9 Random Effects Model for Clustered Categorical Data

In Chapter 8, we introduced methods for comparing categorical responses for two samples when each sample has the same subject or when a natural pairing exists between each subject in one sample and a subject from the other sample. Highly stratified data often come from a design with cluster sampling. These are designs where two or more observations are made on each primary sampling unit or cluster.

Example: Common examples of such data are paired observations such as fraternal twins, litter mates, before-and-after outcomes from the same subject, or two occasions for expressing an opinion.

Tests were developed to compare the marginal distributions for matched-pair data. We also introduced subject-specific models with a separate parameter for each individual. We then introduced a conditional logistic regression model for matched pairs.

9.1 Correlated, Clustered Responses

We obtain correlated responses of the response variable in various types of studies.

- In longitudinal studies the response of each subject is observed repeatedly over time.
- Correlated responses can also occur when we observe matched sets of subjects. For instance, we might sample families in an obesity study and observe the children in each family.
- The matched set of observations is referred to as a cluster. Observations within a cluster are typically correlated. Even though analyses that ignore correlation can estimate parameters well, the estimated standard errors can be badly biased.
- In Chapter 9, Agresti contrasts marginal models with conditional models for clustered binary responses. Suppose that we have T observations per cluster. We write the observations in a cluster as $(Y_1, \ldots, Y_T)$.

Example: A longitudinal study is carried out to compare a new drug with a standard drug for treating subjects suffering mental depression. Subjects were classified into two groups according to initial severity (mild or severe) of depression. In each groups, subjects were assigned to one of two drugs. The response measure was each subject's extent of suffering from mental depression (normal or abnormal) at 1 week, 2 weeks, and 4 weeks. Here $T = 3$ and we observe (Y_1, Y_2, Y_3) for each subject.

- Here we have $2 \times 2 = 4$ treatment groups. For each group, the data consist of counts of the multivariate responses $(Y_1, Y_2, Y_3) = (0,0,0), (0,0,1), \ldots, (1,1,1)$ for each of the 4 groups where $Y_j = 1$ if the subject was classified as normal at time j.

		Response at Three Times							
Severity	Treatment	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

- In this study, we might be interested in the marginal proportions for each group: $P(Y_1 = 1|\text{Treatment Group}),\ P(Y_2 = 1|\text{Treatment Group}), P(Y_3 = 1|\text{Treatment Group})$. The following table shows the proportion of normal responses from each treatment group:

		Sample Proportion		
Severity	Treatment	Week 1	Week 2	Week 3
Mild	Standard	0.51	0.59	0.68
Mild	New Drug	0.53	0.79	0.97
Severe	Standard	0.21	0.28	0.46
Severe	New Drug	0.18	0.50	0.83

- In a marginal model, we wish to model the logits of the T success probabilities $\{P(Y_1 = 1), \ldots, P(Y_T = 1)\}$ as functions of the explanatory variables.
- We can model the logit of the t^{th} response as a function of severity ($s = 1$ for severe and $s = 0$ for mild), drug ($d = 1$ for new and $d = 0$ for standard), and time t using the main effects model

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

- This model assumes that the time effect is the same for all treatment groups. An alternative model includes an interaction term:

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

- We need to use appropriate units for the time effect in the model. The interaction term enables the different groups to have different rates of improvement for the different treatments.
- Maximum likelihood fitting of marginal models is difficult. In Chapter 9, Agresti uses the generalized estimating equation (GEE) approach to estimate parameters for a marginal model.

- The marginal models describe how the probability of a normal response at time t depends on severity, drug, and time for a randomly selected subject.
- In a conditional model, the probabilities can differ from subject to subject. We let Y_{it} denote the response of subject i at time t. A subject-specific model is given by

$$\text{logit}[P(Y_{it} = 1)] = \alpha_i + \beta_1 s + \beta_2 d + \beta_3 t + \beta_4 (d \times t).$$

  Each subject has his/her own intercept α_i which provides variability among the subjects who share the same setting (s, d, t) of the explanatory variables.
- This model has a subject (or cluster) specific interpretation, whereas the marginal models have a population-averaged interpretation.
- The problem with the conditional model is that there is a parameter α_i for each cluster. Thus, the number of parameters would be extremely large in many studies.
- In Chapter 10, Agresti develops the generalized linear mixed model to enable us to model the cluster effects as random effects. We consider $\{\alpha_i\}$ to be a random sample from a normal population. This reduces the number of parameters from the number of clusters down to a mean and variance for a normal distribution.

9.2 The Generalized Linear Mixed Model (GLMM)

The models that we have studied previously all have contained only fixed effects.

When we have clusters, a convenient approach to modeling stipulates a random effect specific to each cluster (or subject). The subjects are assumed to be a sample of subjects from a large population of subjects. Instead of having a parameter for each subject, the random effects model describes the population of potential subjects using a few parameters.

