TIBCO Statistica®

Logistic Regression Formula Guide

Version 14.0.1
July 2022

Contents

Overview

Logistic regression is used for modeling binary outcome variables such as credit default or warranty claims.

It is assumed that the binary response, Y, takes on the values of 0 and 1 with 0 representing failure and 1 representing success.

Overview of Logistic Regression Model

The logistic regression function models the probability that the binary response is as a function of a set of predictor variables $\boldsymbol{X} = [X_1, X_2, ..., X_p]^T$ and regression coefficients $\boldsymbol{\beta} = [\beta_0, \beta_1, ..., \beta_p]^T$ as given by:

$$\pi(\mathbf{X}) = \left[\frac{e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p}}{1 + e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p}} \right]$$

In practice, the regression coefficients are unknown and are estimated by maximizing the likelihood function. Note that the w_i below are case weights and are assumed to be positive. All observations that have a case weight less than or equal to zero are excluded from the analysis and all subsequent results.

$$L = \prod_{i=1}^{N} \pi_i(\boldsymbol{X})^{w_i y_i} (1 - \pi_i(\boldsymbol{X}))^{w_i (1 - y_i)}$$

where,

$\pi_i(\boldsymbol{X}) =$ probability for i^{th} case

The maximization of the likelihood is achieved by an iterative method called Fisher scoring. Fisher scoring is similar to the Newton-Raphson procedure except that the hessian matrix (matrix of second order partial derivatives) is replaced with its expected value.

The Fisher scoring update formula for the regression coefficients is given by:

$$\hat{\boldsymbol{\beta}}_{k+1} = \hat{\boldsymbol{\beta}}_k + \left[I(\hat{\boldsymbol{\beta}}_k) \right]^{-1} s(\hat{\boldsymbol{\beta}}_{\boldsymbol{k}})$$

where,

$\hat{\beta}_k =$ estimate of β based on k^{th} iteration

The algorithm completes when the convergence criterion is satisfied or when the maximum number of iterations has been reached. Convergence is obtained when the difference between the log-likelihood function from one iteration to the next is small. By default, the convergence criterion is 1e-7, thus convergence is obtained when:

$$|ln\, L_{k+1} - ln\, L_k| \leq 1e-7$$

where,

$L_k =$ likelihood evaluated at $\hat{\beta}_k$

In the equations below:

$\hat{\pi}_i =$ estimated probability of i^{th} case

The score vector is given by:

$$s(\hat{\boldsymbol{\beta}}) = \frac{\partial l}{\partial \hat{\boldsymbol{\beta}}} = \begin{bmatrix} \sum_{i=1}^{N} w_i(y_i - \hat{\pi}_i) \\ \sum_{i=1}^{N} w_i(y_i - \hat{\pi}_i)X_{i1} \\ \vdots \\ \sum_{i=1}^{N} w_i(y_i - \hat{\pi}_i)X_{ip} \end{bmatrix}$$

The information matrix is given by:

$$I(\hat{\boldsymbol{\beta}}) = -E\left[\frac{\partial^2 l}{\partial \hat{\beta}_i \partial \hat{\beta}_j}\right] = \begin{bmatrix} \sum_{i=1}^{N} w_i(\hat{\pi}_i)(1-\hat{\pi}_i) & \cdots & \sum_{i=1}^{N} w_i(\hat{\pi}_i)(1-\hat{\pi}_i)X_{ip} \\ \sum_{i=1}^{N} w_i(\hat{\pi}_i)(1-\hat{\pi}_i)X_{i1} & \cdots & \sum_{i=1}^{N} w_i(\hat{\pi}_i)(1-\hat{\pi}_i)X_{i1}X_{ip} \\ \vdots & \ddots & \vdots \\ \sum_{i=1}^{N} w_i(\hat{\pi}_i)(1-\hat{\pi}_i)X_{ip} & \cdots & \sum_{i=1}^{N} w_i(\hat{\pi}_i)(1-\hat{\pi}_i)X_{ip}X_{ip} \end{bmatrix}$$

The asymptotic estimated covariance matrix of the estimated coefficients is given by:

$$\hat{\Sigma}_{\hat{\beta}} = I(\hat{\boldsymbol{\beta}})^{-1}$$

Dispersion

The default dispersion parameter for logistic regression is 1.

However, this may not be adequate for the given data. Statistica® offers two methods of estimating the dispersion parameter, ϕ.

- The dispersion parameter can be estimated using the deviance goodness of fit

statistic, D (see Deviance in Goodness of Fit).

$$\hat{\phi}_D = \mathrm{D} / (N - p)$$

where,
N = Number of observations
p = Number of parameters

- The dispersion parameter can be estimated using the Pearson χ^2 goodness of fit statistic, Pearson χ^2 (see Pearson in Goodness of Fit).

$$\hat{\phi}_P = \frac{\text{Pearson } \chi^2}{(N - p)}$$

where,
N = Number of observations
p = Number of parameters

Parameterization

Effects of categorical predictor variables can be coded in the design matrix using different types of Parameterization, specifically overparameterized, sigma-restricted, or reference coding.

Sigma-Restricted Model

In a sigma-restricted model, the categories can be assigned any values corresponding to group membership to facilitate interpretation of the regression coefficient associated with the single categorical predictor variable. The values on the resulting predictor variable represent a quantitative contrast between the categories.

For a categorical variable with k levels, Statistica® creates k-1 indicator variables with the reference level coded as -1. For example, if the categorical variable has 3 levels, say, X, Y, and Z, with Z being the reference level, then the indicator variable would be coded as follows:

	Column X	Column Y
X	1	0
Y	0	1
Z	-1	-1

The values used to represent group membership, that is, 1 and -1, sum to zero. This parameterization leads to the interpretation that each coefficient estimates the difference between each level and the average of the other 2 levels, that is, the coefficient for X is the estimate of the difference between level X and the average of levels of Y and Z.

Overparameterized Model

An overparameterized model uses the indicator variable approach to represent the effects for categorical predictor variables. In this method, a separate predictor variable is coded for each group identified by a categorical predictor variable.

