

# Categorical Data Analysis

## Chapter 12

Deyuan Li  
School of Management  
Fudan University

Fall 2015

## 12. Random Effects: Generalized Linear Mixed Models for Categorical Responses

GLM  $\Rightarrow$  GLMM

We use **cluster-level** terms in the model. These terms take the same value for each observation in a cluster but different values for different clusters.

They are **unobserved** and, when treated as varying randomly among clusters, are called **random effects** (随机效应) .

The models have conditional interpretations, referred to as *subject-specific* when each cluster is a subject.

This contrasts with marginal models, which have *population-averaged* interpretations.

# Outline

- 1 Random effects modeling of clustered categorical data
- 2 Binary responses: logistic-normal model
- 3 Examples of random effects models for binary data
- 4 Random effects models for multivariate data
- 5 \* Multivariate random effects models for binary data

## 12.1 Random effects modeling of clustered categorical data

Parameters that describe a factor's effects in ordinary linear models are called **fixed effects**. They apply to all categories of interest, such as genders, age groupings, or treatments.

By contrast, **random effects** usually apply to a sample. For a study using a sample of clinics, for example, the model treats observations from a given clinic as a cluster, and it has a random effect for each clinic.

The **generalized linear mixed model** (GLMM, 广义线性混合模型) is a further extension that permits random effects as well as fixed effects in the linear predictor.

## 12.1.1 Generalized linear mixed model

$y_{it}$ : observation  $t$  in cluster  $i$ ,  $t = 1, \dots, T_i$ ,

$\mathbf{x}_{it}$ : a column vector of values of explanatory variables,  
for fixed effect model parameters  $\beta$ ;

$\mathbf{u}_i$ : the vector of random effect values for cluster  $i$   
(common to all observations in the cluster);

$\mathbf{z}_{it}$ : a column vector of their explanatory variables  
(normally, the random effect is univariate).

Conditional on  $\mathbf{u}_i$ , let  $\mu_{it} = E(Y_{it}|\mathbf{u}_i)$ . The GLMM is

$$g(\mu_{it}) = \mathbf{x}_{it}'\beta + \mathbf{z}_{it}'\mathbf{u}_i \quad (12.1)$$

for link function  $g$ . The random effect vector  $\mathbf{u}_i$  is assumed to have a multivariate normal distribution  $N(0, \Sigma)$ .

## 12.1.1 Generalized linear mixed model

Conditional on  $\mathbf{u}_i$ , the model treats  $y_{it}$  as independent over  $i$  and  $t$ .

Consider the special case with univariate random effect and  $z_{it} = 1$ . With  $u_i$  replaced by  $u_i^* \sigma$ , where  $\{u_i^*\}$  are  $N(0, 1)$ , the GLMM has the form

$$g(\mu_{it}) = \mathbf{x}'_{it}\boldsymbol{\beta} + u_i^* \sigma.$$

This has the form of an ordinary GLM with unobserved values  $\{u_i^*\}$  of a particular covariate.

Thus, random effects models relate to methods of dealing with unmeasured predictors and other forms of missing data.

## 12.1.1 Generalized linear mixed model

Random effects also sometimes represent **random measurement error** in the explanatory variables.

If we replace a particular predictor  $x_{it}$  by  $x_{it}^* + \epsilon_i$ , with  $x_{it}^*$  the true value and  $\epsilon_i$  the measurement error, then  $\epsilon$  times the regression parameter can be absorbed in the random effects term.

Related to these motivations, random effects also provide a mechanism for explaining **overdispersion** in basic models not having those effects.

## 12.1.2 Logit GLMM for binary matched pairs

Cluster  $i$  consists of the responses  $(y_{i1}, y_{i2})$  for matched pair  $i$ . Let  $y_{it} = 1$  (success) or 0 (failure),  $t = 1, 2$ .

In Section 10.2.2 we introduced the model

$$\text{logit}P(Y_{it} = 1) = \alpha_i + \beta x_t, \quad (12.2)$$

where  $x_1 = 0$ ,  $x_2 = 1$ . Here  $\alpha_i$  is treated as a fixed effect. An equivalent representation is

$$\text{logit}P(Y_{i1} = 1 | u_i) = \alpha + u_i, \quad \text{logit}P(Y_{i2} = 1 | u_i) = \alpha + \beta + u_i, \quad (12.3)$$

where  $u_i = \alpha_i - \alpha$ .

$u_i$  is the random effect for cluster  $i$ .  $u_i$  iid  $N(0, \sigma^2)$  with unknown  $\sigma$ . Conditionally on  $u_i$ ,  $y_{i1}$  and  $y_{i2}$  are independent.



## 12.1.2 Logit GLMM for binary matched pairs

The model (12.3) is the special case of GLMM (12.1), in which

$$\begin{aligned}\mu_{it} &= P(Y_{it} = 1 | u_i), \quad g \text{ is the log link} \\ \beta' &= (\alpha, \beta), \\ \mathbf{x}'_{i1} &= (1, 0) \text{ and } \mathbf{x}'_{i2} = (1, 1) \text{ for all } i, \\ z_{it} &= 1 \text{ for all } i, t.\end{aligned}$$

The univariate random effect adjusts the intercept but does not modify the fixed effect.

A GLMM with random effect of this form is called a **random intercept** model (随机截距模型). Instead of the usual fixed intercept  $\alpha$ , it has a random intercept  $\alpha + u_i$ .

## 12.1.2 Logit GLMM for binary matched pairs

Let  $Y_1 = \sum_i y_{i1}$ ,  $Y_2 = \sum_i y_{i2}$ .

Marginally,  $Y_1$  is binomial with  $n$  trials and parameter  $E\{\exp(\alpha + U)/[1 + \exp(\alpha + U)]\}$ ,  $Y_2$  is binomial with parameter  $E\{\exp(\alpha + \beta + U)/[1 + \exp(\alpha + \beta + U)]\}$ , where  $U \sim N(0, \sigma^2)$ .

