Categorical Data
and
Estimating Equations

Objectives:

- Generalized Estimating Equations
- Model specification and interpretation
- GEE Extensions

Motivation

- Vaccine preparedness study (VPS), 1995-1998.
  - 5,000 subjects with high-risk for HIV acquisition.
  - Feasibility of phase III HIV vaccine trials.
  - Willingness, knowledge?

- Informed Consent Substudy (IC)
  - 20% selected to undergo mock informed consent.
  - Understanding of key items at 6mo, 12mo, 18mo.

- **Reference**: Coletti et al. (2003) *JAIDS*

Simple Example: VPS IC Analysis

To develop methods to assure that participants in future HIV vaccine trials understand the implications and potential risks of participating, the HIVNET developed a prototype informed consent process for a hypothetical future HIV vaccine efficacy trial. A 20% random subsample of the 4,892 Vaccine Preparedness Study (VPS) cohort was enrolled in a mock informed consent process at month 3 of the study (between the enrollment visit and the scheduled follow-up visit at month 6). Knowledge of 10 key HIV concepts and willingness to participate in future vaccine efficacy trials among these participants were compared with knowledge and willingness levels of participants not randomized to the informed consent procedure.

Motivation

Items:

- **Q4SAFE** – "We can be sure that the HIV vaccine is safe once we begin phase III testing"

- **NURSE** – "The study nurse decides whether placebo or active product is given to a participant"

EDA – time cross-sectional

BASELINE

```
ICgroup |q4safe0
        |0      |1      |RowTotl|
 -------+-------+-------+-------+
 0      |218    |282    |500    |
        |0.44   |0.56   |0.5    |
 -------+-------+-------+-------+
 1      |216    |284    |500    |
        |0.43   |0.57   |0.5    |
 -------+-------+-------+-------+
```

EDA – time cross-sectional

```
ICgroup |q4safe6
        |0      |1      |RowTotl|
 -------+-------+-------+-------+
 0      |226    |274    |500    |
        |0.45   |0.55   |0.5    |
 -------+-------+-------+-------+
 1      |180    |320    |500    |
        |0.36   |0.64   |0.5    |
 -------+-------+-------+-------+

ICgroup |q4safe12
        |0      |1      |RowTotl|
 -------+-------+-------+-------+
 0      |208    |292    |500    |
        |0.42   |0.58   |0.5    |
 -------+-------+-------+-------+
 1      |177    |323    |500    |
        |0.35   |0.65   |0.5    |
 -------+-------+-------+-------+
```

Regression Models

Q: Is there an intervention effect? If so what is it?

Q: Does the intervention effect "wane"?

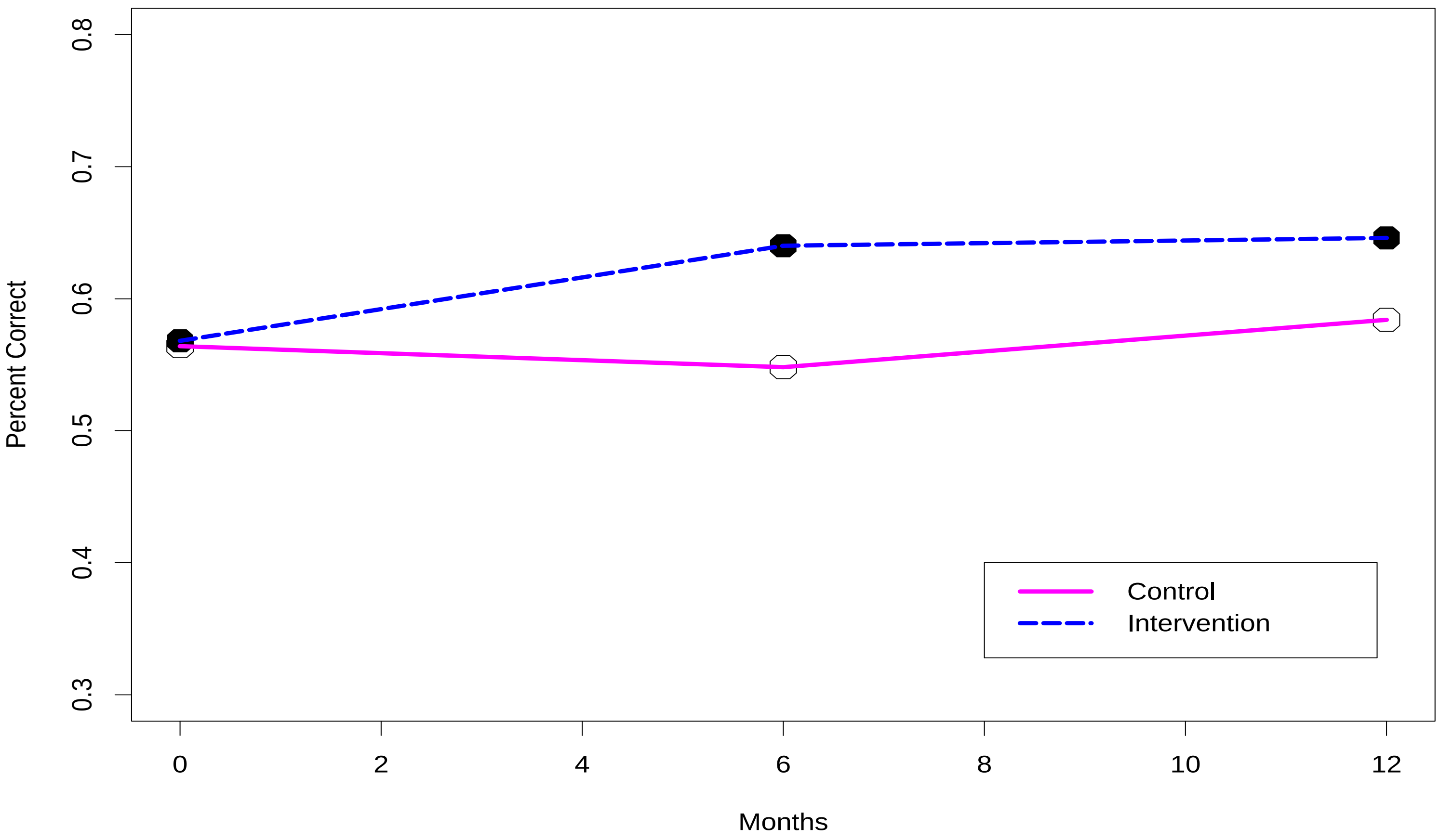
Regression Models:

$$Y_{ij} = \text{response at time } j \text{ for subject } i$$

$$\mu_{ij} = E(Y_{ij} \mid X_{ij})$$

$$\begin{aligned} \text{logit}(\mu_{ij}) = \; & \beta_0 \; + \; \beta_1 \cdot (\text{Tx}) \; + \\ & \beta_2 \cdot (\text{Time=6}) \; + \; \beta_3 \cdot (\text{Time=12}) \; + \\ & \beta_4 \cdot (\text{Time=6} \cdot \text{Tx}) \; + \; \beta_5 \cdot (\text{Time=12} \cdot \text{Tx}) \end{aligned}$$

HIVNET IC – Percent by Time and Group



Regression Models

Analysis Options:

- Cross-sectional analyses at 0, 6, and 12 month.

⋆ Semi-parametric methods (GEE)

- "Random effects" models. / Transition models.

GEE-1 Liang and Zeger (1986)

Q: We've seen that WLS and the LMM assuming multivariate normality can be used for semi-parametric and likelihood based estimation with continuous response variables. What about models/methods for discrete response variables such as binary data?

Answer: There are semi-parametric approaches (GEE) and likelihood based methods (GLMMs and other models).

$\boxed{\star\star\star}$ Let's consider GEE first:

- Focus on a generalized linear model regression parameter that characterizes systematic variation across covariate levels: $\boldsymbol{\beta}$.

- Correlation structure is a **nuisance** feature of the data.

GEE1 - Simple Illustration / WLS Review

Consider vectors, $\boldsymbol{Y}_i = (Y_{i1}, Y_{i2}, ..., Y_{in})$, (length $n \;\; \forall \; i$) observed on N subjects. Let $\boldsymbol{X}_i = (\boldsymbol{X}_{i1}, \boldsymbol{X}_{i2}, \ldots, \boldsymbol{X}_{in})$ be corresponding covariates.

Q: What happens if we fit the linear model ignoring the clustering?

Mean Model

$$E[Y_{ij} \mid \boldsymbol{X}_{ij}] = \boldsymbol{X}_{ij}\boldsymbol{\beta}$$

OLS Estimation

$$
\begin{aligned}
\widehat{\boldsymbol{\beta}} &= \left(\sum_i \boldsymbol{X}_i^T \boldsymbol{X}_i\right)^{-1} \sum_i \boldsymbol{X}_i^T \boldsymbol{Y}_i \\
E[\widehat{\boldsymbol{\beta}}] &= \left(\sum_i \boldsymbol{X}_i^T \boldsymbol{X}_i\right)^{-1} \sum_i \boldsymbol{X}_i^T E[\boldsymbol{Y}_i] \\
&= \left(\sum_i \boldsymbol{X}_i^T \boldsymbol{X}_i\right)^{-1} \sum_i \boldsymbol{X}_i^T \boldsymbol{X}_i \boldsymbol{\beta} \\
&= \boldsymbol{\beta}
\end{aligned}
$$

Therefore, ignoring the clustering still provides a valid point estimate for $\boldsymbol{\beta}$

GEE1 - Simple Illustration

Q: What about the estimate for var$(\widehat{\boldsymbol{\beta}})$?

$$\begin{aligned}
\text{var}(\widehat{\boldsymbol{\beta}}) &= \text{var}\left\{ \left(\sum_i \boldsymbol{X}_i^T \boldsymbol{X}_i \right)^{-1} \sum_i \boldsymbol{X}_i^T \boldsymbol{Y}_i \right\} \\
&= \left(\sum_i \boldsymbol{X}_i^T \boldsymbol{X}_i \right)^{-1} \sum_i \boldsymbol{X}_i^T V[\boldsymbol{Y}_i] \boldsymbol{X}_i \left(\sum_i \boldsymbol{X}_i^T \boldsymbol{X}_i \right)^{-1} \\
&? \quad \sigma^2 \left(\sum_i \boldsymbol{X}_i^T \boldsymbol{X}_i \right)^{-1}
\end{aligned}$$

Therefore, we could be wrong if we used the variance obtained by assuming independence...