For a categorical variable with k levels, Statistica® creates k indicator variables. For example, if the categorical variable has 3 levels, X, Y, and Z, then the indicator variables would be as follows:

	Column X	Column Y	Column Z
X	1	0	0
Y	0	1	0
Z	0	0	1

Reference Coding

The Ref option is selected to compute the design matrix for categorical predictors in the models using reference coding. For a categorical variable with k levels, Statistica® creates k-1 indicator variables with the reference level coded as 0. The parameter estimates of the reference-coded categorical predictor estimates the difference of effect between the specific level and the reference level. For example, for a categorical variable with 3 levels, X, Y and Z, with Z being the reference level, the indicator variable would be as follows:

	Column X	Column Y
X	1	0
Y	0	1
Z	0	0

Model Summary - Wald Statistic

The Wald statistic is used to test the linear hypotheses about the regression coefficients, and it is based on large sample sizes, like the large-sample normality of parameter estimates.

Testing for Single Parameter βk

To test the null hypothesis that the single parameter estimate equals 0, the Wald statistic is given by:

$$\text{Wald} = \left(\frac{\hat{\beta}}{\hat{\sigma}_{\hat{\beta}_i}}\right)^2$$

where,

$\hat{\sigma}_{\hat{\beta}_i}$ = Estimated standard error of the i^{th} estimated coefficient

The Wald statistic is asymptotically distributed as χ^2 with 1 degree of freedom. The estimated standard error of the ith estimated coefficient, $\hat{\sigma}_{\hat{\beta}_i}$, is the square root of the ith diagonal element of the estimated covariance matrix $\hat{\Sigma}_{\hat{\beta}}$, that is, $\hat{\sigma}_{ii}$.

Testing for several βk

When β is k-dimensional, and asymptotic is normal, the hypothesis test is given by the following quadratic form:

$$\text{Wald} = (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta_0})^T \hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\beta}}}^{-1} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta_0})$$

where,

$\hat{\Sigma}_{\hat{\boldsymbol{\beta}}}$ = Estimated variance-covariance matrix of $\hat{\beta}$

This statistic is asymptotically distributed as χ^2 with degrees of freedom equal to the number of parameters estimated for a given effect and can be used to test the hypothesis that all parameters estimated for a given effect are equal to 0.

Likelihood Ratio Test

In general, the likelihood ratio test is used to compare the fit of two models, one of which is nested within the other.

This is typically performed to determine if a simpler model can be used to adequately model the data. The test is based on a comparison of full and reduced models where both models are fitted to the data and their log-likelihoods are calculated.

Let the full model (F) have p parameters and the reduced model (R) have q parameters such that q < p.

Full Model

$$\text{Logit}[\pi(\boldsymbol{X})] = ln\left(\frac{\pi(\boldsymbol{X})}{1-\pi(\boldsymbol{X})}\right) = \beta_0+\beta_1 X_1+\cdots+\beta_{q-2}X_{q-2}+\beta_{q-1}X_{q-1}+\beta_q X_q+\beta_{q+1}+\cdots+\beta_{p-1}$$

Reduced Model

$$\text{Logit}[\pi(\boldsymbol{X})] = ln\left(\frac{\pi(\boldsymbol{X})}{1-\pi(\boldsymbol{X})}\right) = \beta_0+\beta_1 X_1+\cdots+\beta_{q-2}X_{q-2}+\beta_{q-1}X_{q-1}$$

Let L(F) denote the maximized log-likelihood of the full model and L(R) represent the maximized log-likelihood of the reduced model. The null and alternative hypotheses with respect to this test are shown below.

$$H_0 : \beta_q = \beta_{q+1} = \cdots = \beta_{p-1} = 0$$
$$H_a : \text{Not } H_0$$

The test statistic is given by:

$$LR = -2[L(F) - L(R)]$$

This LR statistic is asymptotically distributed as χ^2 with p-q degrees of freedom.

Type 1 Likelihood Ratio Test

The Type 1 Likelihood Ratio (LR) test provides sequential tests for the effects in the model.

Specifically, the Type 1 LR test reports the log-likelihood for the model that includes a particular effect (shown in the respective row of the spreadsheet) and all effects that precede it (shown in the previous rows of the spreadsheet). The incremental χ^2 statistic then provides a test of the increment in the log-likelihood, attributable to the respective (current) effect.

Type 3 Likelihood Ratio Test

The Type 3 Likelihood Ratio test reports the log-likelihood for the model that includes all effects except for the current effect (shown in the respective row of the spreadsheet).

The incremental χ^2 statistic for that model, and the full model (that includes all effects), then provides a test of the increment in the log-likelihood, attributable to the respective (current) effect, while controlling for all other effects.

When the model is not full-rank, that is, redundant columns are detected during the evaluation of the design matrix, some difficulties arise when computing the Wald statistic for the overall model, and when attempting to compute Type 3 LR test of effects.

Specifically, because of the redundancy of some of the parameters in the model, independent tests of effects, controlling for all other parameters in the model (not belonging to the effect under consideration), cannot be computed. Therefore, the Summary of all effects and Type 3 LR test buttons are not available in that case.

Cell Statistics

The Cell statistics output gives basic statistics at each level of the selected grouping factors.

The basic mean, standard deviation, standard error, and confidence limits are computed for each covariate at each level of the factors in the model. An additional output spreadsheet gives the frequency of each level of Y, broken down by the factors in the model.

Mean

The Mean is computed for each covariate over the total data set and broken down by levels of each factor and combinations of factors depending on the effects specified in the model.

The mean for a covariate, x, is given by:

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} w_i x_i$$

where,

$x_i = i^{th}$ observation

$w_i =$ Weight for i^{th} observation

$N =$ Sample size over the appropriate set of cases,
i.e. entire dataset or subset defined by specific factor combinations

Standard Deviation

Standard deviation is computed for each covariate over the total data set and broken down by levels of each factor in the model.

The standard deviation for the covariate, x, is given by:

$$s_x = \sqrt{\frac{\sum_{i=1}^{N} w_i (x_i - \bar{x})^2}{N-1}}$$

where,

$x_i = i^{th}$ observation of the covariate

$w_i =$ Weight of the i^{th} observation

$N =$ Sample size over the appropriate set of cases,
i.e. entire dataset or subset defined by specific factor combinations

$\bar{x} =$ Mean of covariate as calculated above.

