Categorical Data Analysis

Chapter 11

Deyuan Li
School of Management
Fudan University

Fall 2015

11. Analyzing Repeated Categorical Response Data

Repeated categorical response data:

(1) response variable for each subject is measured repeatedly (at several times or under various conditions); dependence among the repeated responses.

For example, blood measures of one person at several times.

(2) responses of subjects in a set or cluster; dependence in the set or cluster.

For example, survival response for each fetus in a litter for a sample of pregnant mice.

Outline

11.1 Comparing marginal distributions: multiple responses

Usually, the multivariate dependence among repeated responses is of less interest than their marginal distributions.

For instance, in treating a chronic condition with some treatment, the primary goal might be to study whether the probability of success increases over the T weeks of a treatment period.

In Sections 10.2.1 and 10.3 we compared marginal distributions for matched pairs ($T = 2$) using models that apply directly to the marginal distributions.

In this section we extend this approach to $T > 2$.

11.1.1 Binary marginal models and marginal homogeneity

Denote T binary responses by $(Y_1, Y_2, ..., Y_T)$. The marginal logit model is

$$\text{logit}[P(Y_t = 1)] = \alpha + \beta_t, \quad t = 1, 2, ..., T, \qquad (11.1)$$

with $\beta_T = 0$. In particular, $\beta_1 = \beta_2 = ... = \beta_T$ implies marginal homogeneity (MH).

For the possible outcome $\mathbf{i} = (i_1, ..., i_T)$ with $i_t = 1$ or 0, let the joint mass probability be

$$\pi_{\mathbf{i}} = P(Y_1 = i_1, Y_2 = i_2, ..., Y_T = i_T),$$

and the set of mass probabilities be $\boldsymbol{\pi} = \{\pi_{\mathbf{i}}\}$. Of course, $\#\boldsymbol{\pi} = 2^T$.

11.1.1 Binary marginal models and marginal homogeneity

Let $n_{\mathbf{i}}$ denote the sample count of outcome $\mathbf{i}$. The kernel of the log likelihood is

$$L(\pi) = \sum_{\mathbf{i}} n_{\mathbf{i}} \log \pi_{\mathbf{i}}.$$

To test marginal homogeneity, the likelihood-ratio test uses

$$-2[L(\hat{\pi}^{MH}) - L(\mathbf{p})] = 2\sum_{\mathbf{i}} n_{\mathbf{i}} \log(p_{\mathbf{i}}/\hat{\pi}^{MH}_{\mathbf{i}}) \sim \chi^2_{df},$$

where $df = T - 1$, $\hat{\pi}^{MH}$ and $\mathbf{p}$ are the ML estimators under and without MH assumption, respectively.

11.1.2 Crossover drug comparison example

Table 11.1: each subject used each of three drugs for treatment of a chronic condition at three times (first drug A, then B and C). Responses are binary (favorable or unfavorable), $2^3 = 8$ cases.

TABLE 11.1 Responses to Three Drugs in a Crossover Study

	Drug A Favorable		Drug A Unfavorable	
	B Favorable	B Unfavorable	B Favorable	B Unfavorable
C Favorable	6	2	2	6
C Unfavorable	16	4	4	6

Source: Reprinted with permission from the Biometric Society (Grizzle et al. 1969).

The sample proportion favorable was $(0.61, 0.61, 0.35)$ for drugs A, B, C. The likelihood-ratio statistic for testing marginal homogeneity is 5.95 ($df = 2$).

For more analysis details, see the book.

11.1.3 Modeling margins of a multicategory response

Multinomial response. With baseline-category logits for I outcome categories, the saturated model is

$$\log[P(Y_t = j)/P(Y_t = I)] = \beta_{tj}, \quad t = 1, ..., T, \quad j = 1, ..., I-1.$$

Marginal homogeneity (MH): $\beta_{1j} = \beta_{2j} = ... = \beta_{Tj}$, $j = 1, ..., I-1$. The likelihood-ratio test for MH is similar but $df = (T-1)(I-1)$.

Ordinal response. One unsaturated model:

$$\text{logit}[P(Y_t \leq j)] = \alpha_j + \beta_t, \quad t = 1, ..., T, \quad j = 1, ..., I-1.$$

with $\beta_T = 0$ and $\alpha_1 < \alpha_2 < \cdots < \alpha_{I-1}$.

