NCSU ST 503 Discussion 8

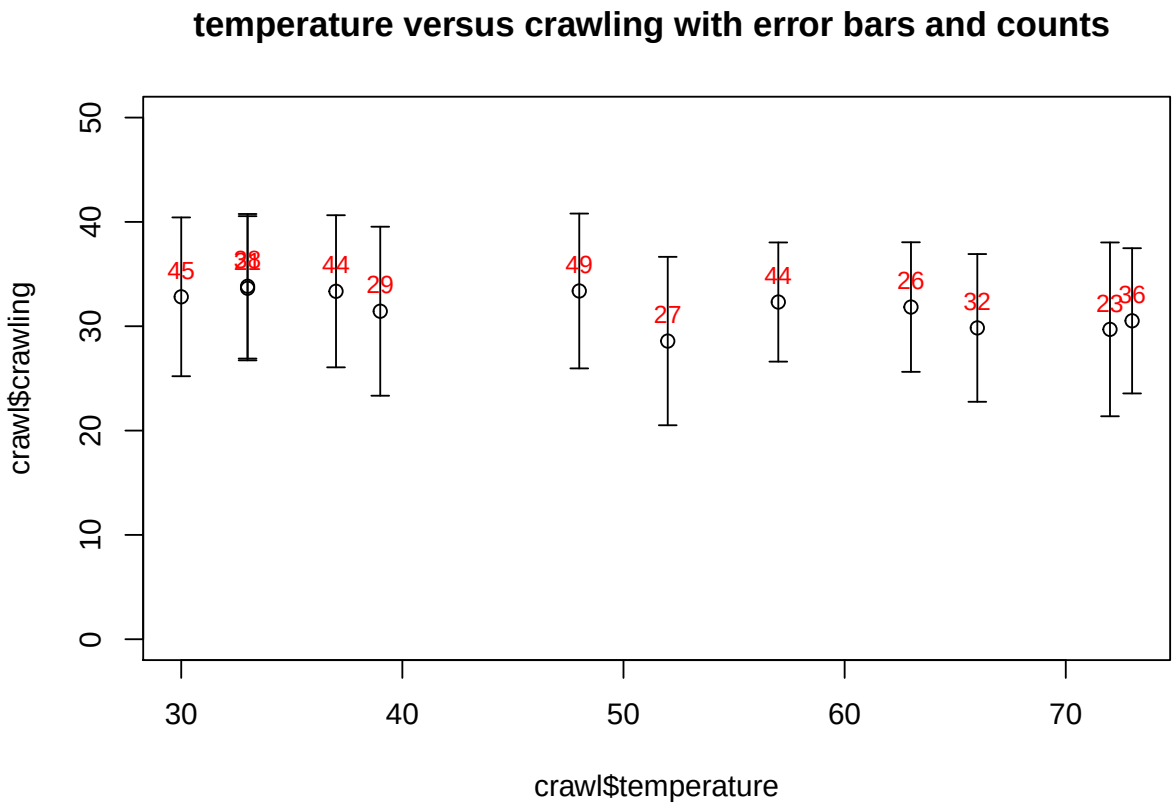
Probem 8.7 Faraway, Julian J. Linear Models with R CRC Press.

Bruce Campbell

8.7 crawl data analysis

The crawl dataset contains data on a study looking at the age when babies learn to crawl as a function of ambient temperatures. There is additional information about the number of babies studied each month and the variation in the response. Make an appropriate choice of weights to investigate the relationship between crawling age and temperature.