GEE1 - Simple Illustration

Q: What if we used WLS (weighted least squares)?

$$
\begin{aligned}
\widehat{\boldsymbol{\beta}} &= \left(\sum_i \boldsymbol{X}_i^T \boldsymbol{W}_i \boldsymbol{X}_i \right)^{-1} \sum_i \boldsymbol{X}_i^T \boldsymbol{W}_i \boldsymbol{Y}_i \\
E[\widehat{\boldsymbol{\beta}}] &= \boldsymbol{\beta}
\end{aligned}
$$

and...

$$
\begin{aligned}
\text{var}(\widehat{\boldsymbol{\beta}}) &= \text{var}\left\{ \left(\sum_i \boldsymbol{X}_i^T \boldsymbol{W}_i \boldsymbol{X}_i \right)^{-1} \sum_i \boldsymbol{X}_i^T \boldsymbol{W}_i \boldsymbol{Y}_i \right\} \\
&= \boldsymbol{A}^{-1} \sum_i \boldsymbol{X}_i^T \boldsymbol{W}_i V[\boldsymbol{Y}_i] \boldsymbol{W}_i \boldsymbol{X}_i \;\; \boldsymbol{A}^{-1} \\
? & \quad \left(\sum_i \boldsymbol{X}_i^T \boldsymbol{W}_i \boldsymbol{X}_i \right)^{-1} = \boldsymbol{A}^{-1}
\end{aligned}
$$

method	estimator	minimization solves
OLS	$\widehat{\boldsymbol{\beta}}_I = \left(\sum_i \boldsymbol{X}_i^T \boldsymbol{X}_i\right)^{-1} \sum_i \boldsymbol{X}_i^T \boldsymbol{Y}_i$	$\sum_i \boldsymbol{X}_i^T \{\boldsymbol{Y}_i - \boldsymbol{X}_i \boldsymbol{\beta}\} = \mathbf{0}$
WLS	$\widehat{\boldsymbol{\beta}}_W = \left(\sum_i \boldsymbol{X}_i^T \boldsymbol{W}_i \boldsymbol{X}_i\right)^{-1} \sum_i \boldsymbol{X}_i^T \boldsymbol{W}_i \boldsymbol{Y}_i$	$\sum_i \boldsymbol{X}_i^T \boldsymbol{W}_i \{\boldsymbol{Y}_i - \boldsymbol{X}_i \boldsymbol{\beta}\} = \mathbf{0}$

method		variance expression
OLS	model based	$\sigma^2 \left(\sum_i \boldsymbol{X}_i^T \boldsymbol{X}_i\right)^{-1}$
	general	$\left(\sum_i \boldsymbol{X}_i^T \boldsymbol{X}_i\right)^{-1} \sum_i \boldsymbol{X}_i^T V[\boldsymbol{Y}_i] \boldsymbol{X}_i \quad \left(\sum_i \boldsymbol{X}_i^T \boldsymbol{X}_i\right)^{-1}$
WLS	model based	$\left(\sum_i \boldsymbol{X}_i^T \boldsymbol{W}_i \boldsymbol{X}_i\right)^{-1}$
	general	$\left(\sum_i \boldsymbol{X}_i^T \boldsymbol{W}_i \boldsymbol{X}_i\right)^{-1} \sum_i \boldsymbol{X}_i^T \boldsymbol{W}_i V[\boldsymbol{Y}_i] \boldsymbol{W}_i \boldsymbol{X}_i \quad \left(\sum_i \boldsymbol{X}_i^T \boldsymbol{W}_i \boldsymbol{X}_i\right)^{-1}$

GEE1 - Notation

Data:

$Y_{i1}, Y_{i2}, \ \ldots, \ Y_{ij}, \ \ldots, \ Y_{in_i}$	response variables
$\boldsymbol{X}_{i1}, \boldsymbol{X}_{i2}, \ \ldots, \boldsymbol{X}_{ij}, \ \ldots, \ \boldsymbol{X}_{in_i}$	covariate vectors

$i \in [1, N]$: index for cluster / subject

$j \in [1, n_i]$: index for measurement within cluster

Assumptions:

- Measurements are independent across clusters (can be relaxed for time and space).

- Measurements may be correlated within cluster.

Mean Model: (primary focus of analysis)

$$
\begin{aligned}
E[Y_{ij} \mid \boldsymbol{X}_{ij}] &= \mu_{ij} \\
g(\mu_{ij}) &= \boldsymbol{X}_{ij}\boldsymbol{\beta}
\end{aligned}
$$

Marginal Mean

Mean Model: (primary focus of analysis)

$$\begin{aligned} E[Y_{ij} \mid \boldsymbol{X}_{ij}] &= \mu_{ij} \\ g(\mu_{ij}) &= \boldsymbol{X}_{ij}\boldsymbol{\beta} \end{aligned}$$

This can be any generalized linear model. For example,

$$\begin{aligned} P[Y_{ij} = 1 \mid \boldsymbol{X}_{ij}] &= \pi_{ij} \\ \log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) &= \boldsymbol{X}_{ij}\boldsymbol{\beta} \end{aligned}$$

Q: Why is this a **marginal** mean?

Answer: There's no extra variable(s) that we condition on (like in some other models for multivariate data).

Other (non-marginal) Mean Models

- **Log-linear models**:

$$E[\ Y_{ij} \mid Y_{ik},\ \ k \neq j,\ \ \boldsymbol{X}_{ij}]$$

- **Transition models**:

$$E[\ Y_{ij} \mid Y_{ik},\ \ k < j,\ \ \boldsymbol{X}_{ij}]$$

- **Latent variable models**:

$$E[Y_{ij} \mid \ b_{ij},\ \ \boldsymbol{X}_{ij}]$$

GEE - covariance

Q: But what about the fact that data are clustered?

Answer: Choose a Correlation Model: (nuisance)

$$\begin{aligned}
\text{var}(Y_{ij} \mid \boldsymbol{X}_i) &= V_{ij} \\
\boldsymbol{S}_i(\boldsymbol{\mu}_i) &= \text{diag}(V_{ij})
\end{aligned}$$

$$\begin{aligned}
\text{corr}(Y_{ij}, Y_{ik} \mid \boldsymbol{X}_i) &= \rho_{ijk}(\boldsymbol{\alpha}) \\
\boldsymbol{R}_i(\boldsymbol{\alpha}) &= \text{correlation matrix} \\
\boldsymbol{V}_i(\boldsymbol{\alpha}) &= \text{cov}(\boldsymbol{Y}_i \mid \boldsymbol{X}_i) \\
&= \boldsymbol{S}_i^{1/2} \boldsymbol{R}_i(\boldsymbol{\alpha}) \boldsymbol{S}_i^{1/2}
\end{aligned}$$

GEE1 - Common Correlation Models

Independence:

$$\boldsymbol{R}_i = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Exchangeable / equicorrelation:

$$\boldsymbol{R}_i(\alpha) = \begin{bmatrix} 1 & \alpha & \alpha & \alpha \\ \alpha & 1 & \alpha & \alpha \\ \alpha & \alpha & 1 & \alpha \\ \alpha & \alpha & \alpha & 1 \end{bmatrix}$$

Unstructured:

$$\boldsymbol{R}_i(\boldsymbol{\alpha}) = \begin{bmatrix} 1 & \alpha_{12} & \alpha_{13} & \alpha_{14} \\ \alpha_{21} & 1 & \alpha_{23} & \alpha_{24} \\ \alpha_{31} & \alpha_{32} & 1 & \alpha_{34} \\ \alpha_{41} & \alpha_{42} & \alpha_{43} & 1 \end{bmatrix}$$

GEE1 - Common Correlation Models

AR-1:

$$\boldsymbol{R}_i(\alpha) = \begin{bmatrix} 1 & \alpha^1 & \alpha^2 & \alpha^3 \\ \alpha^1 & 1 & \alpha^1 & \alpha^2 \\ \alpha^2 & \alpha^1 & 1 & \alpha^1 \\ \alpha^3 & \alpha^2 & \alpha^1 & 1 \end{bmatrix}$$

Stationary m-dependent ($m = 2$):

$$\boldsymbol{R}_i(\boldsymbol{\alpha}) = \begin{bmatrix} 1 & \alpha_1 & \alpha_2 & 0 \\ \alpha_1 & 1 & \alpha_1 & \alpha_2 \\ \alpha_2 & \alpha_1 & 1 & \alpha_1 \\ 0 & \alpha_2 & \alpha_1 & 1 \end{bmatrix}$$

Non-stationary m-dependent ($m = 2$):

$$\boldsymbol{R}_i(\boldsymbol{\alpha}) = \begin{bmatrix} 1 & \alpha_{12} & \alpha_{13} & 0 \\ \alpha_{21} & 1 & \alpha_{23} & \alpha_{24} \\ \alpha_{31} & \alpha_{32} & 1 & \alpha_{34} \\ 0 & \alpha_{42} & \alpha_{43} & 1 \end{bmatrix}$$

GEE1 - semiparametric model

Q: Does specification of a mean model, $\mu_{ij}(\boldsymbol{\beta})$, and a correlation model, $\boldsymbol{R}_i(\boldsymbol{\alpha})$, identify a complete probability model for $\boldsymbol{Y}_i$?

- **NO**. If further assumptions can be made then a probability model can be identified. In general, for categorical data this is a difficult task.

- The model $\{\mu_{ij}(\boldsymbol{\beta}), \boldsymbol{R}_i(\boldsymbol{\alpha})\}$ is *semiparametric* since it only specifies the first two multivariate moments of $\boldsymbol{Y}_i$.

Q: Without a likelihood function how can we estimate $\boldsymbol{\beta}$ (and possibly $\boldsymbol{\alpha}$) and perform valid statistical inference that takes the dependence into consideration?

Answer: Construct an unbiased estimating function.

GEE1 - estimation

Define:

$$\boldsymbol{D}_i(\boldsymbol{\beta}) = \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}}$$

$$\boldsymbol{D}_i(j,k) = \frac{\partial \mu_{ij}}{\partial \beta_k}$$

$$\boldsymbol{V}_i(\boldsymbol{\beta},\boldsymbol{\alpha}) = \boldsymbol{S}_i(\boldsymbol{\mu}_i)^{1/2}\boldsymbol{R}_i(\boldsymbol{\alpha})\boldsymbol{S}_i(\boldsymbol{\mu}_i)^{1/2}$$

Define:

$$U(\boldsymbol{\beta}) = \sum_{i=1}^{N} \boldsymbol{D}_i^T(\boldsymbol{\beta})\boldsymbol{V}_i^{-1}(\boldsymbol{\beta},\boldsymbol{\alpha})\left\{\boldsymbol{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta})\right\}$$

Note:

- $U(\boldsymbol{\beta})$ is called an estimating function.
- $U(\boldsymbol{\beta})$ also depends on the model/value for $\boldsymbol{\alpha}$.

GEE1 - estimation

Estimating Equations: solution to the following system of equations defines an estimator $\widehat{\boldsymbol{\beta}}$

$$
\begin{aligned}
\mathbf{0} &= U(\widehat{\boldsymbol{\beta}}) \\
&= \sum_{i=1}^{N} \boldsymbol{D}_i^T(\boldsymbol{\beta}) \boldsymbol{V}_i^{-1}(\boldsymbol{\beta}, \boldsymbol{\alpha}) \left\{ \boldsymbol{Y}_i - \boldsymbol{\mu}_i(\widehat{\boldsymbol{\beta}}) \right\}
\end{aligned}
$$

Note: use $\boldsymbol{D}_i$, and $\boldsymbol{V}_i(\boldsymbol{\alpha})$ to denote $\boldsymbol{D}_i(\boldsymbol{\beta})$ and $\boldsymbol{V}_i(\boldsymbol{\beta}, \boldsymbol{\alpha})$.

GEE1 - estimation

Q: What are the properties of $\widehat{\boldsymbol{\beta}}$, the regression estimate?

Robustness Property:

- The regression coefficient estimate, $\widehat{\boldsymbol{\beta}}$, will be correct (in large samples) **even if** you choose the **wrong** dependence model.

