Module 4: Generalized Estimating Equations

BIOS 526

Reading

- A useful reference: Halekoh, U., S. Hojsgaard, and J. Yan. The R Package geepack for Generalized Estimating Equations. *Journal of Statistical Software*. 2006.
- Informal overview: https://rlbarter.github.io/Practical-Statistics/2017/05/10/generalized-estimating-equations-gee/

Concepts

- Weighted and generalized least-squares.
- Estimating equation.
- Marginal correlation structures.
- Robust standard error.

Motivating Example

Consider data y_1, y_2, and y_3 from

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}$$

- y_1 and y_2 come from the same *cluster* and may be correlated ($R \neq 0$).
- y_1, y_2, and y_3 have the same expectation μ. The average of y_i will give an unbiased estimator.

$$\hat{\mu}_1 = \frac{y_1}{3} + \frac{y_2}{3} + \frac{y_3}{3}$$

The above estimator has variance

$$[1/3, 1/3, 1/3] \ \Sigma \ \begin{bmatrix} 1/3 \\ 1/3 \\ 1/3 \end{bmatrix} = \frac{3\sigma^2 + 2R}{9}.$$

R increases $\rightarrow$ variance increases.

Motivating Example

We wish to find a set of weights $\mathbf{w} = (w_1, w_2, w_3)$ such that

$$\hat{\mu} = w_1 y_1 + w_2 y_2 + w_3 y_3.$$

We set $w_1 + w_2 + w_3 = 1$, so $\hat{\mu}$ is unbiased.

The variance for $\hat{\mu}$ is:

When $R = 0$, minimum variance when $w_1 = w_2 = w_3$ (the simple sample average).

Motivating Example

Consider the following sets of weights:

1. Simple average:

$$\mathbf{w} = (1/3, 1/3, 1/3)$$

2. Average y_1 and y_2 first; then average with y_3:

$$\mathbf{w} = (1/4, 1/4, 1/2)$$

3. Use only one of y_1 or y_2:

$$\mathbf{w} = (1/2, 0, 1/2)$$

4. Secret: assuming R and σ^2 are known,

$$\mathbf{w} = \left(\frac{1}{3 + R/\sigma^2}, \frac{1}{3 + R/\sigma^2}, \frac{1 + R/\sigma^2}{3 + R/\sigma^2} \right)$$

Motivating Example

Estimator weights	$Var[\hat{\mu}]$	R $= 0$	R $= 0.5\sigma^2$	R $= \sigma^2$
$(1/3, 1/3, 1/3)$	$\frac{3\sigma^2+2R}{9}$	$\frac{1}{3}\sigma^2$	$\frac{4}{9}\sigma^2$	$\frac{5}{9}\sigma^2$
$(1/4, 1/4, 1/2)$	$\frac{3\sigma^2+R}{8}$	$\frac{3}{8}\sigma^2$	$\frac{7}{16}\sigma^2$	$\frac{1}{2}\sigma^2$
$(1/2, 0, 1/2)$	$\frac{\sigma^2}{2}$	$\frac{1}{2}\sigma^2$	$\frac{1}{2}\sigma^2$	$\frac{1}{2}\sigma^2$
$\left(\frac{1}{3+R/\sigma^2}, \frac{1}{3+R/\sigma^2}, \frac{1+R/\sigma^2}{3+R/\sigma^2}\right)$	$\left(\frac{1+R/\sigma^2}{3+R/\sigma^2}\right)\sigma^2$	$\frac{1}{3}\sigma^2$	$\frac{3}{7}\sigma^2$	$\frac{1}{2}\sigma^2$

The secret (optimal) weights

- have the smallest standard error for the complete range of R values.
- become a simple average when $R = 0$ (independence).
- gives weights $1/4$ to y_1 and y_2 when $R = \sigma^2$.

Estimating Equation

All four estimators of $\hat{\mu}$ are solutions to the following estimating equation. We will define an equation to produce a class of unbiased estimators of $\hat{\mu}$.

We know $E(y_1 - \mu) = 0$, $E(y_2 - \mu) = 0$, and $E(y_3 - \mu) = 0$.

Combining the three equations, an estimator $\hat{\mu}$ can be obtained by solving:

$$\alpha_1(y_1 - \hat{\mu}) + \alpha_2(y_2 - \hat{\mu}) + \alpha_3(y_3 - \hat{\mu}) = 0. \tag{1}$$

The above gives:

$$\hat{\mu} = \frac{\alpha_1 y_1 + \alpha_2 y_2 + \alpha_3 y_3}{\alpha_1 + \alpha_2 + \alpha_3}.$$

Therefore the weights for $i = 1, 2, 3$ are

$$w_i = \frac{\alpha_i}{\alpha_1 + \alpha_2 + \alpha_3}.$$

An estimating equation is a very general approach. Equation (1) is an example where we set up an equation, set it to zero, and solve for $\hat{\mu}$. Methods of moments and maximum likelihood (setting the first derivative of the log-lik to zero) are examples of estimating equations.

Regression with heteroscedastic errors

Now consider the linear regression problem with heteroscedastic errors:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \qquad \boldsymbol{\epsilon} \sim N(\mathbf{0}, \mathbf{V}).$$

Here we assume the residual covariance matrix $\mathbf{V}$ is known and diagonal,

$$\mathbf{V} = \begin{bmatrix} v_1 & 0 & \dots & 0 \\ 0 & v_2 & \dots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & \dots & v_n \end{bmatrix}.$$

What is the least square estimate of $\boldsymbol{\beta}$?

Let $\mathbf{W}$ be a diagonal matrix with $diag(\mathbf{W}) = (v_1^{-1/2}, v_2^{-1/2}, \dots, v_n^{-1/2})$.
Note that $\mathbf{WW} = \mathbf{V}^{-1}$.

Consider the transformed $\mathbf{y}$:

$$\mathbf{y}^* = \mathbf{Wy}.$$

- $Var(y_i^*) = Var(w_i y_i) = w_i^2 v_i = v_i^{-1} v_i = 1$,
- $E(\mathbf{y}^*) = \mathbf{WX}\beta$

Weighted Least Squares

Let $\mathbf{X}^* = \mathbf{W}\mathbf{X}$, we now have a regression problem with equal variance:

$$\mathbf{y}^* = \mathbf{X}^*\boldsymbol{\beta} + \boldsymbol{\epsilon}^*, \qquad \boldsymbol{\epsilon}^* \sim N(\mathbf{0}, \mathbf{I}).$$

We have

The above equation is known as the weighted least squares estimate.

Weighted Least Squares, cont.

Weighted Least Squares

The weighted least squares estimate is unbiased (for any $\mathbf{V}$).

$$\begin{aligned} E(\hat{\boldsymbol{\beta}}_{wls}) &= E(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}) \\ &= (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}E(\mathbf{y}) \\ &= (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\boldsymbol{\beta} = \boldsymbol{\beta}. \end{aligned}$$

What if we used the regular ordinary least squares estimate?

$$\begin{aligned} E(\hat{\boldsymbol{\beta}}_{ols}) &= E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}] \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E(\mathbf{y}) \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\boldsymbol{\beta} = \boldsymbol{\beta}. \qquad \text{Still unbiased!} \end{aligned}$$

Note on OLS estimator

The OLS estimator of $\boldsymbol{\beta}$ is unbiased. It is also consistent

$$\lim_{n \to \infty} P\left\{ |\hat{\boldsymbol{\beta}}_{ols} - \boldsymbol{\beta}| > \epsilon \right\} = 0.$$

So what is the problem?

