Chapter 12: Clustered Categorical Data

Dipankar Bandyopadhyay

Department of Biostatistics,
Virginia Commonwealth University

BIOS 625: Categorical Data & GLM

Example of repeated measures:

- Data are comprised of several repeated measurements on the same individual over time, e.g. Y_{ij} might indicate an acne outbreak for patient i in month j.
- 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.

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

Let T_i binary responses $\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{iT_i})$ come from the i^{th} cluster (individual, litter, clinic, etc.) Let $\boldsymbol{\mu}_i = (\mu_{i1}, \ldots, \mu_{iT_i})$ where $\mu_{ij} = E(Y_{ij})$.
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

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

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

Others are

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

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

You can also specify a fixed, known **R** as well as m-dependent 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? i.e. $T_i = T$ for all i? Not sure.
- 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.
- Repeated measures are accounted for via REPEATED statement.
- The order of $(Y_{i1}, \ldots, Y_{iT})$ makes a difference with some $\mathbf{R}(\alpha)$. 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 $(Y_{i1}, \ldots, Y_{iT})$.
- CORRW in the REPEATED statement gives the final working correlation matrix estimate.
- Elements of β are interpreted as usual, but *averaged over clusters*. This is a *marginal* interpretation.

12.3.4 GEE Methodology: Technical Details

Let $\mu_{ij} = g^{-1}(\mathbf{x}'_{ij}\boldsymbol{\beta})$ be the *marginal* mean. We assume 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:

- $\mu_{ij} = E(Y_{ij}) = b'(\theta_{ij})$ and $v(\mu_{ij}) = \text{var}(Y_{ij}) = b''(\theta_{ij})\phi$.
- $\mathbf{R}(\boldsymbol{\alpha})$ 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).
- $\mathbf{B}_i = \text{diag}(b''(\theta_{i1}), \ldots, b''(\theta_{iT_i}))$ 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.
- Note: $\mathbf{V}_i = cov(\mathbf{Y}_i)$ if $\mathbf{R}(\boldsymbol{\alpha})$ is the true correlation matrix.

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

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

The generalized estimating equations (GEE) are

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

These correspond to quasi-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 $[\mathbf{y}_i - \boldsymbol{\mu}_i(\hat{\boldsymbol{\beta}})][\mathbf{y}_i - \boldsymbol{\mu}_i(\hat{\boldsymbol{\beta}})]'$.
- 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).
- The purpose of the sandwich estimator is to use the data's empirical evidence about covariation to adjust the standard errors in case the true covariance differs substantially from the working guess.

To reiterate, the ingredients for the marginal GEE approach are

- A marginal model where Y_{ij} is binomial, Poisson, normal, gamma, etc. with mean $\mu_{ij} = g^{-1}(\mathbf{x}'_{ij}\boldsymbol{\beta})$.
  Note that often for repeated measures, $\mathbf{x}_{ij} = \mathbf{x}_i$ for $j = 1, \ldots, T_i$; e.g. gender and weight are not apt to change over a 6 month study.
- An assumption on how the elements of $\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{iT_i})$ are correlated, $\text{corr}(\mathbf{Y}_i) = \mathbf{R}(\boldsymbol{\alpha})$.
- With binary data, the correlation may not be the best way to express with-cluster association because $E(Y_{ij}Y_{is}) = P(Y_{ij} = 1, Y_{is} = 1)$ depends on $P(Y_{ij} = 1)$ and $P(Y_{is} = 1)$. One can consider alternating logistic regression (Fitzmaurice et al 1993, Lipsitz et al 1991 and Carey et al 1993). See book page 470.

Table 12.1 (p. 456) houses data from a longitudinal study comparing a new drug with a standard drug for treatment of subjects suffering mental depression. $n = 340$ Patients were 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 from mental depression 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 $T = 3$:

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

```
data depress;
  infile "[Insert directory information here]/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=ind corrw;
```

Fit of independence model to get initial estimate of β:

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	.	.

Results under exchangeable model with: *repeated subject=case / type=exch corrw;*

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 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.

Which correlation structure to use $\mathbf{R}(\alpha)$?