## 12.1.2 Logit GLMM for binary matched pairs

**Proof.** Note that

$$P(Y_1 = k) = E[P(Y_1 = k | (u_1, u_2, \dots, u_n))]$$

and

$$\begin{aligned} & P(Y_1 = k | (u_1, \dots, u_n)) \\ &= P(Y_{i1} = 1, i = 1, \dots, k; Y_{j1} = 0, j = k + 1, \dots, n) | (u_1, \dots, u_n) \\ &+ \dots + P(Y_{i1} = 0, i = 1, \dots, n - k; Y_{j1} = 1, j = n - k + 1, \dots, n) \\ &= \prod_{i=1}^k \frac{e^{\alpha + u_i}}{1 + e^{\alpha + u_i}} \prod_{j=k+1}^n \frac{1}{1 + e^{\alpha + u_j}} + \dots \quad (C_n^k \text{ items in total}). \end{aligned}$$

## 12.1.2 Logit GLMM for binary matched pairs

So,

$$\begin{aligned} E[Y_1 = k | (u_1, \dots, u_n)] &= C_n^k \prod_{i=1}^k E\left[\frac{e^{\alpha+u_i}}{1 + e^{\alpha+u_i}}\right] \prod_{j=k+1}^n E\left[\frac{1}{1 + e^{\alpha+u_j}}\right] \\ &= C_n^k \left(E\left[\frac{e^{\alpha+U}}{1 + e^{\alpha+U}}\right]\right)^k \left(1 - E\left[\frac{e^{\alpha+U}}{1 + e^{\alpha+U}}\right]\right)^{n-k} \end{aligned}$$

and hence

$$Y_1 \sim \text{Bin}\left(n, E\left[\frac{e^{\alpha+U}}{1 + e^{\alpha+U}}\right]\right).$$

Similarly,

$$Y_2 \sim \text{Bin}\left(n, E\left[\frac{e^{\alpha+\beta+U}}{1 + e^{\alpha+\beta+U}}\right]\right).$$

## 12.1.2 Logit GLMM for binary matched pairs

The model implies a nonnegative correlation between  $Y_1$  and  $Y_2$ , with greater association resulting from greater heterogeneity (i.e., larger  $\sigma$ ).

Clusters with a large positive  $u_i$  have a relatively large  $P(Y_{it} = 1|u_i)$  for each  $t$ , whereas clusters with a large negative  $u_i$  have a relatively small  $P(Y_{it} = 1|u_i)$  for each.

For this model,  $Y_1$  and  $Y_2$  are independent only if  $\sigma = 0$ .

A  $2 \times 2$  table, Let  $(y_{i1}, y_{i2})$  take value of  $(1, 1)$ ,  $(1, 0)$ ,  $(0, 1)$  or  $(0, 0)$ . Let  $\{n_{ab}\}$  denote the counts of the four values.

## 12.1.2 Logit GLMM for binary matched pairs

Let  $\{\hat{\mu}_{ab}\}$  denote marginal fitted values for model (12.3). Then

$$\hat{\beta} = \log(\hat{\mu}_{21}/\hat{\mu}_{12}).$$

When the sample log odds ratio  $\log(n_{11}n_{22}/n_{12}n_{21}) \geq 0$ , then  $\{\hat{\mu}_{ab} = n_{ab}\}$  and  $\hat{\beta} = \log(n_{21}/n_{12})$ . This is the same as the conditional ML estimate (Section 10.2.3)

The model implies that the true log odds ratio is at least 0.

When  $\log(n_{11}n_{22}/n_{12}n_{21}) < 0$ , however, then  $\hat{\sigma} = 0$  and the fitted values  $\{\hat{\mu}_{ab} = n_{a+}n_{+b}/n\}$  satisfy independence. Then,  $\hat{\beta}$  is identical to the estimate for the marginal model (10.6) by which  $\beta$  is the difference between logits for the two marginal distributions, namely  $\hat{\beta} = \log[(n_{2+}n_{+1})/(n_{1+}n_{+2})]$ .

## 12.1.3 Ratings of Prime Minister revisited

**TABLE 12.1 Rating of Performance of Prime Minister**

First Survey	Second Survey		Total
	Approve	Disapprove	
Approve	794	150	944
Disapprove	86	570	656
Total	880	720	1600

For Table 12.1, the ML fit of model (12.3), treating  $\{u_i\}$  as normal, yields  $\hat{\beta} = \log(85/150) = -0.556$  ( $SE = 0.135$ ), with  $\hat{\sigma} = 5.16$ . This is identical to (10.10), but with standard error  $[(1/86) + (1/500)]^{1/2} \approx 0.12 \neq 0.135$ .

For a given subject, the estimated odds of approval at the second survey equal  $\exp(-0.556) = 0.57$  times those at the first survey. The large  $\hat{\sigma}$  reflects the very strong association between the two responses, with sample odds ratio 35.1.

## 12.1.4 Extension: Rasch model and item response models

Extension of model (12.3) for  $T > 2$  observations in each cluster. The random intercept model then has form

$$\text{logit}[P(Y_{it} = 1 | u_i)] = u_i + \beta_t, \quad (12.4)$$

where  $\{u_i\}$  iid  $N(0, \sigma^2)$ . Equivalently, the model can add an intercept  $\alpha$  or let  $E(u_i) = \alpha$ , but then identifiability requires a constraint such as  $\beta_T = 0$ .

The model (12.4) is called the **Rasch model** (*item-response model*, 项目反映模型), which is an early application of GLMM in psychometrics (心理测量学).



## 12.1.4 Extension: Rasch model and item response models

The  $\{\beta_t\}$  in the Rasch model differ from those in corresponding marginal models such as (11.1), since the effects are subject specific.

Rasch model refers to a  $T \times 2 \times n$  table of observation by outcome by subject, whereas the marginal model refers to the  $T \times 2$  observation-by-outcome table of the  $T$  marginal distributions, collapsed over subjects.

For observations  $s$  and  $t$  for a given subject  $i$  with model (12.4),

$$\beta_s - \beta_t = \text{logit}[P(Y_{is} = 1 | u_i)] - \text{logit}[P(Y_{it} = 1 | u_i)],$$

which is a log odds ratio conditional on the subject.

## 12.1.4 Extension: Rasch model and item response models

By contrast, the corresponding population-averaged effect in marginal model (11.1) is

$$\beta_s - \beta_t = \text{logit}[P(Y_{hs} = 1)] - \text{logit}[P(Y_{it} = 1)],$$

with subject  $h$  randomly selected for observation  $s$  and subject  $i$  randomly selected for observation  $t$  (i.e.,  $h$  and  $i$  are independent observations).

### Extension of Rasch Model

- Rasch model.

$$P(Y_{it} = 1) = \frac{\exp\{a(u_i + \beta_t)\}}{1 + \exp\{a(u_i + \beta_t)\}}, \quad a = 1 \text{ or else,}$$

## 12.1.4 Extension: Rasch model and item response models

- two parameter logisitic model

$$P(Y_{it} = 1) = \frac{\exp\{a_t(u_i + \beta_t)\}}{1 + \exp\{a_t(u_i + \beta_t)\}},$$

where  $a_t$  is discriminant parameter for the  $t$ -th question.