Note on OLS estimator

The covariance of $\hat{\boldsymbol{\beta}}_{ols}$ is

$$\begin{aligned} Cov(\hat{\boldsymbol{\beta}}_{ols}) &= Cov[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}] \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' Cov(\mathbf{y})\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \end{aligned}$$

However, our estimate of the covariance is

$$\widehat{Cov}(\hat{\boldsymbol{\beta}}_{ols}) = (\mathbf{X}'\mathbf{X})^{-1}\hat{\sigma}^2_{ols}.$$

Then if $\mathbf{V} \neq \sigma^2\mathbf{I}$, $E\left\{\widehat{Cov}(\hat{\boldsymbol{\beta}}_{ols})\right\} \neq (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$.

Two issues with OLS when heteroscedasticity

The OLS estimate of the SE of $\hat{\boldsymbol{\beta}}_{ols}$ can lead to invalid inference.

Under the null, $\hat{\beta}_k/\widehat{SE}_{ols}(\hat{\beta}_k)$ does not follow a t distribution.

Moreover, the true variance of $\hat{\beta}_{ols}$ is larger than $\hat{\beta}_{wls}$ – OLS is inefficient and WLS is more accurate.

Summarize:

1. OLS inference is invalid when the errors are not iid.
2. We can do a better job of estimating the coefficients with WLS.

Example: Fixed Effect Meta Analysis

Let $\hat{\beta}_i$ be the effect size from study i with estimated variance $\hat{\sigma}_i^2$. We will assume the model

$$\hat{\beta}_i = \mu + \epsilon_i \qquad \epsilon_i \sim N(0, \hat{\sigma}_i^2).$$

This can be expressed as

$$\mathbf{y} = [\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_n]'$$

$$\mathbf{X} = [1, 1, \ldots, 1]' \quad \text{(i.e., intercept only)}$$

$$\mathbf{V} = diag(\hat{\sigma}_1^2, \hat{\sigma}_2^2, \ldots, \hat{\sigma}_n^2)$$

Therefore

- $\mathbf{V}^{-1} = diag(1/\hat{\sigma}_1^2, 1/\hat{\sigma}_2^2, \ldots, 1/\hat{\sigma}_n^2)$
- $\mathbf{X}'\mathbf{V}^{-1} = [1/\hat{\sigma}_1^2, 1/\hat{\sigma}_2^2, \ldots, 1/\hat{\sigma}_n^2]'$
- $\mathbf{X}'\mathbf{V}^{-1}\mathbf{X} = \sum_{i=1}^n 1/\hat{\sigma}_i^2$
- $\mathbf{X}'\mathbf{V}^{-1}\mathbf{y} = \sum_{i=1}^n \hat{\beta}_i/\hat{\sigma}_i^2$

$$\mu_{wls} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y} = \frac{\sum_{i=1}^n \hat{\beta}_i/\hat{\sigma}_i^2}{\sum_{i=1}^n 1/\hat{\sigma}_i^2}$$

... the inverse-variance weighted average!

Generalized Least Squares

If $\mathbf{V}$ is not diagonal, the weighted least squares estimate still holds.

For non-diagonal $\mathbf{V}$, this is often called generalized least squares:

$$\hat{\boldsymbol{\beta}}_{gls} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y},$$

$$Cov(\hat{\boldsymbol{\beta}}_{gls}) = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}.$$

This is because $\mathbf{V}$ is a covariance matrix (symmetric and positive definite). So we can aways find a matrix $\mathbf{W} = \mathbf{V}^{-1/2}$, i.e.,

$$\mathbf{W}\mathbf{W}' = \mathbf{V}^{-1}.$$

E.g., $\mathbf{W}$ is from EVD. Also could use Cholesky decomposition.

OLS estimator $\hat{\beta}$ is unbiased and consistent, but the estimate of the standard errors are incorrect $\rightarrow$ wrong confidence intervals, p-values, and inference.

It can be shown that the weighting scheme using $\mathbf{V}$ in $\hat{\boldsymbol{\beta}}_{gls}$ gives the smallest variances (optimally efficient) among unbiased estimators if we know $\mathbf{V}$.

Generalized Least Squares: homoscedastic

For normal (Gaussian) regression model,

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \qquad \boldsymbol{\epsilon} \sim N(\mathbf{0}, \mathbf{V}).$$

we often assume the observations are homoscedastic (equal variance).

$$\mathbf{V} = \sigma^2 \mathbf{R}$$

$$\begin{aligned}
\hat{\boldsymbol{\beta}}_{gls} &= (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y} \\
&= (\mathbf{X}'\sigma^{-2}\mathbf{R}^{-1}\mathbf{X})^{-1}\mathbf{X}'\sigma^{-2}\mathbf{R}^{-1}\mathbf{y} = (\mathbf{X}'\mathbf{R}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{R}^{-1}\mathbf{y}
\end{aligned}$$

$$\begin{aligned}
Cov(\hat{\boldsymbol{\beta}}_{gls}) &= (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1} \\
&= (\mathbf{X}'\sigma^{-2}\mathbf{R}^{-1}\mathbf{X})^{-1} = \sigma^2(\mathbf{X}'\mathbf{R}^{-1}\mathbf{X})^{-1}
\end{aligned}$$

Later on, this is called the “naive” SE in `gee::gee`

Generalized Least Squares

Back to the motivating example (assume $\sigma^2 = 1$),

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} \sim N\left([\boldsymbol{\mu}],\ \begin{bmatrix} 1 & R & 0 \\ R & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \right).$$

Here

$$\mathbf{X} = [1, 1, 1]'$$

$$\mathbf{V}^{-1} = \begin{bmatrix} 1/(1-R^2) & -R/(1-R^2) & 0 \\ -R/(1-R^2) & 1/(1-R^2) & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Weighted Least Squares

$$(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}) = [1,1,1]\begin{bmatrix} 1/(1-R^2) & -R/(1-R^2) & 0 \\ -R/(1-R^2) & 1/(1-R^2) & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix}1\\1\\1\end{bmatrix}$$

$$= \left[\frac{1-R}{1-R^2}, \frac{1-R}{1-R^2}, 1\right]\begin{bmatrix}1\\1\\1\end{bmatrix} = \frac{3-2R-R^2}{1-R^2}$$

$$= \frac{(3+R)(1-R)}{(1-R)(1+R)} = \frac{3+R}{1+R}.$$

$$\mathbf{X}'\mathbf{V}^{-1} = [1,1,1]\begin{bmatrix} 1/(1-R^2) & -R/(1-R^2) & 0 \\ -R/(1-R^2) & 1/(1-R^2) & 0 \\ 0 & 0 & 1 \end{bmatrix} = \left[\frac{1}{1+R}, \frac{1}{1+R}, 1\right]$$

$$(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1} = \frac{1+R}{3+R}\left[\frac{1}{1+R}, \frac{1}{1+R}, 1\right]$$

$$= \left[\frac{1}{3+R}, \frac{1}{3+R}, \frac{1+R}{3+R}\right] = \text{the secret weights}.$$

Robust Regression

If we know $\mathbf{V}$ and $\mathbf{V} \neq \sigma^2\mathbf{I}$,

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \qquad \boldsymbol{\epsilon} \sim N(\mathbf{0}, \mathbf{V}).$$

the generalized least squares estimate $\hat{\boldsymbol{\beta}}_{gls}$

- accounts for the correlation between observations to result in an estimator with lower variance.
- the variance $(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}$ allows for valid inference.