- However, the **variance** of the regression estimate must capture the correlation in the data, either through choosing the correct correlation model, or via an alternative variance estimate.

- Choosing a "wise" (approximately correct) correlation model will make the regression estimate $\widehat{\boldsymbol{\beta}}$ more efficient in the extraction of information (ie. $\widehat{\boldsymbol{\beta}}$ has the smallest variance if you select the correct correlation model).

GEE1 - estimation

Q: What are the properties of $\widehat{\boldsymbol{\beta}}$, the estimating equations solution?

Property:

The root $U(\widehat{\boldsymbol{\beta}}) = \mathbf{0}$ is consistent and asymptotically normal.

$$\sqrt{N}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \rightarrow \mathcal{N}(\ \mathbf{0},\ \lim_{N} N \cdot \boldsymbol{H}_N\ )$$

where

$$\boldsymbol{H}_N = (\mathcal{I}_M^{-1}) \left(\sum_{i=1}^{N} E[U_i(\boldsymbol{\beta}) U_i(\boldsymbol{\beta})^T] \right) (\mathcal{I}_M^{-1})$$

$$\mathcal{I}_M = \sum_{i=1}^{N} \boldsymbol{D}_i^T \boldsymbol{V}_i(\boldsymbol{\alpha})^{-1} \boldsymbol{D}_i$$

GEE1 - estimation

$$\sum_{i=1}^{N} E[U_i(\boldsymbol{\beta})U_i(\boldsymbol{\beta})^T] \quad = \quad \sum_{i=1}^{N} \boldsymbol{D}_i^T \boldsymbol{V}_i^{-1}(\boldsymbol{\alpha})\text{cov}(\boldsymbol{Y}_i)\boldsymbol{V}_i^{-1}(\boldsymbol{\alpha})\boldsymbol{D}_i$$

(Proof: we've already done this for a general estimating function!)

- This is the familiar "sandwich" form with

$$\begin{aligned} \boldsymbol{A}_N &= \frac{1}{N}\mathcal{I}_M \\ \boldsymbol{B}_N &= \frac{1}{N}\sum_i \text{var}[\boldsymbol{U}_i(\boldsymbol{\beta})] \end{aligned}$$

Define: The **model based variance** estimate for $V[\widehat{\boldsymbol{\beta}}]$ is given by $\mathcal{I}_M^{-1}$.

GEE1 - estimation

If the correlation model is correct then...

$$\begin{aligned}
\sum_i E[U_i U_i^T] &= \sum_i \boldsymbol{D}_i^T \boldsymbol{V}_i^{-1}(\boldsymbol{\alpha}) \mathsf{cov}(\boldsymbol{Y}_i) \boldsymbol{V}_i^{-1}(\boldsymbol{\alpha}) \boldsymbol{D}_i \\
&= \sum_i \boldsymbol{D}_i^T \boldsymbol{V}_i^{-1}(\boldsymbol{\alpha}) \boldsymbol{V}_i(\boldsymbol{\alpha}) \boldsymbol{V}_i^{-1}(\boldsymbol{\alpha}) \boldsymbol{D}_i \\
&= \sum_i \boldsymbol{D}_i^T \boldsymbol{V}_i^{-1}(\boldsymbol{\alpha}) \boldsymbol{D}_i \\
&= \mathcal{I}_M
\end{aligned}$$

$$\begin{aligned}
\boldsymbol{H}_N &= (\mathcal{I}_M)^{-1} (\mathcal{I}_M) (\mathcal{I}_M)^{-1} \\
&= \mathcal{I}_M^{-1}
\end{aligned}$$

GEE1 - estimation

If the correlation model is **not** correct then...

$$\boldsymbol{H}_N = (\mathcal{I}_M)^{-1} \left(\sum_i E[U_i U_i^T] \right) (\mathcal{I}_M)^{-1}$$

and, the middle term may not be correctly specified by $\mathcal{I}_M$.

Q: Can we use the data to estimate the middle of the "sandwich"?

An empirical variance estimate can be used in the middle of the variance "sandwich":

$$\begin{aligned}\widehat{\boldsymbol{H}}_N &= (\mathcal{I}_M)^{-1}\left(\sum_i U_i U_i^T\right)(\mathcal{I}_M)^{-1} \\ &= (\mathcal{I}_M)^{-1}\left(\sum_i \boldsymbol{D}_i^T \boldsymbol{V}_i^{-1}\{\boldsymbol{Y}_i - \boldsymbol{\mu}_i\}\{\boldsymbol{Y}_i - \boldsymbol{\mu}_i\}^T \boldsymbol{V}_i^{-1}\boldsymbol{D}_i\right)(\mathcal{I}_M)^{-1}\end{aligned}$$

Define: The **empirical variance** estimate for $V[\widehat{\boldsymbol{\beta}}]$ is given by $\widehat{\boldsymbol{H}}_N$.

NOTE: Use of an empirical variance estimator requires sufficient independent replication, $i = 1, 2, \ldots, N$, (ie. N large).

GEE1 - sandwich estimator

Note:

$$\widehat{\boldsymbol{V}}_i = \{\boldsymbol{Y}_i - \boldsymbol{\mu}_i\}\{\boldsymbol{Y}_i - \boldsymbol{\mu}_i\}^T$$

is a poor estimate for cov$(\boldsymbol{Y}_i)$.

However, we don't need a good estimate for each i but rather a good estimate of the average (total) covariance:

$$\frac{1}{N}\mathcal{I}_R = \frac{1}{N}\sum_i \boldsymbol{D}_i^T \boldsymbol{V}_i^{-1} \text{cov}(\boldsymbol{Y}_i) \boldsymbol{V}_i^{-1} \boldsymbol{D}_i$$

$$\frac{1}{N}\widehat{\mathcal{I}_R} = \frac{1}{N}\sum_i \boldsymbol{D}_i^T \boldsymbol{V}_i^{-1} \{\boldsymbol{Y}_i - \boldsymbol{\mu}_i\}\{\boldsymbol{Y}_i - \boldsymbol{\mu}_i\}^T \boldsymbol{V}_i^{-1} \boldsymbol{D}_i$$

which can be well estimated with sufficient independent replication (law of large numbers for non-iid).

GEE1 - α unknown

In the previous discussion we have assumed that $\boldsymbol{\alpha}$, the correlation parameters, are fixed and known.

Q: How can estimates for $\boldsymbol{\alpha}$ be obtained and what is the impact on the estimation of $\boldsymbol{\beta}$?

Simple Moment Estimators Liang and Zeger (1986):

$$\text{Define} \quad : \quad r_{ij} = \frac{\{Y_{ij} - \mu_{ij}(\widehat{\boldsymbol{\beta}})\}}{\hat{V}_{ij}^{1/2}}$$

$$\text{General} \quad R(j,k) = \sum_{i=1}^{N} r_{ij} r_{ik} / (N - p)$$

$$p = \text{dim}(\boldsymbol{\beta})$$

Example 1 : 1-dependence

$$\hat{\alpha} = \sum_{t=1}^{n} \hat{\alpha}_t / (n-1)$$

$$\hat{\alpha}_t = \sum_{i=1}^{N} r_{it} r_{it+1} / (N-p)$$

Example 2 : exchangeable

$$\hat{\alpha} = \sum_{i=1}^{N} \sum_{j \neq k} r_{ij} r_{ik} / (\sum_{i} n_i(n_i - 1) - p)$$

GEE1 - α unknown

NOTE: For count data and continuous data the GLM also includes a scale parameter, ϕ, such that

$$V[Y_{ij} \mid \boldsymbol{X}_i] \;=\; \phi \cdot \nu(\mu_{ij})$$

and this can also be estimated through simple moment equations:

$$\hat{\phi} \;=\; \left(\sum_{i=1}^{N} \sum_{j=1}^{n_i} \frac{\{Y_{ij} - \hat{\mu}_{ij}\}^2}{\hat{\nu}_{ij}} \right) / (\sum_i n_i - p)$$

Theorem 2 (Liang and Zeger, 1986):
Under mild regularity conditions (references later), and

- $\widehat{\boldsymbol{\alpha}}$ is $N^{1/2}$ consistent given $\boldsymbol{\beta}$ and ϕ
- $\hat{\phi}$ is $N^{1/2}$ consistent given $\boldsymbol{\beta}$

- $|\partial\hat{\alpha}(\beta,\phi)/\partial\phi| \leq H(Y,\beta)$ which is $O_p(1)$

Then, $\widehat{\boldsymbol{\beta}} \sim N(\,\boldsymbol{\beta},\ \boldsymbol{H}_N\,)$.

NOTE: These assumptions may not be satisfied with certain moment-estimators for correlation model parameters when misspecified - see Crowder (1995)

GEE1 Estimation

To obtain a GEE1 estimate of $\boldsymbol{\beta}$ we solve:

$$\mathbf{0} = \sum_{i=1}^{n} \boldsymbol{D}_i^T \boldsymbol{V}_i^{-1}(\boldsymbol{\alpha})\{\boldsymbol{Y}_i - \mu_i(\boldsymbol{\beta})\} \quad (1)$$

$$\widehat{\boldsymbol{\alpha}} = \text{simple function of } r_{ij} \quad (2)$$

- Given an estimate for $\boldsymbol{\beta}$, the moment estimator of $\boldsymbol{\alpha}$ can be calculated.
- Given an estimate for $\boldsymbol{\alpha}$, the estimating equation can be solved using a Fisher scoring algorithm:

$$\boldsymbol{\beta}^{(j+1)} = \boldsymbol{\beta}^{(j)} + \left(\sum_i \boldsymbol{D}_i^T \boldsymbol{V}_i^{-1} \boldsymbol{D}_i \right)^{-1} \sum_i \boldsymbol{D}_i^T \boldsymbol{V}_i^{-1} \{\boldsymbol{Y}_i - \boldsymbol{\mu}_i\}$$

- These two estimation steps are iterated until convergence is achieved.

Comments

- The estimating function $U(\boldsymbol{\beta})$ is unbiased for any choice of $\boldsymbol{V}_i(\boldsymbol{\alpha})$:

$$\begin{aligned} E[\sum_i \boldsymbol{D}_i^T \boldsymbol{V}_i^{-1}(\boldsymbol{\alpha})\{\boldsymbol{Y}_i - \mu_i(\boldsymbol{\beta})\}] &= \\ \sum_i \boldsymbol{D}_i^T \boldsymbol{V}_i^{-1}(\boldsymbol{\alpha})\{E[\boldsymbol{Y}_i - \mu_i(\boldsymbol{\beta})]\} &= \mathbf{0} \end{aligned}$$

- This unbiasedness implies that $\widehat{\boldsymbol{\beta}}$ obtained as the root of the estimating function is a consistent estimator, for any $\boldsymbol{V}_i(\boldsymbol{\alpha})$.

$$\frac{1}{N} U(\boldsymbol{\beta}) \to E[U(\boldsymbol{\beta})] = 0$$

- The model selected for $\boldsymbol{R}_i(\boldsymbol{\alpha})$ is called the "working correlation" since this need not be the true correlation to obtain a valid point estimate $\widehat{\boldsymbol{\beta}}$.