First we plot the dat along with the error and count information



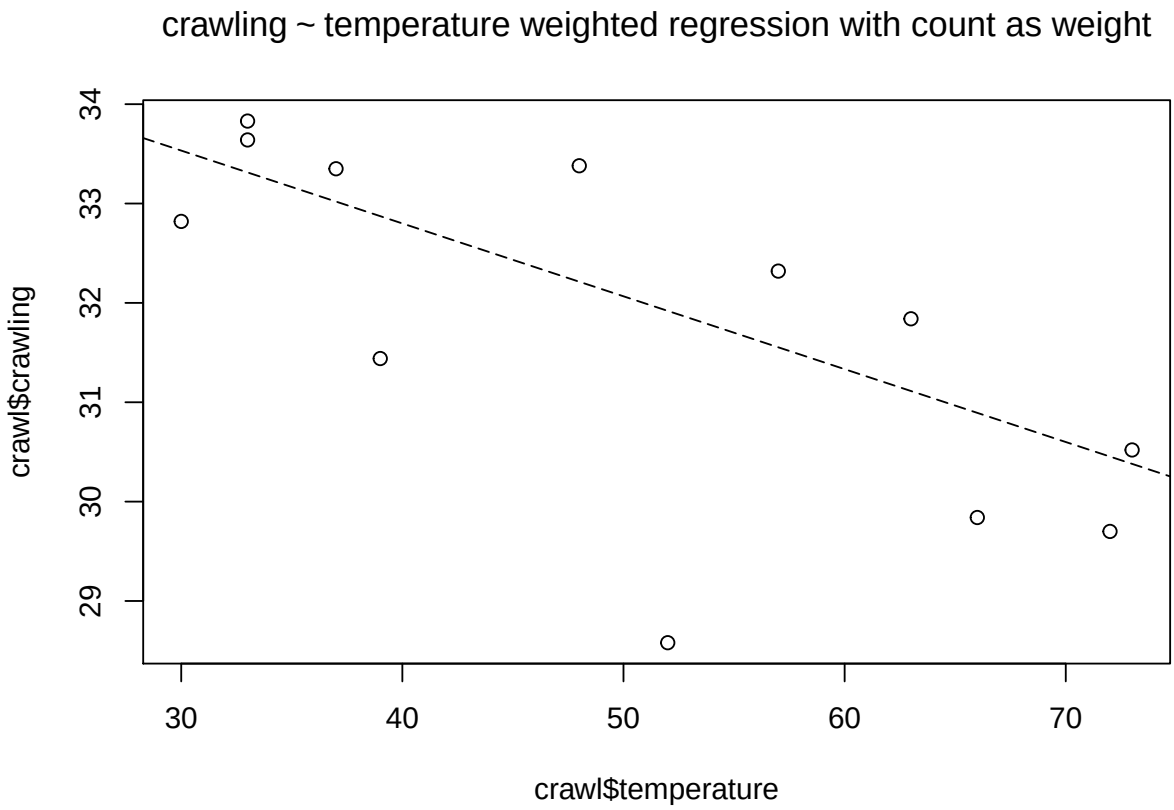
We fit a weighted least squares model with `lm` using weights $w_i = \frac{n_i}{SD_i^2}$.