- three parameter logisitic model

$$P(Y_{it} = 1) = c_t + (1 - c_t) \frac{\exp\{a_t(u_i + \beta_t)\}}{1 + \exp\{a_t(u_i + \beta_t)\}},$$

where  $c_t$  is guess parameter for the  $t$ -th question.

- Normal ogive model

$$P(Y_{it} = 1) = c_t + (1 - c_t)\Phi(a_t(u_i + \beta_t)).$$

# Outline

- 1 Random effects modeling of clustered categorical data
- 2 Binary responses: logistic-normal model**
- 3 Examples of random effects models for binary data
- 4 Random effects models for multivariate data
- 5 \* Multivariate random effects models for binary data

## 12.2 Binary responses: logistic-normal model

Model (12.4) is a special case of the logistic-normal models for binary data (below), with univariate random effect,

$$\text{logit}P(Y_{it} = 1|u_i) = \mathbf{x}'_{it}\boldsymbol{\beta} + u_i \quad (12.5)$$

where  $\{u_i\}$  iid  $N(0, \sigma^2)$ .

For such models,  $Y_{is}$  and  $Y_{it}$  are conditionally independent (*assumed*) but are marginally nonnegatively correlated.

Let  $\Phi$  denote the cdf. For  $s \neq t$ ,

$$\begin{aligned} \text{cov}(Y_{is}, Y_{it}) &= E[\text{cov}(Y_{is}, Y_{it}|u_i)] + \text{cov}[E(Y_{is}|u_i), E(Y_{it}|u_i)] \\ &= 0 + \text{cov}[\Phi(\mathbf{x}'_{is}\boldsymbol{\beta} + u_i), \Phi(\mathbf{x}'_{it}\boldsymbol{\beta} + u_i)] > 0. \end{aligned}$$

For common predictor value  $\mathbf{x}$  at each  $t$ , the joint distribution for the model is exchangeable. This is often plausible for clustered data.

## 12.2 Binary responses: logistic-normal model

Usually, the main focus in using a GLMM is inference about the fixed effects.

The random effects part of the model is a mechanism for representing how the positive correlation occurs between observations within a cluster.

The estimate  $\hat{\sigma}$  of the standard deviation of a random intercept may be a useful summary of the degree of heterogeneity of a population.

## 12.2.1 Interpreting heterogeneity in logistic-normal models

When  $\sigma > 0$ , how can we interpret the variability in effects? Consider observations  $(y_{it}, \mathbf{x}_{it})$  and  $(y_{hs}, \mathbf{x}_{hs})$ . Their log odds ratio is

$$\text{logit}[P(Y_{it} = 1 | u_i)] - \text{logit}[P(Y_{hs} = 1 | u_h)] = (\mathbf{x}_{it} - \mathbf{x}_{hs})' \boldsymbol{\beta} + (u_i - u_h).$$

We cannot observe  $(u_i - u_h)$ , which is  $N(0, 2\sigma^2)$  random. But,  $1 - \alpha$  of those log odds ratios fall within

$$(\mathbf{x}_{it} - \mathbf{x}_{hs})' \boldsymbol{\beta} \pm z_{\alpha/2} \sqrt{2}\sigma.$$

When  $\sigma > 0$ ,  $(\mathbf{x}_{it} - \mathbf{x}_{hs})' \boldsymbol{\beta}$  is the log odds ratio for two observations in the same cluster ( $h = i$ ) or with the same random effect value.

## 12.2.1 Interpreting heterogeneity in logistic-normal models

Suppose that  $\mathbf{x}_{it} = \mathbf{x}_{hs}$  for observations from different clusters.

Then, using  $z_{0.25} = 0.674$ , the middle 50% of the log odds ratios fall within  $\pm 0.674\sqrt{2}\sigma = \pm 0.95\sigma$ .

Hence, the median odds ratio between the observation with higher random effect and the observation with lower random effect equals  $\exp(0.95\sigma)$ .

With a single predictor and  $x_{it} - x_{hs} = 1$ , the median such odds ratio equals  $\exp(\beta + 0.95\sigma)$ .



# Outline

- 1 Random effects modeling of clustered categorical data
- 2 Binary responses: logistic-normal model
- 3 Examples of random effects models for binary data**
- 4 Random effects models for multivariate data
- 5 \* Multivariate random effects models for binary data

## 12.3.1 Small-area estimation of binomial proportions

**Small-area estimation** refers to estimation of parameters for a large number of geographical areas when each has relatively few observations.

For instance, one might want county-specific estimates of characteristics such as the unemployment rate or the proportion of families having health insurance coverage. With a national or statewide survey, some counties may have few observations. Then, sample proportions in the counties may poorly estimate the true countrywide proportions.

Random effects models that treat each county as a cluster can provide improved estimates. In assuming that the true proportions vary according to some distribution, the fitting process "**borrowing from the whole**"—it uses data from all the counties to estimate the proportion in any given one.

## 12.3.1 Small-area estimation of binomial proportions

$\pi_i$ : the true proportion in area  $i$ ,  $i = 1, 2, \dots, n$ ;

$\{y_i\}$ : independent  $\text{Bin}(T_i, \pi_i)$  variates; that is  $y_i = \sum_{t=1}^{T_i} y_{it}$ ,  
 where  $\{y_{it}\}$  iid with  $P(Y_{it} = 1) = \pi_i$  and  $P(Y_{it} = 0) = 1 - \pi_i$

$\{p_i = y_i/T_i\}$ : sample proportions, ML estimators for  $\{\pi_i\}$  for the saturated model

$$\text{logit}(\pi_i) = \alpha + \beta_i, \quad i = 1, 2, \dots, n.$$

For small  $\{T_i\}$ ,  $\{p_i\}$  have large standard errors. Thus,  $\{p_i\}$  may display much more variability than  $\{\pi_i\}$ , especially when  $\{\pi_i\}$  are similar.

Then, it is helpful to shrink  $\{p_i\}$  toward their overall mean.

