Macro STDLOGOR and STDLOGOR_INT User's Guide

6 January 2020

Statistical Methods

For a univariate logistic regression model with a binary response variable and covariate z having sample standard deviation σ_z and regression parameter estimate $\hat{\beta}$ with estimated variance $\hat{\sigma}_\varepsilon^2$, the standardized log odds ratio, that is, the change in the log odds ratio associated with a one standard deviation increase in the covariate value, is $\hat{\beta}_\sigma = \sigma_z \hat{\beta}$. The absolute value of $\hat{\beta}_\sigma$ can also be expressed as the sample standard deviation of the values $z_i\hat{\beta},\ i = 1, 2, \ldots, n$, where the z_i are the sample covariate values, that is,

$$|\hat{\beta}_\sigma| = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left\{\left(z_i - \overline{z}\right)\hat{\beta}\right\}^2}$$

and $\overline{z} = (1/n)\sum_{i=1}^{n} z_i$.

Generalizing to a multivariate Cox regression model with covariate p-vector $\mathbf{z}_i = \begin{pmatrix} z_{i1} & z_{i2} & \cdots & z_{ip} \end{pmatrix}^{\mathrm{T}}$ and regression parameter estimate vector $\hat{\boldsymbol{\beta}} = \begin{pmatrix} \hat{\beta}_1 & \hat{\beta}_2 & \cdots & \hat{\beta}_p \end{pmatrix}^{\mathrm{T}}$ with $p \times p$ estimated covariance matrix $\hat{\mathbf{V}}$, the corresponding standardized absolute log odds ratio B_σ can be estimated consistently by the sample standard deviation of the inner products $\mathbf{z}_i^{\mathrm{T}}\hat{\boldsymbol{\beta}}$, that is,

$$\hat{B}_\sigma = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left\{\left(\mathbf{z}_i^{\mathrm{T}} - \overline{\mathbf{z}}^{\mathrm{T}}\right)\hat{\boldsymbol{\beta}}\right\}^2}$$

where $\overline{\mathbf{z}} = (1/n)\sum_{i=1}^{n} \mathbf{z}_i$. If we think of $\mathbf{z}^{\mathrm{T}}\hat{\boldsymbol{\beta}}$ as a risk score with a population distribution, then the standardized absolute log odds ratio $\hat{B}_\sigma$ is an estimate of the absolute value of the shift in the log odds associated with a one standard deviation change in the risk score. Using a matrix formulation, this is equivalent to $\hat{B}_\sigma = \sqrt{\hat{\boldsymbol{\beta}}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{z}}\hat{\boldsymbol{\beta}}}$, where $\boldsymbol{\Sigma}_{\mathbf{z}} = \sum_{i=1}^{n}(\mathbf{z}_i - \overline{\mathbf{z}})(\mathbf{z}_i - \overline{\mathbf{z}})^{\mathrm{T}} \big/ (n-1)$ is the sample covariance matrix of the covariate vector. A consistent estimator of the absolute standardized odds ratio $\exp(B_\sigma)$ is $\exp\left(\hat{B}_\sigma\right)$.

A size α test of the point null hypothesis $H_0 : B_\sigma = 0$ can be formed by rejecting H_0 when $\Psi = \hat{\boldsymbol{\beta}}^{\mathrm{T}}\hat{\mathbf{V}}^{-1}\hat{\boldsymbol{\beta}} > \chi_p^2(1-\alpha;0)$, where $\hat{\mathbf{V}}$ is the estimated covariance matrix of $\hat{\boldsymbol{\beta}}$, $\chi_p^2(1-\alpha;0)$ denotes the $1-\alpha$ quantile of the central chi-square distribution with degrees of freedom p equal to the number of covariates in the model. Similarly, we can form a conservative size α test of the interval null hypothesis H_0: $B_\sigma^2 \le \eta^2$ versus the alternative H_1: $B_\sigma^2 > \eta^2$ by rejecting H_0 when $\Psi > \chi_p^2\left(1-\alpha;\nu(\eta^2)\right)$, where $\chi_p^2(1-\alpha;\nu)$ denotes the $1-\alpha$ quantile of the noncentral chi-square distribution with p degrees of freedom and noncentrality parameter ν, $\nu(\eta^2) = \eta^2 \big/ \min\left\{\lambda_1, \lambda_2, \ldots, \lambda_p\right\}$, and $\lambda_1, \lambda_2, \ldots, \lambda_k$ are the eigenvalues of $\boldsymbol{\Sigma}_{\mathbf{z}}^{1/2}\hat{\mathbf{V}}\boldsymbol{\Sigma}_{\mathbf{z}}^{1/2}$. Here $\boldsymbol{\Sigma}_{\mathbf{z}}^{1/2}$ is a symmetric square root of $\boldsymbol{\Sigma}_{\mathbf{z}}$, that is, a matrix satisfying $\boldsymbol{\Sigma}_{\mathbf{z}}^{1/2}\boldsymbol{\Sigma}_{\mathbf{z}}^{1/2} = \boldsymbol{\Sigma}_{\mathbf{z}}$.