We now develop a model where each response has a single predictor and the intercept for each subject (or cluster) is a random effect.

Let Y_{it} denote response t within cluster i. The value of the corresponding predictor is x_{it}. We let $\mu_{it} = E(Y_{it}|u_i)$ be the conditional mean response for the given value of the random effect u_i. The GLMM has the form

$$g(\mu_{it}) = u_i + \beta x_{it}, \qquad i = 1, \ldots, n, \qquad t = 1, \ldots, T,$$

where $g(\cdot)$ is the link function. We assume that $\{u_i\}$ forms a random sample from a $N(\alpha, \sigma^2)$ distribution with unknown parameters.

An alternative way of writing the model is

$$g(\mu_{it}) = u_i + \alpha + \beta x_{it}, \qquad i = 1, \ldots, n, \qquad t = 1, \ldots, T,$$

where $\{u_i\}$ forms a random sample from a $N(0, \sigma^2)$ distribution.

- We call the variance σ^2 a *variance component*.

Maximum likelihood produces estimates of the fixed effect $(\hat{\beta})$ and the variance component $(\hat{\sigma}^2)$ and predictions of the random effects $(\hat{u}_i)$.

9.2.1 Logistic GLMM for Binary Matched Pairs

Recall the logistic model for binary matched pairs (Y_{i1}, Y_{i2}):

$$\text{logit}[P(Y_{i1} = 1)] = \alpha_i + \beta, \qquad \text{logit}[P(Y_{i2} = 1)] = \alpha_i.$$

Here the fixed effect β was a log odds ratio given the cluster. We used conditional ML to estimate β after eliminating $\{\alpha_i\}$ from the likelihood.

To form a random effects model, we replace α_i with $u_i + \alpha$:

$$\text{logit}[P(Y_{i1} = 1)] = u_i + \alpha + \beta, \qquad \text{logit}[P(Y_{i2} = 1)] = u_i + \alpha.$$

- This is a special case of a GLMM with logit link, $T = 2$ and x_{it} an indicator function for observation within the pair. We call this a *logistic-normal model* since the random effect has a normal distribution.

- This model implies a nonnegative correlation among observations within a cluster. Clusters with large positive values of u_i have a large value of $P(Y_{it} = 1)$ for each t. Clusters with large (in magnitude) negative values of u_i have a small value of $P(Y_{it} = 1)$ for each t.

- We will have greater association when the spread of the u_i values is large; i.e, when σ is large. This results in a greater range of values of $P(Y_{it} = 1)$ among the clusters.

Example: **Rating of Performance of the Prime Minister**

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

- Odds ratio for marginal model:

$$\exp(\hat{\beta}) = \frac{944 \times 720}{656 \times 880} = 1.177$$

- Odds ratio for conditional model:

$$\exp(\hat{\beta}) = 150/86 = 1.744$$

- ML estimates for GLLM: $\hat{\beta} = 0.556,\ \hat{\sigma} = 5.16$

- The GLLM has the same ML estimate of the odds ratio as the conditional ML estimate. The large value of σ indicates a large variation among the random effects resulting in a strong correlation within clusters.

Example: **Sacrifices for the Environment (Agresti, Chapters 8 and 10)**

Pay Higher	Cut Living Standards		
Taxes	Yes	No	Total
Yes	227	132	359
No	107	678	785
Total	334	810	1144

- Odds ratio for marginal model:

$$\exp(\hat{\beta}) = \frac{359/785}{334/810} = 1.11$$

- Odds ratio for conditional model:

$$\exp(\hat{\beta}) = 132/107 = 1.23$$

- ML estimates for GLLM: $\hat{\beta} = 0.210,\ \hat{\sigma} = 2.85$

- The GLLM has the same ML estimate of the odds ratio as the conditional ML estimate. The large value of σ indicates a large variation among the random effects resulting in a strong correlation within clusters.

9.3 Differing Effects in Conditional Models and Marginal Models

The parameters in GLMMs and marginal models have different interpretations. The effects in marginal models are averaged over all clusters and do not refer to a fixed value of the random effect. In contrast, the effects in a GLMM have a conditional interpretation, given the value of the random effect in the particular cluster.