Marginal homogeneity (MH): $\beta_1 = \beta_2 = ... = \beta_T$. The likelihood-ratio test for MH has $df = (T-1)$.

11.1.4 Wald and Generalized CMH Score Tests of Marginal Homogeneity

Wald Test.
$p_j(t)$: sample proportion in category j for response Y_t. Define

$$\bar{p}_j = \sum_t p_j(t)/T, \quad d_j(t) = p_j(t) - \bar{p}_j.$$

$\mathbf{d}$: the vector of $\{d_j(t), t = 1, ..., T-1, j = 1, ..., I-1\}$,
$\hat{\mathbf{V}}$: the estimated covariance matrix of $\sqrt{n}\mathbf{d}$.
Bhapkar (1973). proposed the Wald statistic

$$W = n\mathbf{d}'\hat{\mathbf{V}}^{-1}\mathbf{d},$$

which converges to $\chi^2(df)$ with $df = (I-1)(T-1)$.

Outline

11.2.1 Longitudinal mental depression example

Mental Depression Example:

- Compare a new drug with a standard drug for treatment of subjects suffering mental depression.
- Subjects were classified into two initial diagnosis groups according to whether severity of depression was mild or severe.
- In each group, subjects were randomly assigned to one of the two drugs.
- Following 1 week, 2 weeks, and 4 weeks of treatment, each subject was classified as normal or abnormal.

11.2.1 Longitudinal mental depression example

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

		Response at Three Times[a]							
Diagnosis	Treatment	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

[a] N, normal; A, abnormal.

$T = 3$. 2^3 table for each group;

12 marginal distributions for three repeated observations for each of the four groups.

11.2.1 Longitudinal mental depression example

Denote
$Y_t = 1$ for normal and 0 for abnormal;
$s = 1$ for the severity of the initial diagnosis and 0 for mild;
$d = 1$ for the new drug and 0 for standard;
$t = 1, 2, 4$ for the time of measurement.

TABLE 11.3 Sample Marginal Proportions of Normal Response for Depression Data of Table 11.2

		Sample Proportion		
Diagnosis	Treatment	Week 1	Week 2	Week 4
Mild	Standard	0.51	0.59	0.68
	New drug	0.53	0.79	0.97
Severe	Standard	0.21	0.28	0.46
	New drug	0.18	0.50	0.83

11.2.1 Longitudinal mental depression example

For instance,

$$\frac{16+13+9+3}{16+13+9+3+14+4+15+6} = 0.51.$$

The sample proportion of normal responses

(1) increased over time for each group;

(2) increased at a faster rate for the new drug than the standard, for each fixed initial diagnosis;

(3) was higher for the mild than the severe initial diagnosis, for each treatment at each occasion.

11.2.1 Longitudinal mental depression example

The marginal logit model

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

has the main effects of the explanatory variables (severity of initial diagnosis and drug) and of the variable (time) that specifies the different components of the multivariate response.

The natural sampling assumption is multinomial for the eight cells in the 2^3 cross-classification of the three responses, independently for the four groups.

11.2.1 Longitudinal mental depression example

A check of model fit compares the 32 cell counts to their ML fitted values.

Since the model describes 12 marginal logits using four parameters, residual $df = 8$. The deviance $G^2 = 34.6$.
$\Rightarrow$ Poor fit but not surprising.

The model assumes a common rate of improvement β_3, but the sample shows a higher rate for the new drug.

New drug: $0.53 \rightarrow 0.79 \rightarrow 0.97$ increases quickly;
Old drug: $0.51 \rightarrow 0.59 \rightarrow 0.68$ increases slowly.

11.2.1 Longitudinal mental depression example

A more realistic model permits the time effect to differ by drug,

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

Estimators are below. $\hat{\beta}_1 = -1.29$ (SE=0.14),
$\hat{\beta}_2 = -0.06$ (SE=0.22), $\hat{\beta}_3 = 0.48$ (SE=0.12),
$\hat{\beta}_4 = 1.01$ (SE=0.18), $\hat{\beta}_3 + \hat{\beta}_4 = 1.49$ (SE=0.14).

The severity of initial diagnosis estimate $\hat{\beta}_1 = -1.29$ implies that for each drug-time combination, the estimated odds of a normal response when the initial diagnosis was severe equal $\exp(-1.29) = 0.27$ times the estimated odds when the initial diagnosis was mild. Explanations for other estimator are similar.