By inverting the interval hypothesis tests, we can derive the $100(1-\alpha)\%$ confidence interval (η_L, η_U) for the absolute standardized log odds ratio B_σ, where $\eta_L^2 = \sup\left\{\eta^2 : \Psi > \chi_p^2\left(1-\alpha/2; \nu^{(+)}(\eta^2)\right)\right\}$, $\nu^{(+)}(\eta^2) = \eta^2 \big/ \min\left\{\lambda_1, \lambda_2, \ldots, \lambda_p\right\}$, $\eta_U^2 = \inf\left\{\eta^2 : \Psi < \chi_p^2\left(\alpha/2; \nu^{(-)}(\eta^2)\right)\right\}$, and $\nu^{(-)}(\eta^2) = \eta^2 \big/ \max\left\{\lambda_1, \lambda_2, \ldots, \lambda_p\right\}$. A confidence interval for the absolute standardized odds ratio is then (e^{η_L}, e^{η_U}).

The interval hypothesis test just described (but not the point null hypothesis test) is conservative in the sense that the true size of the test is less than or equal to α because it uses the estimated noncentrality parameter $\nu(\eta^2) = \max\limits_{\boldsymbol{\gamma}^{\mathrm{T}}\boldsymbol{\gamma} \le \eta^2} \sum_{k=1}^{p} \left(\mathbf{e}_k^{\mathrm{T}}\boldsymbol{\gamma}\right)^2 \big/ \lambda_k = \eta^2 \big/ \min\left\{\lambda_1, \lambda_2, \ldots, \lambda_p\right\}$, the absolute maximum over all possible directions in p-space of $\boldsymbol{\gamma} = \boldsymbol{\Sigma}_{\mathbf{z}}^{1/2}\boldsymbol{\beta}$, the transformed true regression parameter vector. Some of these directions may be quite unlikely given the estimate $\hat{\boldsymbol{\gamma}}$. To produce a less conservative test, we can restrict attention to composite null hypotheses about $B_\sigma^2 = \boldsymbol{\gamma}^{\mathrm{T}}\boldsymbol{\gamma}$ in the subset of vectors $\boldsymbol{\gamma}$ that have a direction that is reasonably close to the direction of $\hat{\boldsymbol{\gamma}}$. Transforming $\hat{\mathbf{c}} = \mathbf{E}^{\mathrm{T}}\hat{\boldsymbol{\gamma}}$, where $\mathbf{E}$ is the matrix of eigenvectors of $\boldsymbol{\Sigma}_{\mathbf{z}}^{1/2}\hat{\mathbf{V}}\boldsymbol{\Sigma}_{\mathbf{z}}^{1/2}$, we consider a vector of the same length as $\hat{\mathbf{c}}$ to be "reasonably close" to $\hat{\mathbf{c}}$ if it falls in the p-dimensional

