Module 4: GEEs

Monday, October 16, 2023 13:33

BIOS526_M4_GEEs_c...

Module 4: Generalized Estimating Equations

BIOS 526

Pivoting to marginal models as opposed to mixed effects
- make inferences about pop. AVERAGES
- no subject effects
- analogous to cross-sectional
  - don't need joint distributional assumptions

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,
WLS,
GLS
- special case
Robust regression
combines robust reg. w/ GLS → GEE

Agenda

2 obs correlated
1 obs indep.

Motivating 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.

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

The above estimator has $\boxed{\text{variance}}$

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

Assigns equal weights
$a = (\frac{1}{3}\ \frac{1}{3}\ \frac{1}{3})^T$
$y = (y_1, y_2, y_3)^T$
$Cov(a'y) = a'[Cov(y)]\,a$
$[\frac{1}{3}\ \frac{1}{3}\ \frac{1}{3}]\begin{bmatrix} \sigma^2 & R & 0 \\ R & \sigma^2 & 0 \\ 0 & 0 & \sigma^2 \end{bmatrix}\begin{bmatrix} 1/3 \\ 1/3 \\ 1/3 \end{bmatrix} = [\ \dots\ ]$

R increases → variance increases.

Motivating Example

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

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

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

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

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

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

best/ smallest

Now see variance is higher

means we have this covariance matrix
$\begin{bmatrix} \sigma^2 & \sigma^2 & 0 \\ \sigma^2 & \sigma^2 & 0 \\ 0 & 0 & \sigma^2 \end{bmatrix}$
aka $y_1 = y_2$

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

best one

best ones

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 $R = \sigma^2$.

So how does this work?

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:

parameter of interest

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

Choose a $\hat{\mu}$ such that we get 0 using this equation

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

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

i.e. $\sum w_i = 1$

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

moving towards weighted least squares

Now consider the linear regression problem with heteroscedastic errors:

$$\mathbf{y} = \mathbf{X}\beta + \epsilon, \qquad \epsilon \sim N(\mathbf{0}, \mathbf{V}).$$ → $\epsilon = \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{bmatrix}$

We start w/ this example

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

Regression with heteroscedastic errors

→ moving towards weighted least squares

Now consider the linear regression problem with heteroscedastic errors:

$$\mathbf{y} = \mathbf{X}\beta + \epsilon, \qquad \epsilon \sim N(\mathbf{0}, \mathbf{V}). \rightarrow \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

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

obs indep, can have diff variances

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

DEFINE $W = \begin{bmatrix} v_1^{-1/2} & 0 & \cdots & 0 \\ 0 & \ddots & & 0 \\ 0 & \cdots & 0 & v_n^{-1/2} \end{bmatrix}$

we start w/ this example bc easy to invert v matrix

What is the least square estimate of β?

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 $\mathbf{y}$:

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

$cov(y^*) = cov(Wy)$

$W\,cov(y)\,W = WVW = (WV^{1/2})(V^{1/2}W) = I \cdot I = I$

Just saying diagonals are equal to 1

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

$= E[WX\beta + W\varepsilon] = WX\beta + 0$

Weighted Least Squares

In this class, WLS means V is diagonal with diff. variances

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

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

→ homoskedastic → therefore

$y^* = (x^{*\prime}x^*)^{-1}x^{*\prime}y^*$

We have $\quad Wy = WX\beta + W\varepsilon$

$$\begin{aligned} \hat\beta_{wls} &= (\mathbf{X}^{*\prime}\mathbf{X}^*)^{-1}\mathbf{X}^{*\prime}\mathbf{y}^* \\ &= [(\mathbf{WX})'(\mathbf{WX})]^{-1}(\mathbf{WX})'(\mathbf{Wy}) \\ &= [\mathbf{X}'\mathbf{W}'\mathbf{WX}]^{-1}\mathbf{X}'\mathbf{W}'\mathbf{Wy} \\ &= (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}. \end{aligned}$$

OLS: $V = \sigma^2 I$

$(X'[\sigma^2 I]^{-1}X)^{-1}X'[\sigma^2 I]^{-1}Y$

$\sigma^2(X'X)^{-1}X'(\sigma^2)^{-1}Y$

$(X'X)^{-1}X'Y$ ✓

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

Weighted Least Squares, cont.

