21

Repeated Measures and Related Designs

21.1 INTRODUCTION

In this chapter we discuss the application of methods for longitudinal data to closely related study designs. In these settings individuals have multiple commensurate measurements made under different circumstances and possibly also at different times. However, the major interest of the analysis is not in changes in the response over time, but in the effect of different circumstances of measurement and/or the effects of covariates on the responses.

The first design that we will consider is the classical repeated measures design. In this setting each subject is measured under a fixed number of different conditions, often corresponding to different treatments. Interest centers on comparing the effects of the different experimental conditions on the outcome. Similar to time in a longitudinal study, experimental condition is a within-subject factor and the conditions are compared using within-subject contrasts. Such designs are popular because subject-to-subject differences in outcome are accounted for in the design. Since each subject acts as his or her own control, comparisons of the outcome under different experimental conditions are estimated free of any between-subject variation in the outcome. As a result the design is potentially very efficient relative to comparing the different experimental conditions on different groups of subjects.

The second design is one that produces what we refer to as "multiple source" data. In this setting the primary outcome of interest is measured by more than one instrument or rater. This frequently happens when the outcome is difficult to measure and is thus determined under multiple different circumstances. In Section 21.3 we describe an example in the context of measuring psychopathology in children. In

Applied Longitudinal Analysis, Second Edition. By Garrett M. Fitzmaurice, Nan M. Laird, and James H. Ware.

this context there may be some interest in comparing the different measures, but ordinarily the main focus of the analysis is the effects of subject-specific covariates on the outcome. Hence, unlike a typical longitudinal study and also unlike a classical repeated measures design, the main interest centers on the effects of subject-specific variables on response, and possibly also their interaction with the multiple sources. In many settings the fact that there are multiple sources could be regarded as a "nuisance" feature of the study design.

The distinction we make between repeated measures and multiple source data is based on what is of primary interest in the analysis. Both share the same analytical methods for regression models with correlated data. Sometimes, however, this distinction is blurred. For example, Hernández et al. (2000) report on a validity study of a new questionnaire designed to measure physical activity and inactivity in school children in Mexico. A self-reported questionnaire was administered both to the mother and the child on two different occasions; in addition a 24-hour recall (considered the best measure, but only limited to a single day) was administered on each occasion. The average of the two 24-hour recalls was considered the "gold standard" for the analysis. Here the objective of the analysis was two-fold: to compare mean responses of the two child and two mother assessments to the average of the two 24-hour recalls, and to look at the correlations between these measures. This is clearly a multiple source data set, since the mother and child questionnaires were both intended to measure the child's average activity levels over the period. However, the primary focus of the analysis was comparing the multiple reports to each other and to the "gold standard" and examining the correlations; here the subject-specific variables, age, gender, and socioeconomic level, were used to adjust the means and the correlations for between-subject differences. Thus, in this example, the analytic goal of comparing means is closer to a repeated measures analysis than a multiple source analysis, although comparing correlations is more typical in the multiple source analysis.

We will first describe the main features of these two designs in greater detail, and then provide some examples to illustrate the application of regression methods for correlated data to repeated measures and multiple source data.

21.2 REPEATED MEASURES DESIGNS

Repeated measures designs are frequently encountered in applications. In the experimental context the repeated measures design is also sometimes called a randomized block design. In the simplest setting subjects each receive n treatments or experimental conditions, and the outcome is recorded for each condition. Thus each subject has a vector of n measurements, $Y_i = (Y_{i1}, \ldots, Y_{in})'$. The treatments may be given sequentially in a randomly assigned order, but in some settings they can be given simultaneously. An example of the latter is a classic study of topical treatments for leprosy, where each patient was given four different treatments simultaneously at four different locations on their body. After a number of days, the skin lesions were recorded at each of the four locations. Letting Y_{ij} denote the response to the j^{th}

treatment ($j = 1, ..., 4$), the primary interest is in comparing the four treatments. However, the analysis must account for the fact that the observations on different treatments, $Y_i = (Y_{i1}, ..., Y_{i4})'$, are correlated.