GEE and Standard Error Estimates

GEE Specification

(1) A flexible regression model for the mean response (linear, logistic).

(2) A correlation model (independence, exchangeable).

Q: What if the selected correlation model is not correct?

Answer: GEE also computes a **sandwich variance** estimator.

$\Rightarrow$ a.k.a. "empirical variance"

$\Rightarrow$ a.k.a. "robust variance"

$\Rightarrow$ a.k.a. "Huber-White correction"

$\boxed{\star}$ The empirical variance gives valid standard errors for the estimated regression coefficients even if the correlation model was wrong.

- The empirical variance is valid in "large samples" – this means it can be used with data sets that contain at least 40 subjects.

GEE1 - efficiency

Q: Is there much advantage to using "covariance weighting" (ie. non-diagonal working correlation models) with estimating equations?

- Permits use of a model based variance estimate.
- May yield improved efficiency for $\widehat{\boldsymbol{\beta}}$.

Independence Est. Eqs.	Weighted Est. Eqs.
$\boldsymbol{R}_i \ = \ \boldsymbol{I}_{n_i \times n_i}$	$\boldsymbol{R}_i \ = \ \boldsymbol{R}_i(\boldsymbol{\alpha})$
$\widehat{\boldsymbol{\beta}}_I$	$\widehat{\boldsymbol{\beta}}_W$
$V[\widehat{\boldsymbol{\beta}}_I] = \boldsymbol{B}_1^{-1}\boldsymbol{B}_2\boldsymbol{B}_1^{-1}$	$V[\widehat{\boldsymbol{\beta}}_W] = \boldsymbol{C}_1^{-1}\boldsymbol{C}_2\boldsymbol{C}_1^{-1}$

GEE1 - efficiency

Define: $\boldsymbol{S}_i = \mathsf{diag}(V_{ij})$; $\boldsymbol{W}_i = \left(\boldsymbol{S}_i^{1/2}\boldsymbol{R}_i(\boldsymbol{\alpha})\boldsymbol{S}_i^{1/2}\right)^{-1}$

$$
\begin{aligned}
V[\widehat{\boldsymbol{\beta}}_I] &= \boldsymbol{B}_1^{-1}\boldsymbol{B}_2\boldsymbol{B}_1^{-1} \\
\boldsymbol{B}_1 &= \sum_i \boldsymbol{D}_i^T\boldsymbol{S}_i^{-1}\boldsymbol{D}_i \\
\boldsymbol{B}_2 &= \sum_i \boldsymbol{D}_i^T\boldsymbol{S}_i^{-1}\mathsf{cov}(\boldsymbol{Y_i})\boldsymbol{A}_i^{-1}\boldsymbol{D}_i
\end{aligned}
$$

$$
\begin{aligned}
V[\widehat{\boldsymbol{\beta}}_W] &= \boldsymbol{C}_1^{-1}\boldsymbol{C}_2\boldsymbol{C}_1^{-1} \\
\boldsymbol{C}_1 &= \sum_i \boldsymbol{D}_i^T\boldsymbol{W}_i\boldsymbol{D}_i \\
\boldsymbol{C}_2 &= \sum_i \boldsymbol{D}_i^T\boldsymbol{W}_i\mathsf{cov}(\boldsymbol{Y_i})\boldsymbol{W}_i\boldsymbol{D}_i
\end{aligned}
$$

Published Comments

- Liang and Zeger (1986) – "little difference when correlation is moderate"

- McDonald (1993) – the independence estimator "may be recommended for practical purposes"

- Zhao, Prentice, and Self (1992) – assuming independence "can lead to important losses of efficiency"

- Fitzmaurice, Laird, and Rotnitzky (1993) – "important to obtain a close approximation to cov$(\boldsymbol{Y}_i)$ in order to achieve high efficiency"

GEE1 - efficiency

Fitzmaurice (1995):
Compares asymptotic relative efficiency of $\widehat{\boldsymbol{\beta}}_I$ and $\widehat{\boldsymbol{\beta}}_W$ to the MLE estimator assuming a Bahadur (1961) representation for the multivariate binary response. (ie. ARE $= V[\widehat{\boldsymbol{\beta}}_{mle}]/V[\widehat{\boldsymbol{\beta}}_W]$)

Mean Model:

$$\text{logit}(\mu_{it}) = \beta_0 + \beta_1 x_{it} + \beta_2 (t-2) \quad t = 1, 2, 3$$

where $x_{it} = 0/1$ denotes a "group" covariate.

Scenario 1: x_{it} is a "cluster-level" covariate:

$$x_{i1} = x_{i2} = x_{i3}$$

Scenario 2: x_{it} is a "time-varying" covariate.

Conclusions:

1. For "cluster-level" covariate:
   - β_2: ARE close to 1.0
   - β_1: GEE results in loss of efficiency
   - However, GEE-pairwise $\approx$ GEE-independence.
2. For "time-varying" covariate:
   - β_2: ARE close to 1.0
   - β_1: GEE results in possible large loss of efficiency
   - GEE-independence is worse than GEE-pairwise.

"some attempt should generally be made to model the association between responses, even when the association is regarded as a nuisance"

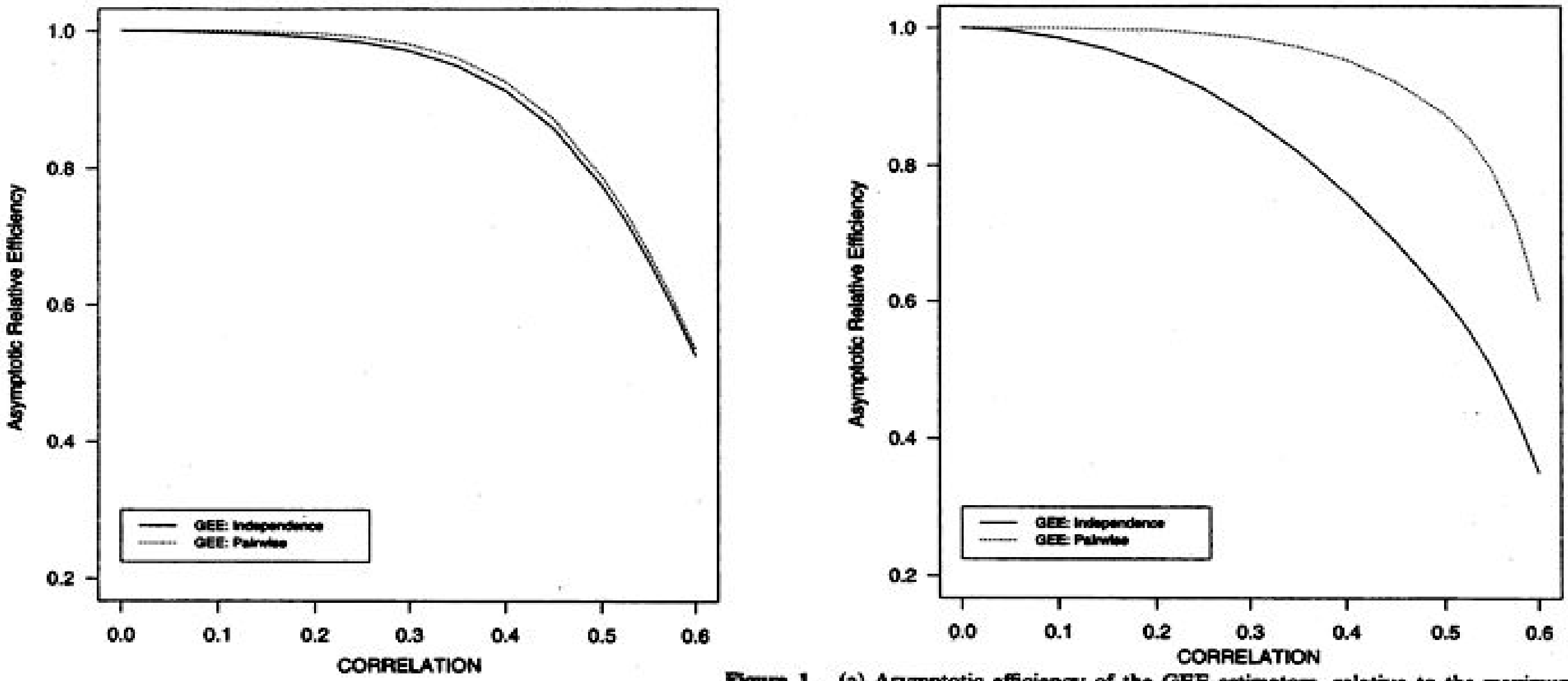


Figure 1. (a) Asymptotic efficiency of the GEE estimators, relative to the maximum likelihood estimator, when the true underlying joint distribution has a Bahadur representation (cluster-level covariate). (b) Asymptotic efficiency of the GEE estimators, relative to the maximum likelihood estimator, when the true underlying joint distribution has a Bahadur representation (within-cluster covariate).

GEE – Summary

Models

- Mean model = general regression model. Focus of analysis.
- Correlation model = simple choices. Nuisance.

Estimates

- Regression estimate, $\widehat{\boldsymbol{\beta}}$.
  - Valid estimate regardless of correlation choice.
  - Correlation choice wrong $\Rightarrow$ $\widehat{\boldsymbol{\beta}}$ still o.k.

- Standard error estimates.
  - Model-based standard errors.
    - ⋆ If correlation choice is correct $\Rightarrow$ valid.
  - Empirical standard errors.
    - ⋆ If correlation choice is <u>incorrect</u> $\Rightarrow$ still valid!

Example: Informed Consent Analysis

- Compare intervention groups, IC=yes to IC=no, separately at month 0, month 6, and month 12.

  $\Rightarrow$ Repeat cross-sectional analyses.

- Use GEE to analyze all follow-up times.

- Consider the question of treatment "waning".

  $\Rightarrow$ compare effects at 6mo and 12mo.