$= Cov((X'V^{-1}X)^{-1}X'V^{-1}Y)$ ↗ same as y?

$$\begin{aligned} Cov(\hat\beta_{wls}) &= (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}Cov(\mathbf{y})[(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}]' \\ &= (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\cancel{\mathbf{V}^{-1}}\cancel{\mathbf{V}}[\mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}] \\ &= (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1} \\ &= (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1} \end{aligned}$$

e.g. OLS:

$[X'(\sigma^2 I)^{-1}X]^{-1} = \sigma^2(X'X)^{-1}$

Weighted Least Squares

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

$$\begin{aligned} E(\hat\beta_{wls}) &= E[\overbrace{(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}}^{\text{all this is fixed}}\mathbf{y}] \\ &= (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}E(\mathbf{y}) \\ &= (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\beta = \beta. \end{aligned}$$

except for y

but then $E(Y) = E(X\beta + \varepsilon) = X\beta$

What if we used the regular ordinary least squares estimate?

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

✓ Still unbiased!

Note on OLS estimator

The OLS estimator of β is unbiased. It is also consistent

→ true generating model i.e. if $y = X\beta + \varepsilon$, $\varepsilon \sim N(0, V)$

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

$\hat\beta_{OLS}$ will converge to the true β

So what is the problem? (why r we worrying)

① Inference: p-values can be wrong

② Efficiency: can define an estimator w/ lower variance

Note on OLS estimator

The covariance of $\hat\beta_{ols}$ is → $E([\hat\beta_{OLS} - E(\hat\beta_{OLS})][\hat\beta_{OLS} - E(\hat\beta_{OLS})]')$

$$\begin{aligned} Cov(\hat\beta_{ols}) &= Cov[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}] \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'Cov(\mathbf{y})\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{VX}(\mathbf{X}'\mathbf{X})^{-1} \end{aligned}$$

$\neq (X'X)^{-1}\hat\sigma^2_{OLS}$

$\neq$ expected value of this either

However, our estimate of the covariance is

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

→ we want $(X'X)^{-1}\hat\sigma^2_{OLS}$ to converge to $(X'X)^{-1}VX(X'X)^{-1}$ but it doesn't

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

I.e. inference is incorrect

no longer a good estimate of covariance, only works if V has same everywhere

Two issues with OLS when heteroscedasticity

① The OLS estimate of the SE of $\hat\beta_{ols}$ can lead to invalid inference. ↗ when $V \neq \sigma^2 I$

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

→ the R shiny project!!!

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

$$\mathbf{y} = [\hat\beta_1, \hat\beta_2, \dots, \hat\beta_n]'$$
$$\mathbf{X} = [1, 1, \dots, 1]' \quad \text{(i.e., intercept only)}$$
$$\mathbf{V} = \text{diag}(\hat\sigma_1^2, \hat\sigma_2^2, \dots, \hat\sigma_n^2)$$

Therefore

- $\mathbf{V}^{-1} = \text{diag}(1/\hat\sigma_1^2, 1/\hat\sigma_2^2, \dots, 1/\hat\sigma_n^2)$
- $\mathbf{X}'\mathbf{V}^{-1} = [1/\hat\sigma_1^2, 1/\hat\sigma_2^2, \dots, 1/\hat\sigma_n^2]'$
- $\mathbf{X}'\mathbf{V}^{-1}\mathbf{X} = \sum_{i=1}^n 1/\hat\sigma_i^2$
- $\mathbf{X}'\mathbf{V}^{-1}\mathbf{y} = \sum_{i=1}^n \hat\beta_i/\hat\sigma_i^2$

all the weights sum to 1

inverse variance weight

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

... the inverse-variance weighted average!

Generalized Least Squares

correlations between obs & heteroskedasticity

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:

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

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

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

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

↳ exists for a symmetric positive definite matrix (matrices of positive integers?)

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

↳ eigenvalue value decomposition $U \Lambda U'$ e.g. $U \Lambda^{-1/2} U' = W$

OLS estimator $\hat\beta$ is unbiased and consistent, but the estimate of the standard errors are incorrect → 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 $\mathbf{V}$.