This model fits much better, with $G^2 = 4.2$ ($df = 7$).

11.2.2 Modeling a repeated multinomial response

At observation t, the marginal response distribution has $I-1$ logits. With nominal responses, baseline-category logit models describe the odds of each outcome relative to a baseline. For ordinal responses, one might use cumulative logit models.

For a particular marginal logit, a model has the form

$$\text{logit}_j(t) = \alpha_j + \beta_j' \mathbf{x}_t, \quad j = 1, ..., I-1, \ t = 1, 2, ...$$

For an ordinal response, perhaps $\text{logit}_j(t) = \text{logit}[P(Y_t \leq j)]$. Then, β_j may simplify to β, in which case the model takes the proportional odds form with the same effects for each logit.

11.2.3 Insomnia example

Table 11.4 shows results of a randomized, double-blind clinical trial comparing an active hypnotic drug（催眠药）with a placebo in patients who have insomnia problems（失眠症）.

The response is the patient's reported time (in minutes) to fall asleep after going to bed. Patients responded before and following a two-week treatment period.

The two treatments, active and placebo, form a binary explanatory variable.

The subjects receiving the two treatments were independent samples.

11.2.3 Insomnia example

TABLE 11.4 Time to Falling Asleep, by Treatment and Occasion

		Time to Falling Asleep			
		Follow-up			
Treatment	Initial	< 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

TABLE 11.5 Sample Marginal Distributions of Table 11.4

		Response			
Treatment	Occasion	< 20	20–30	30–60	> 60
Active	Initial	0.101	0.168	0.336	0.395
	Follow-up	0.336	0.412	0.160	0.092
Placebo	Initial	0.117	0.167	0.292	0.425
	Follow-up	0.258	0.242	0.292	0.208

11.2.3 Insomnia example

From the initial to follow-up occasion, time to falling asleep seems to shift downward for both treatments. The degree of shift seems greater for the active treatment, indicating possible interaction.

Consider the proportional odds model

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

$t =$ occasion (0=initial and 1=follow-up); $x =$ treatment (0= placebo and 1=active).

For ML model fitting, $G^2 = 8.0$ ($df = 6$). The ML estimates are $\hat{\beta}_1 = 1.074$ (SE=0.162), $\hat{\beta}_2 = 0.046$ (SE=0.236), $\hat{\beta}_3 = 0.662$ (SE=0.244).

11.2.3 Insomnia example

This shows evidence of interaction.

At the initial observation, the estimated odds that time to falling asleep for the active treatment is below any fixed level equal $\exp(0.046) = 1.04$ times the estimated odds for the placebo treatment; at the follow-up observation, the effect is $\exp(0.046 + 0.662) = 2.03$.

In other words, initially the two groups had similar distributions, but at the follow-up those with the active treatment tended to fall asleep more quickly.

11.2.4 Comparisons that control for initial response

Let Y_2 denote the follow-up response, for treatment x with initial response y_1. In the model

$$\text{logit}[P(Y_2 \leq j)] = \alpha_j + \beta_1 x + \beta_2 y_1, \qquad (11.7)$$

β_1 compares the follow-up distributions for the treatments, controlling for initial observation. This is an analog of an analysis-of-covariance model, with ordinal rather than continuous response.

It is an example of a *transitional model*, discussed in the final section of this chapter.

Outline

11.3.1 GEE methodology: basic idea

Multivariate response $(Y_1, Y_2, ..., Y_T)$, where T sometimes varies by subject.

In the univariate case, the quasilikelihood method specifies a variance function $v(\mu)$ describing how $var(Y)$ depends on $\mu = E(Y)$.

We requires a working guess for the correlation structure among $\{Y_t\}$. The estimates are solutions of quasi-likelihood equations called generalized estimating equations（广义估计方程）. The method is often referred to as the GEE method.

Liang and Zeger (1986) proposed it for marginal modeling with GLMs. We outline concepts here and give more details in Section 11.4.

11.3.1 GEE methodology: basic idea

The GEE approach utilizes an assumed covariance structure for $(Y_1, Y_2, ..., Y_T)$, specifying a variance function and a pairwise correlation pattern, without assuming a particular multivariate distribution.