R/S+ Program

```
#
# hivnet-CDA-gee.q
#
# ------------------------------------------------------------
#
# PURPOSE:  analysis of IC data
#
# AUTHOR:  P. Heagerty
#
# DATE:  00/05/09
#
# ------------------------------------------------------------
#
data <- read.table( "HivnetWide.dat", header=F )
#
vps.data <- data.frame(
  id = data[,1],
  risk.group = factor( data[,2], levels=1:4,
                       labels=c("MSM","MaleIDU","WAHR","WAHR+IDU") ),
  education = factor( data[,3], levels=1:6,
                       labels=c("<HS","HS","some college","college",
                                "some post","grad/prof") ),
  age = data[,4],
  cohort = data[,5],
  ICgroup = data[,6],
  will0 = data[,7],
```

```
  know0 = data[,8],
  q4safe0 = data[,9],
  q4safe6 = data[,10],
  q4safe12 = data[,11],
  nurse0 = data[,12],
  nurse6 = data[,13],
  nurse12 = data[,14] )
#
##################################################
#  Separate Analyses at 0, 6, 12                 #
##################################################
#
crosstabs( ~ q4safe0 + ICgroup, data = vps.data )
fit0 <- glm( q4safe0 ~ ICgroup, family=binomial, data=vps.data )
summary( fit0 )
#
crosstabs( ~ q4safe6 + ICgroup, data = vps.data )
fit6 <- glm( q4safe6 ~ ICgroup, family=binomial, data=vps.data )
summary( fit6 )
#
crosstabs( ~ q4safe12 + ICgroup, data = vps.data )
fit12 <- glm( q4safe12 ~ ICgroup, family=binomial, data=vps.data )
summary( fit12 )
#
##################################################
#  Longitudinal data                             #
##################################################
y0 <- vps.data$q4safe0
```

```
y6 <- vps.data$q4safe6
y12 <- vps.data$q4safe12
nsubjects <- length(y0)
#
##### stacked data for regression
#
vps.stacked <- data.frame(
  y = as.vector( rbind( y0, y6, y12 ) ),
  visit = rep( c(0,1,2), nsubjects ),
  ICgroup = as.vector( rbind( vps.data$ICgroup, vps.data$ICgroup,
                              vps.data$ICgroup ) ),
  id = rep( 1:nsubjects, rep(3,nsubjects) )
)
vps.stacked$visit6 <- as.integer( vps.stacked$visit==1 )
vps.stacked$visit12 <- as.integer( vps.stacked$visit==2 )
vps.stacked$post <- as.integer( vps.stacked$visit > 0 )
#
##################################################
#  Pre/Post Analysis of month 0 and month 6      #
##################################################
#
##### gee analysis
#
library(gee)
options( contrasts="contr.treatment" )
#
fit1a <- gee( y ~ visit6 + ICgroup + visit6*ICgroup,
                  family=binomial,
```

```
                  id=id,
                  corstr="independence",
                  subset=(visit<=1), data=vps.stacked )
summary( fit1a )
#
fit1b <- gee( y ~ visit6 + ICgroup + visit6*ICgroup,
                  family=binomial,
                  id=id, corstr="exchangeable",
                  subset=(visit<=1), data=vps.stacked )
summary( fit1b )
#
##################################################
#  Longitudinal Analysis of 0, 6, and 12         #
##################################################
#
fit2a <- gee( y ~ visit6 + visit12 + ICgroup +
                  visit6*ICgroup + visit12*ICgroup,
             family=binomial,
             id=id, corstr="unstructured",
             data=vps.stacked )
summary( fit2a )
#
fit2b <- gee( y ~ post + visit12 + ICgroup +
                  post*ICgroup + visit12*ICgroup,
             family=binomial,
             id=id, corstr="unstructured",
             data=vps.stacked )
summary( fit2b )
```

```
#
##### Adjusted for demographics / risk
#
vps.stacked$risk.group <-  factor( rep(data[,2],rep(3,nsubjects)),
                             levels=1:4,
                             labels=c("MSM","MaleIDU","WAHR","WAHR+IDU") )
vps.stacked$education <- factor( rep(data[,3],rep(3,nsubjects)),
                             levels=1:6,
                             labels=c("<HS","HS","some college","college",
                                "some post","grad/prof") )
vps.stacked$age <- rep( data[,4], rep(3,nsubjects) )
vps.stacked$cohort <- rep( data[,5], rep(3,nsubjects) )
#
fit2c <- gee( y ~ post + visit12 + ICgroup +
                  post*ICgroup + visit12*ICgroup + age + education +
                  cohort + risk.group,
             family=binomial,
             id=id, corstr="unstructured",
             data=vps.stacked )
summary( fit2c )
#
#
# end-of-file...
```

Cross-sectional Results | Baseline

```
Call:
crosstabs( ~ q4safe0 + ICgroup, data = vps.data)
1000 cases in table

q4safe0|ICgroup
       |0      |1      |RowTotl|
-------+-------+-------+-------+
0      |218    |216    |434    |
       |0.50   |0.50   |0.43   |
       |0.44   |0.43   |       |
       |0.22   |0.22   |       |
-------+-------+-------+-------+
1      |282    |284    |566    |
       |0.50   |0.50   |0.57   |
       |0.56   |0.57   |       |
       |0.28   |0.28   |       |
-------+-------+-------+-------+
ColTotl|500    |500    |1000   |
       |0.5    |0.5    |       |
-------+-------+-------+-------+
Test for independence of all factors
        Chi^2 = 0.01628373 d.f.= 1 (p=0.8984594)
        Yates' correction not used
```

Cross-sectional Results Baseline

```
LOGISTIC REGRESSION:

Call: glm(formula = q4safe0 ~ ICgroup, family = binomial, data = vps.data)
Deviance Residuals:

Coefficients:
                 Value Std. Error   t value
(Intercept) 0.25741188 0.09017643 2.8545360
    ICgroup 0.01628375 0.12759574 0.1276199

(Dispersion Parameter for Binomial family taken to be 1 )
    Null Deviance: 1368.819 on 999 degrees of freedom
Residual Deviance: 1368.803 on 998 degrees of freedom
```

Cross-sectional Results Month 6

```
Call:
crosstabs( ~ q4safe6 + ICgroup, data = vps.data)
1000 cases in table

q4safe6|ICgroup
       |0      |1      |RowTotl|
-------+-------+-------+-------+
0      |226    |180    |406    |
       |0.56   |0.44   |0.41   |
       |0.45   |0.36   |       |
       |0.23   |0.18   |       |
-------+-------+-------+-------+
1      |274    |320    |594    |
       |0.46   |0.54   |0.59   |
       |0.55   |0.64   |       |
       |0.27   |0.32   |       |
-------+-------+-------+-------+
ColTotl|500    |500    |1000   |
       |0.5    |0.5    |       |
-------+-------+-------+-------+
Test for independence of all factors
        Chi^2 = 8.774112 d.f.= 1 (p=0.003055358)
        Yates' correction not used
```

Cross-sectional Results Month 6

```
LOGISTIC REGRESSION:

Call: glm(formula = q4safe6 ~ ICgroup, family = binomial, data = vps.data)

Coefficients:
                Value Std. Error  t value
(Intercept) 0.1925931 0.08985519 2.143372
    ICgroup 0.3827386 0.12929701 2.960150

(Dispersion Parameter for Binomial family taken to be 1 )
    Null Deviance: 1350.739 on 999 degrees of freedom
Residual Deviance: 1341.95 on 998 degrees of freedom
```

Cross-sectional Results Month 12

```
Call:
crosstabs( ~ q4safe12 + ICgroup, data = vps.data)
1000 cases in table

q4safe12|ICgroup
       |0      |1      |RowTotl|
-------+-------+-------+-------+
0      |208    |177    |385    |
       |0.54   |0.46   |0.39   |
       |0.42   |0.35   |       |
       |0.21   |0.18   |       |
-------+-------+-------+-------+
1      |292    |323    |615    |
       |0.47   |0.53   |0.61   |
       |0.58   |0.65   |       |
       |0.29   |0.32   |       |
-------+-------+-------+-------+
ColTotl|500    |500    |1000   |
       |0.5    |0.5    |       |
-------+-------+-------+-------+
Test for independence of all factors
        Chi^2 = 4.058706 d.f.= 1 (p=0.04394417)
        Yates' correction not used
```

Cross-sectional Results Month 12

```
LOGISTIC REGRESSION:

Call: glm(formula = q4safe12 ~ ICgroup, family = binomial, data = vps.data)

Coefficients:
                Value Std. Error  t value
(Intercept) 0.3392149 0.09070791 3.739639
    ICgroup 0.2622439 0.13011397 2.015494

(Dispersion Parameter for Binomial family taken to be 1 )
    Null Deviance: 1332.918 on 999 degrees of freedom
Residual Deviance: 1328.856 on 998 degrees of freedom
```

GEE Results for month 0 and month 6 Independence

```
 GEE:  GENERALIZED LINEAR MODELS FOR DEPENDENT DATA
 gee S-function, version 4.4 modified 96/09/27 (1996)

Model:
 Link:                      Logit
 Variance to Mean Relation: Binomial
 Correlation Structure:     Independent

Call:
gee(formula = y ~ visit6 + ICgroup + visit6 * ICgroup, id = id, data =
        vps.stacked, subset = (visit <= 1), family = binomial, corstr =
        "independence")

Coefficients:
                  Estimate Naive S.E.    Naive z Robust S.E.   Robust z
   (Intercept)  0.25741201 0.09018456  2.8542802  0.09018456  2.8542802
        visit6 -0.06481890 0.12730934 -0.5091449  0.09854411 -0.6577653
       ICgroup  0.01628382 0.12760882  0.1276073  0.12760882  0.1276073
visit6:ICgroup  0.36648722 0.18176628  2.0162552  0.14438852  2.5382019

Working Correlation                        Estimated Scale Parameter:  1
     [,1] [,2]
[1,]    1    0
[2,]    0    1
```

GEE Results for month 0 and month 6 | Exchangeable

```
 GEE:  GENERALIZED LINEAR MODELS FOR DEPENDENT DATA
 gee S-function, version 4.4 modified 96/09/27 (1996)

Model:
 Link:                      Logit
 Variance to Mean Relation: Binomial
 Correlation Structure:     Exchangeable

Call:
gee(formula = y ~ visit6 + ICgroup + visit6 * ICgroup, id = id, data =
        vps.stacked, subset = (visit <= 1), family = binomial, corstr =
        "exchangeable")

Coefficients:
                  Estimate Naive S.E.    Naive z Robust S.E.   Robust z
   (Intercept)  0.25741201 0.09018456  2.8542802  0.09018456  2.8542802
        visit6 -0.06481890 0.10107597 -0.6412890  0.09854411 -0.6577653
       ICgroup  0.01628382 0.12760882  0.1276073  0.12760882  0.1276073
visit6:ICgroup  0.36648722 0.14432204  2.5393711  0.14438852  2.5382019

Working Correlation                          Estimated Scale Parameter:  1
          [,1]      [,2]
[1,] 1.0000000 0.3696619
[2,] 0.3696619 1.0000000
```

GEE Results for months 0, 6, 12 Unstructured

```
 GEE:  GENERALIZED LINEAR MODELS FOR DEPENDENT DATA
 gee S-function, version 4.4 modified 96/09/27 (1996)

Model:    Link:                      Logit
          Variance to Mean Relation: Binomial
          Correlation Structure:     Unstructured
Call:
gee(formula = y ~ visit6 + visit12 + ICgroup + visit6 * ICgroup + visit12 *
        ICgroup, id = id, data = vps.stacked, family = binomial, corstr =
        "unstructured")

Coefficients:
                   Estimate Naive S.E.    Naive z Robust S.E.   Robust z
    (Intercept)  0.25741201 0.09018456  2.8542802  0.09018456  2.8542802
         visit6 -0.06481890 0.10107593 -0.6412892  0.09854411 -0.6577653
        visit12  0.08180371 0.10900000  0.7504928  0.11093456  0.7374051
        ICgroup  0.01628382 0.12760882  0.1276073  0.12760882  0.1276073
 visit6:ICgroup  0.36648722 0.14432198  2.5393721  0.14438852  2.5382019
visit12:ICgroup  0.24600305 0.15540323  1.5829983  0.15535967  1.5834421

Working Correlation                           Estimated Scale Parameter:  1
          [,1]      [,2]      [,3]
[1,] 1.0000000 0.3696624 0.2740313
[2,] 0.3696624 1.0000000 0.3901932
[3,] 0.2740313 0.3901932 1.0000000
```

