GEEs Summary
Saturday, November 25, 2023   17:14

MARGINAL MODEL FEATURES

link function

$$g(\mu_{ij}) = X_{ij}\,\beta \quad ①$$

mean

- Marginal *expectation* of Y_{ij}
- μ_{ij}

$$\phi\, Var(\mu_{ij}) = Var(Y_{ij}) \quad ②$$

scale parameter

- Marginal *variance* of Y_{ij}

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

Use GEE when:
① Observations in a cluster may be correlated
② Observations in separate clusters are indep.
③ A monotone transformation of the expectation is linearly related to explanatory vars
④ variance is a function of expectation (doesn't have to hold if using robust)

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
- can be less powerful than mixed model

use MLE / use quasi-likelihood (defined by mean & variance) → Parameter estimates derived by SCORE EQUATIONS of quasi-likelihood ⟹ aka ESTIMATING EQUATIONS

(no AIC/BIC w/ quasi-likelihood but QIC will)

INFERENCE:
WALD is most common.
no LRT for GEE

INTERPRETATIONS: Population. For "pigs" NOT "a pig".
- aka no subject-specific predictions (no random effects)

WORKING CORRELATION STRUCTURE

↳ if you choose the wrong one,
β will still be consistent but → How to choose?
SE may be off. To fix this,
use ROBUST estimator for SE's   QIC
aka SANDWICH ESTIMATOR   (even if $\hat V$ is misspecified, SE is valid)

Essentially, add some structure on V in robust regression model

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

$$Cov(\hat\beta_{GLS}) = (X'\hat V^{-1} X)^{-1} X'\hat V^{-1}\left([\,y - X\hat\beta_{GLS}\,][\,y - X\hat\beta_{GLS}\,]'\right)\hat V^{-1} X (X'\hat V^{-1} X)^{-1}$$

variance in GLS / provides robustness by allowing heteroskedasticity

- $\beta_0 = 15.098$: 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.

Common Correlation Structures
Example using cluster of the Y_i.

INDEPENDENT
$cor(y_{ij}, y_{ij'}) = 0$ ← ICC
$n = 0$ aka 0 parameters
Robust SE = robust regression
- no within-group correlation
- no clustering structure

$$R_i(\alpha) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

EXCHANGEABLE
$cor(y_{ij}, y_{ij'}) = \alpha$ ← ICC = τ^2
$n = 1$ aka 1 parameter
- same correlation structure implied by random intercept model

$$R_i(\alpha) = \begin{pmatrix} 1 & \alpha & \alpha & \alpha \\ \alpha & 1 & \alpha & \alpha \\ \alpha & \alpha & 1 & \alpha \\ \alpha & \alpha & \alpha & 1 \end{pmatrix}$$

UNSTRUCTURED
$cor(y_{ij}, y_{ij'}) = \alpha_{jj'}$ ← ICC
$n = $ all lists of parameters
- can cause convergence issues

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

AUTO REGRESSIVE ORDER 1 (AR-1)
$cor(y_{ij}, y_{ij'}) = \alpha^{|j-j'|}$ ← ICC
for $\alpha \leq 1$
- correlation DECAYS w/ time
- visits closer in time are more correlated
- only for equally spaced time pts.

$$R_i(\alpha) = \begin{pmatrix} 1 & \alpha & \alpha^2 & \alpha^3 \\ \alpha & 1 & \alpha & \alpha^2 \\ \alpha^2 & \alpha & 1 & \alpha \\ \alpha^3 & \alpha^2 & \alpha & 1 \end{pmatrix}$$

Based on AR-1,
$\varepsilon_t = \alpha\varepsilon_{t-1} + v_t$
$v_t \sim N(0, \sigma^2)$
Then, $Corr(\varepsilon_t, \varepsilon_{t-s}) = \alpha^s$