- 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 $\text{cov}(\hat{\beta})$, this GEE is robust to misspecification of $\mathbf{R}(\alpha)$. 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

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

  where $\hat{\Omega} = \sum_{i=1}^{n} \mathbf{D}_i'\mathbf{V}_i\mathbf{D}_i$; see Pan (2001).
- A SAS macro for obtaining the QIC is at `http://support.sas.com/ctx/samples/index.jsp?sid=1686`, but QIC is automatically included in version 9.2 and above.

Example (data from SAS documentation): 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 $\mu_{ij} = P(Y_{ij} = 1) = E(Y_{ij})$ is modeled

$$\text{logit } P(Y_{ij} = 1) = \beta_0 + \beta_1 \text{city}_i + \beta_2 \text{age}_j + \beta_3 \text{smoke}_{ij1} + \beta_4 \text{smoke}_{ij2},$$

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

S_{ij}	s_{ij1}	s_{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:

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

```
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
 3 kingston  9 0 1  10 0 0  11 1 0  12 1 0
 4 portage   9 0 0  10 0 1  11 0 1  12 1 0
 5 kingston  9 0 0  10 1 0  11 1 0  12 1 0
 6 portage   9 0 0  10 1 0  11 1 0  12 1 0
 7 kingston  9 1 0  10 1 0  11 0 0  12 0 0
 8 portage   9 1 0  10 1 0  11 1 0  12 2 0
 9 portage   9 2 1  10 2 0  11 1 0  12 1 0
10 kingston  9 0 0  10 0 0  11 0 0  12 1 0
11 kingston  9 1 1  10 0 0  11 0 1  12 0 1
12 portage   9 1 0  10 0 0  11 0 0  12 0 0
13 kingston  9 1 0  10 0 1  11 1 1  12 1 1
14 portage   9 1 0  10 2 0  11 1 0  12 2 1
15 kingston  9 1 0  10 1 0  11 1 0  12 2 1
16 portage   9 1 1  10 1 1  11 2 0  12 1 0
;
proc genmod data=six;
   class case city smoke;
   model wheeze = city age smoke / dist=bin link=logit;
   repeated subject=case / type=exch corrw;
```

Working Correlation Matrix

	Col1	Col2	Col3	Col4
Row1	1.0000	0.1837	0.1837	0.1837
Row2	0.1837	1.0000	0.1837	0.1837
Row3	0.1837	0.1837	1.0000	0.1837
Row4	0.1837	0.1837	0.1837	1.0000

Exchangeable Working Correlation

Correlation 0.1836880264

Analysis Of GEE Parameter Estimates
Empirical Standard Error Estimates

Parameter		Estimate	Standard Error	95% Confidence Limits		Z	Pr > |Z|
Intercept		2.1597	2.8229	−3.3731	7.6926	0.77	0.4442
city	kingston	0.1605	0.6741	−1.1607	1.4817	0.24	0.8118
city	portage	0.0000	0.0000	0.0000	0.0000	.	.
age		−0.2444	0.2736	−0.7806	0.2918	−0.89	0.3716
smoke	0	−0.2163	0.6386	−1.4680	1.0353	−0.34	0.7348
smoke	1	−1.0680	0.8014	−2.6387	0.5027	−1.33	0.1826
smoke	2	0.0000	0.0000	0.0000	0.0000	.	.

Unstructured, `type=un`, crashes the program. Rerun with `type=ar`:

Working Correlation Matrix

	Col1	Col2	Col3	Col4
Row1	1.0000	0.4269	0.1823	0.0778
Row2	0.4269	1.0000	0.4269	0.1823
Row3	0.1823	0.4269	1.0000	0.4269
Row4	0.0778	0.1823	0.4269	1.0000

Analysis Of GEE Parameter Estimates
Empirical Standard Error Estimates

Parameter		Estimate	Standard Error	95% Confidence Limits		Z	Pr > |Z|
Intercept		2.1264	2.6797	−3.1257	7.3784	0.79	0.4275
city	kingston	0.3400	0.6466	−0.9273	1.6073	0.53	0.5990
city	portage	0.0000	0.0000	0.0000	0.0000	.	.
age		−0.2420	0.2622	−0.7559	0.2719	−0.92	0.3561
smoke	0	−0.4130	0.6731	−1.7322	0.9062	−0.61	0.5395
smoke	1	−1.0222	0.7611	−2.5139	0.4696	−1.34	0.1793
smoke	2	0.0000	0.0000	0.0000	0.0000	.	.