What if I don't know $\mathbf{V}$ (as in most cases)?

Initial approach:

1. Use $\hat{\boldsymbol{\beta}}_{ols} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ because it is unbiased and consistent.
2. How about standard errors? Recall

$$Cov(\hat{\boldsymbol{\beta}}_{ols}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$$

So we can estimate $\mathbf{V}$ and plug it in:

$$\widehat{Cov}(\hat{\boldsymbol{\beta}}_{ols}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\hat{\mathbf{V}}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$$

Robust Standard Error

$$\widehat{Cov}(\hat{\boldsymbol{\beta}}_{ols}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\hat{\mathbf{V}}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$$

One estimate of $\hat{\mathbf{V}}$ is simply the empirical residual covariance:

$$\begin{aligned}
\hat{\mathbf{V}} &= [\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{ols}][\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{ols}]' \\
&= \begin{bmatrix} (y_1 - \boldsymbol{x}_1'\hat{\boldsymbol{\beta}}_{ols})^2 & (y_1 - \boldsymbol{x}_1'\hat{\boldsymbol{\beta}}_{ols})(y_2 - \boldsymbol{x}_2'\hat{\boldsymbol{\beta}}_{ols}) & \cdots \\ (y_2 - \boldsymbol{x}_2'\hat{\boldsymbol{\beta}}_{ols})(y_1 - \boldsymbol{x}_1'\hat{\boldsymbol{\beta}}_{ols}) & (y_2 - \boldsymbol{x}_2'\hat{\boldsymbol{\beta}}_{ols})^2 & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix}
\end{aligned}$$

This gives

$$\widehat{Cov}(\hat{\boldsymbol{\beta}}_{ols})_{robust} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\left([\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{ols}][\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{ols}]'\right)\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}.$$

The above is known as the robust estimator or the *sandwich* estimator (Bread′ Meat Bread). This estimator of the covariance is consistent.

Robust Standard Error

The robust standard error provides an estimate of $Cov(\hat{\boldsymbol{\beta}})$:

$$\widehat{Cov}(\hat{\boldsymbol{\beta}}_{ols})_{robust} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\left([\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{ols}][\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{ols}]'\right)\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}.$$

However, note that the empirical estimate of the meat

$$[\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{ols}][\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{ols}]'$$

uses only one observation for each element.

Thus, it is a poor estimator of $\mathbf{V}$, i.e., inaccurate.

We can potentially have a better estimate if we have some idea about what is the structure of $\mathbf{V}$.

Tradeoffs in modeling

When assumptions are true:

- stronger assumptions = more accurate estimates.
- weaker assumptions = less accurate if there is more structure to exploit.

When assumptions are false:

- stronger assumptions = incorrect inference.
- weaker assumptions = more robust.

For example, we see this in non-parametric tests, which are generally less powerful when parametric assumptions hold.

Introducing structure to $\mathbf{V}$

Assume some structure on $\mathbf{V}$.

Call this $\tilde{\mathbf{V}}$. Then,

1. Use $\hat{\boldsymbol{\beta}}_{gls} = (\mathbf{X}'\tilde{\mathbf{V}}^{-1}\mathbf{X})^{-1}\mathbf{X}'\tilde{\mathbf{V}}^{-1}\mathbf{y}$ because it is unbiased and consistent.

2. Use the sandwich estimator for standard errors:
$\widehat{Cov}(\hat{\boldsymbol{\beta}}_{gls})_{robust} = (\mathbf{X}'\tilde{\mathbf{V}}^{-1}\mathbf{X})^{-1}\mathbf{X}'\tilde{\mathbf{V}}^{-1}\left([\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{ols}][\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{ols}]'\right)\tilde{\mathbf{V}}^{-1}\mathbf{X}(\mathbf{X}'\tilde{\mathbf{V}}^{-1}\mathbf{X})^{-1}$.

Idea: even if $\tilde{\mathbf{V}}$ is mis-specified, the standard errors are valid for inference (confidence intervals, hypothesis testing)!

If we get the structure correct, we get a more accurate estimate of β

Marginal Model for Grouped Data

For clustered data, we often assume:

1. observations in the same group are correlated;
2. observations between groups are independent.

For example, consider 2 groups and 2 observations per group:

$$\mathbf{y} = \begin{bmatrix} y_{11} \\ y_{12} \\ y_{21} \\ y_{22} \end{bmatrix}$$

where $\mathbf{V}_1$ and $\mathbf{V}_2$ are two group-specific covariance matrices.

Steps in analysis:

1. decide on the structure of $\mathbf{V}_i$.
2. estimate $\boldsymbol{\beta}$ with generalized least squares.
3. robustify the standard errors of $\hat{\boldsymbol{\beta}}$.

Because our inference focuses on parameters for the mean trend, $\hat{\boldsymbol{\beta}}$ is known as population-averaged or marginal estimates.

Marginal Model for Grouped Data

Denote y_{ij} the jth observation in group i, with $j = 1, 2, \ldots, r_i$. Also let $\mathbf{V}_i = Cov(\mathbf{y}_i)$, where $\mathbf{y}_i = (y_{i1}, y_{i2}, \ldots, y_{ir_i})$. For $i = 1, 2, ...n$,

We often assume the within group correlation has the same structure across groups, $\alpha_i = \alpha_{i'}$ for all i.

$\mathbf{R}_i(\alpha)$ is an $r_i \times r_i$ correlation matrix. $\mathbf{R}_i[j, j'] = cor(y_{ij}, y_{ij'})$.

α is a parameter or vector of parameters that determines the functional form of the correlation.

Marginal Model for Grouped Data

Common choices are: independent, exchangeable, auto-regressive, and unstructured. $\mathbf{R}_i(\alpha)$ is often called the working correlation structure (our best guess).

