Stat 608 Chapter 4

2/14/22 – R. Cleaning / R data-transform

2/15/22 - R. cleaning (covered down to chap4)

2/16 - went over HW # 3, 2/18 went over practice MT

STARTED: 2/23/22 Wednesday (Week 6, Lecture 15)

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).

The Use of Weighted Least Squares

Y_i is the average or the median of n_i observations	Y_i is the sum of n_i observations
• $w_i = n_i$ • Ex: Y_i = Per capita spending on education, Crimes per 100,000 people, average income $$\text{Var}(Y_i) \propto \frac{1}{n_i}$$	• $w_i = 1/n_i$ • Ex: Y_i = GDP, Total spending per department $$\text{Var}(Y_i) \propto n_i$$

- These are the most common uses of WLS. One more: $var(y_i) \propto x_i \Rightarrow w_i = \frac{1}{x_i}$
- The weights w_i must be known.
- WLS is sensitive to outliers; it is not robust at all. * ⇒ if outliers, better to do transformations
- 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.

Examples: Variance Proportional to n and x

Suppose Y_i = Per Capita GDP. Then $Var(y_i) \propto \frac{1}{n_i}$

Then we could take:

$$Var(\sqrt{n_i}\, e_i) = n_i Var(e_i) = n_i \frac{\sigma^2}{n_i} = \sigma^2 \text{ (constant)}$$

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

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

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?

$\text{var}(\underline{e}) = \Sigma, \quad \text{var}(e_i) = \frac{\sigma^2}{w_i}$

$$\Sigma = \sigma^2 \begin{bmatrix} 1/w_1 & & & 0 \\ & \ddots & & \\ & & \ddots & \\ 0 & & & 1/w_n \end{bmatrix}$$

$$W = \begin{bmatrix} w_1 & & & 0 \\ & \ddots & & \\ & & \ddots & \\ 0 & & & w_n \end{bmatrix}$$

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:

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

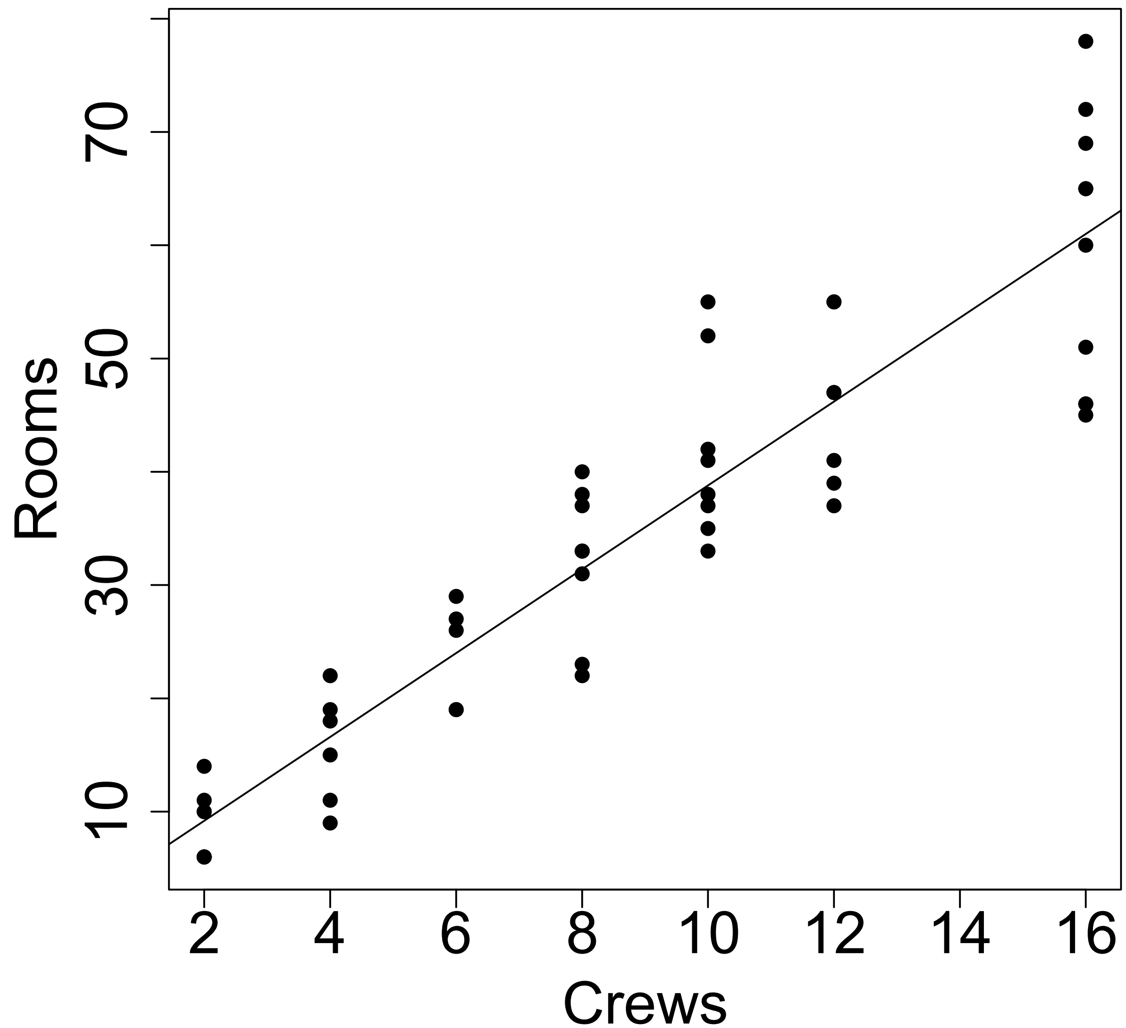
$$-2\mathbf{X}'\mathbf{W}\mathbf{Y} + \mathbf{2}\mathbf{X}'\mathbf{W}\mathbf{X}\hat{\boldsymbol{\beta}} = 0$$

