

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.

• Toy example
• WEIGHTED LEAST SQUARES
• Generalized Least Squares
• Special case of GLS:
• Robust regression
• Generalized estimating equations
(will combine robust regression with GLS)

Motivating Example Toy example

Consider data y_1, y_2 , and y_3 from

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} \sim N\left(\begin{bmatrix} \mu \\ \mu \\ \mu \end{bmatrix}, \Sigma = \begin{bmatrix} \sigma^2 & R & 0 \\ R & \sigma^2 & 0 \\ 0 & 0 & \sigma^2 \end{bmatrix}\right)$$

- 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.

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

assigns equal weight to

$$\mathbf{a} = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3} \right)^T$$

$$\mathbf{y} = (y_1, y_2, y_3)^T$$

The above estimator has variance

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

$$\text{Cov}(\mathbf{a}' \mathbf{y}) = \mathbf{a}' (\text{cov}(\mathbf{y})) \mathbf{a}$$

$$\begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix} \begin{bmatrix} \sigma^2 & R & 0 \\ R & \sigma^2 & 0 \\ 0 & 0 & \sigma^2 \end{bmatrix} \begin{bmatrix} \frac{1}{3} \\ \frac{1}{3} \\ \frac{1}{3} \end{bmatrix}$$

Positive Correlation observations ↑ variance of sample mean
 R increases → variance increases.

Motivating Example

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

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

$w_1, w_2, w_3 > 0$, and

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

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

$$\begin{aligned}\text{Var}(\hat{\mu}) &= [w_1 \ w_2 \ w_3] \begin{bmatrix} \sigma^2 & R & 0 \\ R & \sigma^2 & 0 \\ 0 & 0 & \sigma^2 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} \\ &= \underbrace{w_1 \sigma^2 + w_2 R}_{w_1}, \underbrace{w_1 R + w_2 \sigma^2}_{w_2}, \underbrace{w_3 \sigma^2}_{w_3} \\ &= \underbrace{w_1^2 \sigma^2}_{w_1} + \underbrace{w_1 w_2 R}_{w_1 w_2} + \underbrace{w_1 w_2 R + w_2^2 \sigma^2}_{w_2^2} + \underbrace{w_3^2 \sigma^2}_{w_3} \\ &= (\underbrace{w_1^2 + w_2^2 + w_3^2}_{w_1 w_2}) \sigma^2 + \underbrace{2 w_1 w_2 R}_{w_1 w_2}.\end{aligned}$$

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 :

$$\mathbf{w} = \underbrace{(1/4, 1/4)}_{\text{---}} \underbrace{(1/2)}_{\text{---}}$$

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

$$\begin{bmatrix} \sigma^2 & \sigma^2 & 0 \\ \sigma^2 & \sigma^2 & 0 \\ 0 & 0 & \sigma^2 \end{bmatrix}$$

H_{1212}
 $y_1 = y_2$

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

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 ^{UNBIASED} estimator $\hat{\mu}$ can be obtained by solving:

$$[\alpha_1(y_1 - \hat{\mu}) + \alpha_2(y_2 - \hat{\mu}) + \alpha_3(y_3 - \hat{\mu}) = 0] \quad (1)$$

The above gives:

$$\hat{\mu} = \frac{\alpha_1 y_1 + \alpha_2 y_2 + \alpha_3 y_3}{\alpha_1 + \alpha_2 + \alpha_3}. \quad \text{i.e., } \sum w_i = 1$$

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.

MOVING TOWARDS WEIGHTED LEAST SQUARES

Regression with heteroscedastic errors

Now consider the linear regression problem with heteroscedastic errors:

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

$\boldsymbol{\epsilon} \in \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{bmatrix}$

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

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

DEFINIS

$$\mathbf{W} = \begin{bmatrix} v_1^{-1/2} & \dots & 0 \\ 0 & \ddots & 0 \\ \vdots & \ddots & 0 \\ 0 & \dots & 0 & v_n^{-1/2} \end{bmatrix}$$

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

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

Note that $\mathbf{WW} = \mathbf{V}^{-1}$.