Scheffé 95% confidence ellipsoid for $\mathbf{c} = \mathbf{E}^{\mathrm{T}}\boldsymbol{\gamma}$, that is, $\left(\mathbf{c}-\hat{\mathbf{c}}\right)^{\mathrm{T}} \boldsymbol{\Lambda}^{-1}\left(\mathbf{c}-\hat{\mathbf{c}}\right) \le \chi_p^2(0.95)$, where $\boldsymbol{\Lambda}$ is a diagonal matrix with the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_p$ of $\boldsymbol{\Sigma}_{\mathbf{z}}^{1/2}\hat{\mathbf{V}}\boldsymbol{\Sigma}_{\mathbf{z}}^{1/2}$ on the main diagonal. Finding the vector $\tilde{\mathbf{c}}_{\max} = \left(\tilde{c}_1^{(\max)}, \tilde{c}_2^{(\max)}, \ldots, \tilde{c}_p^{(\max)}\right)^{\mathrm{T}}$ that maximizes the noncentrality parameter $\sum_{k=1}^{p}\left(c_k\right)^2 / \lambda_k$ while remaining within the Scheffé 95% ellipsoid, we can form an asymptotic size α test of the interval null hypothesis $H_0:\ B_\sigma^2 \le \eta^2$ versus the alternative $H_1:\ B_\sigma^2 > \eta^2$ by rejecting H_0 when $\Psi > \chi_p^2\left(1-\alpha; \nu_{Sch}^{(\max)}(\eta^2)\right)$, where $\nu_{Sch}^{(\max)}(\eta^2) = \left(\eta^2 / \hat{B}_\sigma^2\right)\sum_{k=1}^{p}\left(\tilde{c}_k^{(\max)}\right)^2 / \lambda_k$. The *p*-value for this test is given by $P_{1-\alpha} = 1 - F_{\chi_p^2\left(\nu_{Sch}^{(\max)}(\eta^2)\right)}(\Psi)$. We will refer to this as the "Scheffé-aligned" method for constructing interval hypothesis tests.

As with the conservative test, we can invert the Scheffé-alignment test to obtain a confidence interval for the absolute standardized log odds ratio. Finding the vector $\tilde{\mathbf{c}}_{\min} = \left(\tilde{c}_1^{(\min)}, \tilde{c}_2^{(\min)}, \ldots, \tilde{c}_p^{(\min)}\right)^{\mathrm{T}}$ that *minimizes* the noncentrality parameter $\sum_{k=1}^{p}\left(c_k\right)^2 / \lambda_k$ while remaining within the Scheffé 95% ellipsoid, and computing the *minimum* noncentrality parameter $\nu_{Sch}^{(\min)} = \sum_{k=1}^{p}\left(\tilde{c}_k^{(\min)}\right)^2 / \lambda_k$, an asymptotic $100(1-\alpha)\%$ confidence interval for the squared standardized log odds ratio B_σ^2 is given by $\left(\eta_{Sch,L}^2, \eta_{Sch,U}^2\right)$, where $\eta_{Sch,L}^2 = \sup\left\{\eta^2 : \Psi > \chi_p^2\left(1-\alpha/2; \nu_{Sch}^{(\max)}(\eta^2)\right)\right\}$ and $\eta_{Sch,U}^2 = \inf\left\{\eta^2 : \Psi < \chi_p^2\left(\alpha/2; \nu_{Sch}^{(\min)}(\eta^2)\right)\right\}$. A confidence interval for the absolute standardized odds ratio is then $(e^{\eta_{Sch,L}}, e^{\eta_{Sch,U}})$.

It can be shown that the variance of the standardized log odds ratio is estimated consistently by $\mathrm{Var}\left(\hat{B}_\sigma\right) = \mathrm{tr}\left(\boldsymbol{\Sigma}_{\mathbf{z}}\hat{\mathbf{V}}\right)$, where tr denotes the trace of a matrix, that is, the sum of its diagonal elements. This quantity is useful in computing regression-to-the-mean-corrected estimates of the absolute standardized odds ratio in true discovery rate degree of association (TDRDA) analysis (Crager, 2010). However, the best way to form confidence intervals for individual absolute standardized odds ratios is with the noncentral chi-square intervals described above.

In univariate proportional oddss regression, $\hat{\beta}$ is asymptotically unbiased for β. However, to generalize the standardized log odds ratio from the univariate case to the multivariate case, we must work with the absolute value or the square of the standardized odds ratio estimate. In the univariate case, $\mathrm{E}\left|\hat{\beta}\right| > 0$ even if $\beta = 0$. In fact, if $\hat{\beta} \sim N(0, \sigma_\varepsilon^2)$ then $\mathrm{E}\left|\hat{\beta}\right| = 2\sigma_\varepsilon \int_0^\infty \left(1/\sqrt{2\pi}\right) x e^{-(1/2)x^2} dx = 0.80\sigma_\varepsilon$, and $\mathrm{E}\left(\hat{\beta}^2\right) = \sigma_\varepsilon^2$. It can be shown that $\mathrm{E}\left[\sum_{i=1}^n \left\{\left(z_i - \bar{z}\right)\hat{\beta}\right\}^2 \Big/ (n-1)\right] = \sigma_z^2\sigma_\varepsilon^2 + \sigma_z^2\beta^2$, so an asymptotically unbiased estimator for $\left(\sigma_z\beta\right)^2$ is $\hat{\beta}_\sigma^{(2)} = \sum_{i=1}^n \left[\left\{\left(z_i - \bar{z}\right)\hat{\beta}\right\}^2 - \left(z_i - \bar{z}\right)^2 \hat{\sigma}_\varepsilon^2\right] \Big/ (n-1)$.