Output from Version 9.2:

	GEE Fit Criteria	
Independent	QIC	85.5221
	QICu	83.6976
Exchangeable	QIC	85.0896
	QICu	83.7432
AR(1)	QIC	84.8718
	QICu	84.0957

- There are 2 versions, QIC and QIC_u.
- QIC_u replaces $2\text{trace}(\hat{\boldsymbol{\Omega}}\mathbf{V}_G)$ with $2p$ and should only be used to choose among regression models (with fixed working correlation)
- QIC can be used to choose among both regression models and working correlation structure.
- Can use QIC to be safe.

Example: Chapter 12 Problem 6 [pp. 480–481]

Here's my SAS code:

```
data abc1;
 input seq a b c count @@;
 datalines;
 1 0 0 0 0 1 0 0 1 2 1 0 1 0 2 1 0 1 1 9 1 1 0 0 0 1 1 0 1 0 1 1 1 0 1 1 1 1 1 1
 2 0 0 0 2 2 0 0 1 0 2 0 1 0 0 2 0 1 1 9 2 1 0 0 1 2 1 0 1 0 2 1 1 0 0 2 1 1 1 4
 3 0 0 0 0 3 0 0 1 1 3 0 1 0 1 3 0 1 1 8 3 1 0 0 1 3 1 0 1 3 3 1 1 0 0 3 1 1 1 1
 4 0 0 0 0 4 0 0 1 1 4 0 1 0 1 4 0 1 1 8 4 1 0 0 1 4 1 0 1 0 4 1 1 0 0 4 1 1 1 1
 5 0 0 0 3 5 0 0 1 0 5 0 1 0 0 5 0 1 1 7 5 1 0 0 0 5 1 0 1 1 5 1 1 0 2 5 1 1 1 1
 6 0 0 0 1 6 0 0 1 5 6 0 1 0 0 6 0 1 1 4 6 1 0 0 0 6 1 0 1 3 6 1 1 0 1 6 1 1 1 0
;
data abc2; set abc1;
 case=0;
 do i=1 to count;
  case=case+1;
  pattern=4*a+2*b+c;
  y=a; treat=1; output;
  y=b; treat=2; output;
  y=c; treat=3; output;
 end;
proc print;
proc genmod descending; class pattern case treat seq;
 model y=treat seq / dist=bin link=logit;
 repeated subject=case(seq*pattern) / type=exch;
 estimate '3 vs 1' treat -1 0 1 / exp;
 estimate '2 vs 1' treat -1 1 0 / exp;
 estimate '3 vs 2' treat 0 -1 1 / exp;
```

SAS output

GEE Model Information

Correlation Structure	Exchangeable
Subject Effect	case(pattern*seq) (86 levels)
Number of Clusters	86
Correlation Matrix Dimension	3
Maximum Cluster Size	3
Minimum Cluster Size	3

Exchangeable Working Correlation

Correlation -0.04403048

Contrast Estimate Results

Label	Estimate	Standard Error	Alpha	Confidence Limits		Chi-Square	Pr > ChiSq
3 vs 1	2.5076	0.4141	0.05	1.6959	3.3193	36.66	<.0001
Exp(3 vs 1)	12.2750	5.0836	0.05	5.4513	27.6400		
2 vs 1	1.9914	0.3876	0.05	1.2317	2.7511	26.39	<.0001
Exp(2 vs 1)	7.3257	2.8396	0.05	3.4270	15.6599		
3 vs 2	0.5162	0.3158	0.05	-0.1029	1.1352	2.67	0.1022
Exp(3 vs 2)	1.6756	0.5292	0.05	0.9023	3.1118		