## 12.3.1 Small-area estimation of binomial proportions

One can accomplish this with the random effects model

$$\text{logit}[P(Y_{it} = 1 | u_i)] = \alpha + u_i, \quad \{u_i\} \text{ i.i.d. } N(0, \sigma^2).$$

Then

$$\hat{\pi}_i = \frac{\exp(\hat{\alpha} + \hat{u}_i)}{1 + \exp(\hat{\alpha} + \hat{u}_i)}.$$

If  $\hat{\sigma} = 0$ , then all  $\hat{u}_i = 0$ . Then, the random effects estimate of each  $\hat{\pi}_i$  is  $(\sum_{i=1}^n \sum_{t=1}^{T_i} y_{it}) / (\sum_i T_i)$ , the overall sample proportion after pooling all  $n$  samples.

When truly all  $\pi_i$  are equal, this is a much better estimator of that common value than the sample proportion from a single sample.

## 12.3.1 Small-area estimation of binomial proportions

**Estimation of GLMM.** In general, we have

$$\begin{cases} g(\mu_{it}) = x'_{it}\beta + z'_{it}u_i, & u_i \text{ iid } N(0, \Sigma), \\ f(y; \beta, \Sigma) = \int f(y|u, \beta)f(u; \Sigma)du. \end{cases}$$

So, the likelihood function is

$$\begin{aligned} L = \prod_i \left( \int_{-\infty}^{\infty} \prod_t \left[ \frac{\exp\{x'_{it}\beta + z'_{it}u_i\}}{1 + \exp\{x'_{it}\beta + z'_{it}u_i\}} \right]^{y_{it}} \right. \\ \left. \times \left[ \frac{1}{1 + \exp\{x'_{it}\beta + z'_{it}u_i\}} \right]^{1-y_{it}} f(u_i; \Sigma) du_i \right). \end{aligned}$$

By maximizing  $\log L$ , we can obtain  $\hat{\beta}$  and  $\hat{\Sigma}$ .

## 12.3.1 Small-area estimation of binomial proportions

### Steps of estimation for model (12.5)

1. By maximizing the log-likelihood function, get MLE  $\hat{\beta}$  and  $\hat{\sigma}$ .
2. Obtain  $\hat{u}_i$  by least squared method, MCMC (sample from the posterior distribution of  $u_i$ , and get the sample mean), or other methods. For example,

$$\hat{u}_i = \operatorname{argmin}_{u_i} \sum_{t=1}^T \left| y_{it} - \frac{\exp\{x'_{it}\hat{\beta} + u_i\}}{1 + \exp\{x'_{it}\hat{\beta} + u_i\}} \right|^2.$$

3. Obtain  $\hat{\pi}_{it}$ .

## 12.3.2 Modeling repeated binary responses

We extend the Rasch model to incorporate covariates. We illustrate using Table 10.13. The subjects indicated whether they supported legalizing abortion (墮胎合法化) in each of three situations. Table 10.13 also classified the subjects by gender.

**TABLE 10.13 Support for Legalizing Abortion in Three Situations, by Gender**

Gender	Sequence of Responses on the Three Items <sup>a</sup>							
	(1, 1, 1)	(1, 1, 2)	(2, 1, 1)	(2, 1, 2)	(1, 2, 1)	(1, 2, 2)	(2, 2, 1)	(2, 2, 2)
Male	342	26	6	21	11	32	19	356
Female	440	25	14	18	14	47	22	457

<sup>a</sup>Items are (1) if the family has a very low income and cannot afford anymore children, (2) when the woman is not married and does not want to marry the man, and (3) when the woman wants it for any reason. 1, yes; 2, no.

*Source:* Data from 1994 General Social Survey, National Opinion Research Center.

## 12.3.2 Modeling repeated binary responses

Let  $y_{it}$  denote the response for subject  $i$  on item  $t$ , with  $y_{it} = 1$  representing support. Consider the model

$$\text{logit}[P(Y_{it} = 1|u_i)] = u_i + \beta_t + \gamma x_i, \quad (12.10)$$

where  $x_i = 1$  (females) or 0 (males), and  $\{u_i\}$  iid  $N(0, \sigma^2)$ .

Note that model (12.10) implies nonnegative association among responses on the items, one should use items and scales for which this should occur.



## 12.3.2 Modeling repeated binary responses

**TABLE 12.3 Summary of ML Estimates for Random Effects Model (12.10) and ML and GEE Estimates for Corresponding Marginal Model**

Effect	Parameter	GLMM ML		Marginal Model ML		Marginal Model GEE	
		Estimate	SE	Estimate	SE	Estimate	SE
Abortion	$\beta_1 - \beta_3$	0.83	0.16	0.148	0.030	0.149	0.030
	$\beta_1 - \beta_2$	0.54	0.16	0.098	0.027	0.097	0.028
	$\beta_2 - \beta_3$	0.29	0.16	0.049	0.027	0.052	0.027
Gender	$\gamma$	0.01	0.48	0.005	0.088	0.003	0.088
$\sqrt{\text{var}(u_i)}$	$\sigma$	8.6	0.54				

Overall chi-squared statistics comparing the 16 observed and fitted counts are  $G^2 = 23.2$  and  $X^2 = 27.8$  ( $df = 9$ ). These are not that large considering the very large sample size and the few parameters ( $\beta_1, \beta_2, \beta_3, \gamma, \sigma$ ) used to describe the 14 multinomial cell probabilities ( $8 - 1 = 7$  for each gender) in Table 10.13.

## 12.3.2 Modeling repeated binary responses

Here,  $df = 9$  since we are modeling 14 multinomial parameters using five GLMM parameters  $(\beta_1, \beta_2, \beta_3, \gamma, \sigma)$ .

An alternative analysis of these data focuses on the marginal distributions, treating the dependence as a nuisance. A marginal model analog of (12.10) is

$$\text{logit}[P(Y_t = 1)] = \beta_t + \gamma \mathbf{x}.$$

For it, Table 12.3 also shows GEE estimates for the exchangeable working correlation structure and ML estimates. The marginal model fits well, with  $G^2 = 1.1$  ( $df = 2$ ).

## 12.3.2 Modeling repeated binary responses

These population-averaged  $\{\hat{\beta}_t\}$  are much smaller than the subject-specific  $\{\hat{\beta}_t\}$  from the GLMM.

This reflects the very large GLMM heterogeneity ( $\hat{\sigma} = 8.6$ ) and the corresponding strong correlations among the three responses. For instance, the GEE analysis estimates a common correlation of 0.82 between pairs of responses.

Although the GLMM  $\{\hat{\beta}_t\}$  are about five to six times the marginal model  $\{\hat{\beta}_t\}$ , so are the standard errors. The two approaches provide similar substantive interpretations and conclusions.

