CNI 세미나 2015-75

다중응답 종단자료의 효율적 추정방법 세미나

주최 · 주관 : 충남연구원

일시 : 2015년 7월 21일(화) 10:00~12:00

장소 : 충남연구원 1층 회의실

진행순서

<u>10:00~10:10</u>

개회 및 참석자 소개

<u>10:10~11:00</u>

조현근 교수 (Western Michigan University)

Efficient estimation for longitudinal data with multiple responses : application to transportation safety study

<u>11:00~11:50</u> 토론 및 질의응답

<u>11:50~12:00</u> 폐회 및 정리

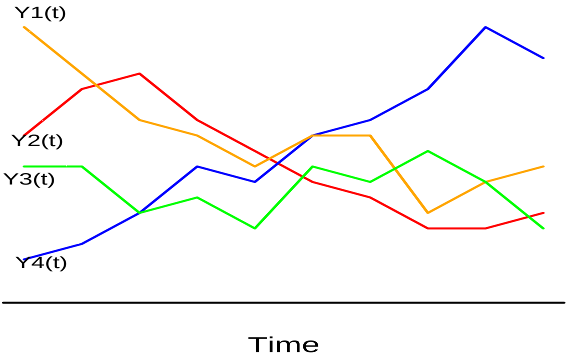
Efficient estimation for longitudinal data with multiple responses: application to transportation safety study

Hyunkeun Cho

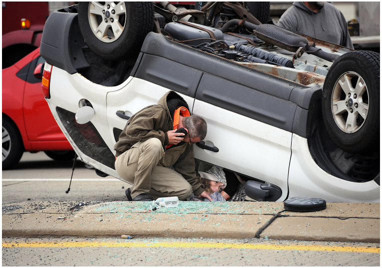
Western Michigan University, USA

July 21, 2015

Longitudinal data



Transportation safety study



Transportation safety study

- 506 midblock segments of arterial roads in Lincoln, Nebraska between 2003 and 2007
- **Two dependent variables** were followed up annually
  - crash frequency
  - presence of crash severity
- Five covariates of interest:
  - the number of through lanes
  - average annual daily traffic
  - presence of median
  - central business district
  - length of segment

Transportation safety study

- Two generalized linear model:

  $\log\{E(Crash)\} = \alpha_0 + \alpha_1 Lane + \alpha_2 AADT + \alpha_3 Med + \alpha_4 CBD + \alpha_5 Length$

  $\text{logit}\{E(Severe)\} = \beta_0 + \beta_1 Lane + \beta_2 AADT + \beta_3 Med + \beta_4 CBD + \beta_5 Length$

- **Goal:** Identify relevant covariates that result in the crash and severity.

Transportation safety study

- Two generalized linear model:

$$\log\{E(Crash)\} = \alpha_0 + \alpha_1 Lane + \alpha_2 AADT + \alpha_3 Med + \alpha_4 CBD + \alpha_5 Length$$

$$\text{logit}\{E(Severe)\} = \beta_0 + \beta_1 Lane + \beta_2 AADT + \beta_3 Med + \beta_4 CBD + \beta_5 Length$$

- **Goal:** Identify relevant covariates that result in the crash and severity.
- The prediction model:

$$Crash = exp(\hat{\alpha}_0 + \hat{\alpha}_1 Lane + \hat{\alpha}_2 AADT + \hat{\alpha}_3 Med + \hat{\alpha}_4 CBD + \hat{\alpha}_5 Length)$$

$$Severe = 1/[1+exp(-\hat{\beta}_0 - \hat{\beta}_1 Lane - \hat{\beta}_2 AADT - \hat{\beta}_3 Med - \hat{\beta}_4 CBD - \hat{\beta}_5 Length)]$$

Marginal model for univariate longitudinal response

- One outcome variable y_{ij} and covariates x_{ij} at time $j = 1, \ldots, m$ for subjects $i = 1, \ldots, n$
- The mean of y_{ij} is specified as

$$\mu_{ij} = E(y_{ij} | x_{ij}) = h(x_{ij}^T \beta),$$

- h is an inverse link function

Generalized estimating equations (GEE)

