Weighted least squares

Patrick Breheny

March 8

Introduction

- As a precursor to fitting generalized linear models, let's first deal with the case of a normally distributed outcome with unequal variances
- Consider the data set `statin` which contains (simulated) records on diabetic patients, collected from 130 practices (hospitals, clinics, etc.) in Pennsylvania
- The outcome variable was the average LDL cholesterol level of the diabetic patients, and the explanatory variable was the percent of diabetic patients at the practice who are on statin drugs
- Only practice-level (*i.e.*, not individual-level) data was available

Ordinary least squares fit

- When we fit a simple linear regression model, we see that the correlation between `Statin` and `LDL` is 0.1, and is not significant ($p = .27$)
- However, this model is treating each practice with equal weight, despite the fact that some practices (such as Philadelphia hospitals) have nearly 2500 patients while others (small rural clinics) have only 11 diabetic patients
- Clearly, the results for the large practices are less variable; it therefore stands to reason that they should have greater weight in the model

Weighted least squares

- Recall that the negative log-likelihood of the model under the assumption of normality is

$$\sum_i \frac{(y_i - \mu_i)^2}{2\sigma_i^2};$$

  note that we are no longer assuming equal variance
- In particular, since the outcome in this case is an average, it is reasonable to assume $\sigma_i^2 = \sigma^2/n_i$, where n_i is the number of patients in clinic i
- Thus, the maximum likelihood estimate will minimize the weighted residual sum of squares,

$$\sum_i \frac{(y_i - \mu_i)^2}{2\sigma^2/n_i} \propto \sum_i w_i(y_i - \mu_i)^2,$$

  where $w_i = n_i$

Solving for the WLS estimate

- In matrix form, we need to minimize

$$(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T \mathbf{W} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}),$$

  where $\mathbf{W}$ is a diagonal matrix of weights
- Taking the derivative with respect to $\boldsymbol{\beta}$, we find that the solution is

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

Sampling distribution of $\hat{\beta}$

- Note that $\hat{\boldsymbol{\beta}}$ is once again a linear combination of $\{y_i\}$, and therefore normally distributed:

$$\hat{\boldsymbol{\beta}} \sim \mathrm{N}\left(\boldsymbol{\beta}, \sigma^2(\mathbf{X}^T\mathbf{W}\mathbf{X})^{-1}\right);$$

  recall that $\mathrm{Var}(\mathbf{y}) = \sigma^2\mathbf{W}^{-1}$

- The subsequent derivation of confidence intervals and tests is similar to what we have seen before
- This model can be fit in `R` using the `weights=` option, or in SAS with a `WEIGHT` statement

Results

Estimated effect of Statin:

	Estimate	SE	t	p
β_{Statin}	-0.19	0.07	2.78	0.01

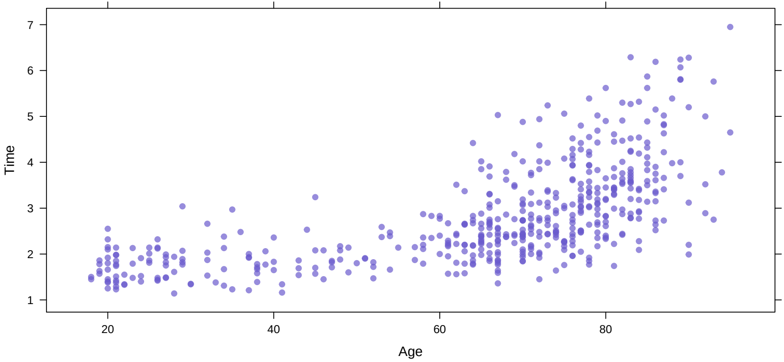
- Taking into account the heteroskedasticity, we now have a highly significant test and more accurate estimate (the data was simulated with a true $\beta_{\text{Statin}} = -0.2$)
- Additionally, the (weighted) correlation increased from 0.1 to 0.24

Introduction: Unknown weights

- The preceding approach (known as *weighted least squares*, or WLS) easily handles unequal variances for problems in which the relative variances of the outcomes are known (note that it was not necessary to know the actual variance, just the relative ratios)
- However, what about situations where the variance seems to increase with the mean?
- Here, we don't know the means ahead of time (otherwise we wouldn't need to fit a model) and therefore we don't know the appropriate weights, either

Hand speed data

For example, consider a study in which the investigator measured hand speed (time it takes to remove a bolt from an S-shaped iron rod) for a number of subjects of various ages:



Remarks

There are a number of interesting things going on in this data set:

- Older individuals take longer to complete the task than younger ones
- However, the trend does not seem to be linear
- As mean hand speed goes up, so does variability

Iterative reweighting

- Roughly speaking, it seems that $\mathrm{SD} \propto \mu$, so $\mathrm{Var}(y) \propto \mu^2$
- Of course, we don't know μ
- Once we fit the model, we have estimates for $\{\mu_i\}$ and therefore, for $\{w_i\}$, where $\hat{w}_i = \hat{\mu}_i^{-2}$
- However, once we change $\{w_i\}$, we change the fit, and thus we change $\{\hat{\mu}_i\}$, which changes $\{w_i\}$, and so on...

Iterative reweighting (cont'd)

- Consider, then, the following approach to model fitting:
  (1) Fit the model, obtaining $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\mu}}$
  (2) Use $\hat{\boldsymbol{\mu}}$ to recalculate $\mathbf{w}$
  (3) Repeat steps (1) and (2) until the model stops changing (*i.e.*, until *convergence*)
- This approach is known as the *iteratively reweighted least squares* (IRLS) algorithm

An iterative loop in R

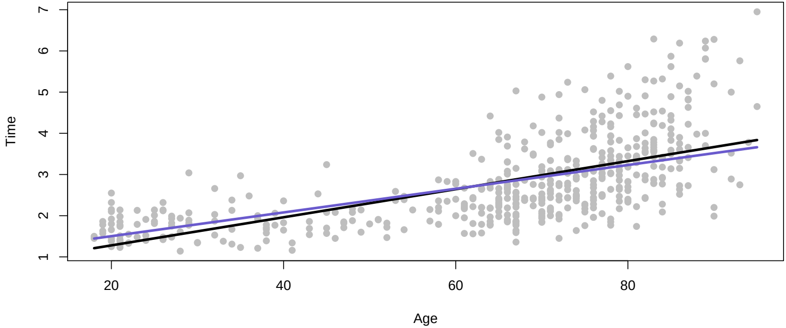
- One way to implement this algorithm is with

```
fit.w <- lm(Time~Age,handspeed)
for (i in 1:20)
  {
    w <- 1/fit.w$fitted.values^2
    fit.w <- lm(Time~Age,handspeed,weights=w)
  }
```

- This implementation assumes that 20 iterations is enough to achieve convergence
- This is fine for this case, but in general, it is better to write a `repeat` or `while` loop which checks for convergence at each iteration and terminates the loop when convergence is reached

OLS vs. WLS

A comparison of the two fits (black=OLS, blue=WLS):



OLS vs. WLS (cont'd)

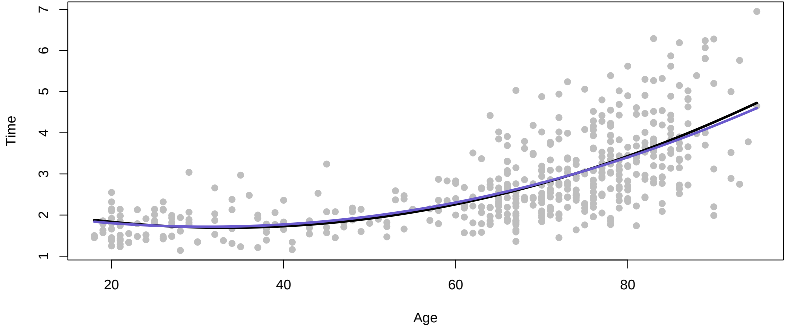
- Note that by taking heteroskedasticity into account, the slope is lowered somewhat, as the regression line is less influenced by the highly variable points on the right
- Furthermore, note that by more accurately modeling the distribution of the outcome, the standard deviation of our estimate is reduced from 0.0018 to 0.0013
- Consequently, we even obtain a larger test statistic, despite the fact that the estimate is closer to zero
- Furthermore, R^2 increases from 0.41 to 0.49

An inappropriate variance model

- Note that this only improves our results if we model the variance accurately
- Suppose we were to iteratively reweight according to $\hat{w}_i = \hat{\mu}_i^2$
- Fitting this model, we more than double our standard error and decrease R^2 from 0.41 to 0.30

Quadratic model

Finally, we might also consider a quadratic fit (black=OLS, blue=WLS):



The fit is actually very similar in this case, although once again we achieve reductions in standard error and an improvement in R^2