For the multivariate case, where asymptotically $\hat{\boldsymbol{\beta}} \sim N(\boldsymbol{\beta}, \mathbf{V})$, an asymptotically unbiased estimator for the square of the standardized log odds ratio is

$$\hat{B}_\sigma^{(2)} = \frac{1}{n-1}\sum_{i=1}^n \left[\left\{\left(\mathbf{z}_i^{\mathrm{T}} - \bar{\mathbf{z}}^{\mathrm{T}}\right)\hat{\boldsymbol{\beta}}\right\}^2 - \left(\mathbf{z}_i^{\mathrm{T}} - \bar{\mathbf{z}}^{\mathrm{T}}\right)\hat{\mathbf{V}}\left(\mathbf{z}_i - \bar{\mathbf{z}}\right)\right]$$

We can transform back to the log odds ratio scale to get a bias-corrected estimate of the absolute standardized log odds ratio $\hat{B}_\sigma^{(1)} = \sqrt{\hat{B}_\sigma^{(2)} \wedge 0}$, where $a \wedge b$ denotes the maximum of a and b. Note that if $\hat{B}_\sigma^{(2)} < 0$, we set $\hat{B}_\sigma^{(1)} = 0$. Using the matrix formulation, we can write the bias-corrected estimate as $\hat{B}_\sigma^{(2)} = \hat{\boldsymbol{\beta}}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{z}}\hat{\boldsymbol{\beta}} - \mathrm{tr}\left(\boldsymbol{\Sigma}_{\mathbf{z}}\hat{\mathbf{V}}\right)$.

The absolute standardized log odds ratio can also be estimated when the data come from a study with a cohort sampling design. In the analysis of such data, patient *i* from the sample essentially represents w_i patients in the full cohort, so a consistent estimator of the absolute standardized log odds ratio in the overall population is given by

$$\hat{B}_\sigma = \sqrt{\frac{1}{W-1}\sum_{i=1}^n w_i \left\{\left(\mathbf{z}_i^{\mathrm{T}} - \bar{\mathbf{z}}^{\mathrm{T}}\right)\hat{\boldsymbol{\beta}}\right\}^2} \tag{4}$$

where $W = \sum_{i=1}^n w_i$ and $\bar{\mathbf{z}} = \left(1/W\right)\sum_{i=1}^n w_i \mathbf{z}_i$. Using the matrix formulation, $\hat{B}_\sigma^2 = \hat{\boldsymbol{\beta}}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{z}}\hat{\boldsymbol{\beta}}$, where now $\boldsymbol{\Sigma}_{\mathbf{z}} = \sum_{i=1}^n w_i\left(\mathbf{z}_i - \bar{\mathbf{z}}\right)\left(\mathbf{z}_i - \bar{\mathbf{z}}\right)^{\mathrm{T}} \Big/ (W-1)$. We can form tests of the point null hypothesis

$H_0 : B_\sigma^2 = 0$ or the interval null hypothesis $H_0 : B_\sigma^2 \le \eta^2$ as in Section 3, replacing $\hat{\mathbf{V}}$ by an estimate of the covariance matrix of $\hat{\boldsymbol{\beta}}$ based on survey sampling methodology.

Sometimes we may wish to assess the degree of association of a subset of the covariates in the proportional odds regression model while controlling for variation due to other covariates in the model. For example, if we are developing a predictive marker for breast cancer recurrence that we expect to have equal predictive value in node-negative and node-positive patients, and our study involves both types of patients, we might want to include nodal status (or number of positive nodes) as a covariate in the analysis. The calculations described previously can easily be adapted to this situation.

