Weighted Least Squares

* The standard linear model assumes that Var$(\varepsilon_i) = \sigma^2$ for $i = 1, \ldots, n$.

* As we have seen, however, there are instances where

$$\text{Var}(Y \mid \boldsymbol{X} = \boldsymbol{x}_i) \;=\; \text{Var}(\varepsilon_i) \;=\; \frac{\sigma^2}{w_i}.$$

* Here $w_1, \ldots, w_n$ are known positive constants.

* Weighted least squares is an estimation technique which weights the observations proportional to the reciprocal of the error variance for that observation and so overcomes the issue of non-constant variance.

Weighted Least Squares in Simple Regression

* Suppose that we have the following model

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \qquad i = 1, \ldots, n$$

where $\varepsilon_i \sim \mathrm{N}(0, \sigma^2/w_i)$ for **known** constants $w_1, \ldots, w_n$.

* The weighted least squares estimates of β_0 and β_1 minimize the quantity

$$S_w(\beta_0, \beta_1) = \sum_{i=1}^{n} w_i (y_i - \beta_0 - \beta_1 x_i)^2$$

* Note that in this weighted sum of squares, the weights are inversely proportional to the corresponding variances; points with low variance will be given higher weights and points with higher variance are given lower weights.

Weighted Least Squares in Simple Regression

* The weighted least squares estimates are then given as

$$
\begin{aligned}
\widehat{\beta}_0 &= \overline{y}_w - \widehat{\beta}_1 \overline{x}_w \\
\widehat{\beta}_1 &= \frac{\sum w_i (x_i - \overline{x}_w)(y_i - \overline{y}_w)}{\sum w_i (x_i - \overline{x}_w)^2}
\end{aligned}
$$

where $\overline{x}_w$ and $\overline{y}_w$ are the weighted means

$$
\overline{x}_w = \frac{\sum w_i x_i}{\sum w_i} \qquad \overline{y}_w = \frac{\sum w_i y_i}{\sum w_i}.
$$

* Some algebra shows that the weighted least squares estimates are still unbiased.

Weighted Least Squares in Simple Regression

* Furthermore we can find their variances

$$\text{Var}(\widehat{\beta}_1) = \frac{\sigma^2}{\sum w_i (x_i - \overline{x}_w)^2}$$

$$\text{Var}(\widehat{\beta}_0) = \left[\frac{1}{\sum w_i} + \frac{\overline{x}_w^2}{\sum w_i (x_i - \overline{x}_w)^2}\right] \sigma^2$$

* Since the estimates can be written in terms of normal random variables, the sampling distributions are still normal.

* The weighted error mean square $S_w(\widehat{\beta}_0, \widehat{\beta}_1)/(n-2)$ also gives us an unbiased estimator of σ^2 so we can derive t tests for the parameters etc.

General Weighted Least Squares Solution

* Let $\boldsymbol{W}$ be a diagonal matrix with diagonal elements equal to $w_1, \ldots, w_n$.

* The the Weighted Residual Sum of Squares is defined by

$$S_w(\boldsymbol{\beta}) = \sum_{i=1}^{n} w_i (y_i - \boldsymbol{x}_i^t \boldsymbol{\beta})^2 = (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})^t \boldsymbol{W} (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}).$$

* Weighted least squares finds estimates of $\boldsymbol{\beta}$ by minimizing the weighted sum of squares.

* The general solution to this is

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}^t \boldsymbol{W} \boldsymbol{X})^{-1} \boldsymbol{X}^t \boldsymbol{W} \boldsymbol{Y}.$$

Weighted Least Squares as a Transformation

* Recall from the previous chapter the model

$$y_i \;=\; \beta_0 + \beta_1 x_i + \varepsilon_i$$

where Var(ε_i) $=$ $x_i^2\sigma^2$.

