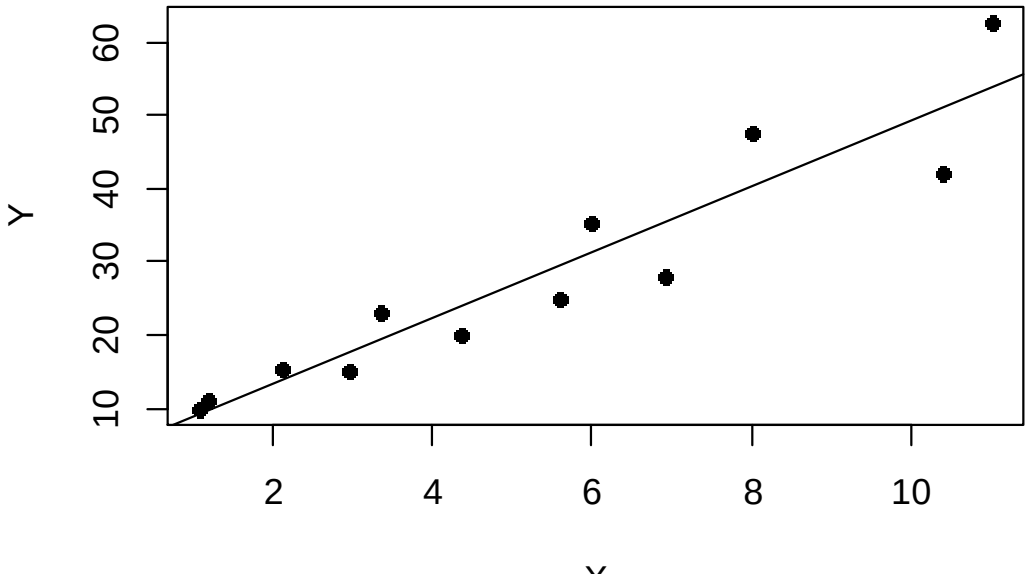
We consider 3 different approaches:

1. One-way ANOVA on the original scale: Residual plot shows obvious megaphone shape. Assumption of equal variance NOT satisfied.
2. One-way ANOVA after square root transformation: Residual plot looks good.
3. One-way ANOVA allowing unequal variances: Notice that plot() function returns a plot of the standardized (not raw) residuals. The residual plot looks good.

In the previous one-way example, we allowed a different variance for each trt group (resulting in 5 means and 5 variances).

In the **Three-way Biomass** example, there are 2 Types (grass, forb), 3 Herbicide treatments (A,B,C) and 2 Water regimes (1, 2) for a total of 2x3x2=12 treatment combinations. Based on the summary statistics and residual plot, it appears that the variance is different for grass versus forbs. We can account for this using varIdent(form = ~1|Type) which gives us 2 different variance components.

2. Weighted Least Squares (WLS)



Weighted least squares (WLS) is a methodology that is used when there is evidence of non-constant variance. It is an alternative to transformation.

In the example plot above, transforming Y might correct non-constant variance, but would ruin an otherwise linear relationship.

Idea of WLS: Do least squares, but decrease the influence of points that have high variance, and increase the influence of points that have low variance.

Select β's to minimize:

$$\sum_{i=1}^{n} w_i \left(y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i) \right)^2 = \sum_{\text{data}} w_i \left(\hat{\varepsilon}_i \right)^2 = \text{SSE}$$

Assume the model: $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$, where $\text{var}(\varepsilon_i) = \sigma_i^2$

For WLS have (ideally): $w_i = \dfrac{1}{\text{var}(\varepsilon_i)}$

In practice, we rarely know σ_i^2, so we can't use the ideal weights. Usually, we assume that the σ_i^2 are related to X in some way:

$$\text{var}(\varepsilon_i) = \sigma^2 \, x_i^{\,r} \quad \text{where r usually is 1,2 or -1.}$$

WLS Example (from Neter and Wasserman):

n = 12 observations relate preparation cost (Y) of a bid to the size of the bid (X). Based on the residual diagnostic plot, there is evidence of increasing variance with X. They are very reluctant to transform to log scale, so we do a WLS.

We try two different weights: (1) 1/X (r = 1) and (2) 1/X^2 (r = 2).

Note that we consider plots of the standardized residuals (not the raw residuals). The first weighted model still shows unequal variance. But the second weighted model looks good!

The WLS regression gives slightly different parameter estimates, because more variable points have not been weighted as heavily as less variable points.

Alternatives to WLS:

1. Iteratively reweighted least squares (IRWLS). This can be used with multiple regression or in simple linear regression when we don't have an *a priori* idea of the appropriate weighting.
2. Robust regression methods (ex: Huber's method).