Consider the transformed \mathbf{y} :

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

$\xrightarrow{\text{Cov}(y^*) = \text{Cov}(Wy)}$
 $\mathbf{W} \text{Cov}(\mathbf{y}) \mathbf{W}$
 $\mathbf{W} \mathbf{V} \mathbf{W}$

- $\text{Var}(y_i^*) = \text{Var}(w_i y_i) = w_i^2 v_i = v_i^{-1} v_i = 1$, $(\mathbf{W} \mathbf{V}^{1/2} \mathbf{V}^{1/2} \mathbf{W})$
 - $E(\mathbf{y}^*) = \mathbf{WX}\boldsymbol{\beta}$
- $E[\mathbf{W}\mathbf{X}\boldsymbol{\beta} + \mathbf{W}\boldsymbol{\epsilon}] = \mathbf{W}\mathbf{X}\boldsymbol{\beta}$

Weighted Least Squares

In this class, WLS means V is diagonal with different variances

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

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

We have

$$W\mathbf{y} = W\mathbf{X}\beta + W\epsilon$$

$$\begin{aligned}\hat{\beta}_{WLS} &= (\mathbf{X}^{*'} \mathbf{X}^*)^{-1} \mathbf{X}^{*'} \mathbf{y}^* \\ &= ([W\mathbf{X}]' W\mathbf{X})^{-1} [W\mathbf{X}]' W\mathbf{y} \\ &= (\mathbf{X}' W' W\mathbf{X})^{-1} \mathbf{X}' W' W\mathbf{y} \\ &= (\mathbf{X}' V^{-1} \mathbf{X})^{-1} \mathbf{X}' V^{-1} \mathbf{y}.\end{aligned}$$

HOMOSCEDASTIC

$$\begin{aligned}OLS: V &= \sigma^2 \mathbf{I} \\ (\mathbf{X}' [\sigma^2 \mathbf{I}]^{-1} \mathbf{X})' \mathbf{X}' [\sigma^2 \mathbf{I}]^{-1} Y \\ &\xrightarrow{\sigma^2} (\mathbf{X}' \mathbf{X})' \mathbf{X}' (\sigma^2)^{-1} Y \\ &\xrightarrow{\sigma^2} (\mathbf{X}' \mathbf{X})' \mathbf{X}' Y\end{aligned}$$

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

Weighted Least Squares, cont.

$$\begin{aligned}
 \text{Cov}(\hat{\beta}_{\text{WLS}}) &= \text{Cov}\left(\underbrace{(X'V^{-1}X)^{-1} X'V^{-1}Y}_{=} \right) \\
 &= (X'V^{-1}X)^{-1} X'V^{-1} \text{Cov}(Y) [(X'V^{-1}X)^{-1} X'V^{-1}]^T \\
 &= (X'V^{-1}X)^{-1} \underbrace{X'V^{-1} V V^{-1} X}_{=} (X'V^{-1}X)^{-1} \\
 &\approx \underbrace{(X'V^{-1}X)^{-1}}_{=I} \underbrace{X'V^{-1}X}_{=} (X'V^{-1}X)^{-1} \\
 &= (X'V^{-1}X)^{-1}
 \end{aligned}$$

$\overbrace{[e.g. \text{ OLS}]^{-1}}^{=} = \sigma^2(X'X)^{-1}$

Weighted Least Squares

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

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

$E(X\beta + \varepsilon) = X\beta$

What if we used the regular ordinary least squares estimate?

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

Still unbiased!

Note on OLS estimator

The OLS estimator of β is unbiased. It is also consistent, i.e., if $Y = X\beta + \varepsilon$,
 $\varepsilon \sim N(0, V)$

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

So what is the problem?