For example,

$$\mathbf{y} = \begin{bmatrix} y_{11} \\ y_{12} \\ y_{21} \\ y_{22} \end{bmatrix}$$

Common Correlation Structures

Example for a cluster of size 4:

Independent $cor(y_{ij}, y_{ij'}) = 0$

$$\mathbf{R}_i(\alpha) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Exchangeable $cor(y_{ij}, y_{ij'}) = \alpha$

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

Unstructured $cor(y_{ij}, y_{ij'}) = \alpha_{jj'}$ and $\alpha_{jj'} = \alpha_{j'j}$

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

Common Correlation Structures

If the data are measured at distinct time-points:

Autoregressive order-1 $cor(y_{ij}, y_{ij'}) = \alpha^{|j-j'|}$ for $\alpha \leq 1$

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

More working correlation structures

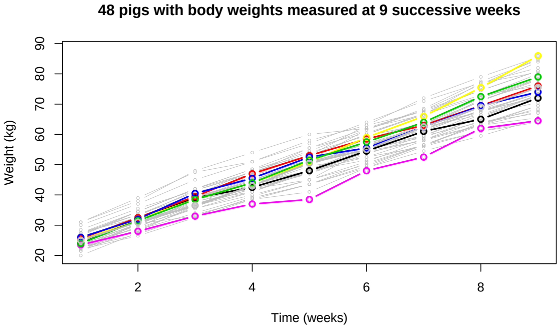
Stationary m-dependent $cor(y_{ij}, y_{ij'}) = \alpha_{|j-j'|}$ if $|j - j'| \leq m$ and 0 otherwise. For $m = 2$,

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

Non-stationary m-dependent $cor(y_{ij}, y_{ij'}) = \alpha_{jj'}$, $|j - j'| \leq m$, else 0:

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

Example: Pig Weight



We need to use a double subscript to denote clustered data with multiple levels. Let y_{ij} be the weight (kg) at the j^{th} week for the i^{th} pig.

For $i = 1, \ldots, 48,\ j = 1, \ldots, 9$, we wish to model weight as a function of week:

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + \epsilon_{ij} \qquad \boldsymbol{\epsilon} \sim N(0, \mathbf{V}).$$

Marginal Model for Pig Weight

For $i = 1, \ldots, 48, \; j = 1, \ldots, 9$

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + \epsilon_{ij} \qquad \boldsymbol{\epsilon} \sim N(0, \mathbf{V}).$$

We assume independence between pigs:

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + \epsilon_i \qquad \boldsymbol{\epsilon}_i \sim N(0, \mathbf{V}_i), \quad \mathbf{V}_i = \sigma^2 \mathbf{R}_i(\alpha).$$

We will estimate β_0 and β_1 using Generalized Estimating Equations (GEE) (marginal) . This is different from the *random effect* approach:

- We do not specify pig-specific intercepts or slopes. We are only concerned about how the within-group data are correlated $\mathbf{R}_i(\alpha)$.
- Advantages:
  - $\mathbf{R}_i(\alpha)$ can be very flexible.
  - Even if we get $\mathbf{R}_i(\alpha)$ wrong, we still have robust standard error.
- Disadvantages:
  - Marginal model does not provide subject-specific predictions – no random effect to balance subject-specific information with population information.
  - Can be less powerful than mixed models.

Fitting GEE in R

```
> library (gee)
> gee
function (formula = formula(data), id = id, data = parent.frame(),
    subset, na.action, R = NULL, b = NULL, tol = 0.001, maxiter = 25,
    family = gaussian, corstr = "independence", Mv = 1, silent = TRUE,
    contrasts = NULL, scale.fix = FALSE, scale.value = 1, v4.4compat = FALSE)
```

- formula: response $\sim$ covariates.
- id: variable name for group id.
- corstr: names of the working correlation structure.
  `"independence", "fixed", "stat_M_dep", "non_stat_M_dep", "exchangeable", "AR-M" and "unstructured"`
- R: user specified structure (with dimension $=$ largest group size) if corstr $=$ fixed.
- Mv: order if required by corstr.
- Scale.fix: assume σ^2 known?

Fitting GEE in R

```
> dat = read.csv ("pig.csv")
> dat[1:3,]
  id weeks weight
1  1     1     24
2  1     2     32
3  1     3     39

> fit = gee (weight~weeks, id = id, data = dat, corstr = "exchangeable")
 GEE:  GENERALIZED LINEAR MODELS FOR DEPENDENT DATA
 gee S-function, version 4.13 modified 98/01/27 (1998)

Model:
 Link:                      Identity
 Variance to Mean Relation: Gaussian
 Correlation Structure:     Exchangeable

Coefficients:
             Estimate Naive S.E.   Naive z Robust S.E. Robust z
(Intercept) 19.355613  0.5983680  32.34734  0.39963854 48.43280
weeks        6.209896  0.0393321 157.88366  0.09107443 68.18485

Estimated Scale Parameter:  19.29006

> round(fit$working.correlation[1:3, 1:3], 2)
     [,1] [,2] [,3]
[1,] 1.00 0.77 0.77
[2,] 0.77 1.00 0.77
[3,] 0.77 0.77 1.00
```

Effects of Working Correlation

Structure	$\hat{\beta}_1$	Naive (GLS) SE	Robust (Sandwich) SE
OLS	6.210	0.082	NA
Independent	6.210	0.082	0.091
Exchangeable	6.210	0.039	0.091
AR-1	6.272	0.079	0.095

- OLS estimate of $\hat{\beta}_1$ is identical to that using GEE with independent $\mathbf{R}_i$ and naive SE.
- Here, GEE with independent and exchangeable $\mathbf{R}_i$ give the same $\hat{\beta}_1$ because the data are balanced and matrix $\tilde{\mathbf{V}}$ gives the same weight in the gls estimate.
- Here, GEE with exchangeable $\mathbf{R}_i$ gives the smaller naive SE compared to independent $\mathbf{R}_i \rightarrow$ accounting for within-group correlation is important.
- Robust SE are similar for any working correlation structure.

Compared to Random Intercept Model

$$y_{ij} = \beta_0 + \theta_i + \beta_1 x_{ij} + \epsilon_{ij} \qquad \epsilon_{ij} \sim N(0, \sigma^2), \quad \theta_i \sim N(0, \tau^2)$$

```
> library (lme4)
> fit.mixed = lmer (weight~weeks + (1|id), data = dat)
> summary (fit.mixed)
Linear mixed model fit by REML

Random effects:
 Groups   Name        Variance Std.Dev.
 id       (Intercept) 15.1418  3.8913
 Residual              4.3947  2.0964

Fixed effects:
            Estimate Std. Error t value
(Intercept) 19.35561    0.60311   32.09
weeks        6.20990    0.03906  158.97
```

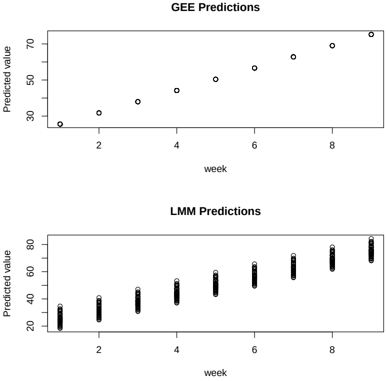
- $\hat{\beta}_1 = 6.210$ with SE $= 0.039$. Here, equal to GEE with exchangeable working correlation matrix and its naive SE.
- The estimated intra-class correlation is $\frac{15.1}{15.1+4.39} = 0.77$, same as the estimated exchangeable working correlation.