Let $\mathbf{z}_{Ai}$ be the observed vectors of covariates for which we wish to estimate the multivariate standardized log odds ratio and let $\mathbf{z}_{Bi}$ be the observed vectors of covariates for which we wish to adjust the analysis using proportional odds, so that the observed complete covariate vectors are given by $\mathbf{z}_i = \left(\mathbf{z}_{Ai}^{\mathrm{T}} \quad \mathbf{z}_{Bi}^{\mathrm{T}} \right)^{\mathrm{T}}$. Similarly decompose the regression parameter estimate vector $\hat{\boldsymbol{\beta}} = \left(\hat{\boldsymbol{\beta}}_A^{\mathrm{T}} \quad \hat{\boldsymbol{\beta}}_B^{\mathrm{T}} \right)^{\mathrm{T}}$ and the estimated covariance matrix

$$\hat{\mathbf{V}} = \begin{bmatrix} \hat{\mathbf{V}}_{AA} & \hat{\mathbf{V}}_{AB} \\ \hat{\mathbf{V}}_{BA} & \hat{\mathbf{V}}_{BB} \end{bmatrix}$$

Then a consistent estimator of the absolute standardized log odds ratio is

$$\hat{B}_{A\sigma} = \sqrt{\frac{1}{W-1} \sum_{i=1}^{n} w_i \left\{ \left(\mathbf{z}_{Ai}^{\mathrm{T}} - \bar{\mathbf{z}}_A^{\mathrm{T}} \right) \hat{\boldsymbol{\beta}}_A \right\}^2}$$

where $W = \sum_{i=1}^{n} w_i$, $\bar{\mathbf{z}} = (1/W) \sum_{i=1}^{n} w_i \mathbf{z}_{Ai}$. Using the matrix formulation, $\hat{B}_\sigma^2 = \hat{\boldsymbol{\beta}}_A^{\mathrm{T}} \boldsymbol{\Sigma}_{\mathbf{z}A} \hat{\boldsymbol{\beta}}_A$, where now $\boldsymbol{\Sigma}_{\mathbf{z}A} = \sum_{i=1}^{n} w_i \left(\mathbf{z}_{Ai} - \bar{\mathbf{z}}_A \right) \left(\mathbf{z}_{Ai} - \bar{\mathbf{z}}_A \right)^{\mathrm{T}} \Big/ (W-1)$. We can form test of the null hypothesis $H_0 : B_\sigma^2 = 0$ or $H_0 : B_\sigma^2 \le \eta^2$ as above, replacing $\hat{\mathbf{V}}$ by an estimate of $\hat{\mathbf{V}}_{AA}$ based on survey sampling methodology.

It may be useful to have standardized estimates of the individual components of the regression parameter $\boldsymbol{\beta}$. These can be computed as the components of the vector $\boldsymbol{\Sigma}_{\mathbf{z}}^{1/2}\hat{\boldsymbol{\beta}}$. The components have a covariance matrix consistently estimated by $\boldsymbol{\Sigma}_{\mathbf{z}}^{\mathbf{1/2}}\hat{\mathbf{V}}\boldsymbol{\Sigma}_{\mathbf{z}}^{\mathbf{1/2}}$. The standard error of the estimated proportional contribution of $\hat{\beta}_k$ to the risk score variance is estimated consistently by

$$\mathrm{SE}\left\{\hat{\pi}_k\right\} = \sqrt{\left(\nabla_{\hat{\boldsymbol{\beta}}}\hat{\pi}_k\right)^{\mathrm{T}} \hat{\mathbf{V}}\left(\nabla_{\hat{\boldsymbol{\beta}}}\hat{\pi}_k\right)},$$

where the l^{th} element of the gradient $\nabla_{\hat{\boldsymbol{\beta}}}\hat{\pi}_k$ is

$$\frac{\partial\hat{\pi}_k}{\partial\hat{\beta}_l} = \frac{\mathrm{diag}_k\left(\boldsymbol{\Sigma}_z^{1/2}\mathbf{D}_l\boldsymbol{\Sigma}_z^{1/2}\right)}{\hat{B}_\sigma^2} - \frac{2\hat{\pi}_k}{\hat{B}_\sigma^2}\sum_{i=1}^{m}\sigma_{li}\hat{\beta}_i.$$

Here diag_k denotes the k^{th} element of the diagonal, $\mathbf{D}_l$ is an $m \times m$ matrix with element $d_{ij}^{(l)} = I_{\{i=l\}}\hat{\beta}_j + I_{\{j=l\}}\hat{\beta}_i$ in row i and column j, and σ_{ij} denotes the element in row i and column j of $\boldsymbol{\Sigma}_{\mathbf{z}}$. For any subset of the m covariates $S \subset \{1, 2, \ldots, m\}$, the standard error of the estimator of the proportional contribution of the subset $\sum_{k\in S}\hat{\pi}_k$ is consistently estimated by

$$SE\left(\sum_{k\in S}\hat{\pi}_k\right) = \sqrt{\left(\sum_{k\in S}\nabla_{\hat{\boldsymbol{\beta}}}\hat{\pi}_k\right)^{\mathrm{T}} \hat{\mathbf{V}}\left(\sum_{k\in S}\nabla_{\hat{\boldsymbol{\beta}}}\hat{\pi}_k\right)}.$$