The GEE estimates of model parameters are valid even if one mis-specifies the covariance structure. Consistency depends on the first moment but not the second.

Three simple choices for the correlation among $\{Y_t\}$.

1. $\{Y_t\}$ are pairwise independent (naive);
2. corr(Y_t, Y_s) are identical for all s, t (exchangeable correlation);
3. corr$(Y_t, Y_s) = \rho^{|t-s|}$ (autoregressive correlation).

11.3.2 Longitudinal mental depression example

TABLE 11.6 Output from Using GEE to Fit Logit Model to Table 11.2

Initial Parameter Estimates			GEE Parameter Estimates Empirical Std Error Estimates		
Parameter	Estimate	Std Error	Parameter	Estimate	Std Error
Intercept	−0.0280	0.1639	Intercept	−0.0280	0.1742
diagnose	−1.3139	0.1464	diagnose	−1.3139	0.1460
drug	−0.0596	0.2222	drug	−0.0596	0.2285
time	0.4824	0.1148	time	0.4824	0.1199
drug*time	1.0174	0.1888	drug*time	1.0174	0.1877

Working Correlation Matrix

	Col1	Col2	Col3
Row1	1.0000	0.0000	0.0000
Row2	0.0000	1.0000	0.0000
Row3	0.0000	0.0000	1.0000

11.3.2 Longitudinal mental depression example

The GEE analysis provides similar results, regardless of the choice of working correlation structure.

GEE estimated slope (on the logit scale) for the standard drug is $\hat{\beta}_3 = 0.48$ (SE=0.12). For the new drug the slope increases by $\hat{\beta}_4 = 1.02$ (SE=0.19). Table 11.6 shows results using the independence working correlations.

The initial estimates and standard errors there are those that apply if the repeated responses are truly independent. They equal those obtained by using ordinary logistic regression with $3 \times 340 = 1020$ independent observations rather than treating the data as three dependent observations for each of 340 subjects.

Outline

11.5 Markov chains: transitional modeling

When Y_t denotes the response at time t, $t = 0, 1, 2, ...$, the indexed family of random variables $(Y_0, Y_1, Y_2, ...)$ is a stochastic process（随机过程）.

The *state space*（状态空间）of the process is the set of possible values for Y_t . The value Y_0 is the *initial state*（初始状态）.

When the state space is categorical and observations occur at a discrete set of times, $\{Y_t\}$ has *discrete state space* and *discrete time*（离散时间离散状态的随机过程）.

11.5.1 Transitional models

Let $f(y_0, ..., y_T)$ denote the joint probability mass function of $(Y_0, ..., Y_T)$ (ignoring, for now, explanatory variables). *Transitional models*（转移模型）use the factorization

$$f(y_0, ..., y_T) = f(y_0)f(y_1|y_0)f(y_2|y_0, y_1) \cdots f(y_T|y_0, y_1, ..., y_{T-1}).$$

Unlike the marginal models, this modeling is conditional on previous responses.

We introduce discrete-time *Markov chains*（马尔可夫链）, a simple stochastic process having discrete state space. Many transitional models have Markov chain structure for at least part of the model.

11.5.2 First order Markov chains

First order Markov chains,

$$f(y_0, ..., y_T) = f(y_0)f(y_1|y_0)f(y_2|y_1)\cdots f(y_T|y_{T-1}).$$

Consequently, kth-order Markov chain is well defined.

Define $\pi_{j|i}(t) = P(Y_t = j|\, Y_{t-1} = i)$. Call $\{\pi_{j|i}(t)\}$ *transition probabilities*.

The $I \times I$ matrix $\{\pi_{j|i}(t) : i = 1, 2, ..., I,\ j = 1, 2, ..., I\}$ is a *transition probability matrix*（转移概率矩阵）.

11.5.2 First order Markov chains

Then, the joint distribution for a Markov chain depends only on one-step transition probabilities and the marginal distribution for the initial state.

It also follows that the joint distribution satisfies loglinear model

$$(Y_0 Y_1, Y_1 Y_2, ..., Y_{T-1} Y_T).$$

Consider ML estimation of transition probabilities. Let $n_{ij}(t)$ denote the number of transitions from state i at time $t-1$ to state j at time t.