- (1) INFEERENCE \rightarrow p-values will be wrong
- (2) EFFICIENCY: can define an estimator with lower variance

Note on OLS estimator

$$E\left(\left[\hat{\beta}_{ols} - E\hat{\beta}_{ols}\right]\left[\hat{\beta}_{ols} - E\hat{\beta}_{ols}\right]'\right)$$

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

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

However, our estimate of the covariance is

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

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

I.e., INFERENCE is incorrect

WE WANT $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{V}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$
 \downarrow
 $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{V}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \xrightarrow{\text{OLS}}$

but it doesn't

Two issues with OLS when heteroscedasticity

when $V \neq \sigma^2 I$

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

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

- (2) 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,
i.e., use our data in a better way

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

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

This can be expressed as

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

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

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

Therefore

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

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

inverse variance
weight



... 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:

correlation between
observations and
heteroscedasticity

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

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

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

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

... exists for symmetric positive definite matrix

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

Eigenvalue value decomposition $\mathbf{U}\mathbf{\Lambda}\mathbf{U}'$, e.g. $\mathbf{U}\mathbf{\Lambda}^{1/2}\mathbf{U}' = \mathbf{W}$

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.

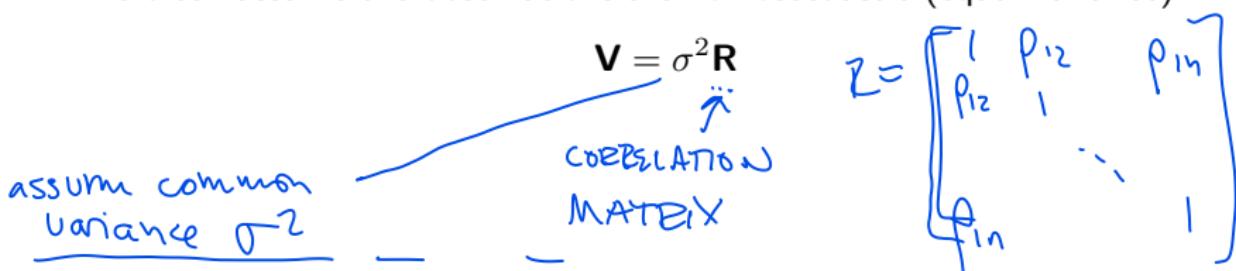
[It can be shown that the weighting scheme using \mathbf{V} in $\hat{\beta}_{gls}$ gives the smallest variances (optimally efficient) among unbiased estimators if we know \mathbf{V} .

Generalized Least Squares: homoscedastic

For normal (Gaussian) regression model,

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

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



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

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

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

Generalized Least Squares

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

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

$$\begin{aligned} (\begin{pmatrix} a & c \\ b & d \end{pmatrix})^{-1} &= \\ (\begin{pmatrix} ad - bc & -b \\ -c & a \end{pmatrix})^{-1} &\begin{bmatrix} d & -b \\ -c & a \end{bmatrix} \\ (1-R^2)^{-1} &\begin{pmatrix} 1 & R \\ -R & 1 \end{pmatrix} \end{aligned}$$

Here

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

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

Generalized

~~Weighted Least Squares~~

$$X = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

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

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

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

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

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

CLS ESTIMATOR

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

Robust Regression a "fix" for OLS estimator inference

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

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

when data not IID

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

- 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:

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

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

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

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

Robust Standard Error

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

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

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

$$= \begin{bmatrix} (y_1 - \mathbf{x}_1' \hat{\beta}_{ols})^2 & (y_1 - \mathbf{x}_1' \hat{\beta}_{ols})(y_2 - \mathbf{x}_2' \hat{\beta}_{ols}) & \dots \\ (y_2 - \mathbf{x}_2' \hat{\beta}_{ols})(y_1 - \mathbf{x}_1' \hat{\beta}_{ols}) & (y_2 - \mathbf{x}_2' \hat{\beta}_{ols})^2 & \dots \\ \vdots & \vdots & \ddots \end{bmatrix}$$

• outer product of the residuals

This gives

$$\widehat{Cov}(\hat{\beta}_{ols})_{robust} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \underbrace{([\mathbf{y} - \mathbf{X}\hat{\beta}_{ols}] [\mathbf{y} - \mathbf{X}\hat{\beta}_{ols}]')}_{\text{outer product of the residuals}} \mathbf{X} (\mathbf{X}'\mathbf{X})^{-1}$$

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

Turns out $\widehat{Cov}(\hat{\beta}_{ols})_{ROBUST} \rightarrow Cov(\hat{\beta}_{OLS})$ for $\varepsilon \sim N(0, V)$