- **Generalized estimating equations** (Liang and Zeger, 1986)

$$\sum_{i=1}^{n} \left(\frac{\partial \mu_i}{\partial \beta} \right)^T V_i^{-1} (y_i - \mu_i) = 0,$$

  - $y_i = (y_{i1}, \ldots, y_{im})^T$ and $\mu_i = (\mu_{i1}, \ldots, \mu_{im})^T$

Generalized estimating equations (GEE)

- **Generalized estimating equations** (Liang and Zeger, 1986)

$$\sum_{i=1}^{n} \left(\frac{\partial \mu_i}{\partial \beta} \right)^T V_i^{-1}(y_i - \mu_i) = 0,$$

  - $y_i = (y_{i1}, \ldots, y_{im})^T$ and $\mu_i = (\mu_{i1}, \ldots, \mu_{im})^T$
  - $V_i^{-1} = A_i^{-1/2} R^{-1} A_i^{-1/2}$
    A_i: diagonal variance matrix of y_i
    R: **working correlation** matrix for all subjects

Generalized estimating equations (GEE)

- **Generalized estimating equations** (Liang and Zeger, 1986)

$$\sum_{i=1}^{n} \left(\frac{\partial \mu_i}{\partial \beta} \right)^T V_i^{-1} (y_i - \mu_i) = 0,$$

  - $y_i = (y_{i1}, \ldots, y_{im})^T$ and $\mu_i = (\mu_{i1}, \ldots, \mu_{im})^T$
  - $V_i^{-1} = A_i^{-1/2} R^{-1} A_i^{-1/2}$
    A_i: diagonal variance matrix of y_i
    R: **working correlation** matrix for all subjects

- **Not efficient** under the misspecified R

Quadratic inference function

- **Approximate** R^{-1},

$$R^{-1} \approx \sum_{j=1}^{b} a_j M_j,$$

  - M_j: a basis matrix
  - a_j: an unknown coefficient

Quadratic inference function

- Example 1: **Exchangeable structure** with a correlation coefficient ρ

$$R = \begin{bmatrix} 1 & \rho & \rho & \cdots & \rho \\ \rho & 1 & \rho & \cdots & \rho \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \rho & \cdots & \rho & 1 & \rho \\ \rho & \cdots & \rho & \rho & 1 \end{bmatrix}_{m\times m}$$

Quadratic inference function

- Example 1: **Exchangeable structure** with a correlation coefficient ρ

$$R = \begin{bmatrix} 1 & \rho & \rho & \cdots & \rho \\ \rho & 1 & \rho & \cdots & \rho \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \rho & \cdots & \rho & 1 & \rho \\ \rho & \cdots & \rho & \rho & 1 \end{bmatrix}_{m\times m}$$

$$\begin{aligned} R^{-1} &= a_0 \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 & 0 \\ 0 & \cdots & 0 & 0 & 1 \end{pmatrix} + a_1 \begin{pmatrix} 0 & 1 & 1 & \cdots & 1 \\ 1 & 0 & 1 & \cdots & 1 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 1 & \cdots & 1 & 0 & 1 \\ 1 & \cdots & 1 & 1 & 0 \end{pmatrix} \\ &= a_0 I_m + a_1 M_1 \end{aligned}$$

Quadratic inference function

- Example 2: **AR-1 structure** with a correlation coefficient ρ

$$R = \begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{m-1} \\ \rho & 1 & \rho & \cdots & \rho^{m-2} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \rho^{m-2} & \cdots & \rho & 1 & \rho \\ \rho^{m-1} & \cdots & \rho^2 & \rho & 1 \end{bmatrix}_{m \times m}$$