11.5.2 First order Markov chains

For fixed t, $\{n_{ij}(t)\}$ form the two-way marginal table for dimensions $t-1$ and t of an I^{T+1} contingency table. For the $n_{i+}(t)$ subjects in category i at time $t-1$, suppose that $\{n_{ij}(t), j = 1, ..., I\}$ have a multinomial distribution with parameters $\{\pi_{j|i}(t)\}$.

Let $\{n_{i0}\}$ denote the initial counts. Suppose that they also have a multinomial distribution, with parameters π_{i0}. If subjects behave independently, the likelihood function for first order Markov chains is proportional to

$$\left(\Pi_{i=1}^{I}\pi_{i0}^{n_{i0}}\right)\left\{\Pi_{t=1}^{T}\Pi_{i=1}^{I}\left[\Pi_{j=1}^{I}\pi_{j|i}(t)^{n_{ij}(t)}\right]\right\}.$$

Then the ML estimator are

$$\hat{\pi}_{j|i}(t) = n_{ij}(t)/n_{i+}(t).$$

11.5.3 Respiratory (呼吸) illness example

Table 11.7 refers to a longitudinal study at Harvard of effects of air pollution on respiratory illness in children. The children were examined annually at ages 9 through 12 and classified according to the presence or absence of wheeze.

Denote the binary response (wheeze (哮喘), no wheeze) by Y_t at age $t = 9, 10, 11, 12$.

TABLE 11.7 Results of Breath Test at Four Ages[a]

Y_9	Y_{10}	Y_{11}	Y_{12}	Count	Y_9	Y_{10}	Y_{11}	Y_{12}	Count
1	1	1	1	94	2	1	1	1	19
1	1	1	2	30	2	1	1	2	15
1	1	2	1	15	2	1	2	1	10
1	1	2	2	28	2	1	2	2	44
1	2	1	1	14	2	2	1	1	17
1	2	1	2	9	2	2	1	2	42
1	2	2	1	12	2	2	2	1	35
1	2	2	2	63	2	2	2	2	572

[a] 1, wheeze; 2, no wheeze.

11.5.2 First order Markov chains

The loglinear model $(Y_9 Y_{10}, Y_{10} Y_{11}, Y_{11} Y_{12})$

$$\log \mu_{ijkl} = \lambda_0 + \lambda_i^{Y_9} + \lambda_j^{Y_{10}} + \lambda_k^{Y_{11}} + \lambda_l^{Y_{12}} + \lambda_{ij}^{Y_9 Y_{10}} + \lambda_{jk}^{Y_{10} Y_{11}} + \lambda_{kl}^{Y_{11} Y_{12}}$$

represents a first-order Markov chain. It fits poorly, with $G^2 = 122.9$ $(df = 8)$.

The model $(Y_9 Y_{10} Y_{11}, Y_{10} Y_{11} Y_{12})$

$$\log \mu_{ijkl} = \lambda_0 + \lambda_i^{Y_9} + \lambda_j^{Y_{10}} + \lambda_k^{Y_{11}} + \lambda_l^{Y_{12}} + \lambda_{ij}^{Y_9 Y_{10}} + \lambda_{jk}^{Y_{10} Y_{11}} + \lambda_{kl}^{Y_{11} Y_{12}} + \lambda_{ik}^{Y_9 Y_{11}} + \lambda_{jl}^{Y_{10} Y_{12}} + \lambda_{ijk}^{Y_9 Y_{10} Y_{11}} + \lambda_{jkl}^{Y_{10} Y_{11} Y_{12}}$$

represents a second-order Markov chain, satisfying conditional independence at ages 9 and 12, given states at ages 10 and 11.

11.5.2 First order Markov chains

This model $(Y_9 Y_{10} Y_{11}, Y_{10} Y_{11} Y_{12})$ also fits poorly, with $G^2 = 23.9$ $(df = 4)$. The poor fits may partly reflect subject heterogeneity, since these analysis ignore possibly relevant covariates such as parental smoking behavior.

The loglinear model $(Y_9 Y_{10}, Y_9 Y_{11}, Y_9 Y_{12}, Y_{10} Y_{11}, Y_{10} Y_{12}, Y_{11} Y_{12})$

$$\begin{aligned}\log \mu_{ijkl} = \lambda_0 &+ \lambda_i^{Y_9} + \lambda_j^{Y_{10}} + \lambda_k^{Y_{11}} + \lambda_l^{Y_{12}} \\ &+ \lambda_{ij}^{Y_9 Y_{10}} + \lambda_{ik}^{Y_9 Y_{11}} + \lambda_{il}^{Y_9 Y_{12}} \\ &+ \lambda_{jk}^{Y_{10} Y_{11}} + \lambda_{jl}^{Y_{10} Y_{12}} + \lambda_{kl}^{Y_{11} Y_{12}}\end{aligned}$$

that permits association at each pair of ages fits well, with $G^2 = 1.5$ $(df = 5)$ (see Table 11.8).

