Sandwich variance estimator for interaction analysis

1. Introduction

We consider a study with n individuals evaluating a discrete or continuous x_1 and a dichotomous or quantitative variable x_2. Let y denote the response. β_3 defines a multiplicative interaction between x_1 and x_2. Explicitly, we assume the analysis model to be

$$g(Ey|x_1, x_2) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2$$

When the true $\beta_3 = 0$, to make type I error is to reject $H_0: \beta_3 = 0$; when the true $\beta_3 \neq 0$, the power of test is the probability to reject $H_0: \beta_3 = 0$. Model-based Type I error rate is calculated by rejecting when p-values of test > 0.05.

2. Notations and Formulas

Next, we consider the sandwich variance estimate in generalized estimating equations (GEE). Let Y_i be the response, and $X_i = [x_{i1}, \dots, x_{ip}]$ be the $1 \times \mathrm{p}$ matrix of covariate values for the *i*th subject (i = 1,…,n). We assume the variance matrix of Y_i depends on the mean of Y_i, i.e. $var(Y|X) = V(\mu_i) =: V_i$, where $\mu_i = g^{-1}(X_i^T \beta) = E(Y_i|X_i)$, $V(\cdot)$ is a known variance function.

Under the independence working assumption, the estimating equations are

$$U(\beta) = \sum_{i=1}^{n} \frac{\partial \mu_i^T}{\partial \beta} V_i^{-1}(Y_i - \mu_i) = 0$$

$\hat{\beta}_I$ is defined as the solution of $U_I(\beta) = 0$. $\hat{\beta}_I$ is consistent estimator of β. Under regularity conditions we have the covariance matrix for $\hat{\beta}_I$

$$V_{model} = \hat{\Omega}^{-1} \left(\sum_i \frac{\partial \mu_i^T}{\partial \beta} \hat{V}_i^{-1} cov(Y_i) \hat{V}_i^{-1} \frac{\partial \mu_i}{\partial \beta} \right) \hat{\Omega}^{-1}$$

The sandwich variance estimator is

$$V_{sandwich} = \hat{\Omega}^{-1} \left(\sum_i \frac{\partial \mu_i^T}{\partial \beta} \hat{V}_i^{-1} \hat{\epsilon}_i \hat{\epsilon}_i^T \hat{V}_i^{-1} \frac{\partial \mu_i}{\partial \beta} \right) \hat{\Omega}^{-1}$$

where $\Omega = \sum_i \partial\mu_i^T / \partial\beta V_i^{-1} \partial\mu_i/\partial\beta$, $\hat{\epsilon}_i = Y_i - \hat{\mu}_i = Y_i - g^{-1}(X_i^T \hat{\beta})$.

Wald test statistics are $\frac{\hat{\beta}_3^2}{\hat{V}_{model}(\hat{\beta}_3)}$(model based) $\frac{\hat{\beta}_3^2}{\hat{V}_{sandwich}(\hat{\beta}_3)}$ (robust), both $\sim\chi_1^2$ under H_0.

For simple linear regression model, we have

$$\hat{V}_{model} = (X^T X)^{-1} \left(\sum_i x_i x_i^T \hat{\sigma}^2 \right) (X^T X)^{-1}$$

$$\hat{V}_{sandwich} = (X^T X)^{-1} (\sum_i x_i x_i^T \left(y_i - x_i^T \hat{\beta}\right)^2)(X^T X)^{-1}$$

where $\hat{\sigma}^2 = \sum_{i=1}^n \left(y_i - x_i^T \hat{\beta}\right)^2 / (n-p)$. A more common notation in textbook and literature for the model-based variance matrix is $\hat{V}_{model} = \hat{\sigma}^2 (X^T X)^{-1}$.

For logistic regression model, let $\hat{\mu}_i = \frac{1}{1+e^{-x_i^T \hat{\beta}}}$, we have

$$\hat{V}_{model} = \left(\sum_i \hat{\mu}_i (1 - \hat{\mu}_i) X_i^T X_i \right)^{-1} \left(\sum_i X_i^T X_i \hat{\mu}_i (1 - \hat{\mu}_i) \right) \left(\sum_i \hat{\mu}_i (1 - \hat{\mu}_i) X_i^T X_i \right)^{-1}$$

$$\hat{V}_{sandwich} = \left(\sum_i \hat{\mu}_i (1 - \hat{\mu}_i) X_i^T X_i \right)^{-1} \left(\sum_i X_i^T X_i (y_i - \hat{\mu}_i)^2 \right) \left(\sum_i \hat{\mu}_i (1 - \hat{\mu}_i) X_i^T X_i \right)^{-1}$$

For score test, we first introduce some new notations. Let data matrix under null hypothesis be $X_{H_0} = (1, X_1, X_2)$ and the corresponding parameters are $\beta_{H_0} = (\beta_0, \beta_1, \beta_2)$. $X_3 = X_1 * X_2$.