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

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

Special Case: Multiple measurements at each value of x

- really straightforward to do weighted regression.



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

$$w_i = \frac{1}{\sigma^2_{Y_i}} \qquad \sigma^2_{Y_i} = \mathrm{var}(Y \mid X = x_i)$$

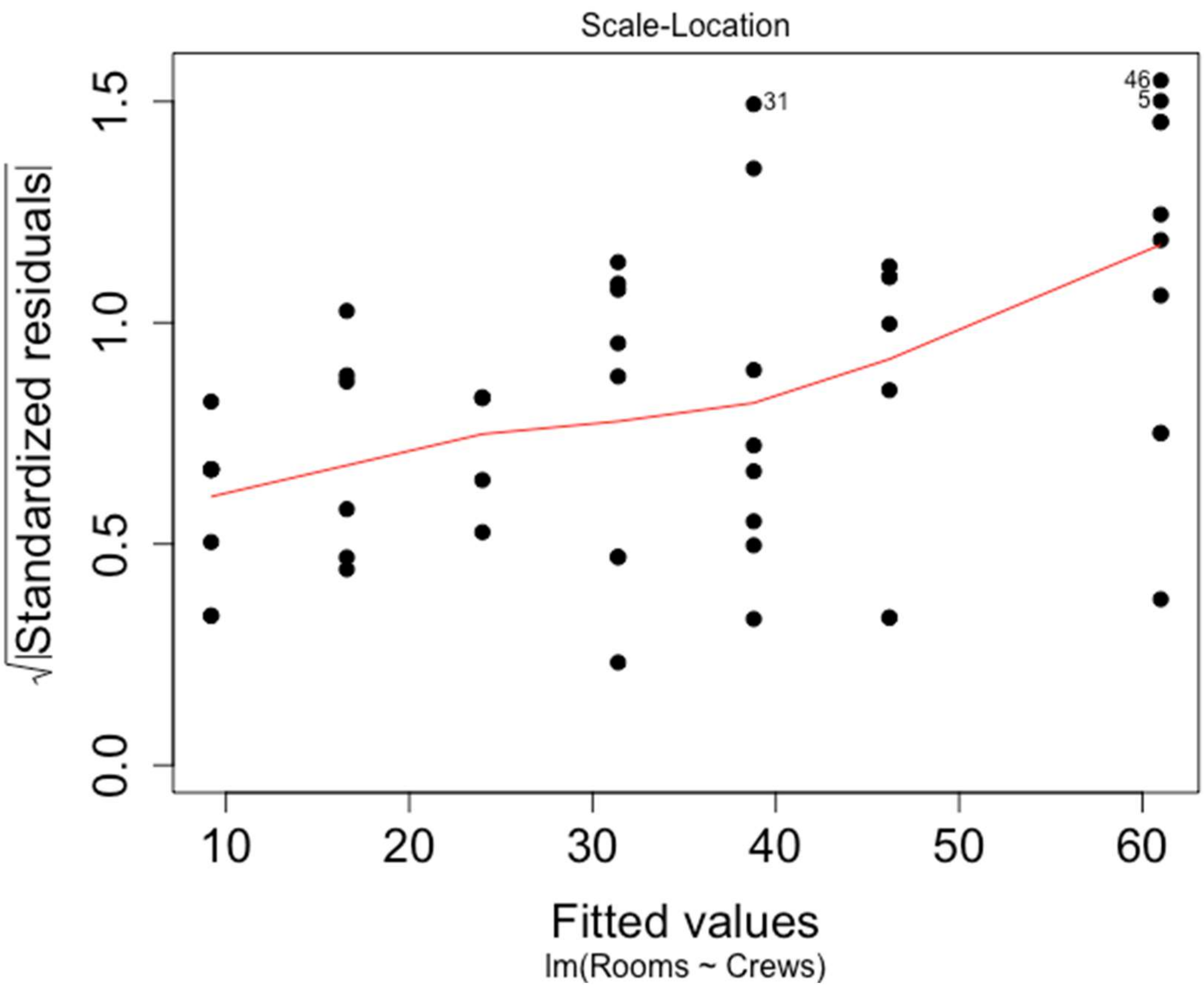
so that the Y_i (given x_i) have variance 1.