So when V has correlations but variance is constant:

Generalized Least Squares: homoscedastic

For normal (Gaussian) regression model,

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

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

assume common variance

$$\mathbf{V} = \sigma^2\mathbf{R} \qquad R = \begin{bmatrix} 1 & \rho_{12} & \cdots & \rho_{1n} \\ \rho_{12} & 1 & & \vdots \\ \vdots & & \ddots & \\ \rho_{1n} & \cdots & & 1 \end{bmatrix}$$

where $\mathbf{R}$ is a correlation matrix describing pair-wise dependence in the observations not explained by $\mathbf{X}\beta$, and σ^2 is the common residual variance. $\mathbf{R} \neq \mathbf{I}$ is called anisotropy.

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

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

Later on, this is called the "naive" SE in gee::gee. → R package GEE "robust" allows departures from homoscedasticity assumption

Generalized Least Squares

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

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

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

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

Here, V matrix

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

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

Generalized Least Squares

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

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

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

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

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

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

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

(back to slide 5)

Robust Regression

patches OLS estimator to fix inference issue when data not IID

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

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

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

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

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:

$\hat\epsilon$

$$\hat{\mathbf{V}} = (\mathbf{y} - \mathbf{X}\hat\beta_{ols})(\mathbf{y} - \mathbf{X}\hat\beta_{ols})'$$

outer product of residuals

$$= \begin{bmatrix} (y_1 - x_1'\hat\beta_{ols})^2 & (y_1 - x_1'\hat\beta_{ols})(y_2 - x_2'\hat\beta_{ols}) & \cdots \\ (y_2 - x_2'\hat\beta_{ols})(y_1 - x_1'\hat\beta_{ols}) & (y_2 - x_2'\hat\beta_{ols})^2 & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix}$$

This gives

$$\boxed{\widehat{Cov}(\hat\beta_{ols})_{robust} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\{(\mathbf{y} - \mathbf{X}\hat\beta_{ols})(\mathbf{y} - \mathbf{X}\hat\beta_{ols})'\}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}.}$$

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

Turns out $\widehat{Cov}(\hat\beta_{OLS})_{ROBUST}$ converges to $Cov(\hat\beta_{OLS})$ for $\varepsilon \sim N(0, V)$

Ideal conditions:

- balanced data
- little or no missingness
- number of obs per subject << number of subjects

↳ performs poorly if lots obs but not many subjects

Robust Standard Error

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

$$\widehat{Cov}(\hat\beta_{ols})_{robust} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\{\underbrace{(\mathbf{y} - \mathbf{X}\hat\beta_{ols})}_{\hat\varepsilon}\underbrace{(\mathbf{y} - \mathbf{X}\hat\beta_{ols})'}_{\hat\varepsilon'}\}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}.$$

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
  (accurate → lower mean squared error)
- 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-parameteric tests, which are generally less powerful when parameteric assumptions hold.

E.g. two-sample t test more powerful than Wilcoxon rank sum when data is normally distributed, but if it is not, t-test ↑ type 1 error

Introducing structure to $\mathbf{V}$

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

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

1. Use $\hat{\beta}_{gls} = (\mathbf{X}'\widehat{\mathbf{V}}^{-1}\mathbf{X})^{-1}\mathbf{X}'\widehat{\mathbf{V}}^{-1}\mathbf{y}$ because it is unbiased and consistent.
↳ GLS estimator for β
2. Use the sandwich estimator for standard errors:

$$\widehat{Cov}(\hat{\beta}_{gls})_{robust} = \underbrace{(\mathbf{X}'\widehat{\mathbf{V}}^{-1}\mathbf{X})^{-1}}_{\text{variance in GLS}}\underbrace{\mathbf{X}'\widehat{\mathbf{V}}^{-1}([\mathbf{y}-\mathbf{X}\hat{\beta}_{gls}][\mathbf{y}-\mathbf{X}\hat{\beta}_{gls}]')\widehat{\mathbf{V}}^{-1}\mathbf{X}}_{\text{provides robustness}}(\mathbf{X}'\widehat{\mathbf{V}}^{-1}\mathbf{X})^{-1}.$$

Idea: even if $\widehat{\mathbf{V}}$ 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 heteroskedasticity

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:

$$\mathbf{y} = \begin{bmatrix} y_{11} \\ y_{12} \\ y_{21} \\ y_{22} \end{bmatrix} \sim N\left(\underset{4\times p}{\mathbf{X}}\underset{p\times 1}{\beta},\ \mathbf{V} = \begin{bmatrix} \mathbf{V}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{V}_2 \end{bmatrix}\right),$$

2x2 covariance matrix

2x2 matrix of 0's

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 β with generalized least squares.
3. robustify the standard errors of $\hat{\beta}$.