GEE Results for months 0, 6, 12 Unstructured

```
 GEE:  GENERALIZED LINEAR MODELS FOR DEPENDENT DATA
 gee S-function, version 4.4 modified 96/09/27 (1996)

Model:   Link:                      Logit
         Variance to Mean Relation: Binomial
         Correlation Structure:     Unstructured
Call:
gee(formula = y ~ post + visit12 + ICgroup + post * ICgroup +
              visit12 * ICgroup,
     id = id, data = vps.stacked, family = binomial, corstr = "unstructured'')

Coefficients:
                   Estimate Naive S.E.    Naive z Robust S.E.   Robust z
    (Intercept)  0.25741201 0.09018456  2.8542802  0.09018456  2.8542802
           post -0.06481890 0.10107593 -0.6412892  0.09854411 -0.6577653
        visit12  0.14662262 0.09972098  1.4703287  0.10356299  1.4157820
        ICgroup  0.01628382 0.12760882  0.1276073  0.12760882  0.1276073
   post:ICgroup  0.36648722 0.14432198  2.5393721  0.14438852  2.5382019
visit12:ICgroup -0.12048417 0.14342582 -0.8400452  0.14323854 -0.8411435

Working Correlation                            Estimated Scale Parameter:  1
          [,1]      [,2]      [,3]
[1,] 1.0000000 0.3696624 0.2740313
[2,] 0.3696624 1.0000000 0.3901932
[3,] 0.2740313 0.3901932 1.0000000
```

GEE Results for months 0, 6, 12 Unstructured

```
 GEE:  GENERALIZED LINEAR MODELS FOR DEPENDENT DATA
 gee S-function, version 4.4 modified 96/09/27 (1996)

Model:
 Link:                      Logit
 Variance to Mean Relation: Binomial
 Correlation Structure:     Unstructured

Call:
gee(formula = y ~ post + visit12 + ICgroup + post * ICgroup +
                  visit12 * ICgroup + age + education + cohort +
                  risk.group, id = id, data = vps.stacked,
        family = binomial, corstr = "unstructured")

Coefficients:
                           Estimate  Naive S.E.    Naive z Robust S.E.
          (Intercept) -0.653812235 0.304047634 -2.1503612 0.302618929
                 post -0.072804832 0.113419402 -0.6419081 0.110246513
              visit12  0.164608064 0.112256885  1.4663516 0.115911691
              ICgroup  0.057407302 0.135750956  0.4228869 0.135383490
                  age  0.002867246 0.006231207  0.4601431 0.006515743
          educationHS  0.532181933 0.176465579  3.0157832 0.171117006
educationsome college  1.078366081 0.190678225  5.6554233 0.191301574
     educationcollege  1.474141590 0.212293846  6.9438734 0.211357932
   educationsome post  1.465868394 0.287874088  5.0920470 0.278359370
   educationgrad/prof  1.719539978 0.240507443  7.1496331 0.240574062
```

```
              cohort -0.176522778 0.104448474 -1.6900465 0.104725704
   risk.groupMaleIDU -0.358384469 0.175537271 -2.0416432 0.169889550
      risk.groupWAHR -0.280944024 0.193319233 -1.4532647 0.199186304
  risk.groupWAHR+IDU -0.579334066 0.218389546 -2.6527555 0.202837378
        post:ICgroup  0.413160363 0.162482356  2.5428014 0.162410485
     visit12:ICgroup -0.135179205 0.161844625 -0.8352406 0.161040205
                        Robust z
         (Intercept) -2.1605133
                post -0.6603822
             visit12  1.4201161
             ICgroup  0.4240347
                 age  0.4400491
         educationHS  3.1100470
educationsome college  5.6369953
     educationcollege  6.9746216
   educationsome post  5.2661004
   educationgrad/prof  7.1476533
              cohort -1.6855726
   risk.groupMaleIDU -2.1095145
      risk.groupWAHR -1.4104585
  risk.groupWAHR+IDU -2.8561504
        post:ICgroup  2.5439267
     visit12:ICgroup -0.8394128

Estimated Scale Parameter:  1.001947
Number of Iterations:  2

Working Correlation
```

```
            [,1]      [,2]      [,3]
[1,] 1.0000000 0.2919778 0.1854306
[2,] 0.2919778 1.0000000 0.3110631
[3,] 0.1854306 0.3110631 1.0000000
```

GEE1 - testing hypotheses

Wald Tests

- $H_o: \quad \beta_j = 0$

$$\hat{\beta}_j / \widehat{\text{s.e.}} \sim N(0,1)$$

- $H_o: \quad \boldsymbol{\gamma} = 0$

$$\boldsymbol{\gamma} = (\beta_{j+1}, \beta_{j+2}, \ldots, \beta_{j+r})$$

$$\widehat{\boldsymbol{\gamma}}^T \boldsymbol{V}_\gamma^{-1} \widehat{\boldsymbol{\gamma}} \sim \chi^2(r)$$

$\boldsymbol{V}_\gamma$ is the empirical variance matrix corresponding to $\widehat{\gamma}$.

GEE1 - testing hypotheses

Score Tests

- Rotnitzky and Jewell (1990)
- $H_o: \quad \boldsymbol{\gamma} = 0$

$$\boldsymbol{\gamma} = (\beta_{j+1}, \beta_{j+2}, \ldots, \beta_{j+r})$$
$$\boldsymbol{\beta} = (\boldsymbol{\gamma}, \boldsymbol{\delta})$$

$$T_s = \frac{1}{N} U_\gamma\{\mathbf{0}, \hat{\boldsymbol{\delta}}\}^T \boldsymbol{\Sigma}_\gamma^{-1} U_\gamma\{\mathbf{0}, \hat{\boldsymbol{\delta}}\}$$

$$T_s \sim \chi^2(r)$$

$$\boldsymbol{\Sigma}_\gamma = \mathcal{I}_{M,\gamma}^{-1} \ \mathcal{I}_\gamma^* \ \mathcal{I}_{M,\gamma}^{-1}$$

Summary

- GEE1 - focus on the marginal mean parameter $\boldsymbol{\beta}$.
- Flexible mean models.
- Choice of "working correlation models".
- Semiparametric since only first (and second) moment model(s).
- "sandwich estimator" for var$(\widehat{\boldsymbol{\beta}})$.
- Caveat: MCAR assumed.
- Caveat: time-dependent covariates and weighting.

- Note: Model versus Estimation versus Software

- Examples:

  HIVNET IC Analysis

  Madras Longitudinal Study of Schizophrenia

GEE EXTENSIONS

GEE1 - What about α?

Recall, for GEE1 we use:

$$\boldsymbol{U}_1(\boldsymbol{\beta}, \boldsymbol{\alpha}) = \sum_{i=1}^{N} \boldsymbol{D}_i^T(\boldsymbol{\beta}) \boldsymbol{V}_i^{-1}(\boldsymbol{\beta}, \boldsymbol{\alpha}) \{ \boldsymbol{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta}) \}$$

- Estimating function depends on unknown $\boldsymbol{\alpha}$.
- $\widehat{\boldsymbol{\alpha}}$ obtained through simple moment estimators.

Q1: Shouldn't we consider the parameter as $(\boldsymbol{\beta}, \boldsymbol{\alpha})$?

Q2: Can't we improve upon the estimation of $\boldsymbol{\alpha}$?

Q3: Would "better" estimation of $\boldsymbol{\alpha}$ help us to "better" estimate $\boldsymbol{\beta}$?

GEE1 - What about α?

Answer1: Depends...

- Is $\boldsymbol{\alpha}$ a nuisance? If the covariance structure is of secondary interest (often the case) then GEE1 is usually fine. However, if the covariance **is** of primary interest then GEE1 is not ideal.
- Are you willing to sacrifice some model robustness in order to let $(\boldsymbol{\beta}, \boldsymbol{\alpha})$ be the target parameter? Note that in GEE1, the estimate $\widehat{\boldsymbol{\beta}}$ is consistent even if the model for $\boldsymbol{\alpha}$ is wrong. Other approaches that treat $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ on equal ground may not have this property.

Answer2: Yes!

- **Model**: We can adopt a more flexible class of covariance models than GEE1 (as implemented in $\mathbf{S+}$ and SAS) currently offers.

- **Model**: We can adopt alternative association (dependence) models that are more suited for categorical data.
- **Estimator**: We can create estimators that are more efficient in estimating $\boldsymbol{\alpha}$ but do not sacrifice $\widehat{\boldsymbol{\beta}}$ robustness (GEE1.5, ALR).
- **Estimator**: We can create estimators that are targeted at $(\boldsymbol{\beta}, \boldsymbol{\alpha})$ jointly and are efficient for both (GEE2, likelihood methods).

Answer3: Depends...

- **Model** for $\boldsymbol{\alpha}$ is important for the efficiency of $\widehat{\boldsymbol{\beta}}$.
- **Estimator** choice may not be important (given a decent model).

Augmented GEE1 (GEE1.5?)

Prentice (1988) - Binary Response

Idea: GEE1 uses an estimating function U_1 based on the centered first moments, $\{\boldsymbol{Y}_i - \boldsymbol{\mu}_i\}$ for the estimation of $\boldsymbol{\beta}$. Let's augment U_1 with a second estimating function based on centered second moments,

$\{(Y_{ij} - \mu_{ij})(Y_{ik} - \mu_{ik}) - \sigma_{ijk}\}$, for the estimation of $\boldsymbol{\alpha}$.

$$
\begin{aligned}
\boldsymbol{U}_1(\boldsymbol{\beta}, \boldsymbol{\alpha}) &= \sum_{i=1}^{N} \boldsymbol{D}_i^T(\boldsymbol{\beta}) \boldsymbol{V}_i^{-1}(\boldsymbol{\beta}, \boldsymbol{\alpha}) \{\boldsymbol{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta})\} \\
\boldsymbol{U}_2(\boldsymbol{\beta}, \boldsymbol{\alpha}) &= \sum_{i=1}^{N} \boldsymbol{E}_i^T(\boldsymbol{\beta}, \boldsymbol{\alpha}) \boldsymbol{W}_i^{-1}(\boldsymbol{\beta}, \boldsymbol{\alpha}) \{\boldsymbol{S}_i - \boldsymbol{\sigma}_i(\boldsymbol{\beta}, \boldsymbol{\alpha})\}
\end{aligned}
$$

$$
\begin{aligned}
\boldsymbol{S}_i &= (\boldsymbol{Y}_i - \boldsymbol{\mu}_i) \otimes (\boldsymbol{Y}_i - \boldsymbol{\mu}_i) \\
&= \text{vec}[(Y_{ij} - \mu_{ij})(Y_{ik} - \mu_{ik})] \\
\boldsymbol{\sigma}_i &= E[\boldsymbol{S}_i] \\
&= \text{vec}(\sigma_{ijk}) \\
\boldsymbol{E}_i &= \partial \boldsymbol{\sigma}_i / \partial \boldsymbol{\alpha} \\
\boldsymbol{W}_i &\approx \text{cov}(\boldsymbol{S}_i)
\end{aligned}
$$

Augmented GEE1 (GEE1.5?)