Quadratic inference function

- Example 2: **AR-1 structure** with a correlation coefficient ρ

$$R = \begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{m-1} \\ \rho & 1 & \rho & \cdots & \rho^{m-2} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \rho^{m-2} & \cdots & \rho & 1 & \rho \\ \rho^{m-1} & \cdots & \rho^2 & \rho & 1 \end{bmatrix}_{m\times m}$$

$$\begin{aligned} R^{-1} &= a_0 I_m + a_1 \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 1 & 0 & 1 & \ddots & 0 \\ 0 & 1 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 & 1 \\ 0 & \cdots & 0 & 1 & 0 \end{pmatrix} + a_2 \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 0 & 0 & 0 \\ 0 & \cdots & 0 & 0 & 1 \end{pmatrix} \\ &= a_0 I_m + a_1 M_2 + a_2 M_3 \\ &\approx a_0 I_m + a_1 M_2 \end{aligned}$$

Quadratic inference function

- **Approximate** R^{-1},

$$R^{-1} \approx \sum_{j=1}^{b} a_j M_j,$$

  - M_j: a basis matrix
  - a_j: an unknown coefficient

- Substitute R^{-1} into GEE,

$$GEE = \sum_{i=1}^{n} \left(\frac{\partial \mu_i}{\partial \beta}\right)^T A_i^{-1/2} \quad R^{-1} \quad A_i^{-1/2}(y_i - \mu_i)$$

Quadratic inference function

- **Approximate** R^{-1},

$$R^{-1} \approx \sum_{j=1}^{b} a_j M_j,$$

  - M_j: a basis matrix
  - a_j: an unknown coefficient

- Substitute R^{-1} into GEE,

$$GEE = \sum_{i=1}^{n} \left(\frac{\partial \mu_i}{\partial \beta} \right)^T A_i^{-1/2} \left(\sum_{j=1}^{b} a_j M_j \right) A_i^{-1/2} (y_i - \mu_i)$$

Quadratic inference function

- Extended score vector:

$$\bar{g}(\beta) = \frac{1}{n}\sum_{i=1}^{n} g_i(\beta) = \frac{1}{n}\begin{pmatrix} \sum \left(\frac{\partial \mu_i}{\partial \beta}\right)^T A_i^{-1/2} M_1 A_i^{-1/2}(y_i - \mu_i) \\ \vdots \\ \sum \left(\frac{\partial \mu_i}{\partial \beta}\right)^T A_i^{-1/2} M_b A_i^{-1/2}(y_i - \mu_i) \end{pmatrix}$$

- **Quadratic inference function** (Qu, Lindsay and Li, 2000)

$$Q(\beta) = n\bar{g}(\beta)^T \bar{C}^{-1} \bar{g}(\beta),$$

where $\bar{C} = \frac{1}{n}\sum_{i=1}^{n} g_i g_i^T$

- No need to specify the likelihood
- Yield efficient and consistent estimates

Crash data

- 506 segments were followed annually for 5 years
- Two dependent variables:
  - crash frequency (Crash)
  - presence of crash severity (Severe)
- Five covariates of interest:
  1) the number of through lanes (Lane)
  2) average annual daily traffic (AADT)
  3) presence of median (Med)
  4) central business district (CBD)
  5) length of segment (Length)

Crash data

- Two generalized linear model:

  $$\log\{E(Crash)\} = \alpha_0 + \alpha_1 Lane + \alpha_2 AADT + \alpha_3 Med + \alpha_4 CBD + \alpha_5 Length$$

  $$\text{logit}\{E(Severe)\} = \beta_0 + \beta_1 Lane + \beta_2 AADT + \beta_3 Med + \beta_4 CBD + \beta_5 Length$$

- Use AR-1 working correlation structure

Table : Estimated coefficients along with p-values from Wald test.

