Stat 608 Chapter 4

Weighted Least Squares

In Chapter 3, we saw that it is sometimes possible to overcome nonconstant error variance by transforming Y and/or X . In this chapter we consider an alternative way of coping with nonconstant error variance, namely weighted least squares (WLS).

Weighted Least Squares Model

Consider the straight line regression model

$$Y_i = \beta_0 + \beta_1 x_i + e_i$$

where the e_i have mean 0 but variance σ^2 / w_i.

w_i large	w_i small
Variance of errors close to 0.	Variance of errors large.
Points close to the line.	Points far from line.
Parameter estimates such that the fitted line at x_i close to y_i. The i^{th} case is taken more into account.	Parameter estimates don't take the values (xi, yi) into account as much.

Weighted Least Squares Model

What does the covariance matrix for the errors look like?

Derivation

To take into account the weights when we estimate the regression parameters, we consider the following weighted version of the residual sum of squares:

$$WRSS = \sum_{i=1}^{n} w_i (y_i - \hat{y}_{w_i})^2 = \sum_{i=1}^{n} w_i (y_i - b_0 - b_1 x_i)^2$$

$$WRSS = (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})' \mathbf{W} (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})$$

We minimize WRSS with respect to the parameters β or b_0 and b_1.

Derivation

$$\frac{\partial WRSS}{\partial \beta} = \frac{\partial (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'\mathbf{W}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}$$

$$= \frac{\partial \mathbf{Y}'\mathbf{W}\mathbf{Y} - \mathbf{2}\mathbf{Y}'\mathbf{W}\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{W}\mathbf{X}\boldsymbol{\beta}}{\partial \boldsymbol{\beta}}$$

$$= -2\mathbf{X}'\mathbf{W}\mathbf{Y} + \mathbf{2}\mathbf{X}'\mathbf{W}\mathbf{X}\boldsymbol{\beta}$$

Then setting the derivative equal to 0, we find the minimum:

$$\frac{\partial WRSS}{\partial \beta} := 0$$

$$-2\mathbf{X}'\mathbf{W}\mathbf{Y} + \mathbf{2}\mathbf{X}'\mathbf{W}\mathbf{X}\hat{\boldsymbol{\beta}} = 0$$

$$\mathbf{X}'\mathbf{W}\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{W}\mathbf{Y}$$

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\mathbf{Y}$$

Special Case: Multiple measurements at each value of x

This image cannot currently be displayed.