Partial Standardized Log Odds Ratio

If we are interested in characterizing the strength of association of a set of variables with the probability of an event *conditioning* on the values of other covariates or stratification variables in the model, we can compute a *partial* standardized log odds ratio. Let $\mathbf{z}_A$ be the vector of variables for which we wish to assess the strength of association and $\mathbf{z}_C$ be the vector of covariates that we wish to condition on. Write the covariance matrix of $\left(\mathbf{z}_A^{\mathrm{T}} \quad \mathbf{z}_C^{\mathrm{T}}\right)^{\mathrm{T}}$ as

$$\begin{bmatrix} \boldsymbol{\Sigma}_{AA} & \boldsymbol{\Sigma}_{AC} \\ \boldsymbol{\Sigma}_{CA} & \boldsymbol{\Sigma}_{CC} \end{bmatrix}$$

Then the partial covariance matrix of $\mathbf{z}_A$ given $\mathbf{z}_C$ is

$$\boldsymbol{\Sigma}_{A|C} = \boldsymbol{\Sigma}_{AA} - \boldsymbol{\Sigma}_{AC}\boldsymbol{\Sigma}_{CC}^{-1}\boldsymbol{\Sigma}_{CA}$$

The partial standardized log odds ratio for $\mathbf{z}_A$ given $\mathbf{z}_C$ is then

$$B_{\sigma}^{2} = \sqrt{\hat{\boldsymbol{\beta}}_{A}^{\mathrm{T}} \boldsymbol{\Sigma}_{A|C} \hat{\boldsymbol{\beta}}_{A}}$$

To compute a partial standardized log odds ratio using STDLOGOR, the variables to condition on must also be specified as adjustment covariates.

Checking for Collinear Covariates

If there is perfect collinearity between any covariate and the rest of the covariates, SAS PROC LOGISTIC (or SURVEYLOGISTIC) will drop that particular covariate from the model. However, if there is strong but not perfect collinearity, SAS will fit the full model, producing very large and unreliable regression parameter estimates for the covariate(s) that are collinear with each other. This results in unreliable (and generally very high) standardized odds ratio estimates. To prevent this, macro STDLOGOR performs a test for collinearity using the "variance inflation factor" (VIF) produced by PROC. The VIF for each covariate is equal to $1/(1-R^2)$, where R is the multiple correlation coefficient for the covariate with all the other covariates. If the maximum VIF over all the covariates is 100 or greater, a GHI note warning is printed in the SAS log and the standardized odds ratio is not computed. A variable containing the maximum VIF is included in the output data set produced by the macro.

Additional information about standardized odds ratios for multivariate Cox regression is found in Crager (2012).

Note: For computing the standardized odds ratio for *interaction* terms a special calculation is needed. The standardized odds ratio for the interaction of treatment with a single covariate is $\hat{\beta}_{I_T z}\sigma_z$, where $\hat{\beta}_{I_T z}$ is the regression coefficient for the term $I_T z$, I_T is the treatment indicator variable and σ_z is the sample standard deviation of z (not $I_T z$). Similarly, the standardized odds ratio for the interaction of a covariate vector $\mathbf{z}$ with treatment is $\sqrt{\hat{\boldsymbol{\beta}}_{I_T\mathbf{z}}^{\mathrm{T}} \boldsymbol{\Sigma}_{\mathbf{z}} \hat{\boldsymbol{\beta}}_{I_T\mathbf{z}}}$ where $\hat{\boldsymbol{\beta}}_{I_T\mathbf{z}}^{\mathrm{T}}$ is the vector of regression parameter estimates for the interaction and is the sample covariance matrix of $\mathbf{z}$ (not $I_T\mathbf{z}$). There are separate macros for calculating main effect standardized odds ratios and interaction standardized odds ratios.

