Hypothesis Testing in GEEs

Recall GEE:

$$\sum_{i=1}^{m} \mathbf{D_i^T V_i^{-1} (Y_i - \boldsymbol{\mu}_i) = 0}$$

$\Rightarrow \hat{\beta} \sim N(\beta, \boldsymbol{\Sigma})$ asymptotically.

$$\begin{aligned}\boldsymbol{\Sigma} &= (\sum_{i=1}^{m} \mathbf{D}_i^T \mathbf{V}_i^{-1} \mathbf{D}_i)^{-1} \{\sum_{i=1}^{m} \mathbf{D}_i^T \mathbf{V}_i^{-1} (\mathbf{Y}_i - \boldsymbol{\mu}_i)(\mathbf{Y_i} - \boldsymbol{\mu}_\mathbf{i})^T \mathbf{V}_i^{-1} \mathbf{D}_i\} \\ & \quad (\sum_{i=1}^{m} \mathbf{D}_i^T \mathbf{V}_i^{-1} \mathbf{D}_i)^{-1}\end{aligned}$$

where $\mathbf{D}_i = \frac{\partial \boldsymbol{\mu}_i}{\partial \beta^T}$, $\mathbf{V}_i = \mathbf{V}_{\boldsymbol{Mi}}^{\frac{1}{2}} \mathbf{R}_i \mathbf{V}_{\boldsymbol{Mi}}^{\frac{1}{2}}$, and $\mathbf{R}_i$ is the working correlation matrix.

Wald Test

Full model:

$$g(\boldsymbol{\mu}) = \mathbf{X}_1^T \beta_1 + \mathbf{X^T}_2 \beta_2$$

Reduced model:

$$g(\boldsymbol{\mu}) = \mathbf{X}_1^T \beta_1$$

Hypothesis:

$$H_0 : \beta_2 = 0 \qquad v.s. \qquad H_1 : \beta_2 \neq 0$$

Wald test:

$$\chi^2_W = \hat{\beta}_2^T \mathbf{\Sigma}_{22}^{-1} \hat{\beta}_2 \to \chi^2_q$$

where $\mathbf{\Sigma} = \begin{pmatrix} \mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12} \\ \mathbf{\Sigma}_{21} & \mathbf{\Sigma}_{22} \end{pmatrix}$, and $\mathbf{\Sigma}(\hat{\beta})$ and $\hat{\beta}$ are obtained under the full model.

Score Test (1)

Let $\mathbf{U} = \begin{pmatrix} \mathbf{U_1} \\ \mathbf{U_2} \end{pmatrix}$, where $\mathbf{U}_1 = \sum_{i=1}^m \mathbf{D}_{1i}^T \mathbf{V}_i^{-1}(\mathbf{Y}_i - \boldsymbol{\mu}_i)$, $\mathbf{D}_{1i} = \frac{\partial \boldsymbol{\mu}_i}{\partial \beta_1^T}$, and $\mathbf{U}_2 = \sum_{i=1}^m \mathbf{D}_{2i}^T \mathbf{V}_i^{-1}(\mathbf{Y}_i - \boldsymbol{\mu}_i)$, $\mathbf{D}_{2i} = \frac{\partial \boldsymbol{\mu}_i}{\partial \beta_2^T}$.

$$
\begin{aligned}
\mathbf{A} &= \sum_i \mathbf{D}_i^T \mathbf{V}_i^{-1} \mathbf{D}_i \\
&= \begin{pmatrix} \sum_i \mathbf{D}_{1i}^T \mathbf{V}_i^{-1} \mathbf{D}_{1i} & \sum \mathbf{D}_{1i}^T \mathbf{V}_i^{-1} \mathbf{D}_{2i} \\ \sum_i \mathbf{D}_{2i}^T \mathbf{V}_i^{-1} \mathbf{D}_{1i} & \sum \mathbf{D}_{2i}^T \mathbf{V}_i^{-1} \mathbf{D}_{2i} \end{pmatrix} \\
&= \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix}
\end{aligned}
$$

Score Test (2)

$$\begin{aligned}\mathbf{B} &= \sum_i \mathbf{D}_i^T \mathbf{V}_i^{-1} (\mathbf{Y}_i - \boldsymbol{\mu}_i)(\mathbf{Y}_i - \boldsymbol{\mu}_i)^T \mathbf{V}_i^{-1} \mathbf{D}_i \\ &= \begin{pmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{B}_{21} & \mathbf{B}_{22} \end{pmatrix}\end{aligned}$$

where $\mathbf{B}_{ik} = \sum \mathbf{D}_{ji}^T \mathbf{V}_i^{-1} (\mathbf{Y}_i - \boldsymbol{\mu}_i)(\mathbf{Y}_i - \boldsymbol{\mu}_i)^T \mathbf{V}_i^{-1} \mathbf{D}_{ki}$ for $j, k = 1, 2$. Then

$$\chi_s^2 = \mathbf{U}_2(\hat{\beta}_1)^T \boldsymbol{\Lambda}^{-1}(\hat{\beta}_1) \mathbf{U}_2(\hat{\beta}_1) \quad \rightarrow \quad \chi_q^2$$

where $\boldsymbol{\Lambda} = \mathbf{B}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{B}_{12} - \mathbf{B}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12} + \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{B}_{11}\mathbf{A}_{11}^{-1}\mathbf{A}_{12}$.