Covariate	log(*Crash*)		logit(*Severe*)	
	Estimator	p-value	Estimator	p-value
intercept	0.0621	0.897	-4.8741	0.002
Lane	-0.0590	0.209	0.1774	0.211
AADT	**0.0002**	**0.000**	0.0000	0.674
Med	-0.1649	0.075	**-0.7435**	**0.011**
CBD	**0.2644**	**0.045**	-0.9451	0.051
Length	**0.6238**	**0.000**	-0.6415	0.099

Transportation safety study

- 506 segments were followed annually for 5 years
- Two dependent variables:
  - crash frequency
  - presence of crash severity

Transportation safety study

- 506 segments were followed annually for 5 years
- Two dependent variables:
  - crash frequency
  - presence of crash severity
- Two types of correlations:
  - repeated measures within the same road
  - **between dependent variables**

Transportation safety study

- 506 segments were followed annually for 5 years
- Two dependent variables:
  - crash frequency
  - presence of crash severity
- Two types of correlations:
  - repeated measures within the same road
  - **between dependent variables**
- **Goal:** Improve estimation efficiency by accommodating the association between responses

Multivariate marginal model

- The kth response variable $y_{i\cdot k} = (y_{i1k}, \ldots, y_{imk})^T$ and covariates $x_i = (x_{i1}, \ldots, x_{im})$ for $k = 1, \ldots, K$
- The mean of y_{ijk} is specified as

$$\mu_{ijk} = E(y_{ijk}|x_{ij}) = h(x_{ijk}^T \beta_k),$$

  - h is an inverse link function
  - $\beta_k = (\beta_{k1}, \ldots, \beta_{kp})^T$ is a parameter vector for the kth response

Multivariate marginal model

- Stack up data as
  - $Y_i = (y_{i \cdot 1}^T, \ldots, y_{i \cdot K}^T)^T$ is an mK-dimensional vector
  - $X_i = (I_K \otimes x_i)$ is a $pK \times mK$-matrix by Kronecker product operator
- The multivariate marginal mean is specified as

$$\boldsymbol{\mu}_i = E(Y_i | X_i) = h(X_i^T \boldsymbol{\beta}),$$

- $\boldsymbol{\beta} = (\beta_1^T, \ldots, \beta_K^T)^T$ is a pK-dimensional parameter vector

Quadratic inference function under multivariate model

- Define the quadratic inference function

$$Q(\boldsymbol{\beta}) = n\bar{g}^T(\boldsymbol{\beta})\bar{C}^{-1}\bar{g}(\boldsymbol{\beta}),$$

where $\bar{C} = \frac{1}{n}\sum_{i=1}^{n} g_i g_i^T$ and $\bar{g}(\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^{n} g_i(\boldsymbol{\beta})$ with

$$g_i(\boldsymbol{\beta}) = \begin{pmatrix} \left(\frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}}\right)^T A_i^{-1/2} M_1 A_i^{-1/2} (Y_i - \boldsymbol{\mu}_i) \\ \vdots \\ \left(\frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}}\right)^T A_i^{-1/2} M_b A_i^{-1/2} (Y_i - \boldsymbol{\mu}_i) \end{pmatrix}$$

- Remind that $R^{-1} \approx \sum_{j=1}^{b} a_j M_j$,
where R is the $mK \times mK$ working correlation matrix of Y_i

Choice of basis matrices

- Assume **responses are correlated** with a correlation coefficient ω

Choice of basis matrices

- Assume **responses are correlated** with a correlation coefficient ω

$$\Omega = \begin{bmatrix} 1 & \omega & \omega & \cdots & \omega \\ \omega & 1 & \omega & \cdots & \omega \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \omega & \cdots & \omega & 1 & \omega \\ \omega & \cdots & \omega & \omega & 1 \end{bmatrix}_{K\times K}$$

Choice of basis matrices

- Assume **responses are correlated** with a correlation coefficient ω

$$\Omega = \begin{bmatrix} 1 & \omega & \omega & \cdots & \omega \\ \omega & 1 & \omega & \cdots & \omega \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \omega & \cdots & \omega & 1 & \omega \\ \omega & \cdots & \omega & \omega & 1 \end{bmatrix}_{K \times K}$$