Recall that the random intercept model induces exchangeable correlation:

$$cov(y_{ij}, y_{ij'}) = \tau^2.$$

Subject-specific predictions

In GEEs, you can only estimate effects averaged across all subjects.



Subject-specific predictions

In GEEs, you can only estimate effects averaged across all subjects.

Subject predictions in LMMs are more accurate.

```
  > # random intercepts improve prediction:
> sum((predict_gee_exchangeable - dat$weight)^2)
[1] 8294.727
> sum((predict_lmer - dat$weight)^2)
[1] 1689.626
```

Generalized Estimating Equations (GEEs)

Definition of GEE

For normal regression, the generalized estimating equation approach corresponds nicely to a generalized regression. However, GEE's are commonly used to analyze count data, binary outcomes, and other data modeled using a distribution from the exponential family.

Generalized Estimating Equations are a general method for analyzing grouped data when

1. observations within a cluster may be correlated;
2. observations in separate clusters are independent;
3. a monotone transformation of the expectation is linearly related to the explanatory variables;
4. the variance is a function of the expectation.

Reference: Halekoh, U., Hojsgaard, S., & Yan, J. (2006). The R package geepack for generalized estimating equations. Journal of Statistical Software, 15(2), 1-11.

GEE and Marginal Covariance

One challenge is that for Poisson or Bernoulli regression, the residual variance depends on the mean structure. For example, consider a logistic regression model: We have

$$y_i \sim \text{Bernoulli}\ (p_i) \qquad p_i = \frac{e^{\boldsymbol{x}_i'\boldsymbol{\beta}}}{1 + e^{\boldsymbol{x}_i'\boldsymbol{\beta}}}$$

$$\mathbf{V} = \begin{bmatrix} \frac{e^{\boldsymbol{x}_1\boldsymbol{\beta}}}{1+e^{\boldsymbol{x}_1\boldsymbol{\beta}}} \times \frac{1}{1+e^{\boldsymbol{x}_1\boldsymbol{\beta}}} & 0 & \cdots & 0 \\ 0 & \frac{e^{\boldsymbol{x}_2\boldsymbol{\beta}}}{1+e^{\boldsymbol{x}_2\boldsymbol{\beta}}} \times \frac{1}{1+e^{\boldsymbol{x}_2\boldsymbol{\beta}}} & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & \cdots & \frac{e^{\boldsymbol{x}_n\boldsymbol{\beta}}}{1+e^{\boldsymbol{x}_n\boldsymbol{\beta}}} \times \frac{1}{1+e^{\boldsymbol{x}_n\boldsymbol{\beta}}} \end{bmatrix}.$$

Even without clustering, the marginal covariance has unequal variances that depend on $\boldsymbol{\beta}$. To see how GEE works for GLMs, we need to examine in more detail about how these models are fitted.

Revisiting GLMs: Estimation for independent data

Normal Regression

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \qquad \boldsymbol{\epsilon} \sim N(\mathbf{0}, \mathbf{V}).$$

$$L(\mathbf{y}; \boldsymbol{\beta}) = (2\pi)^{-n/2} |\mathbf{V}|^{-1/2} \exp\left(-(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'\right)$$

If we assume $\mathbf{V}$ is diagonal, the data likelihood is given by:

$$L(\mathbf{y}; \boldsymbol{\beta}) = (2\pi)^{-n/2} |\mathbf{V}|^{-1/2} \exp\left(-\sum_{i=1}^{n} \frac{1}{v_i} (y_i - \boldsymbol{x}_i' \boldsymbol{\beta})^2\right)$$

Therefore, the maximum likelihood estimate of $\boldsymbol{\beta}$ can be obtained by minimizing the function:

$$U(\boldsymbol{\beta}) = \sum_{i=1}^{n} \frac{1}{v_i} (y_i - \boldsymbol{x}_i' \boldsymbol{\beta})^2.$$

Equivalent to solving a system of linear equations:

$$\frac{\partial U(\boldsymbol{\beta})}{\partial \beta_k} = \sum_{i=1}^{n} x_{ik} \frac{1}{v_i} (y_i - \boldsymbol{x}_i' \boldsymbol{\beta}) = 0$$

The above is an *estimating equation* too.

Logistic Regression Model

$$y_i \sim \text{Bernoulli } (p_i), \qquad \text{with } p_i = \frac{e^{\boldsymbol{x}_i'\boldsymbol{\beta}}}{1+e^{\boldsymbol{x}_i'\boldsymbol{\beta}}}.$$

For a single observation y_i, the log-likelihood is given by

$$\begin{aligned}
\ell(y_i;\boldsymbol{\beta}) &= \log\left[\left(\frac{e^{\boldsymbol{x}_i'\boldsymbol{\beta}}}{1+e^{\boldsymbol{x}_i'\boldsymbol{\beta}}}\right)^{y_i}\left(\frac{1}{1+e^{\boldsymbol{x}_i'\boldsymbol{\beta}}}\right)^{1-y_i}\right] \\
&= y_i\boldsymbol{x}_i'\boldsymbol{\beta} - y_i\log(1+e^{\boldsymbol{x}_i'\boldsymbol{\beta}}) - (1-y_i)\log(1+e^{\boldsymbol{x}_i'\boldsymbol{\beta}}) \\
&= y_i\boldsymbol{x}_i'\boldsymbol{\beta} - \log(1+e^{\boldsymbol{x}_i'\boldsymbol{\beta}})
\end{aligned}$$

The score function is

$$\frac{\partial \ell(y_i;\boldsymbol{\beta})}{\partial \beta_k} = x_{ik}y_i - \frac{x_{ik}e^{\boldsymbol{x}_i'\boldsymbol{\beta}}}{1+e^{\boldsymbol{x}_i'\boldsymbol{\beta}}} = x_{ik}\left(y_i - \frac{e^{\boldsymbol{x}_i'\boldsymbol{\beta}}}{1+e^{\boldsymbol{x}_i'\boldsymbol{\beta}}}\right).$$

Therefore, the MLE is obtained by solving p equations with p unknowns:

Poisson Log-linear Model

$$y_i \sim \text{Poisson } (\lambda_i), \qquad \text{with } \lambda_i = e^{\boldsymbol{x}_i' \boldsymbol{\beta}}.$$

For a single observation y_i, the log-likelihood is given by

$$\begin{aligned} \ell(y_i; \boldsymbol{\beta}) &= \log \left(\frac{e^{-\lambda_i} \lambda_i^{y_i}}{y_i!} \right) \\ &= -\lambda_i + y_i \log \lambda_i - \log y_i! = -e^{\boldsymbol{x}_i' \boldsymbol{\beta}} + y_i \boldsymbol{x}_i' \boldsymbol{\beta} - (\log y_i!). \end{aligned}$$

The score function with respect to β_k is

$$\frac{\partial \ell(y_i; \boldsymbol{\beta})}{\partial \beta_k} = -x_{ik} e^{\boldsymbol{x}_i' \boldsymbol{\beta}} + y_i x_{ik} = x_{ik} \left(y_i - e^{\boldsymbol{x}_i' \boldsymbol{\beta}} \right).$$

Therefore, the MLE is obtained by solving for a system of regressions:

Estimating Equations

What do the above three score equations have in common?

