0.0.1 Introduction

In statistics and optimization, statistical errors and residuals are two closely related and easily confused measures of the deviation of an observed value of an element of a statistical sample from its "theoretical value". The error (or disturbance) of an observed value is the deviation of the observed value from the (unobservable) true function value, while the residual of an observed value is the difference between the observed value and the estimated function value.

The distinction is most important in regression analysis, where it leads to the concept of studentized residuals.

0.0.2 Residual

Residual (or error) represents unexplained (or residual) variation after fitting a regression model. It is the difference (or left over) between the observed value of the variable and the value suggested by the regression model.

The difference between the observed value of the dependent variable (y) and the predicted value () is called the residual (e). Each data point has one residual.

Residual = Observed value - Predicted value

$$e = y - \hat{y}$$

Both the sum and the mean of the residuals are equal to zero. That is, $\sum e = 0$ and e = 0.

0.0.3 Residual

A residual (or fitting error), on the other hand, is an observable estimate of the unobservable statistical error. Consider the previous example with men's heights and suppose we have a random sample of n people. The sample mean could serve as a good estimator of the population mean. Then we have:

The difference between the height of each man in the sample and the unobservable population mean is a statistical error, whereas The difference between the height of each man in the sample and the observable sample mean is a residual. Note that the sum of the residuals within a random sample is necessarily zero, and thus the residuals are necessarily not independent. The statistical errors on the other hand are independent, and their sum within the random sample is almost surely not zero.

Other uses of the word "error" in statistics:

The use of the term "error" as discussed in the sections above is in the sense of a deviation of a value from a hypothetical unobserved value. At least two other uses also occur in statistics, both referring to observable prediction errors:

- Mean square error or mean squared error (abbreviated MSE) and root mean square error (RMSE) refer to the amount by which the values predicted by an estimator differ from the quantities being estimated (typically outside the sample from which the model was estimated).

- Sum of squared errors, typically abbreviated SSE or SSe, refers to the residual sum of squares (the sum of squared residuals) of a regression; this is the sum of the squares of the deviations of the actual values from the predicted values, within the sample used for estimation. Likewise, the sum of absolute errors (SAE) refers to the sum of the absolute values of the residuals, which is minimized in the least absolute deviations approach to regression.

0.0.4   Influence Diagnostics: Basic Idea and Statistics

The general idea of quantifying the influence of one or more observations relies on computing parameter estimates based on all data points, removing the cases in question from the data, refitting the model, and computing statistics based on the change between full-data and reduced-data estimation.

0.0.5 Standardization

A random variable is said to be standardized if the difference from its mean is scaled by its standard deviation. The residuals above have mean zero but their variance is unknown, it depends on the true values of θ. Standardization is thus not possible in practice.

0.0.6 Studentization

Instead, you can compute studentized residuals by dividing a residual by an estimate of its standard deviation.

0.0.7 Internal and External Studentization

If that estimate is independent of the $i-$th observation, the process is termed 'external studentization'. This is usually accomplished by excluding the $i-$th observation when computing the estimate of its standard error. If the observation contributes to the standard error computation, the residual is said to be internally studentized.

Externally studentized residual require iterative influence analysis or a profiled residuals variance.

0.0.8 Computation

The computation of internally studentized residuals relies on the diagonal entries of $\boldsymbol{V}(\hat{\theta})$ - $\boldsymbol{Q}(\hat{\theta})$, where $\boldsymbol{Q}(\hat{\theta})$ is computed as

$$\boldsymbol{Q}(\hat{\theta}) = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{Q}(\hat{\theta})^{-1}\boldsymbol{X})\boldsymbol{X}^{-1}$$

0.0.9 Standardized and studentized residuals

To alleviate the problem caused by inconstant variance, the residuals are scaled (i.e. divided) by their standard deviations. This results in a 'standardized residual'. Because true standard deviations are frequently unknown, one can instead divide a residual by the estimated standard deviation to obtain the 'studentized residual.

0.0.10 Internal and External Studentization

If that estimate is independent of the $i-$th observation, the process is termed 'external studentization'. This is usually accomplished by excluding the $i-$th observation when computing the estimate of its standard error. If the observation contributes to the standard error computation, the residual is said to be internally studentized.

Externally studentized residual require iterative influence analysis or a profiled residuals variance.

0.0.11 Computation

The computation of internally studentized residuals relies on the diagonal entries of $\boldsymbol{V}(\hat{\theta})$ - $\boldsymbol{Q}(\hat{\theta})$, where $\boldsymbol{Q}(\hat{\theta})$ is computed as

$$\boldsymbol{Q}(\hat{\theta}) = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{Q}(\hat{\theta})^{-1}\boldsymbol{X})\boldsymbol{X}^{-1}$$

0.0.12 Pearson Residual

Another possible scaled residual is the 'Pearson residual', whereby a residual is divided by the standard deviation of the dependent variable. The Pearson residual can be used when the variability of $\hat{\beta}$ is disregarded in the underlying assumptions.

0.0.13 Standardized and studentized residuals

Externally studentized residual require iterative influence analysis or a profiled residuals variance.

0.0.14 Computation

The computation of internally studentized residuals relies on the diagonal entries of $\boldsymbol{V}(\hat{\theta})$ - $\boldsymbol{Q}(\hat{\theta})$, where $\boldsymbol{Q}(\hat{\theta})$ is computed as

$$\boldsymbol{Q}(\hat{\theta}) = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{Q}(\hat{\theta})^{-1}\boldsymbol{X})\boldsymbol{X}^{-1}$$

0.0.15 Pearson Residual

Another possible scaled residual is the 'Pearson residual', whereby a residual is divided by the standard deviation of the dependent variable. The Pearson residual can be used when the variability of $\hat{\beta}$ is disregarded in the underlying assumptions.

0.0.16 Covariance Parameters

The unknown variance elements are referred to as the covariance parameters and collected in the vector θ.

0.0.17 Computation

The computation of internally studentized residuals relies on the diagonal entries of $\boldsymbol{V}(\hat{\theta})$ - $\boldsymbol{Q}(\hat{\theta})$, where $\boldsymbol{Q}(\hat{\theta})$ is computed as

$$\boldsymbol{Q}(\hat{\theta}) = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{Q}(\hat{\theta})^{-1}\boldsymbol{X})\boldsymbol{X}^{-1}$$

0.0.18 Studentization

In statistics, a studentized residual is the quotient resulting from the division of a residual by an estimate of its standard deviation. Typically the standard deviations of residuals in a sample vary greatly from one data point to another even when the errors all have the same standard deviation, particularly in regression analysis; thus it does not make sense to compare residuals at different data points without first studentizing. It is a form of a Student's t-statistic, with the estimate of error varying between points.

This is an important technique in the detection of outliers. It is named in honor of William Sealey Gosset, who wrote under the pseudonym Student, and dividing by an estimate of scale is called studentizing, in analogy with standardizing and normalizing: see Studentization.

0.1 Covariance Parameters

The unknown variance elements are referred to as the covariance parameters and collected in the vector θ.

0.1.1 Pearson Residual

Another possible scaled residual is the 'Pearson residual', whereby a residual is divided by the standard deviation of the dependent variable. The Pearson residual can be used when the variability of $\hat{\beta}$ is disregarded in the underlying assumptions.