SAS output

Empirical Standard Error Estimates

Parameter		Estimate	Standard Error	95% Confidence Limits		Z	Pr > |Z|
Intercept		0.9554	0.3282	0.3121	1.5987	2.91	0.0036
treat	1	-2.5076	0.4141	-3.3193	-1.6959	-6.05	<.0001
treat	2	-0.5162	0.3158	-1.1352	0.1029	-1.63	0.1022
treat	3	0.0000	0.0000	0.0000	0.0000	.	.
seq	1	0.5200	0.3907	-0.2459	1.2858	1.33	0.1833
seq	2	0.7775	0.5352	-0.2715	1.8265	1.45	0.1463
seq	3	0.6454	0.3865	-0.1122	1.4029	1.67	0.0950
seq	4	0.5830	0.4230	-0.2460	1.4121	1.38	0.1681
seq	5	0.2384	0.5116	-0.7642	1.2410	0.47	0.6412
seq	6	0.0000	0.0000	0.0000	0.0000	.	.

I am nesting the subject (case) index within both the drug sequence $k = 1, \ldots, 6$ and pattern type $p = 1, \ldots, 8$ for $(0,0,0), (0,0,1), \ldots, (1,1,1)$. The model looks like

$$\text{logit } P(Y_{i(k*p)j} = 1) = \gamma + \alpha_k + \beta_j,$$

where $\beta_3 = \alpha_6 = 0$ correspond to baseline.

12.3.7 Dealing with Missing Data

- Classifications of Missing Data
  - Missing Completely at Random (MCAR): Data are said to be missing completely at random if the failure to observe a value does not depend on any data, either observed or missing.
  - Missing at Random (MAR): Data are said to be missing at random if, conditional on the observed data, the failure to observe a value does not depend on the data which are unobserved.
  - Missing Not at Random (MNAR): The missing data mechanism is said to be nonignorable, or Missing Not at Random (MNAR), if the failure to observe a value depends on the value that would have been observed or other missing values in the dataset.

- In general, Bias can arise in direct GEE estimates when some data are missing unless the data are MCAR.
- Weighted Estimating Equations (WEE) are doubly robust in the sense that, in order to obtain a consistent estimate of the regression parameters, either the missing data mechanism or the score vector for the missing data given the observed data has to be correctly specified, but not both.

12.4: Markov chains for transitional modeling

When j indexes time, $Y_{i1}, Y_{i2}, \ldots, Y_{iT_i}$ is a stochastic process, often termed a *time series*. Let's consider $Y_{ij} = 0, 1$ for now.

The series $Y_{i1}, Y_{i2}, \ldots, Y_{iT_i}$ follows a first-order Markov chain if the distribution of Y_{ij} only cares about the previous value $Y_{i,j-1}$, formally $[Y_{ij}|Y_{i1}, \ldots, Y_{i,j-1}] = [Y_{ij}|Y_{i,j-1}]$.

Time-varying covariates can be included:

$$\text{logit } P(Y_{ij} = 1|Y_{i,j-1}) = \mathbf{x}'_{ij}\boldsymbol{\beta} + \gamma_1 Y_{i,j-1},$$

where γ_1 models the effect of the i^{th} subject's previous observation on the probability of a current (time j) success. e^{γ_1} has a nice interpretation in terms of how success odds changes based on what happened at last time point.

Markov chain

Second-order, and in general t-order, Markov chains can be considered by including the most previous t observations $(Y_{i,j-1}, \ldots, Y_{i,j-t})$:

$$\text{logit } P(Y_{ij} = 1 | Y_{i,j-1}, \ldots, Y_{i,j-t}) = \mathbf{x}'_{ij}\boldsymbol{\beta} + \sum_{s=1}^{t} \gamma_s Y_{i,j-s}.$$

Interactions between covariates $\mathbf{x}_{ij}$ and previous values can also improve model fit.