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

i.e. mean trend & covariance, no random effects

Marginal Model for Grouped Data

Denote y_{ij} the jth 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 $i = 1, 2, \dots n$.

$$\underset{r_i \times r_i}{\mathbf{V}_i} = \underset{\text{scalar}}{\sigma^2}\underbrace{\mathbf{R}_i(\alpha_i)}_{\text{correlation matrix } r_i\times r_i}$$

vector of parameters (sometimes a scalar)

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).

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

For example,

$$\mathbf{y} = \begin{bmatrix} y_{11} \\ y_{12} \\ y_{21} \\ y_{22} \end{bmatrix} \sim N\left(\mathbf{X}\beta,\ \sigma^2\begin{bmatrix} \mathbf{R}_1(\alpha) & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_2(\alpha) \end{bmatrix}\right).$$

2x2 matrix of 0's

Common Correlation Structures

Example for a cluster of size 4:

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

$\alpha = 0$
0 parameters

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

No within group correlation
No clustering structure
0 parameters

Robust SE = Robust Regression

like "Robust vs. Naive"

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

(same correlation no matter the j indices)

Same correlation structures: implied by the random intercept model

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

1 parameter

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

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

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

6 parameters (α's)

Common Correlation Structures

If the data are measured at distinct time points:

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

AR model

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

-AR-1: visits closer in time are more correlated. only suitable for measurements made at EQUALLY SPACED TIME POINTS

- correlation DECAYS with time
- can have AR(k) correlation structures in gee::gee (R package)

capital eta? Some Greek letter

- Based on AR 1 model, $\varepsilon_{it} = \alpha\varepsilon_{it-1} + r_{it}$, $r_t \overset{iid}{\sim} N(0, \varsigma^2)$ Regression of residuals against a lag?
- $Corr(\varepsilon_{it}, \varepsilon_{it-1}) =$
$Corr(\alpha\varepsilon_{it-1} + r_{it}, \varepsilon_{t-1}) = \alpha$

More working correlation structures

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

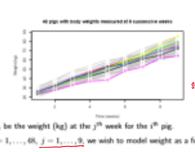
$$\mathbf{R}_i(\alpha) = \begin{bmatrix} 1 & \alpha_1 & \alpha_2 & 0 \\ \alpha_1 & 1 & \alpha_1 & \alpha_2 \\ \alpha_2 & \alpha_1 & 1 & \alpha_1 \\ 0 & \alpha_2 & \alpha_1 & 1 \end{bmatrix}.$$

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

How vit. E & Cu level in feed affect the weights?



Some multivariate normal distributed data

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:

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

working correlation = ...endence

Marginal Model for Pig Weight (Continuous data)

MARGINAL MODEL FEATURES

① $\boxed{g(\mu_{ij}) = X_{ij}\beta}$ = Marginal expectation of Y_{ij} $= \mu_{ij}$

link function; mean

② $\boxed{\phi\, Var(\mu_{ij}) = Var(Y_{ij})}$ - Marginal variance of Y_{ij}

scale parameter

③ Also need WORKING CORRELATION STRUCTURE
aka "within-subject" association (it is a function of the means + other parameters)

use MLE

use quasilikelihood (defined by mean & variance) → Parameter estimates derived by SCORE EQUATIONS of quasi-likelihood] Aka ESTIMATING EQUATIONS

(so AIC/BIC won't work but QIC will)

GLM
- Need to specify data distribution (poisson, etc.)

In both
- linear predictor $X\beta$
- link function $g(\ )$

GEE
- Don't need to specify data distribution
- in GEE, dictates mean structure of data
- Variance
- correlation structure

→ WORKING CORRELATION STRUCTURE

↳ If you choose the wrong one, β will still be consistent but SE may be off. To fix this, use ROBUST estimator for SE's aka SANDWICH ESTIMATOR → How to choose? QIC

INFERENCE:

WALD is most common.
no LRT for GEE

INTERPRETATIONS: Population. For "pigs" NOT "a pig".