Special Case: Multiple measurements at each value of x

Consider the simple linear regression model

$$Y_i = \beta_0 + \beta_1 x_i + e_i$$

where the e_i have mean 0 but variance σ^2/w_i. In this case we take

$$w_i = \frac{1}{\sigma^2_{Y_i}}$$

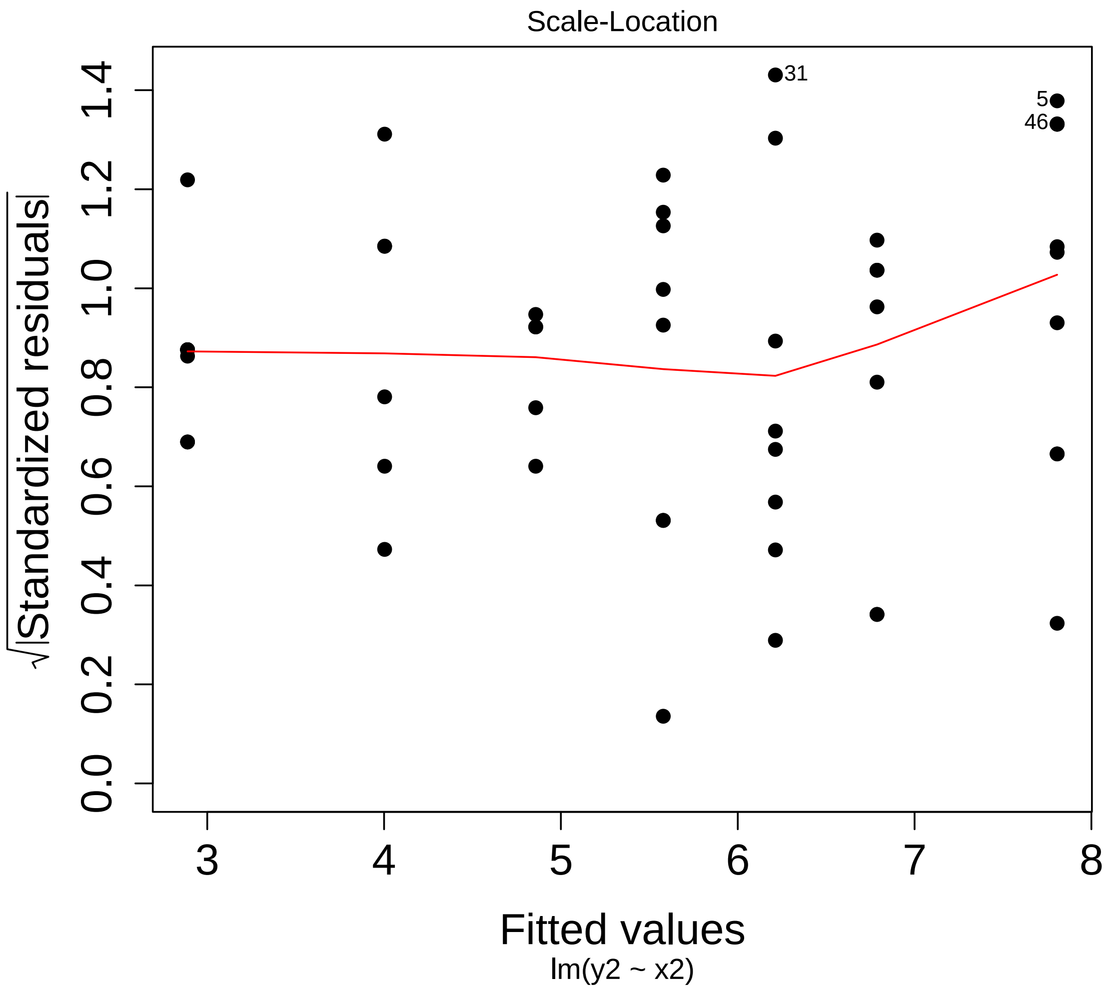
so that the Y_i (given x_i) have variance 1.

In the special case of the cleaning data set, we were able to estimate the variances of Y at each value of x_i because we had multiple measurements at each x_i.