Standard Error

Standard error is the standard deviation of the sampling distribution of a statistic, in this case the mean, and is computed for each covariate over the total data set and broken down by levels of each factor in the model.

The standard error for the covariate, x, is given by:

$$s_{\bar{x}} = \frac{s_x}{N}$$

where,

$s_x =$ Standard deviation of covariate

$N =$ Sample size over the appropriate set of cases,
i.e. entire dataset or subset defined by specific factor combinations

Upper/Lower Confidence Limits for the Mean

The confidence limits for the covariate, x, are given by:

$$\bar{x} \pm t_{(\alpha/2,N-1)} s_{\bar{x}}$$

where,

$\bar{x}$ = Mean of the covariate

$t_{(\alpha/2,N-1)} = 1 - \alpha/2$ quantile from Student t distribution with N-1 degrees of freedom

$s_{\bar{x}}$ = Standard deviation of the covariate

N = Sample size over the appropriate set of cases,
i.e. entire dataset or subset defined by specific factor combinations

Estimated Variance-Covariance Matrix

The asymptotic estimated covariance matrix for parameter estimates is obtained from the Fisher scoring estimation method. Specifically, the asymptotic covariance matrix is given by the inverse of the information matrix. For more information, see Overview of Logistic Regression Model.

Estimated Correlation Matrix

The estimated correlation matrix is the normalized covariance matrix, like each element $\hat{\sigma}_{ij}$ of the estimated variance-covariance matrix $\hat{\Sigma}_{\hat{\beta}}$, is divided by the product of $\hat{\sigma}_{ii}$ and $\hat{\sigma}_{jj}$.

Parameter Estimates

Estimate

The maximum likelihood estimate $\hat{\boldsymbol{\beta}}$ is the value of $\boldsymbol{\beta}$ that maximizes the likelihood function over a valid range of parameter values. The computation of the maximum likelihood estimate requires an iterative computational procedure. The Fisher scoring

algorithm is used, and the parameter estimates are the values obtained from the last iteration of the algorithm. For more information, see Overview of Logistic Regression Model.

Standard Error

The standard error estimates are the square roots of the diagonals of the estimated variance-covariance matrix. For more information, see Estimated Variance-Covariance Matrix.

Wald Statistic

For each parameter estimate, the Wald statistic is computed as the square of the ratio of the parameter estimate over the estimated standard error of the estimated coefficient.

$$\text{Wald} = \left(\frac{\hat{\beta}_i}{\hat{\sigma}_{\hat{\beta}_i}} \right)^2$$

where,

$\hat{\sigma}_{\hat{\beta}_i}$ = Estimated standard error of the i^{th} estimated coefficient

The Wald statistic is asymptotically distributed as χ^2 with one degree of freedom and can be used to test the hypothesis that the parameter is equal to 0. The estimated standard error of the i^{th} estimated coefficient, $\hat{\sigma}_{\hat{\beta}_i}$, is the square root of the i^{th} diagonal element of the estimated covariance matrix, $\hat{\Sigma}_{\hat{\beta}}$, that is, $\hat{\sigma}_{ii}$.

Upper and Lower Confidence Limit for the Estimate

The confidence limits for an individual parameter are given by:

$$\hat{\beta}_i \pm Z_{\alpha/2}\hat{\sigma}_{\hat{\beta}_i}$$

where,

$Z_{\alpha/2}$ = Z-value from the standard normal distribution beyond which the area in the tail is equal to $\alpha/2$

$\hat{\sigma}_{\hat{\beta}_i}$ = Estimated standard error of the i^{th} estimated coefficient

Odds Ratio

The odds ratio for a continuous predictor or categorical predictor with reference or overparameterized coding is obtained by exponentiating the parameter estimate:

$$\text{odds ratio} = e^{\hat{\beta}_i}$$

The odds ratio gives the relative amount by which the odds of the outcome increase (if odds ratio >1) or decrease (if odds ratio <1) when the value of predictor is increased by 1 unit.

The odds ratio for a sigma-restricted coded categorical variable is given by:

$$\text{odds ratio} = e^{\hat{\beta}_i + \sum_{j=1}^{k}\hat{\beta}_j}$$

where,

k = Number of levels of effect - 1

Confidence Limits for Odds Ratios

The confidence limits for the odds ratio for a continuous predictor or categorical predictor with reference or overparameterized coding are given by:

$$e^{\hat{\beta}_i \pm Z_{\alpha/2}\hat{\sigma}_{\hat{\beta}_i}}$$

where,

$Z_{\alpha/2}$ = Z-value from the standard normal distribution beyond which the area in the tail is equal to $\alpha/2$

$\hat{\sigma}_{\hat{\beta}_i}$ = Estimated standard error of the i^{th} estimated coefficient

The confidence limits for the odds ratio of a sigma-restricted coded categorical variable are given by:

$$e^{\hat{\beta}_i + \sum_{j=1}^{k} \hat{\beta}_j \pm \hat{\sigma}_{ln(OR)} Z_{\alpha/2}}$$

where,

$\hat{\sigma}_{ln(OR_i)} = \sqrt{\boldsymbol{l}^T \hat{\Sigma}_{\hat{\beta}} \boldsymbol{l}}$

$\hat{\Sigma}_{\hat{\beta}}$ = Estimated covariance matrix of parameter estimates

$\boldsymbol{l}$ = p by 1 vector of 1's with a 2 in the i^{th} position

Classification of Cases and Odds Ratio

A classification matrix displays the frequencies of the predicted and observed classification of cases and percentage of correct predictions based on the logistic regression model.

An observation is predicted as 1 if the predicted probability is greater than 0.5 else it is predicted as 0. The format of the classification matrix is as follows:

	Predicted 0	Predicted 1
Observed: 0	$f_{0,0}$	$f_{0,1}$
Observed: 1	$f_{1,0}$	$f_{1,1}$

where,

f_{ij} = Frequency of those cases observed as i and predicted by the model as j

$i = 0, 1; j = 0, 1$

The odds ratio of the classification matrix is given by:

$$\text{Odds ratio} = \left(\frac{f_{0,0} * f_{1,1}}{f_{0,1} * f_{1,0}} \right)$$

The log odds ratio is the natural log of the odds ratio as defined above.