The parameters $\tilde{\beta} = (\tilde{\beta}_0, \tilde{\beta}_1, \tilde{\beta}_2)$ are estimated by the estimating equations

$$U_{H_0}(\beta_{H_0}) = \sum_{i=1}^n \frac{\partial \mu_i^T}{\partial \beta_{H_0}} V_i^{-1} (Y_i - \mu_i) = 0$$

Under the independence working assumption, the score statistic for the interaction effect (β_3) is

$$U_3(\hat{\beta}_{H_0}) = \sum_{i=1}^{n} \frac{\partial \mu_i^T}{\partial \beta_3} V_i^{-1}(Y_i - \hat{\mu}_i)$$

where $\hat{\mu}_i$ is estimated under H_0.

Score test statistic: $\frac{U_3^2(\widehat{\beta}_{H_0})}{\widehat{\sigma}_{model}}$ (model based), $\frac{U_3^2(\widehat{\beta}_{H_0})}{\widehat{\sigma}_{sandwich}}$ (robust), both $\sim\chi_1^2$ under H_0.

$$\hat{\sigma}_{model} = \tilde{A}\left(\sum_i \frac{\partial \mu_i^T}{\partial \beta} \hat{V}_i^{-1} \widehat{cov}(Y_i) \hat{V}_i^{-1} \frac{\partial \mu_i}{\partial \beta}\right) \tilde{A}^T$$

$$\hat{\sigma}_{sandwich} = \tilde{A}\left(\sum_i \frac{\partial \mu_i^T}{\partial \beta} \hat{V}_i^{-1} \hat{\epsilon}_i \hat{\epsilon}_i^T \hat{V}_i^{-1} \frac{\partial \mu_i}{\partial \beta}\right) \tilde{A}^T$$

where $\tilde{A} = \left\{-\left[\sum_{i=1}^{n} \frac{\partial \mu_i^T}{\partial \beta_3} \hat{V}_i^{-1} \frac{\partial \mu_i}{\partial \beta_{H_0}}\right]\left[\sum_{i=1}^{n} \frac{\partial \mu_i^T}{\partial \beta_{H_0}} \hat{V}_i^{-1} \frac{\partial \mu_i}{\partial \beta_{H_0}}\right]^{-1}, 1\right\}$.

Explicitly, for quantitative trait, $\tilde{\sigma}^2 = \sum_{i=1}^{n}\left(y_i - [1, X_{i1}, X_{i2}]\tilde{\beta}\right)^2 / (n-p)$

$$U_3(\hat{\beta}_{H_0}) = \sum_{i=1}^{n} x_{i1} x_{i2} \tilde{\sigma}^{-2}\left(Y_i - [1, x_{i1}, x_{i2}]\tilde{\beta}\right)$$

$$\hat{\sigma}_{model} = \tilde{A}\left(\sum_i \tilde{\sigma}^{-2} X_i^T X_i\right) \tilde{A}^T$$

$$\hat{\sigma}_{sandwich} = \tilde{A}\left(\sum_i \left(y_i - x_i^T \tilde{\beta}\right)^2 \tilde{\sigma}^{-2} X_i X_i^T\right) \tilde{A}^T$$

$$\tilde{A} = \left\{-\left[\sum_{i=1}^{n} x_{i1} x_{i2} [1, x_{i1}, x_{i2}]\right]\left[\sum_{i=1}^{n} \begin{bmatrix} 1 \\ x_{i1} \\ x_{i2} \end{bmatrix} [1, x_{i1}, x_{i2}]\right]^{-1}, 1\right\}$$

For dichotomous trait (logit link), let $\tilde{\mu}_i = \frac{1}{1+e^{-x_i^T \tilde{\beta}}}$ we have

$$U_3(\hat{\beta}_{H_0}) = \sum_{i=1}^{n} x_{i1}x_{i2}(y_i - \tilde{\mu}_i)$$

$$\hat{\sigma}_{model} = \tilde{A}\left(\sum_i \tilde{\mu}_i(1-\tilde{\mu}_i)X_i^T X_i\right)\tilde{A}^T$$

$$\hat{\sigma}_{sandwich} = \tilde{A}\left(\sum_i (Y_i - \tilde{\mu}_i)^2 X_i^T X_i\right)\tilde{A}^T$$

$$\tilde{A} = \left\{-\left[\sum_{i=1}^{n} x_{i1}x_{i2}\tilde{\mu}_i(1-\tilde{\mu}_i)[1, x_{i1}, x_{i2}]\right]\left[\sum_{i=1}^{n} \tilde{\mu}_i(1-\tilde{\mu}_i)\begin{bmatrix}1\\ x_{i1}\\ x_{i2}\end{bmatrix}[1, x_{i1}, x_{i2}]\right]^{-1}, 1\right\}$$