Robust Standard Error

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

$$\widehat{Cov}(\hat{\beta}_{ols})_{robust} = \underbrace{(\mathbf{X}'\mathbf{X})^{-1}}_{\hat{\Sigma} \hat{\Sigma}'} \mathbf{X}' (\underbrace{[\mathbf{y} - \mathbf{X}\hat{\beta}_{ols}] [\mathbf{y} - \mathbf{X}\hat{\beta}_{ols}]'}_{\text{the meat}}) \underbrace{\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}}_{\mathbf{V}}$$

However, note that the empirical estimate of the meat

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

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. fewer type II errors
- weaker assumptions = less accurate if there is more structure to exploit.

When assumptions are false:

- stronger assumptions = incorrect inference. more type I errors
- weaker assumptions = more robust.

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

E.g., two-sample t-test more powerful than Wilcoxon rank sum when data are normally distributed, however, if data are not normally dist., t-test has ↑ type I errors

Introducing structure to \mathbf{V}

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

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

1. Use $\hat{\beta}_{gls} = (\mathbf{X}'\tilde{\mathbf{V}}^{-1}\mathbf{X})^{-1}\mathbf{X}'\tilde{\mathbf{V}}^{-1}\mathbf{y}$ because it is unbiased and consistent. *(USE GLS estimator of $\hat{\beta}$)*
2. Use the sandwich estimator for standard errors:
$$\widehat{Cov}(\hat{\beta}_{gls})_{robust} = (\mathbf{X}'\tilde{\mathbf{V}}^{-1}\mathbf{X})^{-1} \underbrace{\mathbf{X}'\tilde{\mathbf{V}}^{-1} ([\mathbf{y} - \mathbf{X}\hat{\beta}_{ols}] [\mathbf{y} - \mathbf{X}\hat{\beta}_{ols}]')}_{\substack{\text{(covariance in GLS)} \\ \text{PROVIDES Robustness!}}} \tilde{\mathbf{V}}^{-1} \mathbf{X} (\mathbf{X}'\tilde{\mathbf{V}}^{-1}\mathbf{X})^{-1}$$

Idea: even if \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

*allows for violations of assumption
of homoscedasticity
[i.e., data can have heteroscedasticity]*

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:

$$\mathbf{y} = \begin{bmatrix} y_{11} \\ y_{12} \\ y_{21} \\ y_{22} \end{bmatrix} \sim N(\mathbf{X}\boldsymbol{\beta}, \mathbf{V})$$

= cluster = subject 2x2 covariance matrix
4x p p x 1
2x2 matrix of Os

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.
i.e., mean trend and covariance
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 j th observation in group i , with $j = 1, 2, \dots, r_i$. Also let $\mathbf{V}_i = Cov(\mathbf{y}_i)$, where $\mathbf{y}_i = (y_{i1}, y_{i2}, \dots, y_{ir_i})$. For $i = 1, 2, \dots, n$,

$$V_i = \sigma^2 R_i(\alpha) \quad \begin{matrix} \text{a vector of parameters} \\ (\text{sometimes a scalar}) \end{matrix}$$

$r_i \times r_i \quad \text{scalar} \quad \text{Correlation matrix } r_i \times r_i$

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

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

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. $R_i(\alpha)$ is often called the working correlation structure (our best guess).

For example,

don't need to get it right, but when we do, it is more efficient

$$\mathbf{y} = \begin{bmatrix} y_{11} \\ y_{12} \\ y_{21} \\ y_{22} \end{bmatrix} \sim N\left(X\beta, \sigma^2 \begin{bmatrix} R_1(\alpha) & \mathbb{O} \\ \mathbb{O} & R_2(\alpha) \end{bmatrix} \right)$$

2x2 matrix of 0s

Common Correlation Structures

Example for a cluster of size 4:

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

- $\alpha = 0$
- 0 parameters
- Robust SE = Robust regression

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

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

"exchange" j indices same correlate

Same correlation structures implied

by random intercept model

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

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

6 parameters

- lots of parameters
- caution: can cause convergence issues in R

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

Common Correlation Structures

If the data are measured at distinct time-points:

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

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

\S^2

- CORRELATION DECAYS WITH TIME
- can have AR(k) correlations in $g_i :: g_j$

BASED on ARI model $\varepsilon_{it} = \alpha \varepsilon_{it-1} + \eta_{it}$, $\eta_{it} \stackrel{\text{iid}}{\sim} N(0, \sigma^2)$

$$\text{Corr}(\varepsilon_{it}, \varepsilon_{it-1}) = \alpha$$

$$\text{Corr}(\alpha \varepsilon_{it-1} + \eta_{it}, \varepsilon_{it-1}) = \alpha$$

More working correlation structures

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

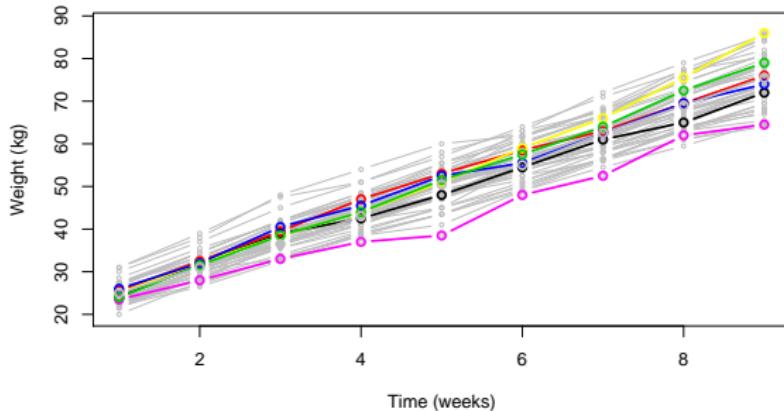
$$\mathbf{R}_i(\alpha) = \begin{bmatrix} 1 & \alpha_1 & \alpha_2 & 0 \\ \alpha_1 & 1 & \alpha_1 & \alpha_2 \\ \alpha_2 & \alpha_1 & 1 & \alpha_1 \\ 0 & \alpha_2 & \alpha_1 & 1 \end{bmatrix}.$$

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

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

Example: Pig Weight

48 pigs with body weights measured at 9 successive weeks



Let y_{ij} be the weight (kg) at the j^{th} week for the i^{th} pig.

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

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

Marginal Model for Pig Weight

For $i = 1, \dots, 48, j = 1, \dots, 9$

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

We assume independence between pigs:

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

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.
*if interested in subject-specific predictions
use mixed model*
 - Can be less powerful than mixed models.

in BLUPs

Note: two Packages

- gee
- geepack

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 ~ covariates.
- id: variable name for group id. *i.e., clustering variable*
- 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. *applies to AR-M,
non-stat_M_dep, stat_M_dep*
- 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: G-LS: $(X' V^{-1} X)^{-1}$

	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
```

extracted from
sandwich:

$$(X' V^{-1} X)^{-1} X' V^{-1} (\gamma - \hat{\beta})(\gamma - \hat{\beta})' V^{-1} X(X' V^{-1} X)^{-1}$$

in OLS, this is like σ^2 ,
residual variance

Here, it's like $\sigma^2 + \tau^2$ is LMM

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

(not in general equal)

- OLS estimate of $\hat{\beta}_1$ is identical to that using GEE with independent R_i and naive SE.
- Here, GEE with independent R_i and exchangeable R_i give the same $\hat{\beta}_1$. Covariate values are the same for each piglet. (This is an uncommon situation.)
- Here, GEE with exchangeable R_i gives the smaller naive SE compared to independent $R_i \rightarrow$ accounting for within-group correlation is important.
- Robust SE are similar for any working correlation structure.

Compared to Random Intercept Model

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

```
> 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

$$\frac{\hat{\tau}^2}{\hat{\tau}^2 + \hat{\sigma}^2} = \frac{15.14}{4.39 + 15.14} \approx 0.77$$

$\hat{\sigma}^2 + \hat{\tau}^2 \approx \text{scale parameter from GEE}$

... & NAIVE SE FROM EXCHANGABLE

- $\hat{\beta}_1 = 6.210$ 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.