For matched-pair data, the cluster-specific model applies to the data when displayed as a separate partial table for each cluster. For each subject in the survey example, the first row has a "1" in the first row if "approve" and a "1" in the second row if "disapprove" for the first survey. Similarly, the second row presents the result on the second survey. There are four possible 2×2 tables:

1	0	1	0	0	1	0	1
1	0	0	1	1	0	0	1

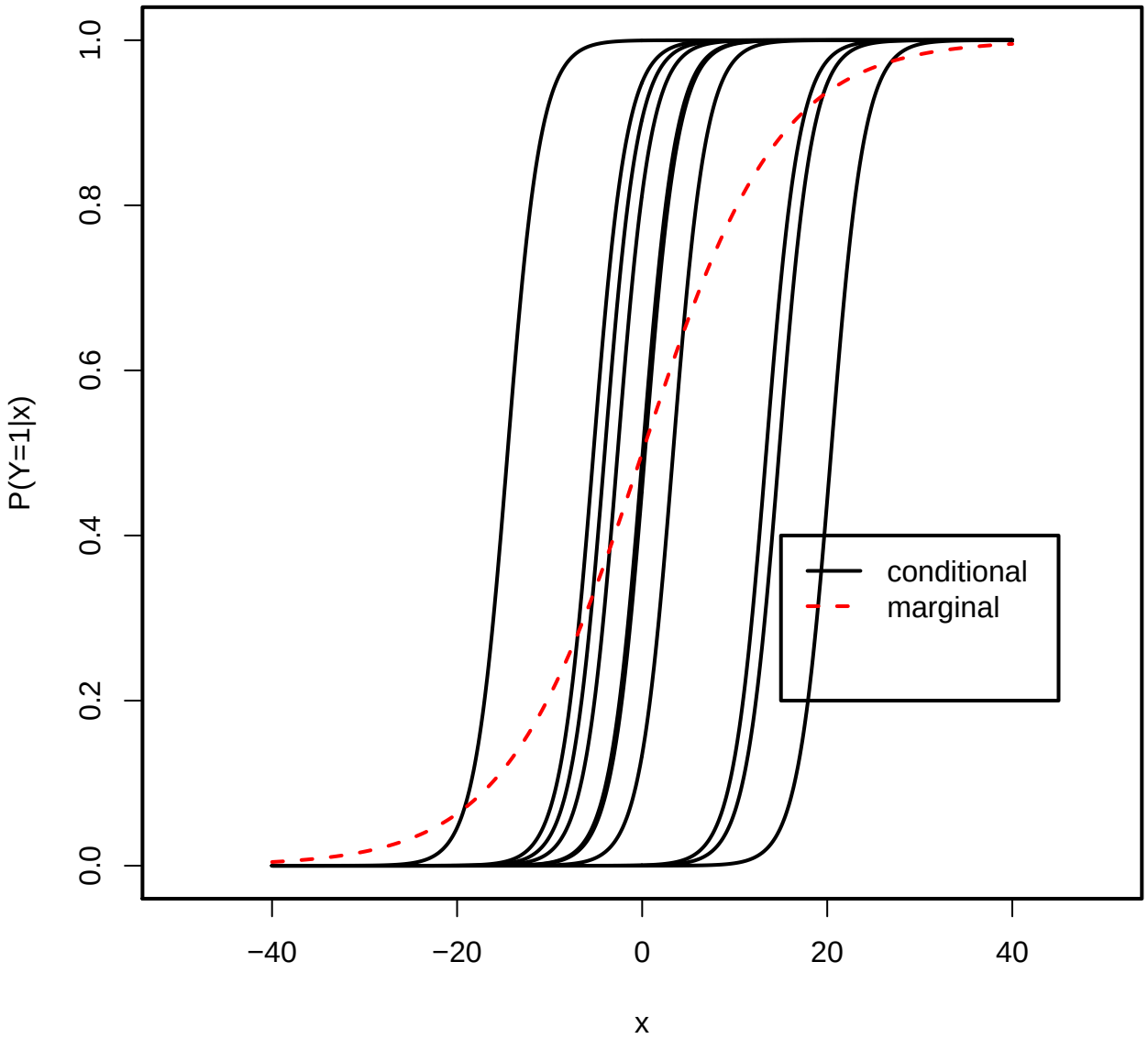
We thus have 3200 observations in a $2 \times 2 \times 1600$ contingency table. If we collapse the table over the 1600 subjects, we obtain the 2×2 table of marginal totals:

944	656
880	720

- For the marginal model, the estimated log odds ratio is $\hat{\beta} = \log(1.177) = 0.163$. For the conditional model, the estimated log odds ratio is $\hat{\beta} = 0.556$. We see that the signs of the coefficients are the same, but the marginal model's coefficient is smaller.
- The odds for an individual subject to say yes on the second survey was $\exp(0.556) = 1.744$ times the odds of the same person saying yes on the first survey. For the marginal model, the odds of a randomly selected subject saying yes on the first survey are only 1.177 times the odds of another randomly selected subject saying yes on the second survey.
- The population-averaged effects in marginal models are typically smaller than the subject-specific effects in GLMMs. Consider a GLMM model with a single numerical predictor x and a random intercept u_i. We picture this by generating several subject-specific logistic regression curves, $P(Y_{it} = 1|u_i)$ for the model

$$\text{logit}[P(Y_{it} = 1)] = u_i + \alpha + \beta x$$

where $u_i \sim N(0, 5.16^2)$, $\alpha = 0$, and $\beta = 0.556$. At any fixed x, there is variability in the conditional means $E(Y_{it}|u_i) = P(Y_{it} = 1|u_i)$. These conditional means have an average of $E(Y_{it})$. The resulting curve will be shallower indicating a smaller marginal effect.