## 12.3.3 Longitudinal mental depression study revisited

**TABLE 11.2 Cross-Classification of Responses on Depression at Three Times by Diagnosis and Treatment**

Diagnosis	Treatment	Response at Three Times <sup>a</sup>							
		NNN	NNA	NAN	NAA	ANN	ANA	AAN	AAA
Mild	Standard	16	13	9	3	14	4	15	6
	New drug	31	0	6	0	22	2	9	0
Severe	Standard	2	2	8	9	9	15	27	28
	New drug	7	2	5	2	31	5	32	6

<sup>a</sup>N, normal; A, abnormal.

$y_t$ : measurement  $t$  on mental depression,

1=normal, 0=abnormal;

$s$ : severity of initial diagnosis, 1=severe, 0=mild;

$d$ : drug treatment, 1=new, 0=standard;

$t$ : time of measurement  $t$ .

## 12.3.3 Longitudinal mental depression study revisited

In section 11.2.1, we used the model

$$\text{logit}[P(Y_t = 1)] = \alpha + \beta_1 \mathbf{s} + \beta_2 \mathbf{d} + \beta_3 \mathbf{t} + \beta_4 \mathbf{dt}$$

to evaluate the marginal distributions.

Now let  $y_{it}$  denote observation  $t$  for subject  $i$ . The model

$$\text{logit}[P(Y_{it} = 1 | u_i)] = \alpha + \beta_1 \mathbf{s} + \beta_2 \mathbf{d} + \beta_3 \mathbf{t} + \beta_4 \mathbf{dt} + u_i$$

has subject-specific rather than population-averaged effects.

## 12.3.3 Longitudinal mental depression study revisited

**TABLE 12.4 Model Parameter Estimates for Marginal and Conditional Logit Models Fitted to Table 11.2**

Parameter	ML Marginal Estimate	Std. Error	GEE Marginal Estimate	Std. Error	Random Effects ML Estimate	Std. Error
Diagnosis	-1.29	0.14	-1.31	0.15	-1.32	0.15
Drug	-0.06	0.22	-0.06	0.23	-0.06	0.22
Time	0.48	0.12	0.48	0.12	0.48	0.12
Drug $\times$ Time	1.01	0.18	1.02	0.19	1.02	0.19

$\hat{\sigma} = 0.07$  shows little heterogeneity among subjects. Based on the model fit, integrating over the  $N(0, 0.07^2)$  random effects distribution yields marginal fitted values of the possible response sequences.

Comparing these to the sample counts in Table 11.2 indicates a relatively good fit. The usual fit statistics  $G^2 = 22.0$  and  $X^2 = 20.8$  ( $df = 28 - 6 = 22$ ).

## 12.3.4 Modeling heterogeneity among multicenter clinical trials

Many applications compare two groups on a response for data stratified on a third variable, such as school or medical clinic.

**TABLE 12.5 Clinical Trial Relating Treatment to Response for Eight Centers**

Center	Treatment	Response		Sample Odds Ratio	Fitted Odds Ratio
		Success	Failure		
1	Drug	11	25	1.19	2.02
	Control	10	27		
2	Drug	16	4	1.82	2.09
	Control	22	10		
3	Drug	14	5	4.80	2.19
	Control	7	12		
4	Drug	2	14	2.29	2.11
	Control	1	16		
5	Drug	6	11	$\infty$	2.18
	Control	0	12		
6	Drug	1	10	$\infty$	2.12
	Control	0	10		
7	Drug	1	4	2.0	2.11
	Control	1	8		
8	Drug	4	2	0.33	2.06
	Control	6	1		

## 12.3.4 Modeling heterogeneity among multicenter clinical trials

Consider random effects models.

For a subject in center  $i$  using treatment  $t$  (1=active drug; 2=control), let  $y_{it} = 1$  for success. Consider the logistic-normal,

$$\text{logit}[P(Y_{i1} = 1|u_i)] = \alpha + \beta/2 + u_i, \quad (12.11)$$

$$\text{logit}[P(Y_{i2} = 1|u_i)] = \alpha - \beta/2 + u_i,$$

where  $\{u_i\}$  iid  $N(0, \sigma^2)$ .

Note that log odds ratio  $\beta$  between treatment and response is constant over centers,  $\sigma$  summarizes center heterogeneity in the success probabilities.



## 12.3.4 Modeling heterogeneity among multicenter clinical trials

A logistic-normal model permitting treatment-by-center interaction is

$$\begin{aligned}\text{logit}[P(Y_{i1} = 1 | u_i, b_i)] &= \alpha + (\beta + b_i)/2 + u_i, \\ \text{logit}[P(Y_{i2} = 1 | u_i, b_i)] &= \alpha - (\beta + b_i)/2 + u_i,\end{aligned}\quad (12.12)$$

where  $\{u_i\}$  iid  $N(0, \sigma_a^2)$ ,  $\{b_i\}$  iid  $N(0, \sigma_b^2)$  and  $\{u_i\} \perp \{b_i\}$ .

The log odds ratio equals  $\beta + b_i$  in center  $i$ . These vary among centers according to a  $N(\beta, \sigma_b^2)$  distribution. That is,  $\beta$  is the expected center-specific log odds ratio between treatment and response, and  $\sigma_b$  describes variability in those log odds ratios.

The model parameters are  $(\alpha, \beta, \sigma_a, \sigma_b)$ .

## 12.3.5 Alternative formulations of random effects models

An equivalent expression for interaction model (12.12) is

$$\text{logit}[P(Y_{it} = 1 | u_i, b_{it})] = \alpha + \beta x_t + b_{it} + u_i,$$

where  $x_t$  is a treatment dummy variable ( $x_1 = 1, x_2 = 0$ ),  $\{u_i\}$  iid  $N(0, \sigma_a^2)$ , and  $\{b_{i1}\}$  and  $\{b_{i2}\}$  iid  $N(0, \sigma^2)$ . Here,  $b_{i1} - b_{i2}$  corresponds to  $b_i$  in (12.12), and  $2\sigma^2$  corresponds to  $\sigma_b^2$ .

Formulating a random effects model requires care about implications of the model expression and the random effects correlation structure.

## 12.3.5 Alternative formulations of random effects models

Suppose that one expressed the interaction model (12.12) as

$$\text{logit}[P(Y_{it} = 1 | u_i, b_i)] = \alpha + (\beta + b_i)x_t + u_i, \quad (12.13)$$

where  $\{b_i\}$  iid  $N(0, \sigma_b^2)$ .

This (i.e. (12.13)) is inappropriate, since the model then imposes greater variability for the logit with the first treatment than the second, since  $x_2 = 0$  and  $\{u_i\}$  and  $\{b_i\}$  are uncorrelated.

