Chapter 12: Population-averaged models for Bernoulli repeated measurements

Example of repeated measures:

- Data are comprised of several repeated measurements on the same individual over time, e.g. $Y_{ij} = 1$ indicates acne outbreak for patient i in month j; $Y_{ij} = 0$ indicates no outbreak.
- Data are recorded in clusters, e.g. Y_{ij} might indicate the presence of tooth decay for tooth j in patient i.
- Data are from naturally associated groups, e.g. Y_{ij} might denote a successful treatment of patient j at clinic i.
- Wheezing data in Chapter 1.

In all of these examples, the repeated measurements are (typically positively) correlated within an individual or group.

Marginal logistic model of multiple 0/1 responses

Let n_i binary responses $\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{in_i})$ come from the i^{th} individual at times $\mathbf{t}_i = (t_{i1}, \ldots, t_{in_i})$. Let $\boldsymbol{\pi}_i = (\pi_{i1}, \ldots, \pi_{in_i})$ where $\pi_{ij} = E(Y_{ij})$. Let $\mathbf{x}_{ij}$ be a $p \times 1$ vector of explanatory variables.

We assume the vectors $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$ are independent, but that elements of $\mathbf{Y}_i$ are correlated. Common choices are

$$\mathbf{R}(\alpha) = \text{corr}(\mathbf{Y}_i) = \begin{bmatrix} 1 & \alpha & \alpha & \cdots & \alpha \\ \alpha & 1 & \alpha & \cdots & \alpha \\ \alpha & \alpha & 1 & \cdots & \alpha \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \alpha & \alpha & \alpha & \cdots & 1 \end{bmatrix}_{n_i \times n_i} \quad \text{exchangeable,}$$

$$\text{and } \mathbf{R}(\alpha) = \text{corr}(\mathbf{Y}_i) = \begin{bmatrix} 1 & \alpha & \alpha^2 & \cdots & \alpha^{n_i-1} \\ \alpha & 1 & \alpha & \cdots & \alpha^{n_i-2} \\ \alpha^2 & \alpha & 1 & \cdots & \alpha^{n_i-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \alpha^{n_i-1} & \alpha^{n_i-2} & \alpha^{n_i-3} & \cdots & 1 \end{bmatrix}_{n_i \times n_i} \quad \text{AR(1).}$$

Others are

$$\mathbf{R}(\boldsymbol{\alpha}) = \text{corr}(\mathbf{Y}_i) = \begin{bmatrix} 1 & \alpha_{12} & \alpha_{13} & \cdots & \alpha_{1n} \\ \alpha_{12} & 1 & \alpha_{23} & \cdots & \alpha_{2n} \\ \alpha_{13} & \alpha_{23} & 1 & \cdots & \alpha_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \alpha_{1n} & \alpha_{2n} & \alpha_{3n} & \cdots & 1 \end{bmatrix}_{n \times n} \quad \text{unstructured,}$$

$$\text{and } \mathbf{R} = \text{corr}(\mathbf{Y}_i) = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}_{n_i \times n_i} \quad \text{independence.}$$

You can also specify a fixed, known $\mathbf{R}$ as well as MDEP(m) which yields $\mathbf{R}(\boldsymbol{\alpha})$ as

$$\text{corr}(Y_{ij}, Y_{i,j+t}) = \left\{ \begin{array}{ll} 1 & t = 0 \\ \alpha_t & t = 1, \ldots, m \\ 0 & t > m \end{array} \right\}.$$

- Unstructured most general; often a default choice. Need balance though.
- Exchangeable useful when time is not important and correlations thought to be approximately equal, e.g. repeated measurements on individual in crossover study, measurements across several individuals from clinic i.
- AR(1) useful when serial correlation plausible, e.g. repeated measurements across equally spaced time points on individual.

Comments:

- These correlation matrices are used in a GEE algorithm (sketched below) in PROC GENMOD (all other PROC MIXED covariance structures available in GLIMMIX).
- Repeated measures are accounted for via REPEATED statement.
- The order of $(Y_{i1}, \ldots, Y_{in})$ makes a difference with some $\mathbf{R}(\boldsymbol{\alpha})$. If ordering is different to that defined in the DATA step, one can use the WITHIN subcommand in the REPEATED statement to tell SAS what the ordering is. Also used when missing some measurements in $(Y_{i1}, \ldots, Y_{in})$.
- CORRW in the REPEATED statement gives the final working correlation matrix estimate.
- Elements of $\boldsymbol{\beta}$ are interpreted as usual, but *averaged over clusters*. This is a *marginal* interpretation.