The Percent correct is given as the percentage correct for a given observed category i and predicted category j.

$$\text{Percent correct} = \left(\frac{f_{i,i}}{f_{i,i} + f_{i,j}} * 100 \right)$$

Iteration Results

Iteration results show the parameter estimates and the model log-likelihood at each iteration during Maximum Likelihood Estimation using the Fisher Scoring update formula described in the Overview of Logistic Regression Model. Specifically, each column of the spreadsheet represents one iteration, and the rows show the respective parameter estimates and model log-likelihood at that iteration.

Goodness of Fit

Deviance

The deviance goodness of fit test is based on the likelihood ratio test of the current model against a full or saturated model. The statistic has an asymptotic χ^2 distribution with degrees of freedom equal to the number of observations minus the number of parameters estimated.

$$D = 2\sum_{i=1}^{N}\left[w_i\left(y_i\ ln\left(\frac{y_i}{\hat{\pi}_i}\right) + (1-y_i)\ ln\left(\frac{1-y_i}{1-\hat{\pi}_i}\right)\right)\right]$$

where,

$\hat{\pi}$ = Predicted probability of the i^{th} case

w_i = Weight for i^{th} case

Scaled Deviance

The scaled deviance statistic is the respective statistic divided by the current estimate of the dispersion parameter. The statistic has an asymptotic χ^2 distribution with degrees of freedom equal to the number of observations minus the number of parameters estimated.

$$D_{scaled} = \frac{D}{\hat{\phi}}$$

where,

$\hat{\phi}$ = Current estimate of dispersion parameter

For a review of how the dispersion parameter is estimated see Dispersion.

Pearson χ^2

The Pearson χ^2 statistic has an asymptotic χ^2 distribution with degrees of freedom equal to the number of observations minus the number of parameters estimated.

$$\text{Pearson } \chi^2 = \sum_{i=1}^{N} \frac{w_i(y_i - \hat{\pi}_i)^2}{\hat{\pi}_i(1-\hat{\pi}_i)}$$

where,

$\hat{\pi}_i$ = Predicted probability of the i^{th} case

w_i = Weight for i^{th} case

Scaled Pearson χ^2

The scaled Pearson χ^2 statistic is the respective statistic divided by the current estimate of the dispersion parameter. The statistic has an asymptotic χ^2 distribution with degrees of freedom equal to the number of observations minus the number of parameters estimated.

$$\text{Pearson } \chi^2_{scaled} = \frac{\text{Pearson } \chi^2}{\hat{\phi}}$$

where,

$\hat{\phi}$ = Current estimate of dispersion parameter

For a review of how the dispersion parameter is estimated see Dispersion.

Akaike Information Criterion (AIC)

$$AIC = -2l + 2p$$

where,

l = Log-likelihood of model

p = Number of parameters in the model

Bayesian Information Criterion (BIC)

$$BIC = -2l + p * ln(n)$$

where,

$l =$ Log-likelihood of model

$p =$ Number of parameters in the model

$n =$ Number of observations

Cox-Snell R^2

In linear regression using ordinary least squares, a measure of goodness of fit is R^2, which represents the proportion of variance explained by the model. Using logistic regression, an equivalent statistic does not exist, and therefore several pseudo-R^2 statistics have been developed. The Cox-Snell R^2 is a pseudo-R^2 statistic, and the ratio of the likelihoods reflects the improvement of the full model over the intercept only model with a smaller ratio reflecting greater improvement. It is given by:

$$\text{Cox-Snell } R^2 = 1 - \left[\frac{L(R)}{L(F)}\right]^{2/N}$$

where,

$L(R) =$ Likelihood of intercept only model

$L(F) =$ Likelihood of specified model

$N =$ Number of observations

Nagelkerke R^2

The Nagelkerke R^2 adjusts the Cox-Snell R^2 so the range of possible values extends to one.

$$\text{Nagelkerke } R^2 = \frac{1-\left[\frac{L(R)}{L(F)}\right]^{2/N}}{1-L(R)^{2/N}}$$

where,

$L(R) =$ Likelihood of intercept only model

$L(F) =$ Likelihood of specified model

$N =$ Number of observations

Log-Likelihood

The first spreadsheet generated from Statistica® when clicking the **Goodness of fit** button from the GLZ Results dialog box contains, in addition to the above, the Log-likelihood statistic as shown in the Overview.

Testing Global Null Hypothesis

The second spreadsheet generated from the **Goodness of fit** button from the GLZ Results dialog box is for global null hypothesis tests, which are used to test whether the parameter estimates are significantly different from zero. Each statistic is assumed to have a χ^2 asymptotic distribution with degrees of freedom equal to the number of restrictions imposed on the model under the null hypothesis.

Likelihood Ratio Test

$$\text{LRT} = 2 * \left[\frac{L(R)}{L(F)} \right]$$

where,

$L(R) =$ Model where all coefficients are set to 0, except for the estimated intercept

$L(F) =$ Specified model including all selected predictors and the estimated intercept

Score Test

The general form of the Score test is given by:

$$\textbf{Score} = s^T(\boldsymbol{\beta}_0)\, I^{-1}(\boldsymbol{\beta}_0)\, s(\boldsymbol{\beta}_0)$$

where,

s = the score vector
I = the information matrix.
s and **I** are evaluated at the null hypothesized values of the parameters.

Specifically for testing the global null hypothesis the parameter vector is set to

$$\boldsymbol{\beta_0} = \left[ln\left(\frac{\hat{\pi}_0}{1-\hat{\pi}_0}\right), 0, 0, \cdots, 0 \right]$$

$$\hat{\pi}_0 = \frac{1}{N}\sum_{i=1}^{N} w_i y_i$$

N = Number of observations

Wald Test

The general form of the Wald Test is given by:

$$\textbf{Wald} = \left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta_0}\right)^T \mathbf{I}^{-1}(\hat{\boldsymbol{\beta}}) \left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta_0}\right)$$

where,

$\hat{\boldsymbol{\beta}}$ = Maximum likelihood estimates of the parameters
I = Information matrix evaluated at the maximum likelihood estimates of the parameters

Specifically for testing the global null hypothesis the parameter vector is set to

$$\boldsymbol{\beta_0} = \left[ln\left(\frac{\hat{\pi}_0}{1-\hat{\pi}_0}\right), 0, 0, \cdots, 0 \right]^T$$

$$\hat{\pi}_0 = \frac{1}{N}\sum_{i=1}^{N} w_i y_i$$

N = Number of observations

Hosmer-Lemeshow Test

To calculate the Hosmer-Lemeshow goodness of fit test, the data are sorted first in their increasing order of predicted probability. The observations are divided into groups using the setting for HL Groups, with a default of 10 groups. The groups are constructed based on the percentiles of the estimated probabilities. See Hosmer and Lemeshow (2000) for details.