* We can transform this into a regular least squares problem by taking

$$y_i' \;=\; \frac{y_i}{x_i} \quad x_i' \;=\; \frac{1}{x_i} \quad \varepsilon_i' \;=\; \frac{\varepsilon_i}{x_i}.$$

* Then the model is

$$y_i' \;=\; \beta_1 + \beta_0 x_i' + \varepsilon_i'$$

where Var(ε_i') $=$ σ^2.

Weighted Least Squares as a Transformation

* The residual sum of squares for the transformed model is

$$
\begin{aligned}
S_1(\beta_0, \beta_1) &= \sum_{i=1}^{n} (y_i' - \beta_1 - \beta_0 x_i')^2 \\
&= \sum_{i=1}^{n} \left(\frac{y_i}{x_i} - \beta_1 - \beta_0 \frac{1}{x_i} \right)^2 \\
&= \sum_{i=1}^{n} \left(\frac{1}{x_i^2} \right) (y_i - \beta_0 - \beta_1 x_i)^2
\end{aligned}
$$

* This is the weighted residual sum of squares with $w_i = 1/x_i^2$.

* Hence the weighted least squares solution is the same as the regular least squares solution of the transformed model.

Weighted Least Squares as a Transformation

* In general suppose we have the linear model

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$$

where $\text{Var}(\boldsymbol{\varepsilon}) = \boldsymbol{W}^{-1}\sigma^2$.

* Let $\boldsymbol{W}^{1/2}$ be a diagonal matrix with diagonal entries equal to $\sqrt{w_i}$.

* Then we have $\text{Var}(\boldsymbol{W}^{1/2}\boldsymbol{\varepsilon}) = \sigma^2\boldsymbol{I}_n$.

Weighted Least Squares as a Transformation

* Hence we consider the transformation

$$\boldsymbol{Y}' = \boldsymbol{W}^{1/2}\boldsymbol{Y} \quad \boldsymbol{X}' = \boldsymbol{W}^{1/2}\boldsymbol{X} \quad \boldsymbol{\varepsilon}' = \boldsymbol{W}^{1/2}\boldsymbol{\varepsilon}.$$

* This gives rise to the usual least squares model

$$\boldsymbol{Y}' = \boldsymbol{X}'\boldsymbol{\beta} + \boldsymbol{\varepsilon}'$$

* Using the results from regular least squares we then get the solution

$$\widehat{\boldsymbol{\beta}} = \left(\left(\boldsymbol{X}'\right)^t \boldsymbol{X}'\right)^{-1} \left(\boldsymbol{X}'\right)^t \boldsymbol{Y}' = \left(\boldsymbol{X}^t \boldsymbol{W} \boldsymbol{X}\right)^{-1} \boldsymbol{X}^t \boldsymbol{W} \boldsymbol{Y}.$$

* Hence this is the weighted least squares solution.

Advantages of Weighted Least Squares

* In the transformed model, the interpretation of the coefficient estimates can be difficult. In weighted least squares the interpretation remains the same as before.

* In the transformed model, there will often not be an intercept which means that the F-tests and R-squared values are quite different. In weighted least squares we generally include an intercept retaining the usual interpretation of these quantities.

* Weighted least squares gives us an easy way to remove one observation from a model by setting its weight equal to 0.

* We can also downweight outlier or influential points to reduce their impact on the overall model.

The Weights

* To apply weighted least squares, we need to know the weights $w_1, \ldots, w_n$.

* There are some instances where this is true.

* We may have a probabilistic model for Var$(Y \mid X = x_i)$ in which case we would use this model to find the w_i.

* For example, with Poisson data we may use $w_i = 1/x_i$ if we expect an increasing relationship between Var$(Y \mid X = x)$ and x.

The Weights

* Another common case is where each observation is not a single measure but an average of n_i actual measures and the original measures each have variance σ^2.

* In that case, standard results tell us that

$$\text{Var}(\varepsilon_i) = \text{Var}\left(\overline{Y}_i \mid \boldsymbol{X} = \boldsymbol{x}_i\right) = \frac{\sigma^2}{n_i}$$

* Thus we would use weighted least squares with weights $w_i = n_i$.