As mentioned earlier, repeated measures designs are popular because subject differences in outcome are accounted for in the design. By removing between-subject variation in the outcome from the comparisons of different experimental conditions, the repeated measures design can be very efficient relative to comparing the different experimental conditions on different groups of subjects. To illustrate this point, consider a simple repeated measures design with N subjects receiving $n = 2$ treatment conditions (producing a total of $2N$ observations). Let Y_{i1} denote the response for the i^{th} subject under the first condition and Y_{i2} denote the response under the second condition. A natural estimate of the effect of treatment on the mean response is

$$\widehat{\delta} = \widehat{\mu}_1 - \widehat{\mu}_2,$$

where

$$\widehat{\mu}_j = \frac{1}{N}\sum_{i=1}^{N} Y_{ij}.$$

The variability of this estimator of the effect of treatment is given by

$$\text{Var}(\hat{\delta}) = \text{Var}\left\{\frac{1}{N}\sum_{i=1}^{N}(Y_{i1} - Y_{i2})\right\} = \frac{1}{N}\left(\sigma_1^2 + \sigma_2^2 - 2\,\sigma_{12}\right),$$

where $\sigma_j^2 = \text{Var}(Y_{ij})$, $\sigma_{12} = \text{Cov}(Y_{i1}, Y_{i2}) = \rho\,\sigma_1\sigma_2$, and $\rho = \text{Corr}(Y_{i1}, Y_{i2})$. To simplify, we assume that treatment may have an impact on the mean response, but not on the variance; this assumption can be justified when treatments are allocated within subjects randomly by time. Then, $\sigma_1^2 = \sigma_2^2 = \sigma^2$, and the variability of this estimator simplifies to

$$\text{Var}(\hat{\delta}) = \frac{2}{N}\left\{\sigma^2(1-\rho)\right\}.$$

Next consider a two-group study design, where the treatment conditions are randomized to $2N$ subjects drawn from the population, with N subjects allocated to one condition, and the remaining N subjects allocated to the other condition (producing a total of $2N$ observations). The effect of treatment on the mean response can be estimated by comparing the mean response in the two groups of subjects. The natural estimate of treatment effect is the same as before:

$$\widehat{\gamma} = \widehat{\mu}_1 - \widehat{\mu}_2,$$

where

$$\widehat{\mu}_j = \frac{1}{N}\sum_{i=1}^{N} Y_{ij}$$

is the sample mean response in the group of subjects that were randomized to the j^{th} treatment condition. The variability of this estimator of the effect of treatment is

given by

$$\text{Var}(\hat{\gamma}) = \frac{1}{N}\left(\sigma_1^2 + \sigma_2^2\right),$$

where $\sigma_j^2 = \text{Var}(Y_{ij})$. Again, if we assume that treatment may have an impact on the mean response, but not on the variance, then $\sigma_1^2 = \sigma_2^2 = \sigma^2$, and

$$\text{Var}(\hat{\gamma}) = \frac{2\sigma^2}{N}.$$

The potential gain in efficiency of the repeated measures design is quantified by taking the ratio of the variances of the two estimators of the effect of treatment:

$$\frac{\text{Var}(\hat{\delta})}{\text{Var}(\hat{\gamma})} = (1 - \rho).$$

Provided that the correlation among the repeated measures is positive, the repeated measures design provides a more precise estimate of the effect of treatment. For example, when $\rho = 0.75$, it is four times more efficient than the two-group design or, put another way, the repeated measures design only requires a quarter of the number of observations to attain the same level of precision as the two-group design.

There are many variations on the repeated measures design. If, for example, treatments are given sequentially in a random order, it is common to use a restricted randomization scheme to balance the treatments over time. This would allow equal numbers of each treatment to be assigned at each time point, for example, with two treatments denoted A and B, half the subjects would be assigned to AB, and half to BA. Thus systematic time or sequence differences will be eliminated, and the design is potentially more efficient. In this case the number of subjects is a multiple of n, the number of treatment conditions. A special case of these designs is the crossover design, where each subject receives each treatment in a random order.