The Hosmer-Lemeshow statistic is distributed as a χ^2 distribution with g-2 degrees of freedom given the null hypothesis.

The Hosmer-Lemeshow goodness of fit test statistic is given by:

$$\chi^2_{HL} = \sum_{j=1}^{g} \frac{(O_j - N_j \bar{\pi}_j)^2}{(N_j \bar{\pi}_j (1 - \bar{\pi}_j))}$$

where,

g = Number of groups

N_j = Number of observations in the jth group

0_j = Number of responses in the jth group

$\bar{\pi}_j$ = Average of the predicted probability for the jth group

Aggregation

When the Aggregation option is selected, the statistics are computed in terms of predicted frequencies. In models with categorical response variables, statistics can be computed in terms of the raw data or for aggregated frequency counts. In the binomial case, the response variable has two possible values: 0 or 1. Accordingly, predicted values should be computed that fall in the range from 0 to 1. If the Aggregation check box is selected, Statistica® considers the aggregated (tabulated) data set. In that case, the response variable is a frequency count, reflecting the number of observations that fall into the respective categories. As an example consider the following data:

x1	x2	Y

1.5	2	1
2.5	1	0
2.5	1	1
1.5	2	0
2.5	1	1
1.5	2	1

After aggregation, the data is represented as follows:

x1	x2	y1	y0	total
1.5	2	2	1	3
2.5	1	2	1	3

Selecting the Aggregation check box also affects the computation (and display) of predicted and residual values.

Overdispersion

In models with binary response variables as in logistic regression, the default dispersion parameter (1.0) for the model may not be adequate. You can select the Overdispersion check box and then select either the Pearson or deviance option button as the estimate of the dispersion parameter.

If you specify deviance, the dispersion parameter is estimated by the deviance divided by its degrees of freedom. If you specify Pearson χ^2, the dispersion parameter is estimated by the Pearson χ^2 statistic divided by its degrees of freedom. The adjustment is reflected in the scale parameter as it is proportional to the dispersion parameter.

Changing the overdispersion parameter affects the computational values of the parameter variances and covariances and the model likelihood, and all related statistics (for example, standard errors, prediction errors).

Residuals

In linear regression, residual analysis is utilized to see if the linear fit of the model is appropriate through identifying poorly fitted values, with a large number of poorly fitted values indicating poor fit of the model. The benefit of residual analysis is instrumental in assessing model fit, and this benefit is needed in logistic regression. However, residuals are not normally distributed and require a more careful analysis. Forms of residuals specific to logistic regression are included in Statistica® and are described below.

Basic Residuals

Raw

$$r_i = y_i - \hat{\pi}_i$$

where,

y_i = Observed binary response of the i^{th} case

$\hat{\pi}_i$ = Predicted probability for the i^{th} case

Pearson

$$r_{i,p} = \frac{\sqrt{w_i}(y_i - \hat{\pi}_i)}{\sqrt{\hat{\pi}_i(1-\hat{\pi}_i)}}$$

where,

w_i = Weight for the i^{th} case

y_i = Observed binary response of the i^{th} case

$\hat{\pi}_i$ = Predicted probability for the i^{th} case

Deviance

$$r_{i,d} = w_i(sign(y_i - \hat{\pi}_i))\sqrt{-2\left[y_i ln(\hat{\pi}_i) + (1-y_i) ln(1-\hat{\pi}_i)\right]}$$

where,

$$\text{sign(x)} = \begin{cases} 1, & \text{if x>1} \\ -1, & \text{if x<0} \\ 0, & \text{otherwise} \end{cases}$$

w_i = Weight for the i^{th} case

y_i = Observed binary response of the i^{th} case

$\hat{\pi}_i$ = Predicted probability for the i^{th} case

Predicted Values

Response

The response of the i^{th} case is the observed value of y_i.

Predicted Value

The predicted value of the i^{th} case is the predicted probability calculated using the specified model.

$$\hat{\pi}_i(\mathbf{X}) = \left[\frac{e^{\boldsymbol{x}_i^T\hat{\boldsymbol{\beta}}}}{1+e^{\boldsymbol{x}_i^T\hat{\boldsymbol{\beta}}}}\right]$$

where,

$\boldsymbol{x_i}$ = vector of predictors for i^{th} case

Linear Predictor

The linear predictor of the i^{th} case is given by:

$$\boldsymbol{x}_i^T\hat{\boldsymbol{\beta}}$$

where $\boldsymbol{x_i}$ = vector of predictors for i^{th} case

Standard Error of Linear Predictor

The standard error of the linear predictor of the i[th] case, $\boldsymbol{x_i^T\hat{\beta}}$, is given by:

$$\sqrt{\boldsymbol{x_i^T}\hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\beta}}}\boldsymbol{x_i}}$$

where,

$\boldsymbol{x_i}$ = Vector of predictors for i^{th} case

$\hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\beta}}}$ = Estimated variance-covariance matrix of the estimated coefficients

Upper/Lower Confidence Limits

The 100*(1-α)% confidence limits for the predicted probability for the i[th] case is given by:

$$\frac{e^{\boldsymbol{x_i^T}\hat{\boldsymbol{\beta}}\pm Z_{\alpha/2}\hat{\sigma}_{\boldsymbol{x_i^T}\hat{\boldsymbol{\beta}}}}}{1+e^{\boldsymbol{x_i^T}\hat{\boldsymbol{\beta}}\pm Z_{\alpha/2}\hat{\sigma}_{\boldsymbol{x_i^T}\hat{\boldsymbol{\beta}}}}}$$

where,

$Z_{\alpha/2}$ = 1-$\alpha/2$ quantile of the standard normal distribution

$\hat{\sigma}_{\boldsymbol{x_i^T}\hat{\boldsymbol{\beta}}} = \sqrt{\sum_{j=0}^{p-1} x_{ij}^2\sigma_{\hat{\beta}_j}^2 + 2\sum_{j=0}^{p-2}\sum_{k=j+1}^{p-1} x_{ij}x_{ik}\sigma_{\hat{\beta}_j,\hat{\beta}_k}}$

Leverage

Leverage values can detect outliers in the design space. Let h_i denote the leverage of the i^{th} case.