- Let U be the correlation structure for measurements within the subject

Choice of basis matrices

- Assume **responses are correlated** with a correlation coefficient ω

$$\Omega = \begin{bmatrix} 1 & \omega & \omega & \cdots & \omega \\ \omega & 1 & \omega & \cdots & \omega \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \omega & \cdots & \omega & 1 & \omega \\ \omega & \cdots & \omega & \omega & 1 \end{bmatrix}_{K \times K}$$

- Let U be the correlation structure for measurements within the subject
- $R = \Omega \otimes U$

Choice of basis matrices

- Assume **responses are correlated** with a correlation coefficient ω

$$\Omega = \begin{bmatrix} 1 & \omega & \omega & \cdots & \omega \\ \omega & 1 & \omega & \cdots & \omega \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \omega & \cdots & \omega & 1 & \omega \\ \omega & \cdots & \omega & \omega & 1 \end{bmatrix}_{K \times K}$$

- Let U be the correlation structure for measurements within the subject
- $R = \Omega \otimes U$
- $R^{-1} = \Omega^{-1} \otimes U^{-1} = (\gamma_0 I_K + \gamma_1 W) \otimes U^{-1}$

$$= \left(\gamma_0 \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 & 0 \\ 0 & \cdots & 0 & 0 & 1 \end{bmatrix} + \gamma_1 \begin{bmatrix} 0 & 1 & 1 & \cdots & 1 \\ 1 & 0 & 1 & \cdots & 1 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 1 & \cdots & 1 & 0 & 1 \\ 1 & \cdots & 1 & 1 & 0 \end{bmatrix} \right) \otimes U^{-1}$$

Choice of basis matrices

- Example 1: U is the **exchangeable structure**

$$
\begin{aligned}
U^{-1} &= \delta_1 \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 & 0 \\ 0 & \cdots & 0 & 0 & 1 \end{pmatrix} + \delta_2 \begin{pmatrix} 0 & 1 & 1 & \cdots & 1 \\ 1 & 0 & 1 & \cdots & 1 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 1 & \cdots & 1 & 0 & 1 \\ 1 & \cdots & 1 & 1 & 0 \end{pmatrix} \\
&= \delta_1 I_m + \delta_2 U_1 .
\end{aligned}
$$

Choice of basis matrices

- Example 1: U is the **exchangeable structure**

$$
\begin{aligned}
U^{-1} &= \delta_1 \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 & 0 \\ 0 & \cdots & 0 & 0 & 1 \end{pmatrix} + \delta_2 \begin{pmatrix} 0 & 1 & 1 & \cdots & 1 \\ 1 & 0 & 1 & \cdots & 1 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 1 & \cdots & 1 & 0 & 1 \\ 1 & \cdots & 1 & 1 & 0 \end{pmatrix} \\
&= \delta_1 I_m + \delta_2 U_1 .
\end{aligned}
$$

$$
\begin{aligned}
R^{-1} &= (\gamma_0 I_K + \gamma_1 W) \otimes U^{-1} \\
&= (\gamma_0 I_K + \gamma_1 W) \otimes (\delta_1 I_m + \delta_2 U_1) \\
&= \gamma_0\delta_1 I_K \otimes I_m + \gamma_1\delta_1 W \otimes I_m + \gamma_0\delta_2 I_K \otimes U_1 + \gamma_1\delta_2 W \otimes U_1 \\
&= a_1 I_{mK} + a_2 M_2 + a_3 M_3 + a_4 M_4
\end{aligned}
$$

Choice of basis matrices

- Example 2: U is an **AR-1** structure

$$
\begin{aligned}
U^{-1} &= \delta_3 I_m + \delta_4 \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 1 & 0 & 1 & \ddots & 0 \\ 0 & 1 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 & 1 \\ 0 & \cdots & 0 & 1 & 0 \end{pmatrix} \\
&= \delta_3 I_m + \delta_4 U_2
\end{aligned}
$$