Also, the model should not depend on the definition of the dummy variable  $x_t$ .

## 12.3.5 Alternative formulations of random effects models

Note, however, that if  $z_t = x_t + c$  for some constant  $c$ , then model (12.13) is equivalently

$$\text{logit}[P(Y_{it} = 1 | u_i, b_i)] = \alpha + (\beta + b_i)(z_t - c) + u_i = \alpha' + (\beta + b_i)z_t + v_i,$$

where  $\alpha' = \alpha - c\beta$  and  $v_i = u_i - cb_i$ .

Thus,  $(v_i, b_i)$  are correlated even if  $(u_i, b_i)$  are not.

In fact, expression (12.13) is sensible only with correlated random effects. It is then equivalent to (12.12) with correlated random effects.

# Outline

- 1 Random effects modeling of clustered categorical data
- 2 Binary responses: logistic-normal model
- 3 Examples of random effects models for binary data
- 4 Random effects models for multivariate data**
- 5 \* Multivariate random effects models for binary data

## 12.4.1 Cumulative logit model with random intercept

Modeling is simpler with ordinal than nominal responses, since often the same random effect and the same fixed effect can apply to each logit.

Let  $y_{it}$  be observation  $t$  in cluster  $i$ ,  $i = 1, 2, \dots, I$ . A GLMM for the cumulative logits has the form

$$\text{logit}[P(Y_{it} \leq j | \mathbf{u}_i)] = \alpha_j + \mathbf{x}'_{it}\boldsymbol{\beta} + \mathbf{z}'_{it}\mathbf{u}_i, \quad j = 1, \dots, I-1. \quad (12.14).$$

## 12.4.2 Insomnia study revisited

Data is from a clinical trial at two occasions comparing a drug with placebo in treating insomnia patients.

**TABLE 11.4 Time to Falling Asleep, by Treatment and Occasion**

Treatment	Time to Falling Asleep				
	Initial	Follow-up			
		< 20	20–30	30–60	> 60
Active	< 20	7	4	1	0
	20–30	11	5	2	2
	30–60	13	23	3	1
	> 60	9	17	13	8
Placebo	< 20	7	4	2	1
	20–30	14	5	1	0
	30–60	6	9	18	2
	> 60	4	11	14	22

## 12.4.2 Insomnia study revisited

For  $y_t$  = time to fall asleep at occasion  $t$ , the marginal model (with interaction between  $t$  and  $x$ ) is

$$\text{logit}[P(Y_t \leq j)] = \alpha_j + \beta_1 t + \beta_2 x + \beta_3 tx$$

Now, let  $y_{it}$  denote the response for subject  $i$  at occasion  $t$ . The random-intercept model is

$$\text{logit}[P(Y_{it} \leq j | u_i)] = u_i + \alpha_j + \beta_1 t + \beta_2 x + \beta_3 tx.$$



## 12.4.2 Insomnia study revisited

**TABLE 12.7** Fits of Cumulative Logit Models to Table 11.4<sup>a</sup>

Effect	Marginal ML	Marginal GEE	Random Effects (GLMM) ML
Treatment	0.046 (0.236)	0.034 (0.238)	0.058 (0.366)
Occasion	1.074 (0.162)	1.038 (0.168)	1.602 (0.283)
Treatment $\times$ occasion	0.662 (0.244)	0.708 (0.244)	1.081 (0.380)

<sup>a</sup>Values in parentheses represent standard errors.

The estimates and standard errors of random effects are about 50% larger.

This reflects the relatively large heterogeneity ( $\hat{\sigma} = 1.90$ ) and the resultant strong association between the responses at the two occasions.

## 12.4.4 Baseline-category logit models with random effects

The general form of the baseline-category logit model with random effects is

$$\log \frac{P(Y_{it} = j)}{P(Y_{it} = l)} = \alpha_j + \mathbf{x}'_{it}\beta_j + \mathbf{z}'_{it}\mathbf{u}_{ij}, \quad j = 1, 2, \dots, l - 1.$$

The fixed effects  $\beta_j$  and the random effects  $\mathbf{u}_{ij}$  depend on  $j$ , since the baseline category is arbitrary.

With nominal responses there is no reason to expect effects to be similar for different  $j$ .

# Outline

- 1 Random effects modeling of clustered categorical data
- 2 Binary responses: logistic-normal model
- 3 Examples of random effects models for binary data
- 4 Random effects models for multivariate data
- 5 **\* Multivariate random effects models for binary data**

## 12.5\* Multivariate random effects models for binary data

In practice, random effects are often univariate, taking the form of random intercepts.

However, we have seen that nominal responses require multivariate random effects and that bivariate random effects are helpful for describing heterogeneity in multicenter clinical trials.

In this section we present other examples in which multivariate random effects are natural.

## 12.5.1 Matched pairs with a bivariate binary response

A sample of schoolboys were interviewed twice, several months apart, and asked about their self-perceived membership in the "leading crowd" and about whether they sometimes needed to go against their principles to belong to that group.

Two binary response variables: membership and attitude, measured at two interview times for each subject.

Table 12.8 labels the categories for attitude as (positive, negative), where "positive" refers to disagreeing with the statement that one must go against his principles.

## 12.5.1 Matched pairs with a bivariate binary response

**TABLE 12.8 Membership and Attitude Toward the “Leading Crowd”**

$(M, A)$ for First Interview	$(M, A)$ for Second Interview <sup>a</sup>			
	(Yes, Positive)	(Yes, Negative)	(No, Positive)	(No, Negative)
Yes, positive	458	140	110	49
Yes, negative	171	182	56	87
No, positive	184	75	531	281
No, negative	85	97	338	554

<sup>a</sup> $M$ , membership;  $A$ , attitude.

Let  $y_{itv}$  be the response of subject  $i$  at interview time  $t$  on variable  $v$ , where  $v = M$  for membership and  $v = A$  for attitude. The logit model

$$\text{logit}[P(Y_{itv} = 1 | u_{iv})] = \beta_{tv} + u_{iv}$$

is a multivariate form of the Rasch-type model (12.4).

## 12.5.1 Matched pairs with a bivariate binary response

Here,  $(u_{iM}, u_{iA})$  is a bivariate random effect that describes subject heterogeneity for (membership, attitude). We assume that the  $\{(u_{iM}, u_{iA})\}$  iid bivariate  $N(0, \Sigma)$ , with possibly different variances and nonzero correlation.