- $\beta_0 = 15.098$: The expected weight at time 0 for pigs with no vitamin E and no copper in their diet.
- $\beta_{time} = 6.94$: The expected increase in pig weight for each one week increase in time, controlling for vitamin E and copper in the diet.
- $\beta_{CuCu035} = -0.765$: The expected different in weight comparing pigs who received a 35 mg/kg dose of copper to pigs who received a 0 mg/kg dose of copper.

EXAMPLES:

① Continuous response variable

$\mu_{ij} = \beta X_{ij}$ (linear predictor)

$Var(Y_{ij}) = \sigma_j^2$ (variance)

$Corr(Y_{ij}, Y_{ik}) = \rho^{|j-k|}$, $0 \le \rho \le 1$ (working corr structure = AR aka autoregressive)

② Count response variable - Poisson

$\log(\mu_{ij}) = \beta X_{ij}$ (linear predictor)

$Var(Y_{ij}) = \underset{\text{overdispersion parameter}}{\phi}\,\mu_{ij}$ (variance)

$Corr(Y_{ij}, Y_{ik}) = \rho$ (working corr structure = compound symmetry)

③ Binary response variable - Bernoulli

$logit(\mu_{ij}) = \log\left(\frac{P(Y_{ij}=1)}{P(Y_{ij}=0)}\right) = \beta X_{ij}$ (linear predictor)

$Var(Y_{ij}) = \mu_{ij}(1 - \mu_{ij})$ (variance)

$corr(Y_{ij}, Y_{ik}) = \rho_{ij}$ (working corr structure = unstructured)

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

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

working correlation = independence correlation structure

Marginal Model for Pig Weight (continuous data)

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

$$y_{ij} = \beta_0 + \beta_1 \underbrace{x_{ij}}_{\text{week number}} + \epsilon_{ij} \qquad \epsilon \sim N(0, \mathbf{V}).$$

We assume independence between pigs:

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + \epsilon_{ij} \qquad \underset{9\times1}{\epsilon_i} \sim N(0, \mathbf{V}_i), \quad \mathbf{V}_i = \sigma^2 \mathbf{R}_i(\alpha).$$

common variance

working correlation

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. (no BLUPs, no shrinkage, etc.) → use mixed models if you need subject specific predictions
  - Can be less powerful than mixed models.

2 R packages:
① gee → Both naive & robust
② geepack → only robust
} Both fine to use

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)
```

(scale.fix = FALSE: default)

- 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?

M4_GEE.R

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   0.3998354  48.43280
weeks        6.209896  0.0393321 157.88388   0.09137443 66.10485

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
```

only 1 grouping level

→ correlation structure

Naive S.E.: $(X'\hat{V}^{-1}X)^{-1}$

Robust S.E.: sandwich $(X'V^{-1}X)^{-1} V^{-1} (Y - X\hat{\beta})(Y - X\hat{\beta})' V^{-1} (X (\ldots$

Estimated Scale Parameter → just 4.392^2 aka σ^2 in OLS (residual variance) → here it's like $\sigma^2 + \tau^2$ in LMM

← ICC

→ usually not equal to each other but here it's a special case: x covariate is the same for all piglets

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

↳ in general use robust

- OLS estimate of $\hat{\beta}_1$ is identical to that using GEE with independent $\mathbf{R}_i$ and naive SE.
- Here, GEE with independent $\mathbf{R}_i$ and exchangeable $\mathbf{R}_i$ give the same $\hat{\beta}_1$. Covariate values are the same for each piglet. (This is an uncommon situation.)
- Here, GEE with exchangeable $\mathbf{R}_i$ gives the smaller naive SE compared to independent $\mathbf{R}_i$ → 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
```

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

↳ sum ≈ scale parameter from GEE ($\hat{\sigma}^2 + \hat{\tau}^2$)

→ naive SE from exchangeable

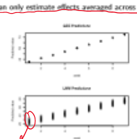
- $\hat{\beta}_1 = 6.210$ with SE $= 0.039$. Here, equal to GEE with exchangeable working correlation matrix and its naive SE.
- The estimated intra-class correlation is $\frac{15.1}{15.1 + 4.39} = 0.77$, same as the estimated exchangeable working correlation.

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.



only get population estimate i.e. marginal mean i.e. averaged across subjects $\hat{\beta}_0 + \hat{\beta}_1 X_{ik}$ (no θ_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$weight)^2)
[1] 8296.737
> sum((predict.lmer - dat$weight)^2)
[1] 1669.636
```