Choice of basis matrices

- Example 2: U is an **AR-1** structure

$$
\begin{aligned}
U^{-1} &= \delta_3 I_m + \delta_4 \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 1 & 0 & 1 & \ddots & 0 \\ 0 & 1 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 & 1 \\ 0 & \cdots & 0 & 1 & 0 \end{pmatrix} \\
&= \delta_3 I_m + \delta_4 U_2
\end{aligned}
$$

$$
\begin{aligned}
R^{-1} &= (\gamma_0 I_K + \gamma_1 W) \otimes U^{-1} \\
&= (\gamma_0 I_K + \gamma_1 W) \otimes (\delta_3 I_m + \delta_4 U_2) \\
&= \gamma_0 \delta_3 I_K \otimes I_m + \gamma_1 \delta_3 W \otimes I_m + \gamma_0 \delta_4 I_K \otimes U_2 + \gamma_1 \delta_4 W \otimes U_2 \\
&= a_1 I_{mK} + a_2 M_2 + a_3 M_3 + a_4 M_4
\end{aligned}
$$

Asymptotic properties

Theorem

Under the regularity conditions, the proposed estimator $\hat{\boldsymbol{\beta}}$ satisfies:

(1) **(Consistency)** $\hat{\boldsymbol{\beta}} \xrightarrow{\text{p}} \boldsymbol{\beta}_0$

(2) **(Asymptotic normality)**

$$\sqrt{n}\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right) \xrightarrow{\text{d}} N\left(0, V\right),$$

where $V = lim_{n\to\infty} 2\ddot{Q}_N(\hat{\boldsymbol{\beta}})^{-1}$ *and* $\ddot{Q}_N(\hat{\boldsymbol{\beta}}) = \frac{\partial^2}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}} Q_N(\boldsymbol{\beta})\Big|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}}$

Asymptotic properties

Theorem

Under the regularity conditions, the proposed estimator $\hat{\boldsymbol{\beta}}$ satisfies:

(1) **(Consistency)** $\hat{\boldsymbol{\beta}} \xrightarrow{\mathrm{p}} \boldsymbol{\beta}_0$

(2) **(Asymptotic normality)**

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \xrightarrow{\mathrm{d}} N(0, V),$$

where $V = lim_{n\to\infty} 2\ddot{Q}_N(\hat{\boldsymbol{\beta}})^{-1}$ *and* $\ddot{Q}_N(\hat{\boldsymbol{\beta}}) = \frac{\partial^2}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}} Q_N(\boldsymbol{\beta})\big|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}}$

Theorem

Under the conditions, $\hat{\boldsymbol{\beta}}$ is **more efficient** *than the estimator $\tilde{\boldsymbol{\beta}}$,*
i.e., $var(b^T\hat{\boldsymbol{\beta}}) \leq var(b^T\tilde{\boldsymbol{\beta}})$ for any constant vector b,
where $\tilde{\boldsymbol{\beta}}$ is the QIF estimator based on the univariate marginal model

Hypothesis test

- Decompose $\boldsymbol{\beta} = (\theta, \vartheta)$
- Test $H_0 : \theta = \theta_0$ versus $H_A : \theta \neq \theta_0$, where θ_0 is a constant vector.
- Test statistic is defined as

$$T_\chi = n\{Q(\theta_0, \check{\vartheta}) - Q(\hat{\theta}, \hat{\vartheta})\},$$

where $\check{\vartheta} = \text{argmin}_\vartheta Q(\theta_0, \vartheta)$ and $(\hat{\theta}, \hat{\vartheta}) = \text{argmin}_{(\theta,\vartheta)} Q(\theta, \vartheta)$