The ML fit yields  $\hat{\beta}_{2M} - \hat{\beta}_{1M} = 0.379$  ( $SE = 0.075$ ) and  $\hat{\beta}_{2A} - \hat{\beta}_{1A} = 0.176$  ( $SE = 0.058$ ). The estimated correlation between the random effects is 0.30. Their estimated standard deviations are  $\hat{\sigma}_1 = 3.1$  for  $\{u_{iM}\}$  and  $\hat{\sigma}_2 = 1.5$  for  $\{u_{iA}\}$ .

Since these are quite different, the relative sizes of membership and attitude effects differ for marginal and conditional models.

## 12.5.1 Matched pairs with a bivariate binary response

The marginal effect is attenuated more for membership, since  $0.176 < 0.379$ .

For this conditional model, the ratio of estimated odds ratios is  $\exp(0.379) / \exp(0.176) = 1.22$ .

For the marginal model, the estimated odds ratios use the marginal distributions of each variable at each time (e.g., it is  $(1392/2006)/(1253/2145) = 1.188$  for membership), and the ratio of estimated odds ratios is  $1.188/1.133 = 1.05$ .



## 12.5.1 Matched pairs with a bivariate binary response

Integrating over the estimated random effects distribution yields fitted values for the 16 possible sequences of responses in Table 12.8. The deviance of  $G^2 = 5.5$  ( $df = 8$ ) compares the 16 observed counts to their fitted values. The model, which describes 15 multinomial probabilities with seven parameters, fits well.

The model constraining the random effects to be uncorrelated fits poorly ( $G^2 = 97.5$ ,  $df = 9$ ). The model constraining the random effects to be perfectly correlated is equivalent to having a single random effect  $u_i$  for each subject. The model is then a Rasch-type model with four items that are the combinations of interviews and variables. That model fits very poorly ( $G^2 = 655.5$ ,  $df = 10$ ).

## 12.5.2 Continuation-ratio logits for cluster ordinal outcomes

For continuation-ratio logit models with ordinal responses, the logits refer to independent binomial variates (Section 7.4.3).

Thus, binary logit random effects models apply to clustered ordinal responses using continuation-ratio logits. For observation  $t$  in cluster  $i$ , let  $w_{ij} = P(Y_{it} = j | Y_{it} \geq j, u_{ij})$ . The continuation-ratio logits are  $\{\text{logit}(w_{ij}), j = 1, 2, \dots, I - 1\}$ .

Let  $n_{ij}$  be the number of subjects in cluster  $i$  making response  $j$ ,  $n_i = \sum_{j=1}^I n_{ij}$  and treat  $(n_{i1}, n_{i2}, \dots, n_{i,I-1})$  as multinomial is equivalent to treating them as a sequential set of independent binomial variates, where  $n_{ij}$  is  $\text{bin}(n_i - \sum_{h < j} n_{ih}, w_{ij})$  for  $j = 1, 2, \dots, I - 1$ .

## 12.5.2 Continuation-ratio logits for cluster ordinal outcomes

**Toxicity study.** To examine the developmental effects of ethylene glycol (EG, 乙二醇) by administering one of four dosages (0, 0.75, 1.50, 3.00 g/kg) to pregnant rodents. The four dose groups had (25, 24, 22, 23) pregnant rodents. The clusters are litters of mice. Three possible outcomes (dead/resorption, malformation, normal) for each fetus. See Table 12.9.

The continuation-ratio logit is natural here since categories are hierarchically related; an animal must survive before a malformation can take place.

## 12.5.2 Continuation-ratio logits for cluster ordinal outcomes

**TABLE 12.9 Response Counts for 94 Litters of Mice on (Number Dead, Number Malformed, Number Normal)**

Dose = 0.00 g/kg	Dose = 0.75 g/kg	Dose = 1.50 g/kg	Dose = 3.00 g/kg
(1, 0, 7), (0, 0, 14)	(0, 3, 7), (1, 3, 11)	(0, 8, 2), (0, 6, 5)	(0, 4, 3), (1, 9, 1)
(0, 0, 13), (0, 0, 10)	(0, 2, 9), (0, 0, 12)	(0, 5, 7), (0, 11, 2)	(0, 4, 8), (1, 11, 0)
(0, 1, 15), (1, 0, 14)	(0, 1, 11), (0, 3, 10)	(1, 6, 3), (0, 7, 6)	(0, 7, 3), (0, 9, 1)
(1, 0, 10), (0, 0, 12)	(0, 0, 15), (0, 0, 11)	(0, 0, 1), (0, 3, 8)	(0, 3, 1), (0, 7, 0)
(0, 0, 11), (0, 0, 8)	(2, 0, 8), (0, 1, 10)	(0, 8, 3), (0, 2, 12)	(0, 1, 3), (0, 12, 0)
(1, 0, 6), (0, 0, 15)	(0, 0, 10), (0, 1, 13)	(0, 1, 12), (0, 10, 5)	(2, 12, 0), (0, 11, 3)
(0, 0, 12), (0, 0, 12)	(0, 1, 9), (0, 0, 14)	(0, 5, 6), (0, 1, 11)	(0, 5, 6), (0, 4, 8)
(0, 0, 13), (0, 0, 10)	(1, 1, 11), (0, 1, 9)	(0, 3, 10), (0, 0, 13)	(0, 5, 7), (2, 3, 9)
(0, 0, 10), (1, 0, 11)	(0, 1, 10), (0, 0, 15)	(0, 6, 1), (0, 2, 6)	(0, 9, 1), (0, 0, 9)
(0, 0, 12), (0, 0, 13)	(0, 0, 15), (0, 3, 10)	(0, 1, 2), (0, 0, 7)	(0, 5, 4), (0, 2, 5)
(1, 0, 14), (0, 0, 13)	(0, 2, 5), (0, 1, 11)	(0, 4, 6), (0, 0, 12)	(1, 3, 9), (0, 2, 5)
(0, 0, 13), (1, 0, 14)	(0, 1, 6), (1, 1, 8)		(0, 1, 11)
(0, 0, 14)			

## 12.5.2 Continuation-ratio logits for cluster ordinal outcomes

For litter  $i$  in dose group  $d$ , let

$\text{logit}(w_{i(d)1})$ : the continuation-ratio logit for the prob. of death;

$\text{logit}(w_{i(d)2})$ : the continuation-ratio logit for the cond. prob. of malformation, given survival ( $i(d)$  represents litter  $i$  nested within dose  $d$ ).