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.

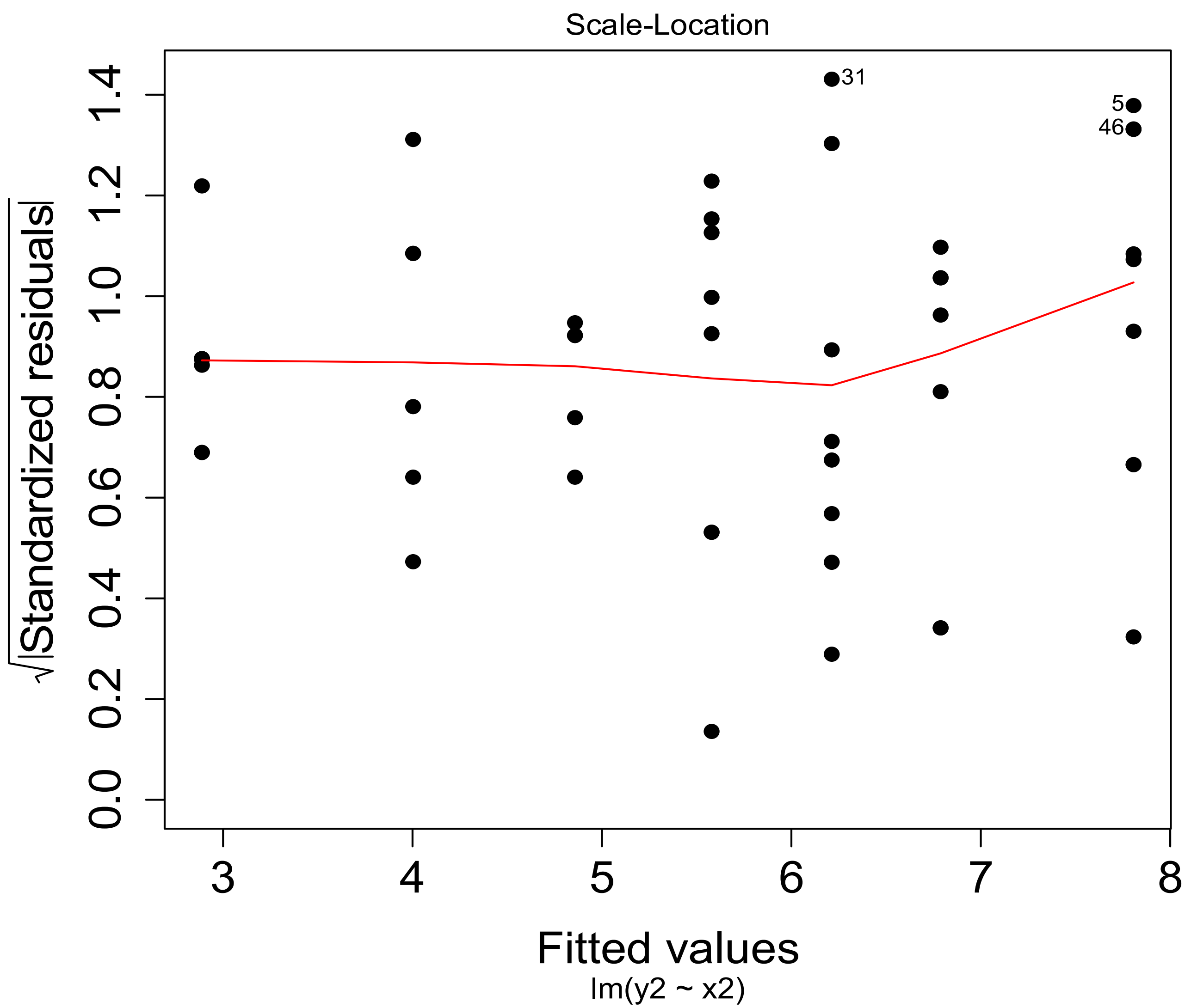
Cleaning Example

Number of Crews	n	Standard deviation
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