Gaussian:

$$\sum_{i=1}^{n} x_{ik} \frac{1}{v_i} (y_i - \boldsymbol{x}_i' \boldsymbol{\beta}) = 0.$$

Logistic:

$$\sum_{i=1}^{n} x_{ik} \left(y_i - \frac{e^{\boldsymbol{x}_i' \boldsymbol{\beta}}}{1 + e^{\boldsymbol{x}_i' \boldsymbol{\beta}}} \right) = 0.$$

Poisson:

$$\sum_{i=1}^{n} x_{ik} \left(y_i - e^{\boldsymbol{x}_i' \boldsymbol{\beta}} \right) = 0.$$

It can be shown that for any generalized linear model, the score equations we need to solve for the maximum likelihood estimate are given by

Estimating Equations

Let's verify it for a single data point:

Gaussian:
$$\frac{\partial E(y_i;\boldsymbol{\beta})}{\partial \beta_k}\frac{1}{V(y_i;\boldsymbol{\beta})} = \frac{\partial \sum_{k=1}^{K} x_{ik}\beta_k}{\partial \beta_k}\frac{1}{v_i} = x_{ik}\frac{1}{v_i}$$

Logistic:
$$\begin{aligned}
\frac{\partial E(y_i;\boldsymbol{\beta})}{\partial \beta_k}\frac{1}{V(y_i;\boldsymbol{\beta})} &= \left[\frac{\partial}{\partial \beta_k}\frac{e^{\boldsymbol{x}_i'\boldsymbol{\beta}}}{1+e^{\boldsymbol{x}_i'\boldsymbol{\beta}}}\right]\frac{\left(1+e^{\boldsymbol{x}_i'\boldsymbol{\beta}}\right)^2}{e^{\boldsymbol{x}_i'\boldsymbol{\beta}}} \\
&= \left[\frac{x_{ik}e^{\boldsymbol{x}_i'\boldsymbol{\beta}}}{1+e^{\boldsymbol{x}_i'\boldsymbol{\beta}}} - \frac{x_{ik}\left(e^{\boldsymbol{x}_i'\boldsymbol{\beta}}\right)^2}{\left(1+e^{\boldsymbol{x}_i'\boldsymbol{\beta}}\right)^2}\right]\frac{\left(1+e^{\boldsymbol{x}_i'\boldsymbol{\beta}}\right)^2}{e^{\boldsymbol{x}_i'\boldsymbol{\beta}}} \\
&= x_{ik}\frac{\left(1+e^{\boldsymbol{x}_i'\boldsymbol{\beta}}\right)e^{\boldsymbol{x}_i'\boldsymbol{\beta}} - \left(e^{\boldsymbol{x}_i'\boldsymbol{\beta}}\right)^2}{\left(1+e^{\boldsymbol{x}_i'\boldsymbol{\beta}}\right)^2}\frac{\left(1+e^{\boldsymbol{x}_i'\boldsymbol{\beta}}\right)^2}{e^{\boldsymbol{x}_i'\boldsymbol{\beta}}} = x_{ik}
\end{aligned}$$

Poisson:
$$\frac{\partial E(y_i;\boldsymbol{\beta})}{\partial \beta_k}\frac{1}{V(y_i;\boldsymbol{\beta})} = \frac{\partial e^{\boldsymbol{x}_i'\boldsymbol{\beta}}}{\partial \beta_k}\frac{1}{e^{\boldsymbol{x}_i'\boldsymbol{\beta}}} = x_{ik}e^{\boldsymbol{x}_i'\boldsymbol{\beta}}\frac{1}{e^{\boldsymbol{x}_i'\boldsymbol{\beta}}} = x_{ik}$$

Estimating Equations for Generalized Linear Model

For a generalized linear model (e.g. logistic, poisson), assuming independent samples, we can estimate $\boldsymbol{\beta}$ by solving

$$\left[\frac{\partial \boldsymbol{\mu}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\right]' V(\mathbf{y};\boldsymbol{\beta})^{-1}[\mathbf{y}-\boldsymbol{\mu}(\boldsymbol{\beta})] = \mathbf{0}$$

where

- $\boldsymbol{\mu}(\boldsymbol{\beta})$ is an $n \times 1$ vector of $E(y_i;\boldsymbol{\beta})$.
- $V(\mathbf{y};\boldsymbol{\beta})$ is an $n \times n$ diagonal matrix.
- $\left[\frac{\partial \boldsymbol{\mu}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\right]$ is an $n \times p$ matrix.

It can be shown that the resulting estimate $\hat{\boldsymbol{\beta}}$ is

- consistent.
- asymptotically normal with covariance

$$Cov(\hat{\boldsymbol{\beta}}) = \left(\left[\frac{\partial \boldsymbol{\mu}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\right]' V(\mathbf{y};\boldsymbol{\beta})^{-1}\left[\frac{\partial \boldsymbol{\mu}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\right]\right)^{-1}.$$

Grouped Data

Grouped Data

GEE for Grouped Data

We now assume the non-Gaussian observations may be correlated.

Let $\mathbf{y}_i = (y_{i1}, y_{i2}, \ldots, y_{ir_i})$ be the vector of observations from group i. We want to model the mean trend with a logistic link and account for dependence.

$$\mathbf{y}_i \sim \text{ Bernoulli } (\boldsymbol{\mu}_i) \qquad Cov(\mathbf{y}_i) = \mathbf{V}_i.$$

We can write

$$\mathbf{V}_i = \mathbf{D}_i^{1/2} \mathbf{R}_i(\alpha) \mathbf{D}_i^{1/2}$$

where $\mathbf{D}_i$ is a diagonal matrix of the marginal variance and $\mathbf{R}_i(\alpha)$ is a working correlation matrix.

In GLMs, all elements of $\mathbf{D}_i$ depend on $\mathbf{X}\beta$ through the link function!

With logistic link, jth obs on ith subj: $\mathbf{D}_{i,[j,j]} = \frac{e^{\boldsymbol{x}'_{ij}\boldsymbol{\beta}}}{\left(1+e^{\boldsymbol{x}'_{ij}\boldsymbol{\beta}}\right)^2}$.

Also, the dependent structure is specified directly on the observations $\mathbf{y}_i$. This gives the marginal interpretation.

GEE for Grouped Data

Denote y_{ij} the jth observation in group i, with $j = 1, 2, \ldots, r_i$. Also let $\mathbf{y}_i = (y_{i1}, y_{i2}, \ldots, y_{ir_i})$. We are interested in the marginal model:

where $g(\cdot)$ is the link function.

In the GEE approach let's forget about the complicated data likelihood induced by the within-group correlation and work directly with the estimating equations.

The GEE estimate of $\boldsymbol{\beta}$ is obtained by solving the system of p equations with p unknowns:

$$\sum_{i=1}^{n} \left[\frac{\partial \boldsymbol{\mu}_i(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \right]' \mathbf{V}_i^{-1} [\mathbf{y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta})] = \mathbf{0} \tag{3}$$

where n is the total number of groups.