9.4 Modelling Overdispersed Categorical Data

In the discussion of Poisson regression and logistic regression, we discussed the effects of overdispersion on the validity of the inferences. Often overdispersion results when there are correlated experimental subjects. One way to account for this correlation in the model is to use a random effect in the model to account for the clustering of subjects.

Example: **Mortality of Cancer Cells**

The data comes from an experiment to measure the mortality of cancer cells under radiation under taken in the Department of Radiology, University of Cape Town. Four hundred cells were placed on a dish, and three dishes were irradiated at a time, or occasion. After the cells were irradiated, the surviving cells were counted. Since cells would also die naturally, dishes with cells were put into the radiation chamber without being irradiated, to establish the natural mortality. Only the zero-dose data are analyzed here.

Analysis of Data:

- A binomial distribution was fit to the 27 counts using a logistic model containing only an intercept. There was substantial lack of fit with $G^2 = 495.6,\ X^2 = 493.0,\ df = 26$ and $AIC = 666.6$.

- A logistic model with a fixed effect due to occasion was fit to the data. This resulted in a better fit with $G^2 = 32.8,\ X^2 = 32.7,\ df = 18$ with $AIC = 219.8$. However, $P(\chi^2_{18} > 32.8) = 0.018$, so there is still some lack of fit using the binomial distribution.

- A logistic model with a random effect for occasion was fit to the data. The fit was apparently not as good with $X^2 = 40.7$ and $X^2/df = 1.57$. The estimated variance of the random effect was $\hat{\sigma}^2 = 0.2249$.

- Since there still seemed to be overdispersion with the occasions, the dishes within occasion were added as a random effect to the model. The fit was better with $X^2 = 26.07$ and $X^2/df = 1.00$. However, the estimates of the variance components that were obtained did not agree with those previously published.

- It is interesting to consider the estimates for the survival proportions. The estimated proportion from the binomial model equals the underlying sample proportion of success in all the trials. Similarly, the estimated proportion for the fixed effects model is the sample proportion within each occasion. The predicted proportions including the random effects can be thought of as *shrinkage estimates* of the proportions for the various occasion. These predicted proportions are closer to the overall proportion of survived cells than are the sample proportions.

9.5 Fitting the Depression Data

Example: A longitudinal study is carried out to compare a new drug with a standard drug for treating subjects suffering mental depression. Subjects were classified into two groups according to initial severity (mild or severe) of depression. In each groups, subjects were assigned to one of two drugs. The response measure was each subject's extent of suffering from mental depression (normal or abnormal) at 1 week, 2 weeks, and 4 weeks. Here $T = 3$ and we observe (Y_1, Y_2, Y_3) for each subject.

- Here we have $2 \times 2 = 4$ treatment groups. For each group, the data consist of counts of the multivariate responses $(Y_1, Y_2, Y_3) = (0, 0, 0), (0, 0, 1), \ldots, (1, 1, 1)$ for each of the 4 groups where $Y_j = 1$ if the subject was classified as normal at time j.
- In this study, we might be interested in the marginal proportions for each group: $P(Y_1 = 1|\text{Treatment Group}),\ P(Y_2 = 1|\text{Treatment Group}), P(Y_3 = 1|\text{Treatment Group})$.
- In a marginal model, we wish to model the logits of the T success probabilities $\{P(Y_1 = 1), \ldots, P(Y_T = 1)\}$ as functions of the explanatory variables.
- We can model the logit of the t^{th} response as a function of severity ($s = 1$ for severe and $s = 0$ for mild), drug ($d = 1$ for new and $d = 0$ for standard), and time t using

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

- Agresti comments that there is apparently little correlation among the pairs of responses for common subjects. Thus, instead of using GEE estimates, one can use ML for ordinary logistic regression.
- Agresti's conclusions for the GEE model follow:
  - There is strong evidence of faster improvement with the new drug.
  - The estimated odds of a a normal response when the initial diagnosis was severe are 0.27 times the estimated odds when the initial diagnosis was mild.
  - The chance of a normal response is similar for the two drugs initially and increases with time (but faster for the new drug).
- A GLLM with a random effect for subject is given by

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

  This is fit to the data using NLMIXED and GLIMMIX.
  - The estimated coefficients and standard errors are virtually the same as those obtain for the marginal model using GEE.
  - The estimated variance component is $\hat{\sigma} = 0.065$. This suggests that there is little heterogeneity among the subjects in their response probabilities.