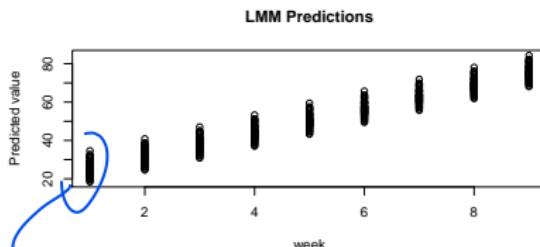
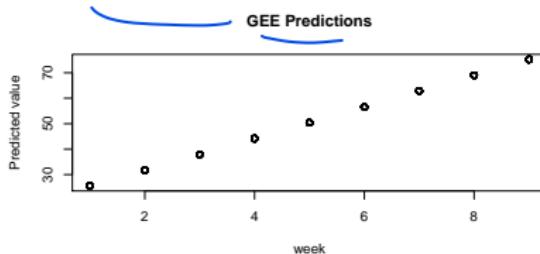
Recall that the random intercept model induces exchangeable correlation:

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

Subject-specific predictions

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

we only get
"marginal" mean,
i.e., averaged
subjects
 $\hat{\beta}_0 + \hat{\beta}_1 x_{ik}$
- no $\hat{\theta}_i$'s



LMMs: BLUPs, i.e.,
 $\hat{\beta}_0 + \hat{\beta}_1 x_{ik} + \hat{\theta}_i$

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  
....
```

$$-\sum (y_{ij} - \hat{y}_{ij})^2$$

LMMs better for subject specific predictions

Generalized Estimating Equations

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; i.e., link function in GLM
- 4. the variance is a function of the expectation; i.e., exponential family

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

↓ Note: Robust SEs allow for departures from this

GLM

~~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

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

$$\text{Var}(y_i) = p_i(1-p_i) = \frac{e^{\mathbf{x}'_i \beta}}{(1 + e^{\mathbf{x}'_i \beta})^2}$$

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

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

Revisit GLMs

Revisiting GLMs: Estimation for independent data

Normal Regression

First, assume y_i independent,

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

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

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

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

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

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

Equivalent to solving a system of linear equations:

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

The above is an *estimating equation* too.

$$\begin{cases} \sum_{i=1}^n x_{1i} \frac{1}{v_i} (y_i - \mathbf{x}'_i \boldsymbol{\beta}) = 0 \\ \vdots \\ \sum_{i=1}^n x_{pi} \frac{1}{v_i} (y_i - \mathbf{x}'_i \boldsymbol{\beta}) = 0 \end{cases}$$

Logistic Regression Model

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

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

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

The score function is

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

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

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

Poisson Log-linear Model

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

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

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

The score function with respect to β_k is

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

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

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

Estimating Equations

What do the above three score equations have in common?

Gaussian:

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

Logistic:

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

Poisson:

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

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

$$\sum_{i=1}^n \frac{\partial E[y_i; \boldsymbol{\beta}]}{\partial \beta_k} V(\underline{y_i; \boldsymbol{\beta}}) \underline{[y_i - E[y_i; \boldsymbol{\beta}]]} = 0$$

Estimating Equations

Let's verify it for a single data point:

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

Logistic:

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

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

Estimating Equations for Generalized Linear Model

For a generalized linear model (e.g. logistic, poisson), assuming independent samples, we can estimate β by solving

e.g. for logistic + Poisson:

$$\text{where } \begin{bmatrix} x_{11} & \cdots & x_{1n} \\ \vdots & & \vdots \\ x_{p1} & \cdots & x_{pn} \end{bmatrix}^P \begin{bmatrix} \frac{\partial \mu(\beta)}{\partial \beta} \\ \vdots \\ \frac{\partial \mu(\beta)}{\partial \beta} \end{bmatrix}^n V(\mathbf{y}; \beta)^{-1} \begin{bmatrix} \mathbf{y} - \mu(\beta) \\ \vdots \\ \mathbf{y} - \mu(\beta) \end{bmatrix}^n = \mathbf{0}^{P \times 1}$$

- $\mu(\beta)$ is an $n \times 1$ vector of $E(y_i; \beta)$.

- $V(\mathbf{y}; \beta)$ is an $n \times n$ **diagonal matrix**.