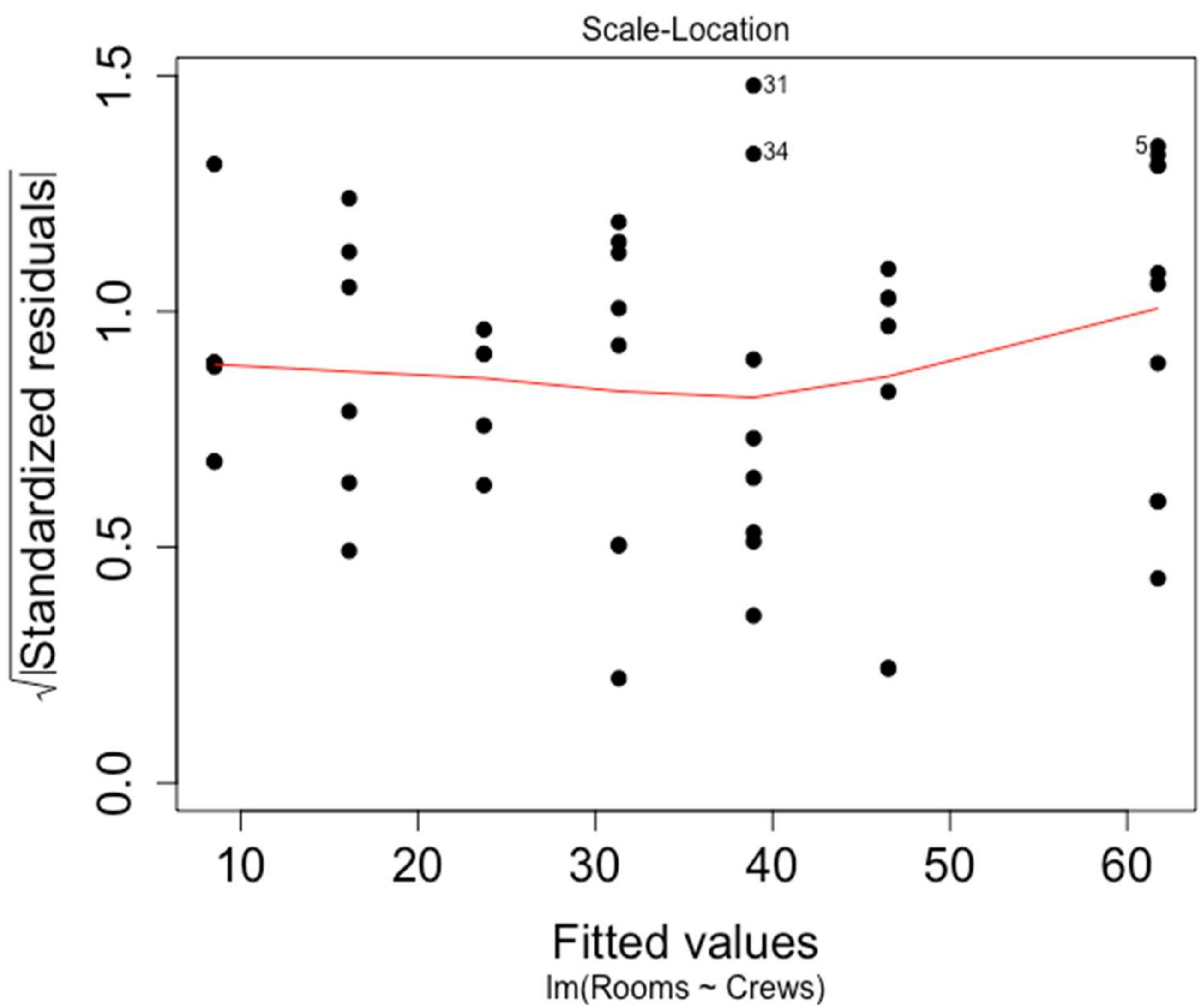
Cleaning Example: Old SLR Model