11.5.2 First order Markov chains

TABLE 11.8 Estimated Conditional Log Odds Ratios for Table 11.7

Association	Estimate	Simpler Structure
Y_9Y_{10}	1.81	1.75
$Y_{10}Y_{11}$	1.65	1.75
$Y_{11}Y_{12}$	1.85	1.75
Y_9Y_{11}	0.95	1.04
Y_9Y_{12}	1.05	1.04
$Y_{10}Y_{12}$	1.07	1.04

The association seems similar for pairs of ages 1 year apart, and somewhat weaker for pairs of ages more than 1 year apart. The simpler model in which

$$\lambda_{ij}^{Y_9Y_{10}} = \lambda_{ij}^{Y_{10}Y_{11}} = \lambda_{ij}^{Y_{11}Y_{12}} \text{ and } \lambda_{ij}^{Y_9Y_{11}} = \lambda_{ij}^{Y_9Y_{12}} = \lambda_{ij}^{Y_{10}Y_{12}}$$

fits well, with $G^2 = 2.3$ ($df = 9$).

11.5.4 Transitional models with explanatory

The joint mass function of T sequential responses is then

$$f(y_1, y_2, ..., y_T; \mathbf{x}) = f(y_1; \mathbf{x}) f(y_2|y_1; \mathbf{x}) \cdots f(y_T|y_1, y_2, ..., y_{T-1}; \mathbf{x}).$$

With binary y, for instance, one might specify a logistic regression model for each term in this factorization, with $y_t = 0, 1$,

$$f(y_t|y_1, ..., y_{t-1}; \mathbf{x}_t) = \frac{\exp[y_t(\alpha + \beta_1 y_1 + ... + \beta_{t-1} y_{t-1} + \boldsymbol{\beta}' \mathbf{x}_t)]}{1 + \exp[y_t(\alpha + \beta_1 y_1 + ... + \beta_{t-1} y_{t-1} + \boldsymbol{\beta}' \mathbf{x}_t)]}.$$

The model treats previous responses as explanatory variables. It is called regressive logistic model.

11.5.4 Transitional models with explanatory

The interpretation and magnitude of $\hat{\beta}$ depends on how many previous observations are in the model. Within-cluster effects may diminish markedly by conditioning on previous responses.

This is an important difference from marginal models, for which the interpretation does not depend on the specification of the dependence structure.

In the special case of first-order Markov structure, the coefficients of $(y_1, ..., y_{t-2})$ equal 0 in the model for y_t.

It may help to allow interaction between x_t and y_{t-1} in their effects on y_t.

11.5.5 Child's respiratory illness and maternal smoking

Table 11.9: Harvard study of air pollution and health. $s = 1$ means maternal smoking at the start of the study, 0 for others. Let y_t denote the response at age $t = 7, 8, 9, 10$.

TABLE 11.9 Child's Respiratory Illness by Age and Maternal Smoking

Child's Respiratory Illness			No Maternal Smoking		Maternal Smoking	
			Age 10		Age 10	
Age 7	Age 8	Age 9	No	Yes	No	Yes
No	No	No	237	10	118	6
		Yes	15	4	8	2
	Yes	No	16	2	11	1
		Yes	7	3	6	4
Yes	No	No	24	3	7	3
		Yes	3	2	3	1
	Yes	No	6	2	4	2
		Yes	5	11	4	7

11.5.5 Child's respiratory illness and maternal smoking

Consider the regressive logistic model

$$\text{logit}[P(Y_t = 1)] = \alpha + \beta_1 s + \beta_2 t + \beta_3 y_{t-1}, \quad t = 8, 9, 10.$$

Each subject contributes three observations to the model fitting. The data set consists of 12 binomials, for the $2 \times 3 \times 2$ combinations of (s, t, y_{t-1}).