$\text{Leverage}_i = h_i = i^{th}$ diagonal element of "Hat Matrix"

where,

Hat Matrix = $V^{1/2}X(X^TVX)^{-1}V^{1/2}$

V = Diagonal matrix with elements $\hat{\pi}(1-\hat{\pi})$

X = Design matrix

Studentized Pearson Residual

$$r_{i,sp} = \frac{\sqrt{w_i}(y_i - \hat{\pi}_i)}{\sqrt{\hat{\pi}_i(1-\hat{\pi}_i)(1-h_i)\hat{\phi}}}$$

where,

w_i = Weight for the i^{th} case

y_i = Observed binary response of the i^{th} case

$\hat{\pi}_i$ = Predicted probability for the i^{th} case

$h_i = i^{th}$ diagonal element of hat matrix

$\hat{\phi}$ = Estimate of dispersion parameter

Studentized Deviance Residual

$$r_{i,sd} = w_i \sqrt{\frac{r_{i,d}}{\hat{\phi}}} \text{ sign } (y_i - \hat{\pi}_i)$$

where,

$$\text{sign(x)} = \begin{cases} 1, & \text{if x>1} \\ -1, & \text{if x<0} \\ 0, & \text{otherwise} \end{cases}$$

w_i = Weight for the i^{th} case

y_i = Observed binary response of the i^{th} case

$\hat{\pi}_i$ = Predicted probability for the i^{th} case

$\hat{\phi}$ = Estimate of dispersion parameter

Likelihood Residual

$$r_{i,L} = \text{sign}(y_i - \hat{\pi}_i)\sqrt{(1-h_i)r_{i,d}^2 + h_i r_{i,p}^2}$$

where,

$$\text{sign(x)} = \begin{cases} 1, & \text{if x>1} \\ -1, & \text{if x<0} \\ 0, & \text{otherwise} \end{cases}$$

$r_{i,d}$ = Deviance residual for i^{th} case
$r_{i,p}$ = Pearson residual for i^{th} case
$\hat{\pi}_i$ = Predicted probability for i^{th} case
h_i = Leverage of the i^{th} case

Other Obs. Stats

General Cook's Distance

General Cook's Distance (D) measures the difference in the residuals of all cases due to removing the i^{th} observation. Cook's D for the i^{th} observation is given by:

$$\text{Cook's D}_i = \left(\frac{1}{p}\right)\left(\frac{h_i}{(1-h_i)^2\hat{\phi}}\right)r_{i,sp}^2$$

where,

p = Number of parameters
h_i = Leverage of the i^{th} case
$r_{i,sp}$ = Studentized Pearson residual of the i^{th} case
$\hat{\phi}$ = Estimate of dispersion parameter

Differential χ^2

The differential χ^2 residual measures the difference in the Pearson χ^2 statistic due to removing the ith observation.

$$r_{i,D\chi^2} = r_{i,sp}^2$$

where,

$r_{i,sp}$ = Studentized Pearson residual of the i^{th} case

Differential Deviance

The differential deviance residuals measure the difference in the deviance statistic due to removing the i[th] observation.

$$r_{i,DD} = r_{i,d}^2 + \hat{\phi} r_{i,sp}^2$$

where,

$\hat{\phi}$ = Dispersion estimate

$r_{i,sp}$ = Studentized Pearson residual for i^{th} case

$r_{i,d}$ = Deviance residual for i^{th} case

Differential Likelihood

The differential likelihood residuals measure the difference in the likelihood due to removing the i^{th} observation.

$$r_{i,DL} = r_{i,D\chi^2} h_i + (1 - h_i) r_{i,DD}$$

where,

$r_{i,D\chi^2}$ = Differential χ^2 residual of the i^{th} case

h_i = Leverage of the i^{th} case

$r_{i,DD}$ = Differential deviance residual of the i^{th} case

ROC Curve

The ROC curve is often used as a measure of goodness-of-fit to evaluate the fit of a logistic regression model with a binary classifier. The plot is constructed using the true positive rate (rate of events that are correctly predicted as events and also called Sensitivity) on the y- axis against the false positive rate (rate of non-events predicted to be events also called 1- Specificity) on the x-axis for the different possible cutoff points based on the model $\hat{\pi}_i(\boldsymbol{x})$.

The range of values for the predicted probabilities $\hat{\pi}_i(\boldsymbol{x})$, Sensitivity and 1-Specificity, is provided in the corresponding spreadsheet. For each i^{th} predicted probability, the algorithm iterates through each case to classify it as events or non-events using the

predicted probability as the threshold value. The Sensitivity and 1-Specificity are calculated by:

$$\text{Sensitivity} = \frac{TP}{TP+FN}$$

$$\text{1-Specificity} = 1 - \frac{TN}{TN+FP}$$

where,

TP (True Positives) = Number of correctly predicted events
FP (False Positives) = Number of Non-events predicted as event
FN (False Negatives) = Number of events predicted as non-events
TN (True Negatives) = Number of correctly predicted events

The area under the ROC curve, sometimes also referred as Area Under Curve (AUC), gives a summary measure of the predictive power of the model, that is, the larger the AUC value (closer to 1), the better is the overall performance of the model to correctly classify the cases.

The AUC is computed using a nonparametric approach known as the trapezoidal rule. The area bounded by the curve and the baseline is divided into a series of trapezoidal regions (intervals) based on (1-Specifity, Sensitivity) points. The area is calculated for each region, and by summing the areas of the region, the total area under the ROC curve is computed.

Lift Chart

The lift chart provides a visual summary of the usefulness of the information provided by the logistic model for predicting a binomial dependent variable. The chart summarizes the utility (gain) that you can expect by using the respective predictive model shown by the Model curve as compared to using baseline information only.