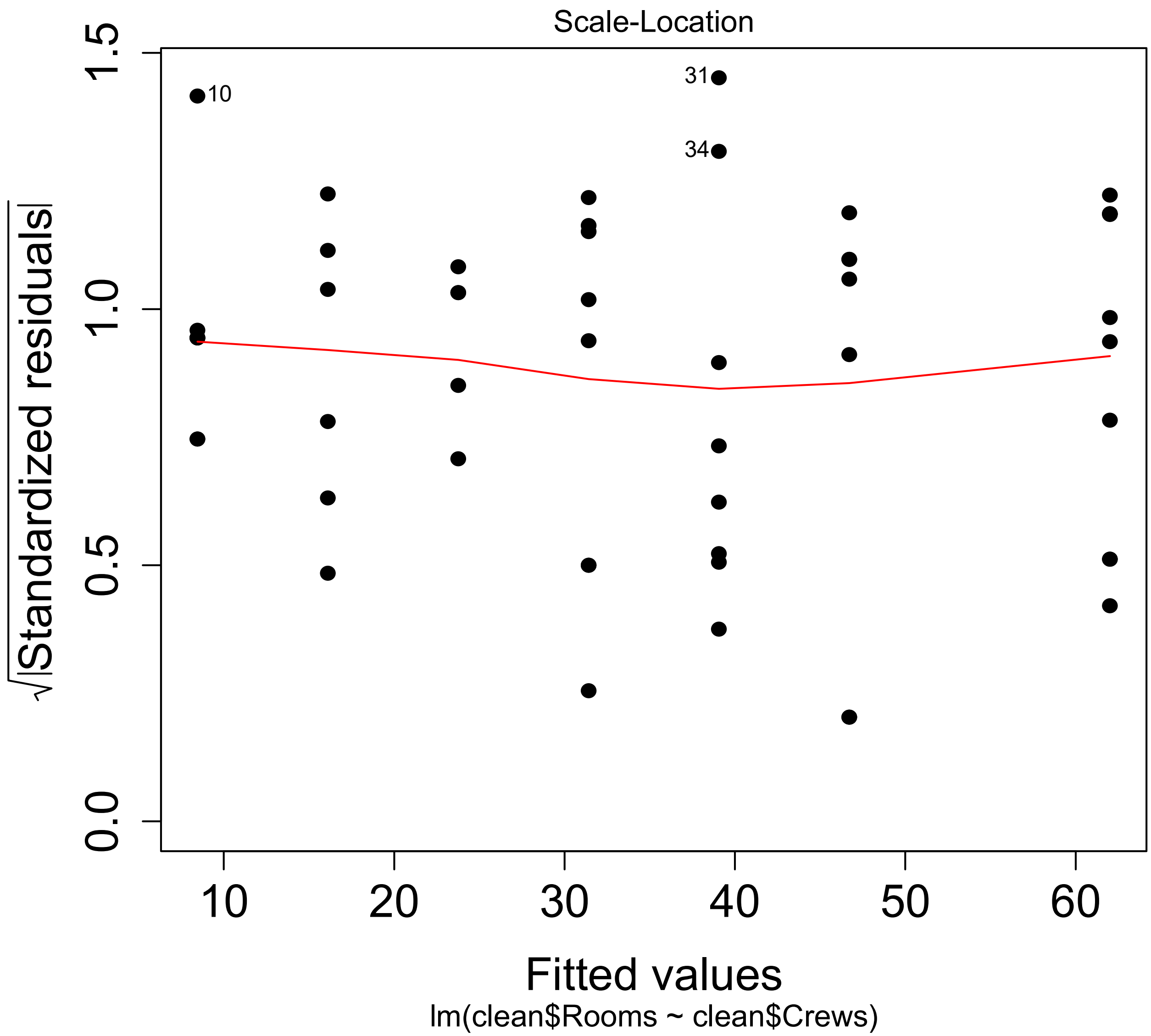
Cleaning Example: Old model with square root transformations