- $\begin{bmatrix} \frac{\partial \mu(\beta)}{\partial \beta} \end{bmatrix}$ is an $n \times p$ matrix, $\begin{bmatrix} \frac{\partial \mu(\beta)}{\partial \beta} \\ \vdots \\ \frac{\partial \mu(\beta)}{\partial \beta} \end{bmatrix}^n$

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

- consistent.
- asymptotically normal with covariance

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

Grouped Data

Grouped Data

GEE for Grouped Data for logistic regression

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

Let $\mathbf{y}_i = (y_{i1}, y_{i2}, \dots, y_{ir_i})$ be the vector of observations from group i .

We want to model the mean trend with a logistic link and account for dependence.

independent subjects

$\mathbf{y}_i \sim \text{Bernoulli}(\mu_i)$

$\text{Cov}(\mathbf{y}_i) = \mathbf{V}_i$

marginal covariance between obs. on same subject

We can write

should be intuitive

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

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

Recall: Correlation = $(\mathbf{D}_i^{-1/2} \mathbf{V}_i \mathbf{D}_i^{-1/2})$

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

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

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

$$g(E(y_i)) = \mathbf{x}_i^\top \boldsymbol{\beta}, \text{Cov}(y_i) = D_i^{Y_2} R_i(\kappa) D_i^{Y_2}$$

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

scalar or vector of dependence parameters

In the GEE approach let's forget about the complicated data likelihood parameters 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:

$$\sum_{i=1}^n \underbrace{\left[\frac{\partial \mu_i(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \right]'}_{p \times r_i} \mathbf{V}_i^{-1} [\mathbf{y}_i - \mu_i(\boldsymbol{\beta})] = \underbrace{\mathbf{0}}_{p \times 1} \quad (3)$$

where n is the total number of groups.

This generalizes (2).

GEE for Grouped Data

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

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

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

The sandwich (robust) covariance of $\hat{\beta}_{gee}$ is given by: *Review for Gaussian case:
 $(\mathbf{x}'\mathbf{v}'\mathbf{x})' \mathbf{x}'\mathbf{v}' (\mathbf{y}-\hat{\mathbf{y}})(\mathbf{y}-\hat{\mathbf{y}})' \mathbf{v}'\mathbf{x}(\mathbf{x}'\mathbf{v})'$*

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

makes robust to specification of working correlation

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

—

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

...

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: electrocardiogram determined normal (0) or abnormal (1).
- Each patient has two binary observations for period 1 or period 2.

Data:

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

$$y_{ij} \sim \text{BERNOULLI}(p_{ij})$$

Example: 2×2 Crossover Trial

Marginal (GEE) :

$$\text{logit}(P(y_{ij}=1)) = \beta_0 + \beta_1 \text{trt}_{ij} + \beta_2 \text{period}_{ij} + \beta_3 \text{trt}_{ij} \cdot \text{period}_{ij}$$

$$\text{Cov}(y_i) = D_i^{1/2} R_i(\alpha) D_i^{1/2}, \quad \text{correlation} = \text{exchangeable}, \quad \begin{bmatrix} 1 & \alpha \\ \alpha & 1 \end{bmatrix}$$

Conditional (Random intercept model) :

GLM

$$\text{logit}(P(y_{ij}=1 | \theta^*)) = \beta_0^* + \theta_i^* + \beta_1^* \text{trt}_{ij} + \beta_2^* \text{period}_{ij} + \beta_3^* \text{trt}_{ij} \cdot \text{period}_{ij}; \quad \theta_i^* \stackrel{\text{iid}}{\sim} N(0, \tau^2)$$

- Cross
- The marginal approach models the exchangeable correlation **directly on the observed binary outcomes**.

- GLM
- 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:

$$\text{logit } E(y_{ij}) = \beta_0 + \beta_1 \text{trt}_{ij} + \beta_2 \text{period}_{ij} + \beta_3 \text{trt}_{ij} * \text{period}_{ij}$$

Conditional (Random intercept model): GLM

$$\text{logit } E(y_{ij} | \theta_i^*) = \underbrace{\beta_0^* + \theta_i^*}_{\text{conditional formulation results in } \beta\text{'s}} + \beta_1^* \text{trt}_{ij} + \beta_2^* \text{period}_{ij} + \beta_3^* \text{trt}_{ij} * \text{period}_{ij}$$

Note: conditional formulation results in β 's that can have very different magnitudes than GEE or GLM

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

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 = 25 )
```