For instance, for the combination $(0, 8, 0)$, $y_8 = 0$ for $237 + 10 + 15 + 4 = 266$ subjects and $y_8 = 1$ for $16 + 2 + 7 + 3 = 28$ subjects. The ML fit is

$$\text{logit}[\hat{P}[(Y_t = 1)] = -0.293 + 0.296s - 0.243t + 2.211y_{t-1},$$

with SE values $(0.846, 0.156, 0.095, 0.158)$.

Not surprisingly, the previous observation has a strong effect (since $\beta_3 = 2.211$).

11.5.5 Child's respiratory illness and maternal smoking

Given that and the child's age, there is slight evidence of a positive effect of maternal smoking: the likelihood-ratio statistic for $H_0 : \beta_1 = 0$ is 3.55 ($df = 1, P = 0.06$).

The model itself does not show any evidence of lack of fit ($G^2 = 3.1,\ df = 8$).

Appendix: SAS Codes

TABLE A.20 SAS Code for Testing Marginal Homogeneity with Crossover Study of Table 11.1

```
data crossover;
input  a  b  c  count  m111  m11p  m1p1  mp11  m1pp  m222  @@;
datalines;
1  1  1  6   1  0  0  0  0  0    1  1  2  16 -1  1  0  0  0  0
1  2  1  2  -1  0  1  0  0  0    1  2  2   4  1 -1 -1  0  1  0
2  1  1  2  -1  0  0  1  0  0    2  1  2   4  1 -1  0 -1  1  0
2  2  1  6   1  0 -1 -1  1  0    2  2  2   6  0  0  0  0  0  1
;
proc genmod;
  model  count = m111  m11p  m1p1  mp11  m1pp  m222 / dist = poi  link = identity;
proc catmod;  weight count;  response marginals;
  model  a*b*c = _response_ /freq;
  repeated drug 3;
```

TABLE A.21 SAS Code for Marginal Modeling of Depression Data in Table 11.2

```
data depress;
input  case  diagnose  drug  time  outcome  @@;   *  outcome = 1 is normal;
datalines;
  1  0  0  0  1    1  0  0  1  1    1  0  0  2  1
...
340  1  1  0  0  340  1  1  1  0  340  1  1  2  0
;
proc genmod descending;  class case;
  model;  outcome = diagnose  drug  time  drug*time / dist = bin  link = logit  type3;
  repeated  subject = case / type = exch  corrw;
proc n1mixed  qpoints = 200;
  parms  alpha = -.03  beta1 = -1.3  beta2 = -.06  beta3 = .48  beta4 = 1.02  sigma = .066;
  eta = alpha + beta1*diagnose + beta2*drug + beta3*time + beta4*drug*time + u;
  p = exp(eta) / (1 + exp(eta));
  model  outcome ~ binary(p);
  random  u ~ normal(0,  sigma*sigma)   subject = case;
```

TABLE A.22 SAS Code for GEE and Random Intercept Cumulative Logit Analysis of Insomnia Data in Table 11.4

```
data francom;
  input  case  treat  time  outcome  @@;
datalines;
  1  1  0  1    1  1  1  1
...
239  0  0  4  239  0  1  4
;
proc genmod;  class case;
  model outcome=treat time treat*time/dist=multinomial
     link=clogit;
  repeated subject=case/type=indep corrw;
proc n1mixed  qpoints=40;
  bounds  i2>0;  bounds i3>0;
  eta1=i1+treat*beta1+time*beta2+treat*time*beta3+u;
  eta2=i1+i2+treat*beta1+time*beta2+treat*time*beta3+u;
  eta3=i1+i2+i3+treat*beta1+time*beta2+treat*time*beta3+u;
  p1=exp(eta)/(1+exp(eta1));
  p2=exp(eta2)/(1+exp(eta2))-exp(eta1)/(1+exp(eta1));
  p3=exp(eta3)/(1+exp(eta3))-exp(eta2)/(1+exp(eta2));
  p4=1-exp(eta3)/(1+exp(eta3));
  11=y1*log(p1)+y2*log(p2)+y3*log(p3)+y4*log(p4);
  model y1~general(11);
  estimate 'interc2' i1+i2; * this is alpha_2 in model, and
     i1 is alpha_1;
  estimate 'interc3' i1+i2+i3; * this is alpha_3 in model;
  random u~normal(0, sigma*sigma) subject=case;
```