A closely related design is the split-plot design. Here there are two treatment factors, a so-called main plot factor and a sub-plot factor. When subjects are the main plots, the main plot factor is a between-subject factor and each subject receives only one level of this factor. The sub-plot factor is a within-subject factor and each subject receives all levels of the within-subject factor. Thus, in the split-plot design, one factor is randomized within subjects, just as in the repeated measures design, and the other factor is randomized between subjects. For example, in a study comparing the effects of different antibiotics (drugs) and topical gels (gels) for the treatment of eye infections, subjects (main plots) are randomized to receive one of three different oral antibiotics. The main plot factor is drug (antibiotics). In addition each subject is required to apply two different topical gels directly to the left and right eye; the two different gels are randomized to the left and right eyes. Here, the sub-plot factor is gels. In the split-plot design, both the main effects and the interaction of the two factors are usually of interest, although this design provides more precise information about the within-subject factor than the between-subject factor.

Much of the literature on repeated measures is in the context of designed experiments where treatments are allocated within subjects randomly by time or location.

This has several ramifications for our discussion. First, because the factors in a randomized design are typically categorical, the analysis of these designs is ordinarily presented in the context of analysis of variance (ANOVA) rather than a general regression model with correlated data. However, the analysis of variance model can be viewed as a special case of the general linear regression model for correlated data presented in Part II. Hence the regression models for correlated data apply quite straightforwardly to the classical repeated measures design.

Second, with randomization, arguments can be made that allow one to simplify the analysis of repeated measures data, especially with balanced and complete designs. There are two main approaches to the analysis of repeated measures data (see Section 3.6), repeated measures analysis by ANOVA and repeated measures analysis by multivariate ANOVA (MANOVA). With the former, one assumes that any contrast between any two repeated measures on the same subject, say $Y_{ij} - Y_{ik}$, has the same variance; that is, $\text{Var}(Y_{ij} - Y_{ik})$ is constant for all choices of i and $j \neq k$. For example, if the covariance matrix of the vector of repeated observations, Y_i, takes on a compound symmetry form, then the requirement for the repeated measures analysis by ANOVA is satisfied. In contrast, the model for repeated measures analysis by MANOVA allows the vector of repeated observations to have an arbitrary covariance structure, but the standard analysis is usually limited to balanced and complete designs. In addition the analysis of repeated measures by ANOVA can be considerably more powerful than the analysis by MANOVA (when the assumptions for the former hold).

If the within-subject factor is randomly allocated to subjects, then randomization arguments can be made to show that the constant variance condition on the contrast does hold. More generally, one can show that $\text{Var}(Y_{ij})$ is constant for all i and j and $\text{Cov}(Y_{ij}, Y_{ik})$ is constant for all i and $j \neq k$. The general approach to a randomization argument involves treating the randomization indicators as random variables, and the observed outcomes as fixed. While the approach provides an attractive justification for using the repeated measures ANOVA in the randomized experiment, it may not be justifiable in the more general setting. In addition the justification relies on a linear model. Note that with constant variance and covariance, we can formulate a repeated measures analysis using mixed effects models with a random subject effect b_i (where $\text{Var}(b_i) = \sigma_b^2$) and independent errors ϵ_{ij} (where $\text{Var}(\epsilon_{ij}) = \sigma_\epsilon^2$). Then $\text{Var}(\widehat{\delta}) = 2\,\sigma_\epsilon^2/N$ and $\text{Var}(\widehat{\gamma}) = 2(\sigma_\epsilon^2 + \sigma_b^2)/N$ because b_i drops out of any within-subject contrast.

Finally, there are many variations on repeated measures designs which are difficult to handle with the classical analytic approaches. For example, missing data lead to unbalanced designs with a variable number of observations on subjects. In addition the outcomes may be count or binary data, neither of which can be handled by classical ANOVA techniques. Using generalized linear models for correlated outcomes enables us to handle these variations in a unified way.