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	<u>1.1096621</u>	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	<u>0.6401548</u>
[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).

$$V_i(\hat{\alpha}) = \hat{D}_i^k P_i(\hat{\alpha}) \hat{D}_i^k$$
$$\begin{bmatrix} e^{x_{i1}\hat{\beta}} & & & \\ & \ddots & & \\ & & \ddots & \\ & & & e^{x_{i1}\hat{\beta}} \end{bmatrix} \begin{bmatrix} 0 & & & \\ & \ddots & & \\ & & 0 & \\ & & & \ddots & 0 \end{bmatrix} \begin{bmatrix} e^{x_{i2}\hat{\beta}} & & & \\ & \ddots & & \\ & & \ddots & \\ & & & e^{x_{i2}\hat{\beta}} \end{bmatrix}$$

Results Comparison

Approach	Log Odds Treatment Effect $\hat{\beta}_1$ ^{in Period = 0}		
	$\hat{\beta}_1$	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

Marginal approaches
conditional approach

(usually will differ)

- 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.
- Note : GEE was easier to optimize

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.
- \underline{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). **EXPOSURE**
- age_{ij} : in years.

Goal

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

ODDS of

ODDS INFECT/No VIT A DEF.
ODDS INFECT/No Deficiency

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
```

For ALL MODELS:

GLM: $y_{ij} \sim \text{BERN}(P(y_{ij}=1))$ Models

$$\text{logit } P(y_{ij} = 1) = \beta_0 + \beta_1 \text{time}_{ij} + \beta_2 \text{age}_{ij} + \beta_3 \text{sex}_i + \beta_4 \text{vit}_{ij}$$

with and without over-dispersion

with overdispersion: family = quasibinomial

Marginal (GEE): $y_{ij} \sim \text{BERN}(P(y_{ij}=1))$

$$\text{logit } P(y_{ij} = 1) = \beta_0 + \beta_1 \text{time}_{ij} + \beta_2 \text{age}_{ij} + \beta_3 \text{sex}_i + \beta_4 \text{vit}_{ij}$$

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

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

over-dispersion parameter, i.e., scale parameter

Conditional (Random intercept model): GLMN $y_{ij} \sim \text{BERN}(P(y_{ij}=1 | \theta_i^*))$

$$\text{logit } P(y_{ij} = 1 | \theta_i^*) = \beta_0^* + \theta_i^* + \beta_1^* \text{time}_{ij} + \beta_2^* \text{age}_{ij} + \beta_3^* \text{sex}_i + \beta_4^* \text{vit}_{ij}$$

Remind us that interpretation
differs $\theta_i^* \sim N(0, \tau^2)$.

R Code

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

### GLM with over-dispersion ("quasi" option) estimates dispersion parameter
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
```

note: you should be able to write this out if I give the working correlation and α

Estimated Correlation Matrix

Exchangeable: $\downarrow \alpha$

	[,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: $\hat{\alpha}^{[2,1]} \hat{\alpha}^{[3,1]}$ $\downarrow \alpha$

	[,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: 15×5

	[,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

in general,
these are equal

	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 ≈ 1
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
GMM	Random intercept	0.633	0.429	$0.633 / 0.429 \approx 1.47$

- ~~Moderate~~ No over-dispersion.
- Marginal point estimates are similar.
- Without controlling for within-child correlation, you would have a significant effect using GLM!

Interpretations

(means we are using GEE)

- 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.

(means we are using GLMM) in a model with a random intercept for child

- Conditional: After controlling for child-specific baseline log odds of respiratory infection, we estimated an OR of $\exp(0.633) = 1.88$ (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

we only get population average in GEE

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.