→ $\Sigma (y_{ij} - \hat{y}_{ij})^2$

↳ LMM 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

just GLM stuff

Reference: Halekoh, U., Højsgaard, 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

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

$$Var(y_i) = p_i(1-p_i) = \frac{e^{x_i'\beta}}{(1+e^{x_i'\beta})^2}$$

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

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

First, assuming independent data (y_i indep),

Normal Regression

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

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

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

index i

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

Therefore, the maximum likelihood estimate of β can be obtained by minimizing the function:

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

Take derivative w/ respect to β_k then set to zero

Equivalent to solving a system of linear equations:

$$\frac{\partial U(\beta)}{\partial \beta_k} = \sum_{i=1}^n x_{ik}\frac{1}{v_i}(y_i - \boldsymbol{x}_i'\beta) = 0 \rightarrow \begin{matrix} \sum_{i=1}^n x_{i1}\frac{1}{v_i}(y_i - x_i'\beta) = 0 \\ \vdots \\ \sum_{i=1}^n x_{ip}\frac{1}{v_i}(y - x_i'\beta) = 0 \end{matrix}$$

The above is an *estimating equation* too.

Logistic Regression Model

A set of p equations

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

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

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

The score function is $y_i \sum_k x_{ik}\beta_k$ (take derivative of this)

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

$y_i - E(y_i)$

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

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

This is the general form

Poisson Log-linear Model

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

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

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

The score function with respect to β_k is

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

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

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

Estimating Equations

What do the above three score equations have in common?

Gaussian:

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

Logistic:

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

Poisson:

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

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

$$\sum_{i=1}^n \frac{\partial E(y_i; \beta)}{\partial \beta_k}\frac{1}{V(y_i; \beta)}[y_i - E(y_i; \beta)] = 0. \qquad (2)$$

Estimating Equations

Let's verify it for a single data point:

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

Logistic:

$$\frac{\partial E(y_i;\beta)}{\partial \beta_k}\frac{1}{V(y_i;\beta)} = \left[\frac{\partial}{\partial \beta_k}\frac{e^{x_i'\beta}}{1+e^{x_i'\beta}}\right]\frac{\left(1+e^{x_i'\beta}\right)^2}{e^{x_i'\beta}}$$

$e^{x_i'\beta}(1+e^{x_i'\beta})^{-1}$

$$= \left[\frac{x_{ik}e^{x_i'\beta}}{1+e^{x_i'\beta}} - \frac{x_{ik}\left(e^{x_i'\beta}\right)^2}{\left(1+e^{x_i'\beta}\right)^2}\right]\frac{\left(1+e^{x_i'\beta}\right)^2}{e^{x_i'\beta}}$$

$$= x_{ik}\frac{\left(1+e^{x_i'\beta}\right)e^{x_i'\beta} - \left(e^{x_i'\beta}\right)^2}{\left(1+e^{x_i'\beta}\right)^2}\frac{\left(1+e^{x_i'\beta}\right)^2}{e^{x_i'\beta}} = x_{ik}$$

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

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: $\begin{bmatrix} x_{11} & \cdots & x_{1n} \\ \vdots & & \vdots \\ x_{p1} & \cdots & x_{pn} \end{bmatrix}$

he switches between E() and μ

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

$\begin{bmatrix} \mu(x_1'\beta) \\ \vdots \\ \mu(x_n'\beta) \end{bmatrix}$

becomes $p\times n$ matrix

where

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

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

- consistent.
- asymptotically normal with covariance

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

Grouped Data

Grouped Data

GEE for Grouped Data

for logistic regression

y_i is a repeated measurement for subject i

We now assume the non-Gaussian observations may be correlated. → specifically, exponential family

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.