Macro STDLOGOR

Macro STDLOGOR takes as input a data set a binary response variable and covariates (one record per patient) and calculates the standardized absolute log odds ratio for a specified set of covariates in a multivariate logistic regression model. The set of covariates for which the standardized absolute odds ratio is computed may be a subset of all the proportional odds covariates used in the model. Weighted analysis for cohort sampling designs is supported. The macro produces as output a data set with one record, or one record per by group if a by variable is specified. The output data set contains the estimate of the standardized absolute odds ratio and the standard error of the estimate, and values of the by group variable, if such a variable is specified.

Do not use macro STDLOGOR to compute the standardized odds ratio for an interaction effect. Use macro STDLOGOR_INT (described below) instead.

Macro STDLOGOR is called as follows:

```
%STDLOGHR(
        /* Input Specification */  indsn=, byvar=, response=, vars=, weight=, sampstrata=,
                                   adjcov=,
                                   var_combo_indsn=,combo_id=,
        /* Analysis Parameters */  partial=,print=, alpha=,intvl_hyp=
        /* Output Specification */ outdsn=, abs_std_log_OR=, var_std_log_OR=,
                                   abs_std_log_OR_correct=,
                                   chi_sq=, df=, p_value=,
                                   p_value_int=, p_value_int_Scheffe=,
                                   min_eigenvalue=, noncentrality_Scheffe=,
                                   abs_std_log_OR_LCL=, abs_std_log_OR_UCL=,
                                   abs_std_log_OR_LCL_Scheffe=, abs_std_log_OR_UCL_Scheffe=,
                                   zb_prefix=, zv_prefix=, predictors=, maxVIF=,
                                   var_combo_outdsn
           );
```

The macro parameters are described in Table 1.

Table 1. Macro STDLOGOR Parameters

Parameter	Type	Required?	Default Value	Description
indsn	$	Yes	(at temporary library)	Libname reference and the input data set name. The dataset name must conform to the rules for SAS names.
byvar	$/#	No	—	List of input data set variables (separated by spaces) that define by groups for the analysis.
response	$/#	Yes	—	The dichotomous dependent variable for the logistic regression analysis.
vars	$/#	Yes	—	List of input data set variables containing the logistic regression model covariates for which the standardized absolute log odds ratio will be computed.
weight	#	No	—	Input data set variable giving the observation's weight in the analysis. If this parameter is set, it is assumed that cohort sampling was used and resulted in the specified weights.
adjcov	#	No	—	Text string giving additional covariates to be included in the logistic regression model but not included in the standardized odds ratio computation.
var_combo_indsn	$	No	—	Optional input data set containing indicator variables for summing variable contributions to the risk score variance. For each record in this input data set, the macro will compute the sum of the contributions of the indicated variables to the risk score variance and the standard error of the sum. The indicator variables must have names ind_<var_name> where <var_name> is the name of the input data set variable included in the model. The parameter contribution_prefix must be specified to use this option.
combo_id	$	No	—	Optional variable combination identifier variable. If specified, this variable must be included in the file var_combo_inds.
sampstrata	$/#	No	—	If a stratified cohort sampling design was used, use this parameter to list the stratification variables.

Table 1. Macro STDLOGOR Parameters

Parameter	Type	Required ?	Default Value	Description
partial	#	No	—	If this parameter is set to a list of variables that is a subset of the set of adjustment covariates, the partial standardized log odds ratio will be calculated conditional on the specified adjustment covariates. If the parameter is not specified,the marginal standardized log odds ratio will be calculated.
print	$	No	yes	If this parameter is set to no, the noprint option will be invoked for PROC LOGISTIC (or PROC SURVEYLOGISTIC).
alpha	#	No	—	If this parameter is set, the macro will compute 100(1-alpha)% confidence intervals for the absolute standardized log odds ratio using both the conservative and Scheffé-alignment methods.
intvl_hyp	#	No	—	If this parameter is set to a numeric value, tests of the interval hypothesis that the absolute log odds ratio is less than or equal to this value will be constructed.
outdsn	$	Yes	(at temporary library)	Libname reference and the output data set name. The dataset name must conform to the rules for SAS names.
abs_std_log_OR	#	No	abs_std_log_OR	Name of output data set variable that will contain the estimate of the absolute standardized log odds ratio.
var_std_log_OR	#	No	var_std_log_OR	Name of the output data set variable that will contain the estimated variance of the estimate of the standardized log odds ratio.
abs_std_log_OR _correct	#	No	abs_std_log_OR_ correct	Name of output data set variable that will contain the bias-corrected estimate of the absolute standardized log odds ratio.
chi_sq	#	No	chi_sq	Name of output data set variable that will contain the chi-square statistic for testing the point null hypothesis that the absolute standardized log odds ratio is 0.