* This situation often occurs in cluster surveys.

Unknown Weights

* In many real-life situations, the weights are not known apriori.

* In such cases we need to estimate the weights in order to use weighted least squares.

* One way to do this is possible when there are multiple repeated observations at each value of the covariate vector.

* That is often possible in designed experiments in which a number of replicates will be observed for each set value of the covariate vector.

* We can then estimate the variance of Y for each fixed covariate vector and use this to estimate the weights.

Pure Error

* Suppose that we have n_i observations at $x = x_j$, $j = 1, \ldots, k$.

* Then a fitted model could be

$$y_{ij} \;=\; \beta_0 + \beta_1 x_j + \varepsilon_{ij} \qquad i = 1, \ldots, n_j;\; j = 1, \ldots, k.$$

* The $(i, j)^{\text{th}}$ residual can then be written as

$$e_{ij} \;=\; y_{ij} - \widehat{y}_{ij} \;=\; (y_{ij} - \overline{y}_j) + (\overline{y}_j - \widehat{y}_{ij}).$$

* The first term is referred to as the pure error.

Pure Error

* Note that the pure error term does not depend on the mean model at all.

* We can use the pure error mean squares to estimate the variance at each value of x.

$$s_j^2 = \frac{1}{n_j - 1} \sum_{i=1}^{n_j} (y_{ij} - \overline{y}_j)^2 \qquad j = 1, \ldots, k.$$

* Then we can use the weights $w_{ij} = 1/s_j^2$ $(i = 1, \ldots, n_j)$ as the weights in a weighted least squares regression model.

Unknown Weights

* For observational studies, however, this is generally not possible since we will not have repeated measures at each value of the covariate(s).

* This is particularly true when there are multiple covariates.

* Sometimes, however, we may be willing to assume that the variance of the observations is the same within each level of some categorical variable but possibly different between levels.

* In that case we can estimate the weights assigned to observations with a given level by an estimate of the variance for that level of the categorical variable.

* This leads to a two-stage method of estimation.

Two-Stage Estimation

* In the two-stage estimation procedure we first fit a regular least squares regression to the data.

* If there is some evidence of non-homogenous variance then we examine plots of the residuals against a categorical variable which we suspect is the culprit for this problem.

* Note that we do still need to have some apriori knowledge of a categorical variable likely to affect variance.

* This categorical variable may, or may not, be included in the mean model.

Two-Stage Estimation

* Let Z be the categorical variable and assume that there are n_j observations with $Z = j$ $(j = 1, \ldots, k)$.

* If the error variability does vary with the levels of this categorical variable then we can use

$$\hat{\sigma}_j^2 = \frac{1}{n_j - 1} \sum_{i: z_i = j} r_i^2$$

as an estimate of the variability when $Z = j$.

* **Note** Your book uses the raw residuals e_i instead of the studentized residuals r_i but that does not work well.

Two-Stage Estimation

* If we now assume that $\sigma_j^2 = c_j\sigma^2$ we can estimate the c_j by

$$\hat{c}_j = \frac{\hat{\sigma}_j^2}{\hat{\sigma}^2} = \frac{\dfrac{1}{n_j - 1}\displaystyle\sum_{i:z_i=j} r_i^2}{\dfrac{1}{n}\displaystyle\sum_{i=1}^{n} r_i^2}$$

* We could then use the reciprocals of these estimates as the weights in a weighted least squares regression in the second stage.

* Approximate inference about the parameters can then be made using the results of the weighted least squares model.

Problems with Two-Stage Estimation

* The method described above is not universally accepted and a number of criticisms have been raised.

* One problem with this approach is that different datasets would result in different estimated weights and this variability is not properly taken into account in the inference.

* Indeed the authors acknowledge that the true sampling distributions are unlikely to be Student-t distributions and are unknown so inference may be suspect.

* Another issue is that it is not clear how to proceed if no categorical variable explaining the variance heterogeneity can be found.