Paired Models:

$$\text{Mean Model} \quad \text{logit}(\mu_{ij}) = \boldsymbol{X}_{ij}\boldsymbol{\beta}$$

$$\text{Correlation Model} \quad g_2(\rho_{ijk}) = \boldsymbol{Z}_{ijk}\boldsymbol{\alpha}$$

- Allows a flexible class of models for dependence.
- No variance model needed for binary response,

$V(Y_{ij}) = \mu_{ij}(1 - \mu_{ij})$.

- Correlation "design" matrix is $n_i(n_i - 1)/2 \times q$ matrix.

Paired Estimating Equations:

$$
\begin{aligned}
0 &= \sum_{i=1}^{N} \boldsymbol{D}_i^T(\boldsymbol{\beta})\boldsymbol{V}_i^{-1}(\boldsymbol{\beta},\boldsymbol{\alpha})\{\boldsymbol{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta})\} \\
0 &= \sum_{i=1}^{N} \boldsymbol{E}_i^T(\boldsymbol{\beta},\boldsymbol{\alpha})\boldsymbol{W}_i^{-1}(\boldsymbol{\beta},\boldsymbol{\alpha})\{\boldsymbol{S}_i - \boldsymbol{\sigma}_i(\boldsymbol{\beta},\boldsymbol{\alpha})\}
\end{aligned}
$$

- $(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\alpha}})$ obtained as solution to these equations.

- $(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\alpha}})$ is consistent and asymptotically normal under correct model specification. Similar to GEE1, $\widehat{\boldsymbol{\beta}}$ is consistent even if the $\boldsymbol{\alpha}$ model is misspecified.

- Both model based and empirical (sandwich) variance estimators are obtained. (see Prentice, 1988 for details)

- Often a simple $\boldsymbol{W}_i$ is used (diagonal matrix) so that only the empirical variance for $\widehat{\boldsymbol{\alpha}}$ is available.

Augmented GEE1 - correlation models?

$\Rightarrow$ correlations for binary data are constrained by their means.

Frechet Inequality:

$$E[Y_1] = \mu_1 \qquad E[Y_2] = \mu_2$$

$$\begin{aligned} \pi_{12} &= E[Y_1 Y_2] \\ \pi_{12} &\leq \min\{\mu_1, \mu_2\} \end{aligned}$$

$$\rho_{12} \leq \min\left\{ \left(\frac{\mu_1/(1-\mu_1)}{\mu_2/(1-\mu_2)} \right)^{1/2}, \left(\frac{\mu_2/(1-\mu_2)}{\mu_1/(1-\mu_1)} \right)^{1/2} \right\}$$

Example:

$$E[Y_1] = 0.3 \qquad E[Y_2] = 0.1$$

$$\rho_{12} \leq 0.26$$

Augmented GEE1 - odds ratios!

$\Rightarrow$ Odds ratios are the "natural" measure of association for a pair of binary variables.

$$\begin{aligned}\Psi_{ijk} &= \frac{P[Y_{ij}=1, Y_{ik}=1]P[Y_{ij}=0, Y_{ik}=0]}{P[Y_{ij}=1, Y_{ik}=0]P[Y_{ij}=0, Y_{ik}=1]} \\ &= \frac{P[Y_{ij}=1 \mid Y_{ik}=1]/P[Y_{ij}=0 \mid Y_{ik}=1]}{P[Y_{ij}=1 \mid Y_{ik}=0]/P[Y_{ij}=0 \mid Y_{ik}=0]}\end{aligned}$$

- Odds ratios are not constrained by the marginal means, $\log \Psi \in (-\infty, \infty)$.

- Odds ratios have a simple interpretation: odds of $Y_{ij}=1$ given that $Y_{ik}=1$ relative to the odds of $Y_{ij}=1$ given that $Y_{ik}=0$.

Note

- The odds ratio, Ψ_{ijk}, and the means, μ_{ij}, μ_{ik} determine the pairwise probability, π_{ijk}, and the correlation, ρ_{ijk}.

$$
\begin{aligned}
\Psi_{ijk} &= \frac{\pi_{ijk}(1 - \mu_{ij} - \mu_{ik} + \pi_{ijk})}{(\mu_{ij} - \pi_{ijk})(\mu_{ik} - \pi_{ijk})} \\
\pi_{ijk} &= \left\{ A - \left[A^2 - 4(\Psi_{ik} - 1)\Psi_{ijk}\mu_{ij}\mu_{ik} \right]^{1/2} \right\} / (2\Psi_{ijk} - 2) \\
A &= 1 - (\mu_{ij} + \mu_{ik})(1 - \Psi_{ijk})
\end{aligned}
$$

- $\{\Psi_{ijk},\ \mu_{ij},\ \mu_{ik}\} \ \rightarrow\ \pi_{ij}\ \rightarrow\ \rho_{ijk}\ \rightarrow\ \boldsymbol{V}_i(\boldsymbol{\beta}, \boldsymbol{\alpha})$

Augmented GEE1 - odds ratios!

Lipsitz, Laird, and Harrington (1991)

<u>Paired Models</u>:

$$\begin{aligned} \text{Mean Model} \quad \text{logit}(\mu_{ij}) &= \boldsymbol{X}_{ij}\boldsymbol{\beta} \\ \text{Dependence Model} \quad \log(\Psi_{ijk}) &= \boldsymbol{Z}_{ijk}\boldsymbol{\alpha} \end{aligned}$$

- Allows a flexible class of models for dependence.
- Pairwise unconstrained Ψ_{ijk} (although multivariate constraints).

Paired Estimating Equations:

$$
\begin{aligned}
0 &= \sum_{i=1}^{N} \boldsymbol{D}_i^T(\boldsymbol{\beta}) \boldsymbol{V}_i^{-1}(\boldsymbol{\beta}, \boldsymbol{\alpha}) \{ \boldsymbol{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta}) \} \\
0 &= \sum_{i=1}^{N} \boldsymbol{E}_i^T(\boldsymbol{\beta}, \boldsymbol{\alpha}) \boldsymbol{W}_i^{-1}(\boldsymbol{\beta}, \boldsymbol{\alpha}) \{ \boldsymbol{S}_i - \boldsymbol{\sigma}_i(\boldsymbol{\beta}, \boldsymbol{\alpha}) \}
\end{aligned}
$$

- Note: centered covariance products are preferable to the uncentered products, $\{Y_{ij}Y_{ik} - \pi_{ijk}\}$, used in LLH (1991).

- $(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\alpha}})$ obtained as solution to these equations.

- $(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\alpha}})$ is consistent and asymptotically normal under correct model specification. Similar to GEE1, $\widehat{\boldsymbol{\beta}}$ is consistent even if the $\boldsymbol{\alpha}$ model is misspecified.

- Both model based and empirical (sandwich) variance estimators are obtained. (see LLH, 1991 for details)

- Often a simple $\boldsymbol{W}_i$ is used (diagonal matrix) so that only the empirical variance for $\widehat{\boldsymbol{\alpha}}$ is available.

Augmented GEE1 - ALR

Carey, Zeger, and Diggle (1993)

Let $\alpha_{ijk} = \log \Psi_{ijk}$.
Consider the pairwise conditional expectation:

$$\text{logit} E[Y_{ij} \mid Y_{ik}, \, \boldsymbol{X}_i] = \Delta_{ijk} + \alpha_{ijk} Y_{ik}$$

$$\Delta_{ijk} = \log \left(\frac{\mu_{ij} - \pi_{ijk}}{1 - \mu_{ij} - \mu_{ik} + \pi_{ijk}} \right)$$

$\Rightarrow$ An estimate for $\boldsymbol{\alpha}$ could be obtained alternating:
(LR 1) A logistic regression of Y_{ij} on X_{ij}
(yields $\boldsymbol{\beta}$)
(LR 2) A logistic regression of Y_{ij} on Y_{ik} with offset Δ_{ijk}
(yields $\boldsymbol{\alpha}$)

- Note: the offset Δ_{ijk} depends on both $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$.
- Alternate (LR 1) and (LR 2) until convergence.
- "Alternating Logistic Regressions"

Augmented GEE1 - ALR

Alternating Logistic Regressions

Formally, this proposal is equivalent to augmenting $U_1(\boldsymbol{\beta}, \boldsymbol{\alpha})$ with the pairwise conditional estimating function:

$$U_3(\boldsymbol{\beta}, \boldsymbol{\alpha}) = \sum_{i=1}^{N} \boldsymbol{F}_i^T(\boldsymbol{\beta}, \boldsymbol{\alpha}) \tilde{\boldsymbol{W}}_i^{-1}(\boldsymbol{\beta}, \boldsymbol{\alpha}) \boldsymbol{T}_i(\boldsymbol{\beta}, \boldsymbol{\alpha})$$

$$\boldsymbol{T}_i = \text{vec}\left(Y_{ij} - \xi_{ijk}\right)$$

$$\xi_{ijk} = E[Y_{ij} \mid Y_{ik}] \qquad \boldsymbol{\xi}_i = \text{vec}(\xi_{ijk})$$

$$\tilde{\boldsymbol{W}}_i = \text{var}\left(Y_{ij} \mid Y_{ik}\right)$$

$$\boldsymbol{F}_i = \partial \boldsymbol{\xi}_i / \partial \boldsymbol{\alpha}$$

Paired Models: same as LLH 1991

Paired Estimating Equations:

$$
\begin{aligned}
0 &= \sum_{i=1}^{N} \boldsymbol{D}_i^T(\boldsymbol{\beta}) \boldsymbol{V}_i^{-1}(\boldsymbol{\beta}, \boldsymbol{\alpha}) \{ \boldsymbol{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta}) \} \\
0 &= \sum_{i=1}^{N} \boldsymbol{F}_i^T(\boldsymbol{\beta}, \boldsymbol{\alpha}) \tilde{\boldsymbol{W}}_i^{-1}(\boldsymbol{\beta}, \boldsymbol{\alpha}) \boldsymbol{T}_i(\boldsymbol{\beta}, \boldsymbol{\alpha})
\end{aligned}
$$

Informed Consent Example: GEE Extensions

Model 1

$$\mu_{ij} = E[Y_{ij} \mid \boldsymbol{X}_{ij}]$$

$$\begin{aligned} \text{logit}(\mu_{ij}) = \; & \beta_0 + \beta_1 \cdot (\text{Tx}) + \\ & \beta_2 \cdot (\text{Post}) + \beta_3 \cdot (\text{Time=12}) + \\ & \beta_4 \cdot (\text{Post} \cdot \text{Tx}) + \beta_5 \cdot (\text{Time=12} \cdot \text{Tx}) \end{aligned}$$

$$\log \Psi_{i(j,k)} = \alpha_{(j,k)}$$

Model 2

$$\log \Psi_{i(j,k)} = \alpha_0 + \alpha_1 \mathbf{1}(\Delta t = 12)$$

Model 1

```
Results for method = GEE1 - invariant
       beta weight = full
      alpha weight = diagonal

Beta estimates and standard errors:
                 Est  rob. s.e.  mod. s.e. rob. z
    (Intercept)  0.2574 0.0901 0.0901    2.854
           post -0.0648 0.0985 0.1006   -0.657
        visit12  0.1466 0.1035 0.0995    1.415
        ICgroup  0.0162 0.1276 0.1276    0.127
   post:ICgroup  0.3664 0.1443 0.1442    2.538
visit12:ICgroup -0.1204 0.1432 0.1438   -0.841

Alpha estimates and standard errors:
        Est rob. s.e.   rob. z
a(1,2) 1.5904 0.1397 11.382
a(1,3) 1.1526 0.1354  8.507
a(2,3) 1.6771 0.1421 11.801
```