Table 1. Macro STDLOGOR Parameters

Parameter	Type	Required ?	Default Value	Description
df	#	No	df	Name of output data set variable that will contain the p-value for testing the point null hypothesis that the absolute standardized log odds ratio is 0.
p_value	#	No	p_value	Name of output data set variable that will contain the degrees of freedom of chi-square statistic for testing the point null hypothesis that the absolute standardized log odds ratio is 0.
p_value_int	#	No	p_value_int	If parameter intvl_hyp is set, p_value_int will contain the p-value for a conservative test of the interval null hypothesis that the absolute standardized log odds ratio is less than or equal to &intvl_hyp.
p_value_int_Scheffe	#	No	p_value_int	If parameter intvl_hyp is set, p_value_int will contain the p-value for a Scheffé-alignment test of the interval null hypothesis that the absolute standardized log odds ratio is less than or equal to &intvl_hyp.
min_eigenvalue	#	No	min_eigenvalue	Name of output data set variable will contain the minimum eigenvalue of the product of (1) the matrix square root of the covariance matrix of the covariate vector with (2) the covariance matrix of the regression parameter estimate vector with (3) the matrix square root of the covariance matrix of the covariate vector. This minimum eigenvalue is needed to construct tests of interval null hypotheses about the absolute standardized log odds ratio.
noncentrality_Scheffe	#	No	noncentrality_Scheffe	Name of output data set variable that will contain the maximum noncentrality parameter consistent with a 95% Scheffé confidence ellipsoid about the transformed parameter estimate vector. This value can be used to construct tests of interval null hypotheses about the absolute standardized odds ratio

Table 1. Macro STDLOGOR Parameters

Parameter	Type	Required ?	Default Value	Description
abs_std_log_OR_LCL	#	No	abs_std_log_OR_LCL	Lower limit of the 100(1-alpha)% confidence interval for the absolute standardized log odds ratio. Computed only if the macro parameter alpha is specified.
abs_std_log_OR_UCL	#	No	abs_std_log_OR_UCL	Upper limit of the 100(1-alpha)% confidence interval for the absolute standardized log odds ratio. Computed only if the macro parameter alpha is specified.
zb_prefix	$	No	—	If this parameter is specified, the vector of standardized log odds ratios for the individual covariates will be stored in variables with the specified prefix and suffixes from 1 to the number of variables. The order of these variables will correspond to the order of the covariates specified in the macro parameter vars.
zv_prefix	$	No	—	If this parameter is specified, the covariance matrix of the standardized log odds ratio estimates for the individual covariates will be stored in variables with the specified prefix and suffixes from 1 to the square of the number of variables. The order of these variables will correspond to the order of the covariates specified in the macro parameter vars: v_{11}, v_{12},..., v_{1p}, v_{21}, v_{22},..., v_{2p},...,v_{p1}, v_{p2},..., v_{pp}.
predictors	$	No	—	If this parameter is specified, the output data set will contain a text variable with the given name that contains the predictors for which the standardized odds ratio was assessed.
maxVIF	#	No	maxVIF	Name of output data set variable will contain the maximum variance inflation factor (VIF) over all the covariates from the screen for multicollinearity. If the maximum VIF is greater than 10, then no standardized log odds ratio will be returned.

Table 1. Macro STDLOGOR Parameters

Parameter	Type	Required?	Default Value	Description
Var_combo_outdsn	$	Yes, if parameter var_combo_indsn specified.		Name of output data set that will contain the estimated contributions of specified combinations of variables together with the standard error of each estimate.

Macro STDLOGOR_INT

This macro is used to compute the standardized log odds ratio for an interaction. It is called in the same way that macro STDLOGOR is called with the exception that there is an additional macro parameter called TMT, which should be the input data set variable that contains the treatment indicator (0 or 1) function. The variables specified by macro parameter VARS are the variables that interact with treatment.

References

Crager MR (2010). Gene identification using true discovery rate degree of association sets and estimates corrected for regression to the mean. *Statistics in Medicine* **29**:33–45. DOI: 10.1002/sim.3789

Crager MR (2012). Generalizing the standardized hazard ratio to multivariate proportional hazards regression, with an application to clinical-genomic studies. *Journal of Applied Statistics* **39**:399–417.