21.3 MULTIPLE SOURCE DATA

Multiple source data usually arise in the context of epidemiological studies, where outcomes and/or risk factors may be difficult to measure. For example, Fitzmaurice et al. (1996) discuss a study involving child psychopathology, which used standardized questionnaires administered to both the child's mother and teacher. Responses to these questionnaires were used to construct, for each informant, dichotomous indicators of whether psychopathology was present in the child. Other informants sometimes used in this setting include the child, the father, and peers. Because informants interact in different settings with the child, the information from different informants reflects different perspectives. However, the primary interest of the study was not to compare informants, but rather to study the effects of covariates on child psychopathology, broadly defined. By including informant as a variable in the correlated data model, and its interaction with other covariates of interest, one can also examine how informants differ overall, and whether covariate effects differ by informant.

In this section we use the term "multiple source" to encompass data that are simultaneously obtained from multiple informants or raters (e.g., self-reports, family members, health care providers, administrators) or via different/parallel instruments or methods (e.g., symptom rating scales, standardized diagnostic interviews, or clinical diagnoses). For example, in psychiatric studies of children, the child's parent is routinely used as a proxy data source; other sources (e.g., self-report, peers, teachers, clinicians, or trained observers) may also be employed, depending on the child's age and the nature of psychopathology under study. Multiple source data have become increasingly common in hospital-based and outpatient-based assessments of the effectiveness of treatments. For example, evaluations of managed care programs for the U.S. Medicaid population require analysis of multiple sources of information, including patient satisfaction with health care, treatment utilization, and appropriate care. Other areas where multiple reports arise include studies of severe mental illness, such as schizophrenia or Alzheimer's disease, where the affected subject is often unable to provide self-report data, family history studies, where many relatives are interviewed about the status of the proband and other family members, and behavioral studies of alcohol/drug use or of eating disorders, where information is obtained from the subject, as well as family members or other sources.

Historically there has been little consensus as to how to analyze multiple source data. Sometimes investigators conduct completely separate univariate analyses for each source. Alternately the multiple source measures may be combined to make a single outcome. For measured responses, investigators may take a mean and for dichotomous responses the "and" or the "or" rules are often used. In the "and" rule binary source data are considered to be positive if *all* of the source data are positive, and negative otherwise; in the "or" rule binary source data are considered to be positive if *any* of the source data are positive, and negative otherwise. All of these ad hoc strategies usually require discarding data when any sources are missing; the separate analysis strategy makes it difficult to compare the results for the two (or more) sources, while the analysis using a combined response can obscure interesting differences. Us-

ing correlated data models similar to those discussed in previous chapters allows one to directly compare source effects and to handle missing data in a unified framework.

21.4 CASE STUDY 1: REPEATED MEASURES EXPERIMENT

In this section we analyze data from a randomized crossover design to illustrate the use of mixed effects models in the repeated measures setting. The study was designed to compare two active drugs and placebo for relief of tension headache. The two analgesic drugs were identical in their active ingredients except that one contained caffeine. The primary comparison of interest was between the two active drugs; the placebo was included for regulatory purposes. What makes the analysis non-standard is that there were three treatments, but only two periods; that is, each subject received only two of the three treatments in a random order. With three treatments and two periods, there are six possible treatment sequences, AB, BA, AP, PA, BP, PB, where A, B, and P denote the two active drugs and placebo. If each sequence is assigned an equal number of subjects, then we have what is known as a balanced incomplete block design. It is balanced because all possible sequences are equally represented, but incomplete because each subject is "missing" a third treatment. In our example the AB and BA sequences were assigned three times as many subjects as the remaining four because of the interest in the A versus B comparison. The descriptive statistics for one measure of pain relief used in the crossover study of analgesics are given in Table 21.1. There were actually two headaches treated within each period, but that feature of the data is ignored here and we use the average of the two measures of pain relief. A few observations were missing, but we have access only to subjects with complete data.

