CHAPTER 9

Modeling Correlated, Clustered Responses

Many studies observe the response variable for each subject repeatedly, at several times (such as in *longitudinal* studies) or under various conditions. Repeated measurement occurs commonly in health-related applications. For example, a physician might evaluate patients at weekly intervals regarding whether a new drug treatment is successful. Repeated observations on a subject are typically correlated.

Correlated observations can also occur when the response variable is observed for *matched sets* of subjects. For example, a study of factors that affect childhood obesity might sample families and then observe the children in each family. A matched set consists of children within a particular family. Children from the same family may tend to respond more similarly than children from different families. Another example is a (survival, nonsurvival) response for each fetus in a litter of a pregnant mouse, for a sample of pregnant mice exposed to various dosages of a toxin. Fetuses from the same litter are likely to be more similar than fetuses from different litters.

We will refer to a matched set of observations as a *cluster*. For repeated measurement on subjects, the set of observations for a given subject forms a cluster. Observations within a cluster are usually positively correlated. Analyses should take the correlation into account. Analyses that ignore the correlation can estimate model parameters well, but the standard error estimators can be badly biased.

9.1 MARGINAL MODELS VERSUS CONDITIONAL MODELS

As with independent observations, with clustered observations models focus on how the probability of a particular outcome (e.g., "success") depends on explanatory variables. For longitudinal studies, one explanatory variable is the time of each observation. For instance, in treating a chronic condition (such as a phobia) with one of two treatments, the model might describe how the probability of success depends on the treatment and on the length of time for which that treatment has been used.

9.1.1 Marginal Models for a Clustered Binary Response

Let T denote the number of observations in each cluster. (In practice, the number of observations often varies by cluster, but it is simpler to use notation that ignores that.) Denote the T observations by $(Y_1, Y_2, \ldots, Y_T)$.

For binary responses, the T success probabilities $\{P(Y_1 = 1), P(Y_2 = 1), \ldots, P(Y_T = 1)\}$ are marginal probabilities of a T-dimensional contingency table that cross classifies the T observations. Marginal models describe how the logits of the marginal probabilities, $\{\text{logit}[P(Y_t = 1)]\}$, depend on explanatory variables. To illustrate the models and questions of interest, let us consider an example to be analyzed in Section 9.2.

9.1.2 Example: Longitudinal Study of Treatments for Depression

Table 9.1 refers to a longitudinal study comparing a new drug with a standard drug for treating subjects suffering mental depression. Subjects were classified into two groups according to whether the initial severity of depression was mild or severe.

Table 9.1. Cross-classification of Responses on Depression at Three Times (N = Normal, A = Abnormal) by Treatment and Diagnosis Severity

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

Source: Reprinted with permission from the Biometric Society (G. G. Koch et al., *Biometrics*, **33**: 133–158, 1977).

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's extent of suffering from mental depression was classified as normal or abnormal.

Table 9.1 shows four groups, the combinations of categories of two explanatory variables – treatment type and severity of depression. Since the study observed the binary response (depression assessment) at $T = 3$ occasions, Table 9.1 shows a $2 \times 2 \times 2$ table for each group. The three depression assessments form a multivariate response with three components, with $Y_t = 1$ for normal and 0 for abnormal at time t. The 12 marginal distributions result from three repeated observations for each of the four groups.

Table 9.2 shows sample proportions of normal responses for the 12 marginal distributions. For instance, from Table 9.1, the sample proportion of normal responses after week 1 for subjects with mild depression using the standard drug was

$$(16 + 13 + 9 + 3)/(16 + 13 + 9 + 3 + 14 + 4 + 15 + 6) = 0.51$$

We see that 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 initial severity of depression, and (3) was higher for the mild than the severe cases of depression, for each treatment at each occasion. In such a study, the company that developed the new drug would hope to show that patients have a significantly higher rate of improvement with it.

Let s denote the initial severity of depression, with $s = 1$ for severe and $s = 0$ for mild. Let d denote the drug, with $d = 1$ for new and $d = 0$ for standard. Let t denote the time of measurement. When the time metric reflects cumulative drug dosage, a logit scale often has an approximate linear effect for the logarithm of time. We use scores (0, 1, 2), the logs to base 2 of the week numbers (1, 2, and 4). Similar substantive results occur using the week numbers themselves.

Let $P(Y_t = 1)$ denote the probability of a normal response at time t for a randomly selected subject. One possible model for how Y_t depends on the severity s, drug d, and the time t is the main effects model,

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

Table 9.2. Sample Marginal Proportions of Normal Response for Depression Data of Table 9.1

Diagnosis Severity	Treatment	Sample Proportion		
		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

This model assumes that the linear time effect β_3 is the same for each group. The sample proportions in Table 9.2, however, show a higher rate of improvement for the new drug. A more realistic model permits the time effect to differ by drug. We do this by including a drug-by-time interaction term,

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

Here, β_3 describes the time effect for the standard drug ($d = 0$) and $\beta_3 + \beta_4$ describes the time effect for the new drug ($d = 1$).

We will fit this model, interpret the estimates, and make inferences in Section 9.2. We will see that an estimated slope (on the logit scale) for the standard drug is $\hat{\beta}_3 = 0.48$. For the new drug the estimated slope increases by $\hat{\beta}_4 = 1.02$, yielding an estimated slope of $\hat{\beta}_3 + \hat{\beta}_4 = 1.50$.