STATIONARY M-DEPENDENT
$cor(y_{ij}, y_{ij'}) = \alpha_{|j-j'|}$
if $|j-j'| \leq m$
and 0 otherwise.   ← ICC
for m=2,

$$R_i(\alpha) = \begin{pmatrix} 1 & \alpha_1 & \alpha_2 & 0 \\ \alpha_1 & 1 & \alpha_1 & \alpha_2 \\ \alpha_2 & \alpha_1 & 1 & \alpha_1 \\ 0 & \alpha_2 & \alpha_1 & 1 \end{pmatrix}$$

NON-STATIONARY M-DEPENDENT
$cor(y_{ij}, y_{ij'}) = \alpha_{jj'}$
if $|j-j'| \leq m$
and 0 otherwise.   ← ICC
for m=2,

$$R_i(\alpha) = \begin{pmatrix} 1 & \alpha_{12} & \alpha_{13} & 0 \\ \alpha_{12} & 1 & \alpha_{23} & \alpha_{24} \\ \alpha_{13} & \alpha_{23} & 1 & \alpha_{34} \\ 0 & \alpha_{24} & \alpha_{34} & 1 \end{pmatrix}$$

Working correlation structure:

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

$$\mathbf{V}_i = \sigma^2 \mathbf{R}_i(\alpha_i)$$

σ^2 → scalar — vector of parameters (sometimes a scalar) / correlation matrix of size $r_i \times r_i$

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

$\mathbf{R}_i(\alpha)$ 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.

e.g. $y = \begin{pmatrix} y_1 \\ y_{12} \\ y_{21} \\ y_{22} \end{pmatrix} \sim N\left(X\beta,\; \sigma^2\begin{pmatrix} R_1(\alpha) & 0 \\ 0 & R_2(\alpha) \end{pmatrix}\right)$

EXAMPLES:

① Continuous response variable
$\mu_{ij} = \beta X_{ij}$ (linear predictor) ①
$Var(Y_{ij}) = \sigma^2_j$ (variance) ②
$Corr(Y_{ij}, Y_{ik}) = \rho^{|j-k|}$ (working corr structure
$0 \leq \rho \leq 1$ = AR-1 aka autoregression) ③

② Count response variable — Poisson
$\log(\lambda_{ij}) = \beta X_{ij}$ (linear predictor) ①
$Var(Y_{ij}) = \phi\lambda_{ij}$ (variance) ②
ϕ → overdispersion parameter
$Corr(Y_{ij}, Y_{ik}) = \rho$ (working corr structure
= compound symmetry) ③

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) ①
$Var(Y_{ij}) = \mu_{ij}(1-\mu_{ij})$ (variance) ②
$Corr(Y_{ij}, Y_{ik}) = \rho_{ij}$ (working corr structure
= unstructured) ③

Writing out the model for binary response var.

$Y_{ij} \stackrel{iid}{\sim} Bernoulli(P_{ij})$

$logit(P(y_{ij}=1)) = \beta_0 + \beta_1 x_{1ij} + \beta_2 \ldots$

$Cov(y_i) = V_i = \phi\, D_i^{1/2} R_i(\alpha) D_i^{1/2}$
ϕ → scale parameter, $R_i(\alpha)$ → working corr structure, $D_i^{1/2}$ → ($\phi=1$ means no overdispersion)

$D_{i,[jj']} = \frac{x_{ij}'\beta}{(1 + e^{x_{ij}'\beta})^2}$ = diagonal matrix of marginal variance

Example:
$V_i(\alpha) = \hat D_i^{1/2} R_i(\alpha) \hat D_i^{1/2}$
$= \begin{pmatrix} \left[\frac{e^{x_{i1}'\beta}}{(1+e^{x_{i1}'\beta})^2}\right]^{1/2} & 0 \\ 0 & \left[\frac{e^{x_{i2}'\beta}}{(1+e^{x_{i2}'\beta})^2}\right]^{1/2} \end{pmatrix}$

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

$Corr(y_{i1}, y_{i4})$ Can be found as this in R:
`alpha.1:8`

3.4 Write out a gee model with outcome `health`, predictors `treatment`, `agegroup`, `treatment*agegroup`, and `time_months`, and exchangeable working correlation. Note: you can adapt this latex code

$$R_i = \begin{pmatrix} 1 & a & b & c \\ d & 1 & e & f \\ g & h & 1 & i \\ j & k & l & 1 \end{pmatrix}$$

If p_{ij} is the probability of good health for subject i and time j, then $y_{ij} \sim$ Bernoulli(p_{ij}) where y_{ij} is the health outcome.