Model 1

```
Results for method = ALR
       beta weight = full
      alpha weight = diagonal (only option with ALR)

Beta estimates and standard errors:
                 Est  rob. s.e.  mod. s.e.  rob. z
    (Intercept)  0.2574 0.0901 0.0901    2.854
           post -0.0648 0.0985 0.1006   -0.657
        visit12  0.1466 0.1035 0.0994    1.415
        ICgroup  0.0162 0.1276 0.1276    0.127
   post:ICgroup  0.3664 0.1443 0.1442    2.538
visit12:ICgroup -0.1204 0.1432 0.1436   -0.841

Alpha estimates and standard errors:
        Est rob. s.e.    rob. z
a(1,2) 1.5880 0.1397 11.366
a(1,3) 1.1537 0.1354  8.515
a(2,3) 1.6844 0.1420 11.858
```

Model 2

```
Results for method = GEE1 - invariant
       beta weight = full
      alpha weight = diagonal

Beta estimates and standard errors:
                 Est  rob. s.e.  mod. s.e. rob. z
    (Intercept)  0.2574 0.0901 0.0901    2.854
           post -0.0648 0.0985 0.0998   -0.657
        visit12  0.1466 0.1035 0.1003    1.415
        ICgroup  0.0162 0.1276 0.1276    0.127
   post:ICgroup  0.3664 0.1443 0.1431    2.538
visit12:ICgroup -0.1204 0.1432 0.1449   -0.841

Alpha estimates and standard errors:
            Est rob. s.e.  rob. z
   (Int)    1.6332 0.1065  15.334
I(dt==12) -0.4806 0.1345  -3.571
```

Model 2

```
Results for method = ALR
       beta weight = full
      alpha weight = diagonal (only option with ALR)

Beta estimates and standard errors:
                    Est  rob. s.e.  mod. s.e. rob. z
    (Intercept)  0.2574 0.0901 0.0901    2.854
           post -0.0648 0.0985 0.0998   -0.657
        visit12  0.1466 0.1035 0.1002    1.415
        ICgroup  0.0162 0.1276 0.1276    0.127
   post:ICgroup  0.3664 0.1443 0.1431    2.538
visit12:ICgroup -0.1204 0.1432 0.1448   -0.841

Alpha estimates and standard errors:
            Est rob. s.e.  rob. z
   (Int)   1.6357 0.1065  15.356
I(dt==12) -0.4819 0.1346  -3.579
```

Summary:

- Paired models
- Paired estimating functions
- $\log \Psi$ to parameterize the covariance matrix

  $\Rightarrow$ interpretation

  $\Rightarrow$ pairwise unconstrained
- Examples:

  Madras Longitudinal Study of Schizophrenia
- Extensions:

  Ordinal response

  Nominal response

GEE2 - Joint Estimating Equations

- Model for first moments (mean) $\Leftrightarrow$ $\boldsymbol{\beta}$
- Model for second moments (covariance) $\Leftrightarrow$ $\boldsymbol{\alpha}$

Q1: $\boldsymbol{\delta} = (\boldsymbol{\beta}, \boldsymbol{\alpha})$?

Q2: Optimal estimating function for $\boldsymbol{\delta}$?

- Add "working assumptions" for the 3rd and 4th moments

  $\Rightarrow$ **GEE2**

- Add assumptions for all moments

  $\Rightarrow$ **Maximum likelihood**

GEE2 - Joint Estimating Equations

Prentice and Zhao (1991)

<u>Paired Models</u>:

$$g_1(\mu_{ij}) = \boldsymbol{X}_{ij}\boldsymbol{\beta}$$

$$g_2(\sigma_{ijk}) = \boldsymbol{Z}_{ijk}\boldsymbol{\alpha}$$

Optimal EF Based on $\boldsymbol{\delta} = (\boldsymbol{\beta}, \boldsymbol{\alpha})$:

$$
\begin{aligned}
U(\boldsymbol{\delta}) &= \sum_{i=1}^{N} \boldsymbol{D}_i^T(\boldsymbol{\delta}) \boldsymbol{V}_i^{-1}(\boldsymbol{\delta}) \boldsymbol{T}_i(\boldsymbol{\delta}) \\
U(\boldsymbol{\delta}) &= \begin{bmatrix} \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}} & \frac{\partial \boldsymbol{\sigma}_i}{\partial \boldsymbol{\beta}} \\ \mathbf{0} & \frac{\partial \boldsymbol{\sigma}_i}{\partial \boldsymbol{\alpha}} \end{bmatrix}^T \begin{bmatrix} \boldsymbol{V}_i(1,1) & \boldsymbol{V}_i(1,2) \\ \boldsymbol{V}_i(2,1) & \boldsymbol{V}_i(2,2) \end{bmatrix}^{-1} \begin{pmatrix} \boldsymbol{Y}_i - \boldsymbol{\mu}_i \\ \boldsymbol{S}_i - \boldsymbol{\sigma}_i \end{pmatrix} \\
\boldsymbol{V}_i(1,1) &= \mathsf{cov}(\boldsymbol{Y}_i) \\
\boldsymbol{V}_i(1,2) &= \mathsf{cov}(\boldsymbol{Y}_i, \boldsymbol{S}_i) \\
\boldsymbol{V}_i(2,2) &= \mathsf{cov}(\boldsymbol{S}_i)
\end{aligned}
$$

$\Rightarrow$ First and second moment models are not enough to obtain $V_i(1,2)$ and $V_i(2,2)$.

GEE2 - Joint Estimating Equations

Prentice and Zhao (1991)

"Working" 3rd/4th Moment Model

- Independence working models

  $\boldsymbol{V}_i(1,2) = \mathbf{0}$

  $\boldsymbol{V}_i(1,2) =$ diagonal matrix

- Gaussian working models

  $\boldsymbol{V}_i(1,2) = \mathbf{0}$

  $\boldsymbol{V}_i(2,2)$: $\text{cov}(S_{ijk}, S_{ilm}) = \sigma_{ijl}\sigma_{ikm} + \sigma_{ijm}\sigma_{ikl}$

- Independence with structural non-zeros

  For binary data:

  $\text{cov}(Y_{ij}, S_{ijk}) = (1 - 2\mu_{ij})\sigma_{ijk}$

  $\text{var}(S_{ijk}) = \text{fnx}(\mu_{ij}, \mu_{ik}, \pi_{ijk})$

all other elements set to 0

Note: GEE2 equations can be derived as the score equations for a Quadratic Exponential Family (QEF) model:

$$\log L_i \;=\; \boldsymbol{\theta}_{i1}^T \boldsymbol{Y}_i + \boldsymbol{\theta}_{2i}^T \boldsymbol{S}_i + \Delta_i + c_i(\boldsymbol{Y}_i)$$

GEE2 - Joint Estimating Equations

Liang, Zeger, and Qaqish (1992)

Paired Models: (for binary data)

$$\text{logit}(\mu_{ij}) = \boldsymbol{X}_{ij}\boldsymbol{\beta}$$

$$\log(\Psi_{ijk}) = \boldsymbol{Z}_{ijk}\boldsymbol{\alpha}$$

Optimal EF Based on $\boldsymbol{\delta} = (\boldsymbol{\beta}, \boldsymbol{\alpha})$:

$$U(\boldsymbol{\delta}) = \sum_{i=1}^{N} \boldsymbol{D}_i^T(\boldsymbol{\delta})\boldsymbol{V}_i^{-1}(\boldsymbol{\delta})\boldsymbol{T}_i(\boldsymbol{\delta})$$

GEE2 - Joint Estimating Equations

"Working" 3rd/4th Moment Model

- Fix marginal 3-way log OR contrasts (at 0)
- Fix marginal 4-way log OR contrasts (at 0)

GEE2 - Estimation / Asymptotic Properties

Estimation: Fisher scoring

$$\begin{aligned}
\boldsymbol{\delta}^{(j+1)} &= \boldsymbol{\delta}^{(j)} + \left(\mathcal{I}_M^{(j)}\right)^{-1} U(\boldsymbol{\delta}^{(j)}) \\
\mathcal{I}_M^{(j)} &= \sum_i \boldsymbol{D}_i^T(\boldsymbol{\delta}^{(j)}) \boldsymbol{V}_i^{-1}(\boldsymbol{\delta}^{(j)}) \boldsymbol{D}_i(\boldsymbol{\delta}^{(j)}) \\
U(\boldsymbol{\delta}^{(j)}) &= \sum_i \boldsymbol{D}_i^T(\boldsymbol{\delta}^{(j)}) \boldsymbol{V}_i^{-1}(\boldsymbol{\delta}^{(j)}) T_i(\boldsymbol{\delta}^{(j)})
\end{aligned}$$

Large sample distn for $\widehat{\boldsymbol{\delta}}$:

$$\widehat{\boldsymbol{\delta}} \sim N(\,\boldsymbol{\delta},\, \mathcal{I}_M^{-1}\,\Sigma_N\,\mathcal{I}_M^{-1}\,)$$

$$\Sigma_N = \sum_{i=1}^{N} E[U_i(\boldsymbol{\delta})U_i(\boldsymbol{\delta})^T]$$

$$\widehat{\Sigma}_N = \sum_{i=1}^{N} U_i(\widehat{\boldsymbol{\delta}})U_i(\widehat{\boldsymbol{\delta}})^T$$

Note:

- "sandwich" estimator protects against the 3rd/4th moment specification.

- $\boldsymbol{\delta} = (\boldsymbol{\beta}, \boldsymbol{\alpha}) \Rightarrow$ joint estimation $\Rightarrow$ consistency of both $\widehat{\beta}$ and $\widehat{\alpha}$ depends on correct model for both <u>mean</u> and <u>covariance</u>. (ouch!)

Comments on GEE2

- $\boldsymbol{V}_i$ has dimension $m_i \times m_i$ where
$$m_i = n_i + n_i(n_i - 1)/2$$

- Requires that $\boldsymbol{V}_i$ be inverted. (ouch!)

- LZQ (1992) solution for 3rd/4th moment restrictions requires solution of higher order polynomial functions. (ouch!)

- Mancl and Prentice (1994)
  Efficiency comparisons of EE2 and EE1 augmented
  Gains depend on correct 3rd/4th specification

Comments on GEE2

- Use of QEF to motivate can also be used for fitting multivariate categorical data but requires recovery of the 2^{n_i} (more generally C^{n_i} probability vector. (ouch!)

- Likelihood methods? – see DHLZ Chapter 11

Summary

- Paired models
- Joint estimating functions
- “working” higher moment model
- $\widehat{\boldsymbol{\beta}}$ consistency?
- Extensions:

  Likelihood methods

References

Generalized Estimating Equations

Liang KY. and Zeger S.L. (1986).

Prentice R.L. (1988).

Rotnitzky and Jewell (1990)

Lipsitz, Laird and Harrington (1991)

Prentice and Zhao (1991)

Liang, Zeger and Qaqish (1992)

Carey, Zeger and Diggle (1993)

Heagerty and Zeger (1996)

Heagerty and Zeger (1998)