Hypothesis test

- Decompose $\boldsymbol{\beta} = (\theta, \vartheta)$
- Test $H_0 : \theta = \theta_0$ versus $H_A : \theta \neq \theta_0$, where θ_0 is a constant vector.
- Test statistic is defined as

$$T_\chi = n\{Q(\theta_0, \check{\vartheta}) - Q(\hat{\theta}, \hat{\vartheta})\},$$

where $\check{\vartheta} = \text{argmin}_\vartheta Q(\theta_0, \vartheta)$ and $(\hat{\theta}, \hat{\vartheta}) = \text{argmin}_{(\theta,\vartheta)} Q(\theta, \vartheta)$

Theorem

Under H_0, $T_\chi \xrightarrow{\mathrm{d}} \chi^2_s$ as $n \to \infty$, where s is a dimension of θ

Crash data

- Two generalized linear model:

$$\log\{E(Crash)\} = \beta_{10}+\beta_{11}Lane+\beta_{12}AADT+\beta_{13}Med+\beta_{14}CBD+\beta_{15}Length$$

$$\text{logit}\{E(Severe)\} = \beta_{20}+\beta_{21}Lane+\beta_{22}AADT+\beta_{23}Med+\beta_{24}CBD+\beta_{25}Length$$

- Two dependent variables:
  - crash frequency (Crash)
  - presence of crash severity (Severe)
- Five covariates of interest:
  - the number of through lanes (Lane)
  - average annual daily traffic (AADT)
  - presence of median (Med)
  - central business district (CBD)
  - length of segment (Length)

Crash data

Table : Estimated coefficients along with p-values from Wald test.

Covariate	log(*Crash*)		logit(*Severe*)	
	Estimator	p-value	Estimator	p-value
intercept	0.0963	0.841	-4.5240	0.004
Lane	-0.0623	0.184	0.1410	0.318
AADT	**0.0002**	**0.000**	0.0000	0.586
Med	-0.1445	0.121	**-0.8046**	**0.005**
CBD	0.1774	0.194	-0.9306	0.052
Length	**0.6075**	**0.000**	-0.5496	0.145

Crash data

Table : Top is from multivariate marginal model and bottom is under univariate marginal model

Covariate	log(*Crash*)		logit(*Severe*)	
	Estimator	*p*-value	Estimator	*p*-value
intercept	0.0963	0.841	-4.5240	0.004
Lane	-0.0623	0.184	0.1410	0.318
AADT	**0.0002**	**0.000**	0.0000	0.586
Med	-0.1445	0.121	**-0.8046**	**0.005**
CBD	0.1774	0.194	-0.9306	0.052
Length	**0.6075**	**0.000**	-0.5496	0.145
intercept	0.0621	0.897	-4.8741	0.002
Lane	-0.0590	0.209	0.1774	0.211
AADT	**0.0002**	**0.000**	0.0000	0.674
Med	-0.1649	0.075	**-0.7435**	**0.011**
CBD	**0.2644**	**0.045**	-0.9451	0.051
Length	**0.6238**	**0.000**	-0.6415	0.099

Three correlated responses

- A marginal regression model:

$$y_{ijk} = \beta_k x_{ij} + e_{ijk}, \quad \text{for } i = 1, \ldots, n,\ j = 1, \ldots, m,\ \text{and } k = 1, \ldots, 3,$$

  - $x_{ij} \sim$ Uniform(0,1)
  - $e_i = (e_{i11}, \ldots, e_{im1}, e_{i12}, \ldots, e_{im2}, e_{i13}, \ldots, e_{im3})^T \sim N(0, R)$, where R is an exchangeable correlation with $\rho = 0.7$

Three correlated responses

- A marginal regression model:

$$y_{ijk} = \beta_k x_{ij} + e_{ijk}, \quad \text{for } i = 1, \dots, n,\ j = 1, \dots, m,\ \text{and } k = 1, \dots, 3,$$

  - $x_{ij} \sim \text{Uniform}(0,1)$
  - $e_i = (e_{i11}, \dots, e_{im1}, e_{i12}, \dots, e_{im2}, e_{i13}, \dots, e_{im3})^T \sim N(0, R)$, where R is an exchangeable correlation with $\rho = 0.7$ $\Rightarrow \text{cor}(y_{ijk}, y_{ij'k}) = \text{cor}(y_{ijk}, y_{ijk'}) = 0.7$, where $j \neq j'$ and $k \neq k'$.