Likelihood Ratio Tests

Constructing a likelihood ratio test for GEE's is very difficult!

Main Reason: There may not be a unique objective function corresponding to the same GEE!

Approximate LR test: Hanfelt and Liang (1995, Biometrika, p461-477)

Pseudo Wald, Score, Likelihood Ratio Tests

(Rotnitzky and Jewell, 1990, Biometrika)

Key Results

- Study the asymptotic distributions of the naive Wald, Score, LR tests constructed by assuming the working correlation (e.g. Independence) is correct.
- The distributions of the naive Wald, Score and LR test statistics are asymptotically mixtures of chisquares.

Useful when the robust tests perform poorly (e.g. large number of observations within a cluster and few clusters)

Naive Wald and Score Tests

Hypothesis: $H_0 : \boldsymbol{\beta}_2 = 0 \, v.s. \, H_1 : \boldsymbol{\beta}_2 \neq 0$

Naive Wald test:

$$\chi^2_{W*} = \hat{\beta}_2^T \boldsymbol{A}_{22}^{-1} \hat{\beta}_2$$

where $\mathbf{A} = (\sum_{i=1}^m \mathbf{D_i^T V_i^{-1} D_i^T})^{-1} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix}$.

Naive Score test:

$$\chi^2_{s*} = \mathbf{U}_2(\hat{\beta}_1)^T \mathbf{A_{22}}(\hat{\beta}_1) \mathbf{U}_2(\hat{\beta}_1)$$

Asymptotic Distribution of Naive Tests

Theorem: Under H_0, χ^2_{w*} and $\chi^2_{s*} \xrightarrow{\mathcal{D}} \sum_{j=1}^{q} c_j \chi^2_j$ where $\chi^2_1, \cdots, \chi^2_q$ are iid $\chi^2(1)$, and $c_1 \geq \cdots \geq c_q$ are the eigenvalues of the limit of the matrix $\mathbf{Q} = \mathbf{Q}_0^{-1}\mathbf{Q}_1$ with

$$\mathbf{Q}_0 = \frac{1}{m}\sum_{i=1}^{m} \widetilde{\mathbf{D}}_i^T \mathbf{V}_i^{-1} \widetilde{\mathbf{D}}_i$$

$$\mathbf{Q}_1 = \frac{1}{m}\sum_{i=1}^{m} \widetilde{\mathbf{D}}_i^T \mathbf{V}_i^{-1} (\mathbf{Y}_i - \boldsymbol{\mu}_i)(\mathbf{Y}_i - \boldsymbol{\mu}_i)^T \mathbf{V}_i^{-1} \widetilde{\mathbf{D}}_i$$

$$\widetilde{\mathbf{D}}_i = \mathbf{D}_i^{(1)} - \mathbf{D}_i^{(2)} (\sum_{i=1}^{m} \mathbf{D}_i^{(2)^T} \mathbf{V}_i^{-1} \mathbf{D}_i^{(2)})^{-1} (\sum_{i=1}^{m} \mathbf{D}_i^{(2)^T} \mathbf{V}_i^{-1} \mathbf{D}_i^{(1)})$$

$$\mathbf{D}_i = \begin{pmatrix} \mathbf{D}_i^{(1)} \\ \mathbf{D}_i^{(2)} \end{pmatrix} \begin{array}{l} n_i \times p \\ n_i \times q \end{array}$$

See Rotnitzky and Jewell Appendix 1 for full proof.

Remarks

1. The generalized Wald test $\chi^2_w = \hat{\beta}_2^T \mathbf{\Sigma}_{22}^{-1} \hat{\beta}_2$ is a little bit more complicated to compute compared to the naive Wald test. However, its critical value is much easier to calculate.
2. The same is true for the score statistic.
3. A special case of the naive Wald and score statistics is the independence working correlation.
4. One can use glm to calculate the naive Wald and score statistics and to calculate the critical values using the mixture of chi-squares.
5. If the working correlation is correct, then the distribution is just usual χ^2_{df}.
6. Rotnitzky and Jewell also derive a naive LRT under working independence.