Analogous lift values (Y-coordinate) that are calculated as the ratio between the results obtained with and without the predicted model can be computed for each percentile of the population sorted in descending order by the predicted value, that is, cases classified into the respective category with the highest classification probability. Each lift value indicates the effectiveness of a predictive model, that is, how many times it is better to use the model than not using one.

Let n be the number of true positives that appear in the top k% of the sorted list.

Let r be the number of true positives that appear if we sort randomly, which is equal to the product of the total number of true positives and k.

Then,

$$\text{Lift Value} = \frac{n}{r};\ \text{for top k\% of the cases}$$

The values for different percentiles can be connected by a line that typically descend slowly and merge with the baseline.

Observed, Unweighted Means

Means of Groups

Observed, unweighted means are available for full factorial designs involving effects of 2 or more categorical factors. When cell frequencies in a multi-factor ANOVA design are equal, no difference exists between weighted and unweighted means. When cell frequencies are unequal, the unweighted means are calculated without adjusting for the differences between the sub-group frequencies. These are computed by averaging the means across the levels and combinations of levels of the factors not used in the marginal means table (or plot), and then dividing by the number of means in the average.

Let k = Number of factors in the model
Without loss of generality, let the given effect consist of the first l factors
Let f_{i_j} denote the i^{th} level of the j^{th} factor

Then, the unweighted mean is computed as

$$\bar{x}_{f_{i_1}, f_{i_2}, \cdots, f_{i_l}} = \frac{1}{r} \sum_{j \in F} \bar{x}_j$$

where,

F = Set of all unique combinations of all k factors that include levels $f_{i_1}, f_{i_2}, \cdots, f_{i_l}$
r = Cardinality of F

Observed, Weighted Means

Means of Groups

The mean of the dependent variable, y, for the j^{th} level of the selected factor is given by:

$$\bar{y} = \frac{1}{N}\sum_{i=1}^{N} w_i y_i$$

where,

$y_i = i^{th}$ observation

$w_i =$ Weight for i^{th} observation

$N =$ Sample size over the appropriate set of cases,
i.e. entire dataset or subset defined by specific factor combinations

Standard Error

The standard error of the mean of the dependent variable, y for the j^{th} level of the selected factor is given by:

$$s_{\bar{y}} = \frac{1}{\sqrt{N}}\sqrt{\frac{\sum_{i=1}^{N} w_i (y_i - \bar{y})^2}{N-1}}$$

where,

$y_i = i^{th}$ observed response

$w_i =$ Weight of the i^{th} observation

$N =$ Sample size over the appropriate set of cases,
i.e. entire dataset or subset defined by specific factor combinations

$\bar{y} =$ Mean of response as calculated above.

Upper and Lower Confidence Limits of the Mean

$$\bar{y} \pm t_{(\alpha/2,N-1)} s_{\bar{y}}$$

where,

$\bar{y}$ = Mean of the response

$t_{(\alpha/2,N-1)} = 1 - \alpha/2$ quantile from Student t distribution with N-1 degrees of freedom

$s_{\bar{y}}$ = Standard error of the mean

N = Sample size over the appropriate set of cases,
i.e. entire dataset or subset defined by specific factor combinations

Predicted Means

The predicted means are the expected value for the generalized linear model. When covariates are present in the model, predicted means are computed from the value of those covariates as set in the Covariate values group box. By default, covariates are set to their respective means.

The computation of the predicted means for each effect in the model is constructed by first generating the L matrix for the given effect. The L matrix is generated exactly in the same manner as in the General Linear Models (GLM) module when computing least square means which is also equivalent to how SAS computes the L matrix. The vector of means on the logit scale for a given effect is then computed as $\boxed{L^T\hat{\beta}}$.

For a given level of an effect, the corresponding row vector *l* from the matrix L is used to compute the predicted mean as given by:

$$\text{Predicted Mean} = \frac{e^{\left(l^T\hat{\beta}\right)}}{1 + e^{\left(l^T\hat{\beta}\right)}}$$

Standard Error

The standard error of $\boldsymbol{l}^T\hat{\boldsymbol{\beta}}$ for a given level of an effect is given by:

$$\hat{\sigma}_{\boldsymbol{l}^T\hat{\boldsymbol{\beta}}} = \sqrt{\boldsymbol{l}^T\hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\beta}}}\boldsymbol{l}}$$

where,

$\hat{\Sigma}_{\hat{\beta}}$ = Covariance matrix of parameter estimates

Upper and Lower Confidence Limits of the Mean

The confidence limits of the predicted mean are given by:

$$\frac{e^{\boldsymbol{l}^T\hat{\boldsymbol{\beta}}\pm\hat{\sigma}_{\boldsymbol{l}^T\hat{\boldsymbol{\beta}}}Z_{\alpha/2}}}{1+e^{\boldsymbol{l}^T\hat{\boldsymbol{\beta}}\pm\hat{\sigma}_{\boldsymbol{l}^T\hat{\boldsymbol{\beta}}}Z_{\alpha/2}}}$$

where,

$\hat{\sigma}_{\boldsymbol{l}^T\hat{\boldsymbol{\beta}}} = \sqrt{\boldsymbol{l}^T\hat{\Sigma}_{\hat{\beta}}\boldsymbol{l}}$

$\hat{\Sigma}_{\hat{\beta}}$ = Estimated covariance matrix of parameter estimates

$Z_{\alpha/2}$ = The 1-α/2 quantile of a standard normal distribution

Model Building Summary

In addition to All Effects, different techniques for automatic model building are available for logistic regression. Specifically, forward stepwise, backward stepwise, forward entry, backward removal, and best-subset search procedures are available in Statistica® and are described below.

Stepwise Logistic Regression

Stepwise Logistic Regression methods, specifically the Forward Stepwise and Backward Stepwise methods, are used to perform a stepwise selection of predictor variables.

During the forward step of stepwise model building, if two or more effects have p-values that are so small as to be virtually indistinguishable from 0, Statistica® selects the effect with the largest score statistic if the degrees of freedom for all effects in question are equal. If the effects differ with respect to the degrees of freedom, the Score statistics are normalized using the Wilson-Hilferty transformation, and the effect with the largest transformed value is entered into the model. For the backward step, if the p-values for two or more effects are virtually indistinguishable from 1, Statistica® removes the effect with the smallest Wald statistic in the case of equal degrees of freedom and the smallest normalized value in the case of unequal degrees of freedom.