Let $\pi_{ij} = g^{-1}(\mathbf{x}'_{ij}\boldsymbol{\beta})$ be the *marginal* mean. In general, Y_{ij} is from an exponential family

$$Y_{ij} \sim f(y_{ij}; \theta_{ij}, \phi) = \exp\{[y_{ij}\theta_{ij} - b(\theta_{ij})]/\phi + c(y_{ij}, \phi)\},$$

where the dispersion ϕ is known. The GEE approach requires some notation:

- $\pi_{ij} = b'(\theta_{ij})$ and $v(\pi_{ij}) = \text{var}(Y_{ij}) = b''(\theta_{ij})\phi$.
- $\mathbf{R}(\boldsymbol{\alpha})$ is "working correlation matrix," reflecting our best guess at the true correlation structure among the elements of $\mathbf{Y}_i$. See the previous slide. Choice of $\mathbf{R}(\boldsymbol{\alpha})$ can be made based on QIC (Pan, 2001).
- $\mathbf{B}_i = \text{diag}(b''(\theta_{i1}), \ldots, b''(\theta_{in}))$ is a diagonal matrix with $\text{var}(Y_{ij})/\phi$ along the diagonal.
- $\mathbf{V}_i = \mathbf{B}_i^{1/2}\mathbf{R}(\boldsymbol{\alpha})\mathbf{B}_i^{1/2}\phi$ is the working covariance matrix.

Let $\mathbf{D}_i = \frac{\partial \boldsymbol{\pi}_i}{\partial \boldsymbol{\beta}} = \mathbf{B}_i \boldsymbol{\Delta}_i \mathbf{X}_i$ be the $n_i \times p$ matrix of first partial derivatives where $\boldsymbol{\pi}_i = \boldsymbol{\pi}_i(\boldsymbol{\beta}) = (g^{-1}(\mathbf{x}'_{i1}\boldsymbol{\beta}), \ldots, g^{-1}(\mathbf{x}'_{in_i}\boldsymbol{\beta}))$,

$\boldsymbol{\Delta}_i = \text{diag}(\frac{\partial \theta_{i1}}{\partial \eta_{in_i}}, \ldots, \frac{\partial \theta_{i1}}{\partial \eta_{in_i}})$, $\eta_{ij} = \mathbf{x}'_{ij}\boldsymbol{\beta}$, and $\mathbf{X}_i = \begin{bmatrix} \mathbf{x}'_{i1} \\ \vdots \\ \mathbf{x}'_{in_i} \end{bmatrix}$.

The generalized estimating equations (GEE) are

$$\mathbf{u}(\boldsymbol{\beta}) = \sum_{i=1}^{n} \mathbf{D}'_i \mathbf{V}_i^{-1} [\mathbf{y}_i - \boldsymbol{\pi}_i(\boldsymbol{\beta})] = \mathbf{0}.$$

These correspond to likelihood (score) equations, but are *not* derived from a proper likelihood. However, the $\hat{\boldsymbol{\beta}}$ that solves them *is consistent*, even when the correlation assumption is *wrong*. Roughly speaking, this is because consistency is a first moment (mean) property.

Liang and Zeger (1986) show $\hat{\boldsymbol{\beta}} \overset{\bullet}{\sim} N_p(\mathbf{0}, \mathbf{V}_G)$ where

$$\mathbf{V}_G = \left[\sum_{i=1}^{n} \mathbf{D}_i' \mathbf{V}_i^{-1} \mathbf{D}_i\right]^{-1} \left[\sum_{i=1}^{n} \mathbf{D}_i' \mathbf{V}_i^{-1} \text{cov}(\mathbf{Y}_i) \mathbf{V}_i^{-1} \mathbf{D}_i\right] \left[\sum_{i=1}^{n} \mathbf{D}_i' \mathbf{V}_i^{-1} \mathbf{D}_i\right]^{-1}.$$