Then the model can be written as:

$$logit(P(y_{ij}=1)) = \beta_0 + \beta_1 Treatment + \beta_2 AgeGroup25to34_{ij} + \beta_3 AgeGroup35plus_{ij} + \beta_4 TimeMonths_{ij} + \beta_5 Treatment*AgeGroup25to34_{ij} + \beta_6 Treatment*AgeGroup35plus_{ij}$$

where we are using an exchangeable working correlation.

$$Cov(y_i) = \phi D_i^{1/2} R_i(\alpha) D_i^{1/2}$$

where ϕ is the scale parameter and $D_{i,[jj']} = e^{x_i'\beta}/(1+e^{x_i'\beta})^2$

β_0 is baseline expected log odds of good health when all other covariates are zero.

β_1 is the log odds ratio associated with treatment while controlling for other covariates. β_2 is the log odds ratio associated with age group 25-34 while controlling for other covariates. β_3 is the log odds ratio associated with age group 34+ while controlling for other covariates. β_4 is the log odds ratio associated with time_months while controlling for other covariates. β_5 is the log odds ratio associated with the interaction of age group 25-34 and treatment while controlling for other covariates. β_6 is the log odds ratio associated with the interaction of age-group 34+ and treatment while controlling for other covariates.

$$R_i(\alpha) = \begin{pmatrix} 1 & \alpha & \alpha & \alpha \\ \alpha & 1 & \alpha & \alpha \\ \alpha & \alpha & 1 & \alpha \\ \alpha & \alpha & \alpha & 1 \end{pmatrix}$$

3.10 Does age group modify the effect of treatment? State the null hypothesis, conduct a statistical test, and state your conclusion. Use the AR1 working correlation structure.

H0: Age group does not modify the effect of treatment. $\beta_5 = \beta_6 = 0$
H1: Age group does modify the effect of treatment.
Test statistic: chi-squared test statistic = 2.06
Alpha = 0.05
P-value = 0.38
Decision: Fail to reject null hypothesis
Conclusion: Age group does not modify the effect of treatment

```
q3_10 <- geeglm(health ~ treatment + agegroup + time_months,
            id = id, corstr = 'ar1', family = "binomial", data = health)

anova(q3_10b, q3_7)

## Analysis of 'Wald Statistic' Table
##
## Model 1 health ~ treatment + agegroup + treatment:agegroup + time_months
## Model 2 health ~ treatment + agegroup + time_months
##    Df   X2 P(>|Chi|)
## 1   2 2.06      0.36
```

Interpretations

GEES • **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.

Mixed Models • **Conditional:** After controlling for child-specific log odds of respiratory infection, we estimated an OR of exp(0.633) = 1.88 (CI95%: 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.

3.9 What is the odds ratio and 95% confidence interval for treatment+intervention in age group 25-34 versus Control in age group 25-34? How does this compare to the GLMM?

GLMM: 1.72 (0.378, 7.76) GEE: 1.38 (0.519, 3.69)

The 95% CI for the GEE is narrower than the 95% CI for GLMM. The OR estimate for GEE is also slightly smaller than GLMM.

```
control <- 1.3267
intervention <- -1.0010

q3_9 <- exp(1.3267 + -1.0010)
q3_9

## [1] 1.38

q3_9var <- vcov(q3_7)
q3_9var2 <- q3_9var["treatmentintervention", "treatmentintervention"] +
            q3_9var["treatmentintervention:agegroup25-34", "treatmentintervention:agegroup25-34"] +
            2 * q3_9var["treatmentintervention", "treatmentintervention:agegroup25-34"]

q3_9 * exp(-1.96 * sqrt(q3_9var2))

## [1] 0.519

q3_9 * exp(1.96 * sqrt(q3_9var2))

## [1] 3.69
```