Likelihood

For a first order Markov-chain with no interaction the likelihood is written

$$\mathcal{L}(\beta) = \prod_{i=1}^{n} f_1(y_{i1}) f_2(y_{i2}|y_{i1}) f_3(y_{i3}|y_{i2}) \cdots f_{T_i}(y_{iT_i}|y_{i,T_i-1}).$$

if we ignore the marginal contribution of the first observation $f_1(y_{i1})$ we get

$$\mathcal{L}(\beta) = \prod_{i=1}^{n} f_2(y_{i2}|y_{i1}) f_3(y_{i3}|y_{i2}) \cdots f_{T_i}(y_{iT_i}|y_{i,T_i-1}).$$

For each subject i we have the product of $T_i - 1$ conditional logistic regression kernels; the transitional model can be fit in PROC LOGISTIC as usual, but for observation Y_{ij}, treating $Y_{i,j-1}$ as an observed predictor!

12.4.4 Respiratory illness and Maternal smoking

Example [p. 476]: Children were evaluated every year on whether they had a respiratory illness. A covariate of interest is whether the child's mom smoked at the beginning of the study; $s_i = 0$ indicates not and $s_i = 1$ indicates a smoker.
Each child has a sequence of 4 indicators $(Y_{i1}, Y_{i2}, Y_{i3}, Y_{i4})$ taken at 7, 8, 9, and 10 years. For each child we have covariates s_i and $t_j = j + 6$. The first order Markov model is fit

$$\text{logit } P(Y_{ij} = 1 | Y_{i,j-1} = y_{i,j-1}) = \beta_0 + \beta_1 s_i + \beta_2 t_j + \beta_3 y_{i,j-1},$$

for $i = 1, \ldots, 537$ and $j = 2, 3, 4$.

SAS code to fit the Markov model

```
data mm1;
 input s y1 y2 y3 y4 count;
 y=y2; yp=y1; sm=s; t=8;  ct=count; output;
 y=y3; yp=y2; sm=s; t=9;  ct=count; output;
 y=y4; yp=y3; sm=s; t=10; ct=count; output;
 datalines;
0 0 0 0 0 237
0 0 0 0 1 10
0 0 0 1 0 15
0 0 0 1 1 4
0 0 1 0 0 16
0 0 1 0 1 2
0 0 1 1 0 7
0 0 1 1 1 3
etc...
1 1 1 0 0 4
1 1 1 0 1 2
1 1 1 1 0 4
1 1 1 1 1 7
;
proc logistic descending;
 freq ct; model y=sm t yp / lackfit;
```

SAS output

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square	Standardized Estimate	Odds Ratio
INTERCPT	1	-0.2926	0.8460	0.1196	0.7295	.	.
SM	1	0.2960	0.1563	3.5837	0.0583	0.077761	1.344
T	1	-0.2428	0.0947	6.5800	0.0103	-0.109336	0.784
YP	1	2.2111	0.1582	195.3589	0.0001	0.450688	9.126

Hosmer and Lemeshow Goodness-of-Fit Test

Goodness-of-fit Statistic = 1.1723 with 6 DF (p=0.9782)

We see both time and whether the child had a respiratory illness the previous year are important predictors. Smoking is *almost* significant at the 5% level (and is significant if we perform a one-sided test). Maternal smoking increases the odds of a respiratory illness by about 34%. As time goes on the child is less likely to have a respiratory illness. If a child had a respiratory illness last year, the odds of having one this year are nine times greater than if the child did not have one last year.

Final comments on GEE (Fitzmaurice, Laird and Ware, Applied Longitudinal Analysis):

- An appealing property of the GEE estimator is that it yields a consistent estimate β even if the assumed model for the covariances among the repeated measures is not correct. It only requires that the model for the mean response be correct.
- The remarkable property of the "sandwich" estimator is that it is robust in the sense that it provides valid standard errors when the assumed model for the covariances among the repeated measures is not correct.
- In general, the closer the "working" covariance matrix approximates the true underlying covariance matrix, the greater the efficiency or precision with which β can be estimated.

- The robustness property of the 'sandwich' estimator is large sample property. It is best suited to balanced longitudinal designs where the number of subjects is relatively large and the number of repeated measures is relatively small.
- If the longitudinal design is severely unbalanced, with each subject having a unique sequence of measurement occasions, the 'sandwich'-based standard errors tends to be biased downward (i.e., too small and underestimate the covariance of $\hat{\beta}$.