Here $\boldsymbol{\beta}$ is replaced by $\hat{\boldsymbol{\beta}}$, ϕ replaced with $\hat{\phi}$ ($\phi = 1$ for binomial and Poisson models), and $\boldsymbol{\alpha}$ replaced by $\hat{\boldsymbol{\alpha}}$. $\text{cov}(\mathbf{Y}_i)$ is replaced by $[\mathbf{y}_i - \boldsymbol{\pi}_i(\hat{\boldsymbol{\beta}})][\mathbf{y}_i - \boldsymbol{\pi}_i(\hat{\boldsymbol{\beta}})]'$.

This *sandwich estimator* sandwiches an empirical estimate between the theoretical (working guess) $\left[\sum_{i=1}^{n} \mathbf{D}_i' \mathbf{V}_i^{-1} \mathbf{D}_i\right]^{-1}$. If we know for certain (we don't) that $\text{corr}(\mathbf{Y}_i) = \mathbf{R}(\boldsymbol{\alpha})$, then we can use this instead (MODELSE in the REPEATED statement).

To reiterate, the ingredients for the marginal GEE approach are

- A marginal model where Y_{ij} is Bernoulli, Poisson, normal, gamma, etc. (see Chapters 11 & 12) with mean $\mu_{ij} = g^{-1}(\mathbf{x}'_{ij}\boldsymbol{\beta})$.
- We are only considering Bernoulli data, so
  $\mu_{ij} = E(Y_{ij}) = \pi_{ij} = g^{-1}(\mathbf{x}'_{ij}\boldsymbol{\beta})$.
  Note that often for repeated measures, $\mathbf{x}_{ij} = \mathbf{x}_i$ for $j = 1, \ldots, n_i$; e.g. gender and weight are not apt to change over a 6 month study.
- An assumption on how the elements of $\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{in_i})$ are correlated, $\text{corr}(\mathbf{Y}_i) = \mathbf{R}(\boldsymbol{\alpha})$.
- This is *not* the case for the wheezing data!

Example (Agresti, 2002, p. 459)

		Response at 1, 2, and 4 weeks							
Diagnosis	Treatment	NNN	NNA	NAN	NAA	ANN	ANA	AAN	AAA
Mild	Standard	16	13	9	3	14	4	15	6
	New	31	0	6	0	22	2	9	0
Severe	Standard	2	2	8	9	9	15	27	28
	New	7	2	5	2	31	5	32	6

Longitudinal study comparing a new drug with a standard drug for treatment of subjects suffering mental depression. $n = 340$ patients either mildly or severely depressed upon admission into the study. At weeks 1, 2, and 4, corresponding to $j = 1, 2, 3$, patient i's suffering Y_{ij} was classified as normal $Y_{ij} = 1$ or abnormal $Y_{ij} = 0$. Let $s_i = 0, 1$ be the severity of the diagnosis (mild, severe) and $d_i = 0, 1$ denote the drug (standard, new).

We treat time as a categorical predictor and fit a marginal logit model with an exchangeable correlation structure; note $n = 3$:

$$\text{corr}(\mathbf{Y}_i) = \text{corr}\left(\begin{bmatrix} Y_{i1} \\ Y_{i2} \\ Y_{i3} \end{bmatrix}\right) = \begin{bmatrix} 1 & \alpha & \alpha \\ \alpha & 1 & \alpha \\ \alpha & \alpha & 1 \end{bmatrix}.$$

The model is:

$$\text{logit}(\pi_{ij}) = \beta_0 + \beta_1 s_i + \beta_2 d_i + \tau_j + d_i \theta_j.$$

```
data depress;
 infile "c:/tim/cat/depress.txt";
 input case diagnose treat time outcome; time=time+1;
proc genmod descending;  class case time;
  model outcome = diagnose treat time treat*time / dist=bin link=logit type3;
  repeated subject=case / type=exch corrw;
```

Fit of independence model to get initial estimate of $\boldsymbol{\beta}$:

Analysis Of Initial Parameter Estimates

Parameter		DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept		1	0.9812	0.1809	0.6267	1.3356	29.43	<.0001
diagnose		1	-1.3116	0.1462	-1.5981	-1.0251	80.50	<.0001
treat		1	2.0429	0.3056	1.4439	2.6420	44.68	<.0001
time	1	1	-0.9600	0.2290	-1.4088	-0.5112	17.58	<.0001
time	2	1	-0.6206	0.2245	-1.0607	-0.1806	7.64	0.0057
time	3	0	0.0000	0.0000	0.0000	0.0000	.	.
treat*time	1	1	-2.0980	0.3893	-2.8610	-1.3351	29.05	<.0001
treat*time	2	1	-1.0961	0.3838	-1.8482	-0.3439	8.16	0.0043
treat*time	3	0	0.0000	0.0000	0.0000	0.0000	.	.

GEE Model Information

Correlation Structure	Exchangeable
Subject Effect	case (340 levels)
Number of Clusters	340
Correlation Matrix Dimension	3

Working Correlation Matrix

	Col1	Col2	Col3
Row1	1.0000	-0.0034	-0.0034
Row2	-0.0034	1.0000	-0.0034
Row3	-0.0034	-0.0034	1.0000

Exchangeable Working Correlation

Correlation -0.003436171

Analysis Of GEE Parameter Estimates
Empirical Standard Error Estimates

Parameter		Estimate	Standard Error	95% Confidence Limits		Z	Pr > |Z|
Intercept		0.9812	0.1841	0.6203	1.3421	5.33	<.0001
diagnose		-1.3117	0.1453	-1.5964	-1.0269	-9.03	<.0001
treat		2.0427	0.3061	1.4428	2.6426	6.67	<.0001
time	1	-0.9601	0.2379	-1.4265	-0.4938	-4.04	<.0001
time	2	-0.6207	0.2372	-1.0855	-0.1559	-2.62	0.0089
time	3	0.0000	0.0000	0.0000	0.0000	.	.
treat*time	1	-2.0975	0.3923	-2.8663	-1.3287	-5.35	<.0001
treat*time	2	-1.0958	0.3900	-1.8602	-0.3314	-2.81	0.0050
treat*time	3	0.0000	0.0000	0.0000	0.0000	.	.

Score Statistics For Type 3 GEE Analysis

Source	DF	Chi-Square	Pr > ChiSq
diagnose	1	70.83	<.0001
treat	1	40.38	<.0001
time	2	15.73	0.0004
treat*time	2	29.52	<.0001

Clearly, there is an important interaction between time and the treatment. The initial diagnosis is also important. Fitting two more models shows that there is no evidence of interaction between diagnosis and treatment or diagnosis and time.

We see a severe diagnosis ($s = 1$) significantly decreases the odds of a normal classification by a factor of $e^{-1.31} = 0.27$. The odds (for normal classification) ratio comparing the new drug to the standard drug changes with time because of the interaction. At 1 week it's $e^{2.04-2.09} = 0.95$, and week 2 it's $e^{2.04-1.10} = 2.6$, and at 4 weeks it's $e^{2.04-0} = 7.7$. The new drug is better, but takes time to work.

Here, the focus is on whole populations of patients at 1, 2, and 4 weeks, and on the new drug versus the standard drug. These interpretations are not within the individual, as one would make for a *conditional*, i.e. random effects, analysis, coming up in Chapter 13.

Look at the estimate of the working correlation matrix. What does this tell you? In fact, if "comment out" the REPEATED statement and assume independent observations across individuals, i.e. Y_{i1}, Y_{i2}, Y_{i3} independent, regression coefficients and standard errors change negligibly.

When to use which correlation structure $\mathbf{R}(\boldsymbol{\alpha})$?

Because GENMOD automatically uses the "sandwich" estimate of the variance, adjusting the working correlation with an empirical (but yet model-based from mean estimates!) estimate of $\text{cov}(\hat{\boldsymbol{\beta}})$, this GEE is robust to misspecification of $\mathbf{R}(\boldsymbol{\alpha})$. However, it's nice to have a formal tool for choosing.