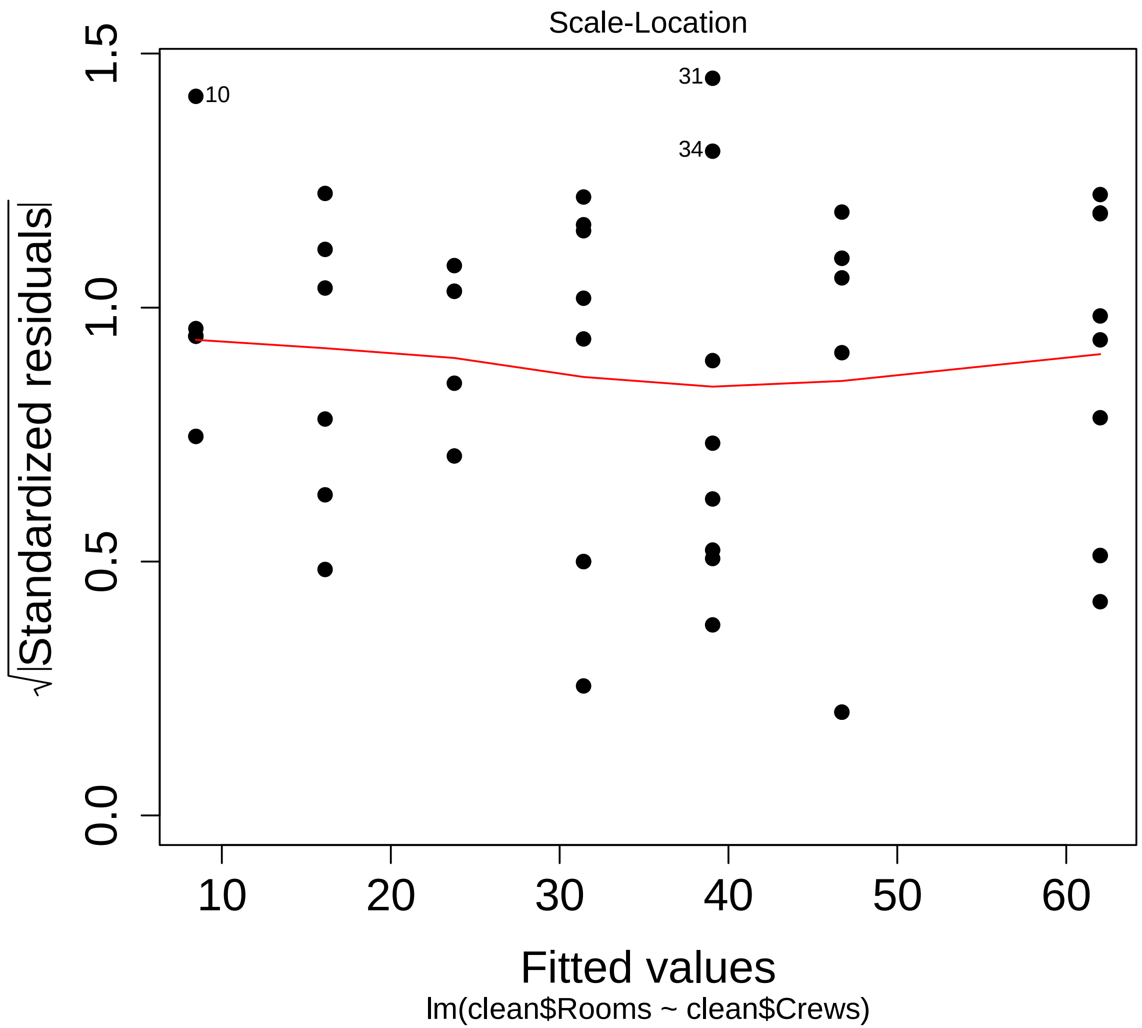
Cleaning Example

Number of Crews	n	Standard deviation (Y_i)
2	9	3.00
4	6	4.97
6	5	4.69
8	8	6.64
10	8	7.93
12	7	7.29
16	10	12.00

Cleaning Example: Old model with square root transformations



Cleaning Example: New weighted model, no transformations



Case 2: X continuous

- In general, we don’t have multiple measurements at every value of X.
- If we have only one observed response at each value of X, how many parameters do we need to estimate using the previous method?

Case 2: X continuous

Calculating weighted least squares:

Consider the least squares regression model:

$$Y_i = \beta_0 + \beta_1 x_i + e_i$$

where the ei have mean 0 but variance σ^2/w_i. If we multiply both sides of the previous equation by $\sqrt{w_i}$ we get:

$$\sqrt{w_i} Y_i = \beta_0 \sqrt{w_i} + \beta_1 \sqrt{w_i} x_i + \sqrt{w_i} e_i$$

where the $\sqrt{w_i}$ ei have mean 0 but variance:

Case 2: X continuous

- We can calculate the least squares fit of the first model by calculating the least squares fit to the second, which is a multiple linear regression model with two predictors and no intercept. Use the following substitutions:

$$Y_{NEWi} = \sqrt{w_i} Y_i \quad x_{1NEWi} = \sqrt{w_i} \quad x_{2NEWi} = \sqrt{w_i} x_i \quad e_{NEWi} = \sqrt{w_i} e_i$$

Then we can rewrite the second model as:

$$Y_{NEWi} = \beta_0 x_{1NEWi} + \beta_1 x_{2NEWi} + e_{NEWi}$$

Case 2: X continuous

In other words, we can rewrite:

Another Parameterization:

Example: Variance Proportional to X

Suppose $\sigma_i^2 \propto x_i$ That is, $Var(e_i) = \sigma^2 x_i$.

Then we could take: $w_i = \frac{1}{x_i}$

$$
\begin{aligned}
Var(\sqrt{w_i e_i}) &= Var\left(\frac{1}{\sqrt{x_i}} e_i\right) \\
&= \frac{1}{x_i} x_i \sigma^2
\end{aligned}
$$

The Use of Weighted Least Squares

- The weighted least squares technique is most commonly used in the special case when Y_i is the average or the median of n_i observations so that:

$$Var(Y_i) \propto \frac{1}{n_i}$$

In this case we take $w_i = n_i$.
Ex: Y_i = GDP, X_i = #children per adult female

- Warning: In many situations, the variance is not constant, and it is not straightforward to determine the correct model for the variance.

- Many people use transformations in these situations.