Let  $x_d$  be the dosage for group  $d$ . Assume random effects  $\mathbf{u}_{i(d)} = (u_{i(d)1}, u_{i(d)2})$  sampled from  $N(0, \Sigma_d)$ , which allows for differing amounts of overdispersion for the probability of death and for the probability of malformation, given survival. A model also permitting different fixed effects for each is

$$\text{logit}(w_{i(d)j}) = u_{i(d)j} + \alpha_j + \beta_j x_d, \quad j = 1, 2, \quad d = 1, \dots, 4. \quad (12.17)$$

## 12.5.2 Continuation-ratio logits for cluster ordinal outcomes

**TABLE 12.10 Comparisons of Log Likelihoods for Multivariate Random Effects Models for Developmental Toxicity Study**

Model	Number of Parameters	Change in Parameters	Change in Log Likelihood
Dose-specific $\Sigma_i$	16	—	—
$\Sigma_i$ , Common $\alpha, \beta$	14	2	28.4
Common $\Sigma$	7	9	7.4
Common $\Sigma, \rho = 0$	6	10	7.4
Univariate $\sigma^2$	5	11	16.7

See the textbook for more details on the discussion of the four models.

## 12.5.2 Continuation-ratio logits for cluster ordinal outcomes

Table 12.10 reports the change in the maximized log likelihood from fitting four special cases of this model:

- Common intercept and slope for the two logits:  $\alpha_1 = \alpha_2$ ,  $\beta_1 = \beta_2$ ;
- Common covariance matrix for the four doses:  $\Sigma_1 = \Sigma_2 = \Sigma_3 = \Sigma_4$ ;
- Common covariance matrix and uncorrelated random effects;
- Univariate common variance component across dose:  $U_{i(d)1} = U_{i(d)2}$ ,  $\sigma_d = \sigma$ .

# Appendix: SAS Codes

**TABLE A.23 SAS Code for Fitting Model (12.3) for Matched Pairs to Table 12.1**

---

```
data matched;
input case occasion response count @@;
datalines;
2 0 1 794    1 1 1 794    2 0 1 150    2 1 0 150
3 0 0 86     3 1 1 86     4 0 0 570    4 1 0 570
;
proc nlmixed;
  eta=alpha+beta*occasion+u;  p=exp(eta) / (1+exp(eta));
  model  response~binary(p);
  random u~normal(0, sigma*sigma) subject=case;
  replicate count;
```

---



**TABLE A.24 SAS Code for GLMM Analysis of Election Data in Table 12.2**

---

```
data vote;
input y n;
case=_n_;
datalines;
  1    5
16   32
...
  1    4
;
proc nlmixed;
  eta=alpha+u;  p=exp(eta) / (1+exp(eta));
  model y~binomial(n,p);
  random u~normal (0, sigma*sigma) subject=case;
  predict p out=new;
proc print data=new;
```

---

**TABLE A.25 SAS Code for GLMM Modeling of Opinions in Table 10.13**


---

```

data new;
input sex poor single any count;
datalines;
1 1 1 1 342
...
2 0 0 0 457
;
data new; set new;
sex=sex-1; case=_n_;
q1=1; q2=0; resp=poor; output;
q1=0, q2=1; resp=single; output;
q1=0; q2=0; resp=any; output;
drop poor single any;
proc nlmixed qpoints=50;
parms alpha=0 beta1=.8 beta2=.3 gamma=0 sigma=8.6;
eta=alpha+beta1*q1+beta2*q2+gamma*sex+u;
p=exp(eta) / (1+exp(eta));
model resp ~ binary(p);
random u ~ normal(0, sigma*sigma) subject=case;
replicate count;

```

---

**TABLE A.26 SAS Code for GLMM for Leading Crowd Data in Table 12.8**


---

```

data crowd;
input mem1 att1 mem2 att2 count;
datalines;
  1  1  1  1 458
...
  0  0  0  0 554
;
data new; set crowd;
  case=_n_;
  x1m=1; x1a=0; x2m=0; x2a=0; var=1; resp=mem1; output;
  x1m=0; x1a=1; x2m=0; x2a=0; var=0; resp=att1; output;
  x1m=0; x1a=0; x2m=1; x2a=0; var=1; resp=mem2; output;
  x1m=0; x1a=0; x2m=0; x2a=1; var=0; resp=att2; output;
drop mem1 att1 mem2 att2;
proc nlmixed data=new;
  eta=beta1m*x1m+beta1a*x1a+beta2m*x2m+beta2a*x2a+um*var+
    ua*(1-var);
  p=exp(eta) / (1+exp(eta));
  model resp ~ binary(p);
  random um ua ~ normal([0,0],[s1*s1, cov12, s2*s2]) subject=case;
  replicate count;
  estimate 'mem change' beta2m-beta1m; estimate 'att change'
    beta2a-beta1a;

```

---

**TABLE A.27 SAS Code for Overdispersion Analysis of Table 4.5**


---

```

data moore;
input  litter  group  n  y  @@;
    z2=0;  z3=0;  z4=0;
    if group=2 then z2=1;  if group=3 then z3=1;  if group=4
        then z4=1;
datalines;
  1  1 10  1      2  1 11  4      3  1 12  9      4  1  4  4
...
55  4 14  1      56  4  8  0      57  4  6  0      58  4 17  0
;
proc logistic;
    model y/n=z2 z3 z4 / scale=williams;
proc logistic;
    model y/n=z2 z3 z4 / scale=pearson;
proc nlmixed  qpoints=200;
    eta=alpha+beta2*z2+beta3*z3+beta4*z4+u;
    p=exp(eta) / (1+exp(eta));
    model y ~ binomial(n,p);
    random u ~ normal(0, sigma*sigma) subject=litter;

```

---

**TABLE A.28 SAS Code for Fitting Models to Murder Data in Table 13.6**


---

```

data new;
input white black other response;
datalines;
1070 119 55 0
    60 16 5 1
...
    1 0 0 6
;
data new; set new; count=white; race=0; output;
    count=black; race=1; output; drop white black other;
data new2; set new; do i=1 to count; output; end; drop i;
proc genmod data=new2;
    model response=race / dist=negbin link=log;
proc genmod data=new2;
    model response=race / dist=poi link=log scale=pearson;
data new; set new; case=_n_;
proc nlmixed data=new qpoints=400;
    parms alpha=-3.7 beta=1.90 sigma=1.6;
    eta=alpha+beta*race+u; mu=exp(eta);
    model response ~ poisson(mu);
    random u ~ normal(0, sigma*sigma) subject=case;
    replicate count;

```

---