IND → as in independent SUBJECTS

$$\underbrace{\mathbf{y}_i}_{r_i\times 1} \sim \text{Bernoulli}\,(\underbrace{\mu_i}_{r_i\times 1}) \qquad Cov(\mathbf{y}_i) = \underbrace{\mathbf{V}_i}_{r_i\times r_i}.$$

← marginal covariance between obs on same subject i

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 $= D_i^{-1/2}(Cov)(D_i^{-1/2})$ ← matrix analog of things from earlier?

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

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

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:

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

where $g(\cdot)$ is the link function. (α:) scalar or vector of dependence parameters

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 β is obtained by solving the system of p equations with p unknowns:

sum across subjects

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

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 from Gaussian case: $(X'V^{-1}X)^{-1}X'V^{-1}(y-E\hat{y})(y-E\hat{y})V^{-1}X(X$ ......

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

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

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

This is the part that gives robustness (makes robust to specification of working correlation)

Scale parameter

by default estimated by these GEE models

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 — overdispersion parameter

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

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 → BINARY response var

"logistic GEE model"

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

In both

Interpreting logistic GEE model parameters

Same interpretation as logistic regression, with the added consideration that these effects are at the population level.

- Exponentiated coefficients are odds ratios

coefs are log ORs!

```
coef(summary(resp_ex))
##                 Estimate  Std.err      Wald  Pr(>|W|)
## (Intercept)   -0.53660  0.62271    0.7426  3.888e-01
## baseline       1.99667  0.32731   37.2127  1.059e-09
## treatactive    1.28288  0.35086   13.3695  2.558e-04
## sexmale       -0.27134  0.42263    0.4122  5.209e-01
## age           -0.01371  0.01334    1.0560  3.041e-01
```

β ; −.537

- $\beta_0 =$: The log odds of poor respiratory status for females aged 0 with good respiratory status at baseline who received placebo (not interpretable)
- $\beta_{baseline} = 1.997$: The log odds ratio of poor respiratory status comparing those with poor respiratory status at baseline to those with good respiratory status at baseline, controlling for age, sex, and treatment.
- $\beta_{treat} = 1.283$: The log odds ratio of poor respiratory status comparing those who received active treatment to those who received placebo, controlling for age, sex, and baseline status

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

In both models, also write $y_{ij} \sim \text{Bernoulli}(p_{ij})$

Marginal (GEE):

$$\text{logit } P(y_{ij} = 1) = \beta_0 + \beta_1 trt_{ij} + \beta_2 period_{ij} + \beta_3 trt_{ij} * period_{ij}$$

matrix $r_i \times 1$

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

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

Conditional (Random intercept model): introduce a random effect

GLMM

$$\text{logit } P(y_{ij} = 1 | \theta_i^*) = \beta_0^* + \theta_i^* + \beta_1^* trt_{ij} + \beta_2^* period_{ij} + \beta_3^* trt_{ij} * period_{ij}$$

$$\theta_i^* \overset{iid}{\sim} N(0, \tau^2).$$

- 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

Compare GLM to GLMER?

OR is comparable → you get very diff. #'s?

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): GLMM

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

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

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)
```

always remember to specify this

this should be 25, not 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 -3.424266
trt          1.1096621  0.5826118  1.904634   0.5738602  1.933714
period       0.8472979  0.5809048  1.433901   0.5820177  1.455794
trt:period  -1.0226607  0.9997117 -1.022946   0.9789063 -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
```

logit link fcn

this is our $R_i(\alpha)$

$\hat{\alpha} = corr(y_{i1}, y_{i2})$

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

If he asks for the formula,

$V_i(\hat{\alpha}) = \hat{D}_i^{1/2} R_i(\hat{\alpha}) \hat{D}_i^{1/2}$

aka

$$\begin{bmatrix} \left[\frac{e^{x_{i1}'\hat{\beta}}}{(1+e^{x_{i1}'\hat{\beta}})^2}\right]^{1/2} & 0 \\ 0 & \left[\frac{e^{x_{i2}'\hat{\beta}}}{(1+e^{x_{i2}'\hat{\beta}})^2}\right]^{1/2} \end{bmatrix}$$

Results Comparison

and period = 0 (period effect has two levels, 0 and 1)

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

marginal approaches

conditional approach (GLMER)

log odds

- The three marginal approaches give the same point estimates. (usually will differ) In this analysis, accounting for between-subject correlation gives very similar standard errors compared to standard glm. → usually not the case
- The conditional effect is larger than the marginal estimates.