Cleaning Example: New weighted model, variance prop. to #Crews



Cleaning Example: New weighted model, estimated variance



Case 1: Multiple measurements at each value of x

- Criticism: For k unique values of X we have to estimate k unique variances $\rightarrow \sigma_1^2 \ldots \sigma_k^2$. We don't take this extra variability into account when computing $var(\hat{\beta}_{WLS})$. ⇒ p-values are too small & CIs too narrow. i.e. there's more variability in our estimation than we're accounting for.
- 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?

↳ n parameters for n variances, but we need at least 2 observations at each X to estimate a variance.

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:

$$var(\sqrt{w_i}\, e_i) = w_i \frac{\sigma^2}{w_i} = \sigma^2$$

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:

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

Then we can rewrite the second model as:

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

$$\begin{bmatrix} \sqrt{w_i}\, y_i \\ \vdots \\ \sqrt{w_n}\, y_n \end{bmatrix} = \begin{bmatrix} \sqrt{w_1} & \sqrt{w_1}\, x_1 \\ \vdots & \vdots \\ \sqrt{w_n} & \sqrt{w_n}\, x_n \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \begin{bmatrix} \sqrt{w_i}\, e_i \\ \vdots \\ \sqrt{w_n}\, e_n \end{bmatrix}$$

Case 2: X continuous

In other words, we can rewrite:

$$y_{new} = W^{1/2} y, \quad e_{new} = W^{1/2} \underline{e}, \quad X_{new} = W^{1/2} X$$

$$\hat{\beta}_{WLS} = (X'_{new} X_{new})^{-1} X'_{new} Y_{new}$$

$$= (X' W^{1/2\prime} W^{1/2} X)^{-1} X W^{1/2\prime} W^{1/2} y$$

$$= (X' W X)^{-1} X' W y$$

Another Parameterization:

$$\bar{x}_w = \frac{\sum w_i x_i}{\sum w_i} \quad , \quad \bar{y}_w = \frac{\sum w_i y_i}{\sum w_i}$$

$$\hat{\beta}_0 = \bar{y}_w - \hat{\beta}_{1,w} \bar{x}_w \quad ; \quad \hat{\beta}_{1w} = \frac{\sum w_i (x_i - \bar{x}_w)(y_i - \bar{y}_w)}{\sum w_i (x_i - \bar{x}_w)^2}$$

$$var(\underline{e}) = \sigma^2 W^{-1} \quad ; \quad W^{-1} = (W^{-1/2})(W^{-1/2}) = (W^{-1})^{1/2} (W^{-1})^{1/2}$$

$$var(\underline{e}_{new}) = var(W^{1/2} \underline{e}) = W^{1/2} \, var(\underline{e}) \, W^{1/2}$$

$$= \sigma^2 W^{1/2} W^{-1} W^{1/2} = \sigma^2 \underline{I}$$

Another Parameterization:

$$E[\hat{\beta}_{WLS} \mid X, W] = E[(X'WX)^{-1}X'Wy \mid X, W]:$$

$$= (X'WX)^{-1}X'W\,E[y \mid X, W] = (X'WX)^{-1}X'WX\beta$$

$$= \beta$$

$$Var(\hat{\beta}_{WLS} \mid X, W) = (X'WX)^{-1}X'W\,Var(Y \mid X, W)\,W'X(X'WX)^{-1}$$

$$Var(y \mid X, W) = \sigma^2 W^{-1}$$

$$= \sigma^2 (X'WX)^{-1} X'W W^{-1} W^{-1} X (X'WX)^{-1}$$

$$= \sigma^2 (X'WX)^{-1}$$

R^2 Rant, continued

- Beware WLS is not at all robust to outliers, so again, R^2 is strongly affected by outliers.

- For models that are improved by weighting, it could be that:

  R^2 is lower for the weighted model.

- For models that are not improved by weighting, it could be that:

  R^2 is higher for the weighted model.

"Sole reliance on [the coefficient of determination, R2] may fail to reveal important data characteristics and model inadequacies."[1]

1. (Kvålseth, T.O. (1985), "Cautionary Note about R2," The American Statistician, 39, 279-285.)

Finished 2/23/22 (Week 6, Lecture 15)