Pan (2001) proposes a measure analogous to AIC for quasi-likelihood termed the QIC. When $\phi = 1$ it reduces to

$$QIC = -2L(\boldsymbol{\pi}(\hat{\boldsymbol{\beta}}); \mathbf{y}_1, \ldots, \mathbf{y}_n) + 2\text{trace}(\hat{\boldsymbol{\Omega}}\mathbf{V}_G),$$

where $\hat{\boldsymbol{\Omega}} = \sum_{i=1}^{n} \mathbf{D}_i'\mathbf{V}_i\mathbf{D}_i$; see Pan (2001).

QIC_u replaces $2\text{trace}(\hat{\boldsymbol{\Omega}}\mathbf{V}_G)$ with $2p$.

Example (Wheezing data from Chapter 1): The data analyzed are from Lipsitz et al. (1994). Binary Y_{ij} is the wheezing status of $n = 16$ children at ages 9, 10, 11, and 12 years ($j = 1, 2, 3, 4$); $Y_{ij} = 1$ for "yes" and $Y_{ij} = 0$ for "no". The mean $\pi_{ij} = P(Y_{ij} = 1) = E(Y_{ij})$ is modeled

$$\text{logit } \pi_{ij} = \beta_0 + \beta_1 I\{c_i = \text{Kingston}\} + \beta_2 I\{s_{ij} = 0\} + \beta_3 I\{s_{ij} = 1\} + t_j \beta_4$$

where the covariates are city of residence, age, and maternal smoking status $s_{ij} = 0, 1, 2$ at the particular age.

s_{ij}	d_{ij1}	d_{ij2}	status
0	1	0	0-9 cigarettes per day
1	0	1	10-19 cigarettes per day
2	0	0	$\geq$ 20 cigarettes per day

If we assume $Y_{i1}, Y_{i2}, Y_{i3}, Y_{i4}$ are equally correlated, we get an exchangeable correlation structure:

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

We can also, e.g., try unstructured

$$\text{corr}(\mathbf{Y}_i) = \begin{bmatrix} 1 & \alpha_{12} & \alpha_{13} & \alpha_{14} \\ \alpha_{12} & 1 & \alpha_{23} & \alpha_{24} \\ \alpha_{13} & \alpha_{23} & 1 & \alpha_{34} \\ \alpha_{14} & \alpha_{24} & \alpha_{34} & 1 \end{bmatrix}.$$

Code:

```
data six;
 input case city$ @@;
 do i=1 to 4; input age smoke wheeze @@; output; end;
datalines;
1 portage   9 0 1  10 0 1  11 0 1  12 0 0
2 kingston  9 1 1  10 2 1  11 2 0  12 2 0
           ...et cetera...
31 kingston 9 1 0  10 . .  11 1 0  12 2 1
32 portage  9 . .  10 1 1  11 1 0  12 1 0
;
run;

proc genmod data=six descending;
 class case city smoke;
 model  wheeze = city smoke age /  dist=bin link=logit type3;
 repeated  subject=case / type=un corrw; * try cs, ar, and un;
run;
```

Working Correlation Matrix

	Col1	Col2	Col3	Col4
Row1	1.0000	0.2011	0.2383	-0.1566
Row2	0.2011	1.0000	0.6267	-0.1366
Row3	0.2383	0.6267	1.0000	0.2368
Row4	-0.1566	-0.1366	0.2368	1.0000

GEE Fit Criteria

QIC	129.8428
QICu	128.7822

Analysis Of GEE Parameter Estimates
Empirical Standard Error Estimates

Parameter		Estimate	Standard Error	95% Confidence Limits		Z	Pr > |Z|
Intercept		1.4511	2.1023	-2.6694	5.5715	0.69	0.4901
city	kingston	0.3407	0.4754	-0.5909	1.2724	0.72	0.4735
city	portage	0.0000	0.0000	0.0000	0.0000	.	.
smoke	0	-0.4426	0.5613	-1.5428	0.6575	-0.79	0.4303
smoke	1	-0.3367	0.7138	-1.7356	1.0623	-0.47	0.6372
smoke	2	0.0000	0.0000	0.0000	0.0000	.	.
age		-0.2005	0.2097	-0.6116	0.2105	-0.96	0.3390

Score Statistics For Type 3 GEE Analysis

Source	DF	Chi-Square	Pr > ChiSq
city	1	0.45	0.5044
smoke	2	0.65	0.7228
age	1	0.76	0.3834