Note: GEE was easier to optimize than mixed model

Example: Vitamin A and Respiratory Infection

→ Another binary example

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

Variables vitamin A deficiency

For child i at time j,

- $response_{ij}$: indicator that the child was suffering from a respiratory infection. → response var, binary
- $time_{ij}$: 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 odds of respiratory infection.

$\left(\dfrac{\text{odds of infection given vit A deficiency}}{\text{odds of infection given no deficiency}} \right)$

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   :0.55   Mean   :0.365   Mean   :3.7
 3rd Qu.:188   3rd Qu.:1.000   3rd Qu.:12.0   3rd Qu.:1.00   3rd Qu.:1.000   3rd Qu.:6.0
 Max.   :250   Max.   :1.000   Max.   :15.0   Max.   :1.00   Max.   :1.000   Max.   :7.0
```

uncommonly clean

Models for vit A def.

$y_{ij} \sim \text{Bernoulli}(P(y_{ij}=1))$ FOR ALL 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

→ with overdispersion: "family = quasibinomial"
without: "family = binomial"

Marginal (GEE):

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

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

overdispersion parameter, scale parameter

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

make sure to specify

$y_{ij} \sim$ Bernoulli ... $| \theta_{ij}$ ...

Conditional (Random intercept model):

GLMM

interpretation differs

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

- $\beta_{baseline}$ = 1.997: The log odds ratio of poor respiratory status comparing those with poor respiratory status at baseline to those with good respiratory status at baseline, controlling for age, sex, and treatment.
- β_{treat} = 1.283: The log odds ratio of poor respiratory status comparing those who received active treatment to those who received placebo, controlling for age, sex, and baseline status

Interpreting logistic GEE model parameters

Same interpretation as logistic regression, with the added consideration that these effects are at the population level.

- Exponentiated coefficients are log-odds ratios
- $e^{\beta_{baseline}}$ = 7.36: The odds of poor respiratory status for those with poor respiratory status at baseline are 7.36 times the odds of poor respiratory status for those with good respiratory status at baseline, controlling for age, sex, and treatment.
- $e^{\beta_{treat}}$ = 3.61: The odds of poor respiratory status for those who received active treatment are 3.61 times the odds of poor respiratory status for those who received placebo, controlling for age, sex, and baseline status

```
broom::tidy(resp_ex, exp = TRUE, conf.int = TRUE) %>%
knitr::kable(digits = 2)
```

family = binomial

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

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

overdispersion parameter, scale parameter

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

make sure to specify

$y_{ij} \sim$ Bernoulli ... $| \theta_{ij}$ ...

Conditional (Random intercept model):

GLMM

interpretation differs

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

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

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
```

estimate dispersion parameter

Note: for exams, we should be able to write this out if he gives us the corr and $\hat{\alpha}$

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:

$\hat{\alpha}^{|2-1|}$ $\hat{\alpha}^{|3-1|}$ $\hat{\alpha}^{|4-1|}$

```
     [,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:

15 α's

```
     [,1] [,2] [,3] [,4] [,5] [,6]
[1,] 1.00 0.56 0.48 0.42 0.50 0.45
[2,] 0.56 1.00 0.58 0.56 0.53 0.56
[3,] 0.48 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, this is true. These are usually equal

scale SE by dispersion but since no overdispersion kinda the same as robust

Log Odds Vitamin Effect (β_4)

Approach	$\hat{\beta}_4$	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		

Note: $z \approx 2.45 > 1.96$

≈ 1

dispersion param slightly = 1

$\rightarrow 0.633 / 0.429 = 1.47$ ... bleh

we start to see differences. So working corr structure is important

$\frac{0.278}{0.223} \approx 1.26 < 1.96$, fail to reject

- No ~~Some~~ 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

he just chose arbitrarily..

- 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.

$e^{0.278 \pm 1.96 (0.223)}$

means we are using GLMM (in a model w/ 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

GEE can't provide this, we only get pop. avg

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).

relaxes assumption about homoscedasticity i.e. if you have hetero, use GEE

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.

Poisson GEE Models:

Coefficients are log rate ratios
Exponentiated coefficients are rate ratios

Same interpretation as Poisson Regression but specify they're at pop level