Weighted Least Square Regression



Definition

Each term in the weighted least squares criterion includes an additional weight, that determines how much each observation in the data set influences the final parameter estimates and it can be used with functions that are either linear or nonlinear in the parameters.

http://www.itl.nist.gov/div898/handbook/index.htm Section 4.4.3.2



Weighted Least Square Regression

One of the common assumptions underlying most process modeling methods, including linear and nonlinear least squares regression, is that each data point provides equally precise information about the deterministic part of the total process variation.

In other words, it is assumed that the standard deviation of the error term is constant over all values of the predictor or explanatory variables. This assumption, clearly does not hold, even approximately, in every modeling application.



Ref: NIST, Section 4.1.4.3

Weighted Least Square Regression

One of the common assumptions underlying most process modeling methods, including linear and nonlinear least squares regression, is that each data point provides equally precise information about the deterministic part of the total process variation.

In other words, it is assumed that the standard deviation of the error term is constant over all values of the predictor or explanatory variables. This assumption, clearly does not hold, even approximately, in every modeling application.



In a weighted fit, less weight is given to the less precise measurements and more weight to more precise measurements when estimating the unknown parameters in the model.

Using weights that are inversely proportional to the variance at each level of the explanatory variables yields the most precise parameter estimates possible.

Ref: NIST, Section 4.1.4.3



In a weighted fit, less weight is given to the less precise measurements and more weight to more precise measurements when estimating the unknown parameters in the model.

Using weights that are inversely proportional to the variance at each level of the explanatory variables yields the most precise parameter estimates possible.

Ref: NIST, Section 4.1.4.3



8000C, describes variance as the difference between the observed instrument response for the ith calibration standard and the predicted or calculated response for the ith calibration standard.

Weighting the sum of the squares of the differences may significantly improve the ability of the least square regression to fit the linear model to the data.

Ref: SW846, 8000C, Section 11.5.2, Revision 3, March 2003



8000C, describes variance as the difference between the observed instrument response for the ith calibration standard and the predicted or calculated response for the ith calibration standard.

Weighting the sum of the squares of the differences may significantly improve the ability of the least square regression to fit the linear model to the data.

Ref: SW846, 8000C, Section 11.5.2, Revision 3, March 2003



$\sum w_i (y_i - y'_i)^2$

where:

- w_i = Weighting factor for the ith calibration standard (w=1 for unweighted least square regression)
- y_i = Observed instrument response for the ith calibration standard
- y'_I = Predicted (or calculated) response for the ith standard
- Σ = The sum of all individual values

Ref: SW846, 8000C, Section 11.5.2,



The mathematics used in unweighted least squares regression has a tendency to favor numbers of larger value over numbers of smaller value. Thus the regression curves that are generated will tend to fit points that are at the upper calibration levels better than those points at the lower calibration levels.

Ref: SW846, 8000C, Section 11.5.2



Examples of weighting factors which can place more emphasis on numbers of smaller value are:

$$w_i = 1/y_i$$
 or $w_i = 1/y_i^2$ where,

- w_i = weighting factor for the ith calibration standard (w_i =1 for unweighted least squares regression).
- y_i = observed instrument response (area or height) for the ith calibration standard.

Ref: SW846, 8000C, Section 11.5.2



Different Types Of Weights

- No Weights: Default higher weighting of higher amounts or signal values
- 1/Amount: Nearly cancels out the weighting of higher amounts
- 1/Amount²: Causes over-proportional weighting of smaller amounts
- 1/Response: Nearly cancels out the weighting of higher signal values

Ref: Chromeleon Manual



Weights Cont.:

1/Response^2: Causes over-proportional weighting of smaller signal values

> Weights signal values with small relative standard deviations more than those with large relative standard deviations

Weights signal values with small relative standard deviations clearly more than those with large relative standard deviation.

Ref: Chromeleon Manual

1/RSD:

1/RSD ²:

Benefits

Weighted least squares is an efficient method that makes good use of small data sets. It also shares the ability to provide different types of easily interpretable statistical intervals for estimation, prediction, calibration and optimization.

The main advantage that weighted least squares enjoys over other methods is the ability to handle regression situations in which the data points are of varying quality.

NIST, Section 4.1.4.3



Benefits

Weighted least squares is an efficient method that makes good use of small data sets. It also shares the ability to provide different types of easily interpretable statistical intervals for estimation, prediction, calibration and optimization.

The main advantage that weighted least squares enjoys over other methods is the ability to handle regression situations in which the data points are of varying quality.

NIST, Section 4.1.4.3



Disadvantages

The biggest disadvantage of weighted least squares, is probably the fact that the theory behind this method is based on the assumption that the weights are known exactly.

The exact weights are almost never known in real applications, so estimated weights must be used instead. The effect of using estimated weights is difficult to assess, but experience indicates that small variations in the weights due to estimation do not often affect a regression analysis or its interpretation.

M Ref: NIST Section 4.1.4.3

Disadvantages

The biggest disadvantage of weighted least squares, is probably the fact that the theory behind this method is based on the assumption that the weights are known exactly.

The exact weights are almost never known in real applications, so estimated weights must be used instead. The effect of using estimated weights is difficult to assess, but experience indicates that small variations in the weights due to estimation do not often affect a regression analysis or its interpretation.

Ref: NIST Section 4.1.4.3

When the weights are estimated from small numbers of replicated observations, the results of an analysis can be very badly and unpredictably affected. This is especially likely to be the case when the weights for extreme values of the predictor or explanatory variables are estimated using only a few observations.

It is important to remain aware of this potential problem, and to only use weighted least squares when the weights can be estimated precisely relative to one another.

When the weights are estimated from small numbers of replicated observations, the results of an analysis can be very badly and unpredictably affected. This is especially likely to be the case when the weights for extreme values of the predictor or explanatory variables are estimated using only a few observations.

It is important to remain aware of this potential problem, and to only use weighted least squares when the weights can be estimated precisely relative to one another.

Weighted least squares regression, is also sensitive to the effects of outliers. If potential outliers are not investigated and dealt with appropriately, they will likely have a negative impact on the parameter estimation and other aspects of a weighted least squares analysis.

If a weighted least squares regression actually increases the influence of an outlier, the results of the analysis may be far inferior to an unweighted least squares analysis.



Weighted least squares regression, is also sensitive to the effects of outliers. If potential outliers are not investigated and dealt with appropriately, they will likely have a negative impact on the parameter estimation and other aspects of a weighted least squares analysis.

If a weighted least squares regression actually increases the influence of an outlier, the results of the analysis may be far inferior to an unweighted least squares analysis.





Unweighted vs. Weighted