Forward Stepwise Method

Let p be the number of total effects considered for the model. Let k be the number of effects currently in the model. Let p1, enter and p2, remove be specified by the user. The Forward Stepwise Method starts with no effects in the model and consists of two alternating steps.

1. Entry step: For each of the remaining p-k effects not in the model, construct the p-k score statistics. See Testing Global Null Hypothesis, for a description of the score statistic. Enter the effect with the smallest p-value less than p1, enter.

2. Removal step: For all of the effects in the model compute the Wald statistic. See Testing Global Null Hypothesis, for a description of the score statistic. The effect with the largest p-value greater than p2, remove is removed from the model.

The procedure continues alternating between both steps until no further effects can either be added or removed or until the maximum number of steps (see the **Max. steps** option on the **Advanced** tab of the GLZ General Custom Design dialog box) has been reached.

Backward Stepwise Method

The Backward Stepwise Method is analogous to the Forward Stepwise Method; the only difference is that the initial model contains all the potential effects and begins attempting to remove effects and then subsequently attempts to enter them.

Forward Entry Method

The Forward Entry Method is the simplified version of Forward Stepwise Method. In this method, the step to test whether a variable already entered into the model should be removed is omitted. Specifically, the Forward Entry Method only allows for effects to be entered into the model and does not allow for the removal of the variable after it is added to the model.

Backward Entry Method

The Backward Entry Method is the simplified version of the Backward Stepwise Method. It is the exact opposite of the Forward Entry Method.

As in the Backward Stepwise Method, this method begins with the model containing all potential predictors. It identifies the variable with the largest p-value, and if the p-value is greater than the p2, remove, the variable is dropped. This method eliminates the step of forward entering, or re-entering of the variable once it has been removed.

Wilson-Hilferty transformation

The Wilson-Hilferty transformation method transforms a variable to the Z-scale so that their p-values are closely approximated. This transformation, therefore, enables the comparison of the statistical significance of the values with different degrees of freedom. The transformation is given by:

$$W(y) = \frac{\left(\frac{Y}{n}\right)^{1/3} - \left(1 - \left(\frac{1}{9}\right)\right)\left(\frac{2}{n}\right)}{\sqrt{\left(\frac{1}{9}\right)\left(\frac{2}{n}\right)}}$$

where,

$Y =$ χ^2 statistic

$n =$ Degrees of freedom

Best Subsets

The best subsets search method can be based on three different test statistics: the score statistic, the model likelihood, and the AIC. Note that, since the score statistic does not require iterative computations, best subset selection based on the score statistic is computationally fastest, while selection based on the other two statistics usually provides more accurate results.

In general, a model containing p possible predictors has 2p-1 possible subsets of predictors available for model consideration. In the best subsets method, one of the aforementioned three test statistics is computed for every possible predictor subset and a spreadsheet is generated such that the subsets are sorted from best to worst by test statistic. Due to the number of possible subsets to consider as they grow exponentially with increasing number of effects, Statistica® only performs a best subset search when the number of effects in the model is 13 or fewer.

TIBCO Documentation and Support Services

For information about this product, you can read the documentation, contact TIBCO Support, and join TIBCO Community.

How to Access TIBCO Documentation

Documentation for TIBCO products is available on the TIBCO Product Documentation website, mainly in HTML and PDF formats.

The TIBCO Product Documentation website is updated frequently and is more current than any other documentation included with the product.

Product-Specific Documentation

The following documentation for this product is available on TIBCO Statistica® Product Documentation page and TIBCO Statistica®- All Servers Product Documentation page:

- *TIBCO Statistica® Release Notes*
- *TIBCO Statistica® Installation*
- *TIBCO Statistica® Quick Reference*
- *TIBCO Statistica® Product Traceability*
- *TIBCO Statistica® Configuration for Windows Server 2019*
- *TIBCO Statistica® Data Entry Administration*
- *TIBCO Statistica® Server Administrator's Guide*
- *TIBCO Statistica® Options Configuration*
- *TIBCO Statistica® R Integration: Features and Options*
- *TIBCO Statistica® Security Guide*
- *TIBCO Statistica® User Guide*
- *TIBCO Statistica® Enterprise Manager Guide*
- *TIBCO Statistica® Statistica Object Model Guide*

- *TIBCO Statistica® Logistic Regression Formula Guide*
- *TIBCO Statistica® Stability Analysis Formula Guide*
- *TIBCO Statistica® Stepwise Model Builder Formula Guide*
- *TIBCO Statistica® Weight of Evidence Formula Guide*

The following documentation for this product is available on TIBCO® Data Science for TIBCO Spotfire® Analyst Product Documentation page:

- *TIBCO® Data Science for TIBCO Spotfire® Analyst Release Notes*
- *TIBCO® Data Science for TIBCO Spotfire® Analyst User Guide*

The following documentation for this product is available on TIBCO® Data Science Service for TIBCO Spotfire® Product Documentation page:

- *TIBCO® Data Science Service for TIBCO Spotfire® Release Notes*
- *TIBCO® Data Science Service for TIBCO Spotfire® User Guide*

How to Contact TIBCO Support

Get an overview of TIBCO Support. You can contact TIBCO Support in the following ways:

- For accessing the Support Knowledge Base and getting personalized content about products you are interested in, visit the TIBCO Support website.
- For creating a Support case, you must have a valid maintenance or support contract with TIBCO. You also need a user name and password to log in to TIBCO Support website. If you do not have a user name, you can request one by clicking **Register** on the website.

How to Join TIBCO Community

TIBCO Community is the official channel for TIBCO customers, partners, and employee subject matter experts to share and access their collective experience. TIBCO Community offers access to Q&A forums, product wikis, and best practices. It also offers access to extensions, adapters, solution accelerators, and tools that extend and enable customers to gain full value from TIBCO products. In addition, users can submit and vote on feature requests from within the TIBCO Ideas Portal. For a free registration, go to TIBCO Community.

Legal and Third-Party Notices