This generalizes (2).

GEE for Grouped Data

$$\sum_{i=1}^{n} \left[\frac{\partial \boldsymbol{\mu}_i(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\right]' \mathbf{V}_i^{-1} [\mathbf{y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta})] = \mathbf{0}$$

The naive covariance (assuming model is correct) of $\hat{\boldsymbol{\beta}}_{gee}$ is given by:

$$\widehat{Cov}(\hat{\boldsymbol{\beta}})_{naive} = \left(\sum_{i=1}^{n} \left[\frac{\partial \boldsymbol{\mu}_i(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\right]' \mathbf{V}_i^{-1} \left[\frac{\partial \boldsymbol{\mu}_i(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\right]\right)^{-1}$$

The sandwich (robust) covariance of $\hat{\boldsymbol{\beta}}_{gee}$ is given by:

$$\begin{aligned}\widehat{Cov}(\hat{\boldsymbol{\beta}})_{robust} = &\left(\sum_{i=1}^{n} \left[\frac{\partial \boldsymbol{\mu}_i(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\right]' \mathbf{V}_i^{-1} \left[\frac{\partial \boldsymbol{\mu}_i(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\right]\right)^{-1} \times \\ &\left(\sum_{i=1}^{n} \left[\frac{\partial \boldsymbol{\mu}_i(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\right]' \mathbf{V}_i^{-1} (\mathbf{y}_i - \boldsymbol{\mu}_i)(\mathbf{y}_i - \boldsymbol{\mu}_i)' \mathbf{V}_i^{-1} \left[\frac{\partial \boldsymbol{\mu}_i(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\right]\right) \times \\ &\left(\sum_{i=1}^{n} \left[\frac{\partial \boldsymbol{\mu}_i(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\right]' \mathbf{V}_i^{-1} \left[\frac{\partial \boldsymbol{\mu}_i(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\right]\right)^{-1}\end{aligned}$$

Scale parameter

When modeling data in a GLM, the marginal variance is completely determined by the mean.

Often this assumption does not hold, e.g., we have more variability in the observed data.

One approach to account for this is by assuming

$$g(E[\mathbf{y}]) = \mathbf{X}\boldsymbol{\beta}, \qquad Cov(\mathbf{y}_i) = \mathbf{V}_i = \phi\, \mathbf{D}_i^{1/2}\mathbf{R}_i(\alpha)\mathbf{D}_i^{1/2}.$$

Parameter ϕ is known as the scale (or dispersion) parameter for a generalized linear model. $\phi = 1$ indicates no excess residual variation.

As before, GEE estimates can be found by solving (3).

Note there is no density generating the estimating equation. Therefore the GEE estimation methodology is also known as a quasi-likelihood approach.

Example: 2×2 Crossover Trial

Data were obtained from a crossover trial on the disease cerebrovascular deficiency. The goal is to investigate the side effects of a treatment drug compared to a placebo.

Design:

- 34 patients: an active drug (A) and followed by a placebo (B)
- 33 patients: a placebo (B) and followed by an active drug (A).
- Outcome: eletrocardiogram determined normal (0) or abnormal (1).
- Each patient has two binary observations for period 1 or period 2.

Data:

	Responses				Period		
Group	(1,1)	(0,1)	(1,0)	(0,0)	1	2	Total
A-B	22	0	6	6	28	22	33
B-A	18	4	2	9	20	22	34

Example: 2×2 Crossover Trial

Marginal (GEE) :

Conditional (Random intercept model):

- The marginal approach models the exchangeable correlation directly on the observed binary outcomes.
- The conditional approach induces an exchangeable correlation on the logit-transformed mean trend. This approach models group-specific baseline log odds explicitly.

Another note on marginal versus conditional

Marginal:

$$\text{logit } E(y_{ij}) = \beta_0 + \beta_1 trt_{ij} + \beta_2 period_{ij} + \beta_3 trt_{ij} * period_{ij}$$

Conditional (Random intercept model):

$$\text{logit } E(y_{ij}|\theta_i^*) = \beta_0^* + \theta_i^* + \beta_1^* trt_{ij} + \beta_2^* period_{ij} + \beta_3^* trt_{ij} * period_{ij}$$

Note:

$$E\left\{\text{logit } E(y_{ij}|\theta_i^*)\right\} \neq \text{logit } E(y_{ij})$$

Conclusion: coefficient estimates and their interpretation differs.

Example: 2×2 Crossover Trial

```
> library (gee)
> library (lme4)
> dat = read.table ("2by2.txt", header = T)

### GLM
> fit.glm = glm (outcome~trt*period,data = dat, family = binomial)

### Marginal GEE model with exch working corr
> fit.exch = gee (outcome~trt*period, id = ID, data = dat,
    family = binomial (link = "logit"),  corstr = "exchangeable")

### Marginal GEE model with ind corr
> fit.ind = gee (outcome~trt*period, id = ID, data = dat,
    family = binomial (link = "logit"),  corstr = "independence")

### GLMM (conditional) model with (exchangeable corr)
> fit.re = glmer (outcome~trt*period+(1|ID), data=dat,
    family = binomial, nAGQ = 2)
```

Example: 2×2 Crossover Trial

```
> summary (fit.exch)

 GEE:  GENERALIZED LINEAR MODELS FOR DEPENDENT DATA
 gee S-function, version 4.13 modified 98/01/27 (1998)

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

Coefficients:
              Estimate Naive S.E.   Naive z Robust S.E.  Robust z
(Intercept) -1.5404450  0.4567363 -3.372723   0.4498677 -3.424218
trt          1.1096621  0.5826118  1.904634   0.5738502  1.933714
period       0.8472979  0.5909040  1.433901   0.5820177  1.455794
trt:period  -1.0226507  0.9997117 -1.022946   0.9789663 -1.044623

Estimated Scale Parameter:  1.030769
Number of Iterations:  1

Working Correlation
          [,1]      [,2]
[1,] 1.0000000 0.6401548
[2,] 0.6401548 1.0000000
```

- $\hat{\phi} = 1.0307 =$ very slight over-dispersion.
- $Cor(y_{ij}, y_{ij'}) = 0.64$ (note this is on the binary outcomes).

Results Comparison

Log Odds Treatment Effect (β_1)

Approach	$\hat{\beta}_1$	Naive SE	Robust SE
GLM	1.11	0.57	NA
GEE, Independent	1.11	0.58	0.57
GEE, Exchangeable	1.11	0.58	0.57
Random intercept	3.60	2.14	NA

- The three marginal approaches give the same point estimates. In this analysis, accounting for between-subject correlation gives very similar standard errors compared to standard glm.
- The conditional effect is larger than the marginal estimates.

Example: Vitamin A and Respiratory Infection

250 children in Indonesia examined in 3-month intervals for 6 visits.