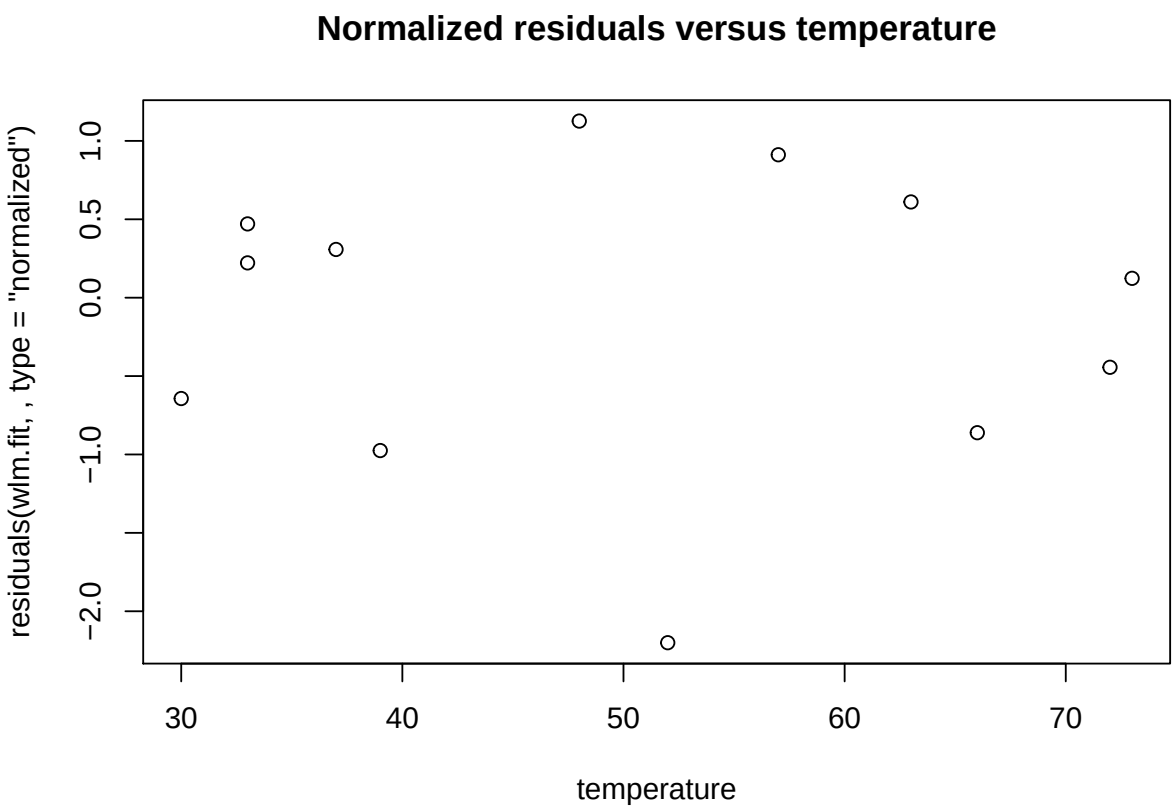
```
##
```

```
## Call:
## lm(formula = crawling ~ temperature, data = crawl, weights = wts)
##
## Weighted Residuals:
##     Min      1Q  Median      3Q     Max
## -2.1504 -0.6817  0.1688  0.4941  1.1009
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 35.73262    1.21153   29.49 4.69e-11 ***
## temperature -0.07332    0.02328   -3.15   0.0103 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9772 on 10 degrees of freedom
## Multiple R-squared:  0.4981, Adjusted R-squared:  0.4479
## F-statistic: 9.923 on 1 and 10 DF,  p-value: 0.01033
```

We fit a weighted least squares model with gls using weights $w_i = \frac{n_i}{SD_i^2}$.

```
## Generalized least squares fit by REML
##   Model: crawling ~ temperature
##   Data: crawl
##        AIC      BIC    logLik
##   48.76397 49.67173 -21.38199
##
## Variance function:
##  Structure: fixed weights
##  Formula: ~SD^2/n
##
## Coefficients:
##                Value Std.Error   t-value p-value
## (Intercept) 35.73262 1.2115286 29.493832  0.0000
## temperature -0.07332 0.0232771 -3.150053  0.0103
##
##  Correlation:
##             (Intr)
## temperature -0.96
##
## Standardized residuals:
##        Min         Q1        Med         Q3        Max
## -2.2007008 -0.6976329  0.1727593  0.5057059  1.1266156
##
## Residual standard error: 0.9771564
## Degrees of freedom: 12 total; 10 residual
```


crawling ~ temperature weighted regression with count as weight
crawl$crawling
34
33
32
31
30
29
30
40
50
60
70
crawl$temperature



This data may be amenable to the lack of fit analysis discussed in the text. We don't have the original $y_i i$ in the formaula for

$$SS_{pe} = \sum_j \sum_i (y_{ji} - \bar{y})^2$$

But we know n is the sum of the counts, that the mean response for each temperature as the crawl variable, and that $SE_j = \frac{1}{n_j} \sum_i (y_{ii} - \bar{y_j})^2$

So

$$\hat{\sigma}^2 = \frac{SS_{pe}}{(\sum_j n_i) - j} = \frac{1}{(\sum_j n_i) - j} \sum_j \sum_i (y_{ji} - \bar{y})^2 = \frac{1}{(\sum_j n_i) - j} \sum_j n_j SE_j$$

Calculating this for our data set we have

Table 1: estimated SD from the repeated predictor values

se
2.718