Three correlated responses

- A marginal regression model:

$$y_{ijk} = \beta_k x_{ij} + e_{ijk}, \quad \text{for } i = 1, \dots, n, \; j = 1, \dots, m, \; \text{and } k = 1, \dots, 3,$$

  - $x_{ij} \sim$ Uniform(0,1)
  - $e_i = (e_{i11}, \dots, e_{im1}, e_{i12}, \dots, e_{im2}, e_{i13}, \dots, e_{im3})^T \sim N(0, R)$, where R is an exchangeable correlation with $\rho = 0.7$ $\Rightarrow \text{cor}(y_{ijk}, y_{ij'k}) = \text{cor}(y_{ijk}, y_{ijk'}) = 0.7$, where $j \neq j'$ and $k \neq k'$.
  - $\boldsymbol{\beta}^T = (\beta_1, \beta_2, \beta_3) = (0.2, 0.4, 0.6)$
  - $n = 25, 100$ and $m = 5, 10$ in 200 simulations

Three correlated responses

Table : Mean squared errors (MSE) for estimators using the QIF under multivariate (MQIF) and univariate (UQIF) models. Exchangeable (EX) and AR1 working correlation structures are applied

	m	Model	n=25	n=100
EX	5	MQIF	0.018	0.004
		UQIF	0.029	0.007
	10	MQIF	0.009	0.002
		UQIF	0.015	0.004
AR1	5	MQIF	0.029	0.006
		UQIF	0.038	0.010
	10	MQIF	0.012	0.003
		UQIF	0.022	0.005

$\text{MSE}(\hat{\boldsymbol{\beta}}) = \frac{1}{600}\sum_{j=1}^{200}\sum_{i=1}^{3}\left(\beta_i - \hat{\beta}_i^{(j)}\right)^2$, where β_i is the true parameter and $\hat{\beta}_i^{(j)}$ is the estimator from the jth simulation.

Three correlated responses

Table : Proportions of times that the null hypothesis ($H_0 : \beta_i = 0$) for $i = 1, 2, 3$ is rejected through a chi-squared test

	Model	n=25					
		m=5			m=10		
		β_1	β_2	β_3	β_1	β_2	β_3
EX	MQIF	0.525	0.960	1.000	0.755	0.995	1.000
	UQIF	0.200	0.620	0.895	0.370	0.870	1.000
AR1	MQIF	0.505	0.950	1.000	0.715	0.990	1.000
	UQIF	0.180	0.500	0.800	0.250	0.720	0.965
	Model	n=100					
		m=5			m=10		
		β_1	β_2	β_3	β_1	β_2	β_3
EX	MQIF	0.945	1.000	1.000	1.000	1.000	1.000
	UQIF	0.710	0.980	1.000	0.945	1.000	1.000
AR1	MQIF	0.910	1.000	1.000	0.990	1.000	1.000
	UQIF	0.540	0.970	1.000	0.810	1.000	1.000

Concluding remarks

- Incorporate correlations both within subjects and different responses
- Estimate all parameters simultaneously
- Apply to correlated discrete as well as continuous responses
- Possess asymptotic properties and yield more efficient estimates
- Propose a statistical inference for hypothesis test

토론

토론

<u>**사회(좌장)**</u>

김원철 책임연구원, 충남연구원

<u>**토론자**</u>

김기용 박사, 교통안전공단

이정범 박사, 대전발전연구원

김형철 박사, 충남연구원

질의응답

감사합니다!

CNI 세미나 2015-75

다중응답 종단자료의 효율적 추정방법 세미나

주최 · 주관 : 충남연구원

2015.7.21