Variables

For child i at time j,

- $response_{ij}$: indicator that the child was suffering from a respiratory infection.
- $time_{ij}$: time in months since initial visit.
- sex_i: sex of the child (female $= 1$).
- vit_{ij}: indicator for whether the child had vitamin A deficiency (0 $=$ no, $1 =$ yes).
- age_{ij}: in years.

Goal

We are interested in determining the association between vitamin A deficiency and the risk of respiratory infection.

Example: Vitamin A and Respiratory Infection

```
> dat = read.table ("Indonesia.txt", header = T)

> dat[1:3,]
  id response time sex vit age
1  1        0    0   1   0   4
2  1        0    3   1   0   4
3  1        0    6   1   0   4

> table (table (dat$id)) ### 6 visits per child ###
  6
250

> table (dat$time) ### Balanced
  0   3   6   9  12  15
250 250 250 250 250 250

##No missing data
> summary (dat)
       id          response            time            sex              vit              age
 Min.   :  1    Min.   :0.000   Min.   : 0.0   Min.   :0.00   Min.   :0.000   Min.   :1.0
 1st Qu.: 63    1st Qu.:0.000   1st Qu.: 3.0   1st Qu.:0.00   1st Qu.:0.000   1st Qu.:2.0
 Median :126    Median :0.000   Median : 7.5   Median :1.00   Median :0.000   Median :4.0
 Mean   :126    Mean   :0.296   Mean   : 7.5   Mean   :0.54   Mean   :0.364   Mean   :3.7
 3rd Qu.:188    3rd Qu.:1.000   3rd Qu.:12.0   3rd Qu.:1.00   3rd Qu.:1.000   3rd Qu.:5.0
 Max.   :250    Max.   :1.000   Max.   :15.0   Max.   :1.00   Max.   :1.000   Max.   :7.0
```

Models

GLM:

$$\text{logit } P(y_{ij}=1) = \beta_0 + \beta_1 time_{ij} + \beta_2 age_{ij} + \beta_3 sex_i + \beta_4 vit_{ij}$$

with and without over-dispersion

Marginal (GEE) :

$$\text{logit } P(y_{ij}=1) = \beta_0 + \beta_1 time_{ij} + \beta_2 age_{ij} + \beta_3 sex_i + \beta_4 vit_{ij}$$

$$Cov(\mathbf{y}_i) = \phi\, \mathbf{D}_i^{1/2} \mathbf{R}_i(\alpha) \mathbf{D}_i^{1/2}$$

$\mathbf{R}_i(\alpha) =$ independent, exchangeable, AR-1 correlation, or unstructured

Conditional (Random intercept model):

$$\text{logit } P(y_{ij}=1|\theta_i^*) = \beta_0^* + \theta_i^* + \beta_1^* time_{ij} + \beta_2^* age_{ij} + \beta_3^* sex_i + \beta_4^* vit_{ij}$$

$$\theta_i^* \sim N(0, \tau^2).$$

R Code

```
### GLM without over-dispersion
fit.glm = glm (response~time+sex+vit+age,family=binomial,data=dat)

### GLM with over-dispersion  ("quasi" option)
fit.glmquasi = glm (response~time+sex+vit+age,family=quasibinomial, data=dat)

fit.ind = gee (response~time+sex+vit+age, id=id, family=binomial,
                                            corstr="independence", data=dat)

fit.exch = gee (response~time+sex+vit+age, id=id, family=binomial,
                                            corstr="exchangeable", data=dat)

fit.ar1 = gee (response~time+sex+vit+age, id=id, family=binomial,
                                            corstr="AR-M", Mv = 1, data=dat)

fit.uns = gee (response~time+sex+vit+age, id=id, family=binomial,
                                            corstr="unstructured", data=dat)

fit.re = glmer (response~time+sex+vit+age + (1|id), family=binomial, data=dat, nAGQ=25)

###  extract correlation matrix
> fit.exch$working.correlation
```

Estimated Correlation Matrix

Exchangeable:

```
      [,1]  [,2]  [,3]  [,4]
[1,] 1.000 0.505 0.505 0.505
[2,] 0.505 1.000 0.505 0.505
[3,] 0.505 0.505 1.000 0.505
[4,] 0.505 0.505 0.505 1.000
```

AR-1:

```
     [,1] [,2] [,3] [,4] [,5] [,6]
[1,] 1.00 0.53 0.28 0.15 0.08 0.04
[2,] 0.53 1.00 0.53 0.28 0.15 0.08
[3,] 0.28 0.53 1.00 0.53 0.28 0.15
[4,] 0.15 0.28 0.53 1.00 0.53 0.28
[5,] 0.08 0.15 0.28 0.53 1.00 0.53
[6,] 0.04 0.08 0.15 0.28 0.53 1.00
```

Unstructured:

```
     [,1] [,2] [,3] [,4] [,5] [,6]
[1,] 1.00 0.56 0.46 0.42 0.50 0.45
[2,] 0.56 1.00 0.58 0.56 0.53 0.56
[3,] 0.46 0.58 1.00 0.56 0.42 0.51
[4,] 0.42 0.56 0.56 1.00 0.43 0.52
[5,] 0.50 0.53 0.42 0.43 1.00 0.51
[6,] 0.45 0.56 0.51 0.52 0.51 1.00
```

Results Comparison

Log Odds Vitamin Effect (β_3)

Approach	$\hat{\beta}_3$	Naive SE	Robust SE	Dispersion
GLM	0.287	0.117		
Quasi-GLM	0.287	0.117		0.998
GEE (independence)	0.287	0.117	0.223	0.998
GEE (exchangeable)	0.278	0.220	0.223	0.998
GEE (AR-1)	0.286	0.180	0.224	0.997
GEE (unstructured)	0.284	0.218	0.220	0.996
Random intercept	0.633	0.429		

- Minor over-dispersion.
- Marginal point estimates are similar.
- Without controlling for within-child correlation, you would have a significant effect using GLM!

Interpretations

- Marginal: Assuming an exchangeable correlation for the respiratory outcomes observed within the same child, we estimated a population-average OR for vitamin deficiency $\exp(0.278) = 1.32$ with a 95% confidence interval of (0.85, 2.04) using a robust standard error. The estimated effect was adjusted for sex, visit number, and age. Therefore, our results suggest no association between vitamin deficiency and the presence of respiratory infection.

- Conditional: After controlling for child-specific baseline log odds of respiratory infection, we estimated an OR of $\exp(0.633) = 1.88$ ($\text{CI}_{95\%}$: 0.80, 4.45) with vitamin deficiency accounting for the child's sex, visit number, and age. Therefore, our results suggest no association between vitamin deficiency and the presence of respiratory infection.

GEE versus GLMM

Use GLMM when you are interested in subject-specific predictions. The GLMM improves predictions using the estimates of the random effects.

GLMMs can also be useful for hierarchical modeling, in which case GEE software has limited flexibility.

GEEs provide more robust inference, which decreases false positives, i.e., you correctly fail to reject a null hypothesis. GLMMs make stronger distributional assumptions (no sandwich estimators).

The marginal interpretation in GEEs may be more intuitive. E.g., if you looked at mortality across different diagnostic categories, the numbers in a GEE will be closer to mortality rates in a demographic table.