An important issue in crossover designs is the possibility of carryover effects. The presence of a carryover effect means that the treatment taken in the first period may influence the treatment effect in the second period, that is, the drug in the first period carries over. In our analysis we will show the results of fitting two different models, one with only treatment and period effects included in the model and one with the carryover effects included. We display the model that includes carryover effects by giving the expected value of the response for each sequence and each period. Letting Y_{ij} denote the treatment response for the i^{th} subject in the j^{th} period, we write

$$E(Y_{ij}) = \beta_1 + \beta_2 X_{2ij} + \beta_3 X_{3ij} + \beta_4 X_{4ij} + \beta_5 X_{5ij} + \beta_6 X_{6ij}, \qquad (21.1)$$

where $X_{1ij} = 1$ for all i and j, X_{2ij} and X_{3ij} are dummy variable indicators of the main effects of treatment A and B (P is the reference condition), X_{4ij} is a dummy variable indicator of period 2 (so for all i, $X_{4i1} = 0$ and $X_{4i2} = 1$) and X_{5ij} and X_{6ij} are dummy variable indicators of the carryover effects for A and B, respectively. For subjects assigned to treatment A in the j^{th} period, $X_{2ij} = 1$, and 0 otherwise; for those assigned B in the j^{th} period, $X_{3ij} = 1$, and 0 otherwise ($X_{2ij} = X_{3ij} = 0$ for subjects assigned to P in the j^{th} period). For subjects in sequences where A is taken first $X_{5i2} = 1$, and $X_{6i2} = 1$ if the first treatment taken is B; for all other values of i, X_{5i2} and X_{6i2} are zero. Likewise, for all i, $X_{5i1} = X_{6i1} = 0$ since, by definition,

Table 21.1 Descriptive statistics (means, standard deviations, and correlation), by sequence, for total pain relief for headache in periods 1 and 2.

		Period 1		Period 2		
Sequence	N	Mean	SD	Mean	SD	ρ
AB	126	10.196	3.347	9.153	3.429	0.20
BA	127	9.581	3.881	10.791	3.530	0.30
AP	43	10.477	3.546	7.273	4.451	0.31
PA	43	8.366	3.777	10.855	3.204	0.47
BP	42	10.333	3.306	8.357	3.944	0.72
PB	42	7.464	4.265	9.911	4.183	0.67

Note: ρ is the correlation between pain relief scores in periods 1 and 2. The estimate of the common correlation (ignoring sequence) is $\widehat{\rho} = 0.38$.

there are no carryover effects in the first period. In Table 21.2 we summarize the interpretations of $\beta_1, ..., \beta_6$ in terms of the mean response for each sequence during each period.

This model includes effects of treatment (β_2 and β_3), period (β_4), and carryover (β_5 and β_6). The active drug comparison is given by $\beta_2 - \beta_3$. In our model the carryover effects are assumed to depend only on the treatments taken in period one. That is, the model assumes that the carryover of one active treatment to the other active treatment is the same as the carryover of that active treatment to the placebo. More complicated models can be used (e.g., see Laird, Skinner, and Kenward, 1992) but are not considered here.

Since subjects are randomized, we will assume $\text{Var}(Y_{i1}) = \text{Var}(Y_{i2})$. A more general covariance structure could easily be accommodated by allowing the variance to depend on period, but we do not do so here. Using the compound symmetry assumption allows us to analyze the data using mixed effects models with a random subject effect. Handling missing responses simply involves removing rows from the design matrix and the response vector (as discussed in Section 5.3).

The balanced incomplete block design has two types of information about treatment. One source of information comes from within-subject contrasts; that is, in the absence of carryover effects, $Y_{i2} - Y_{i1}$ is a treatment contrast with variance $2\sigma^2(1 - \rho)$. There is also information about treatment contrasts from between-subject comparisons; that is, in the absence of carryover effects, $Y_{ij} - Y_{i'j}$ estimates a treatment contrast for two subjects with different treatments in period j. Now, however, the variance of the contrast is $2\sigma^2$. When there are carryover effects in this design, information about treatment effects can be found in both within-subject and between-subject contrasts. If ρ is modest ($\rho < 0.4$), it is often suggested that the

Table 21.2 Regression parameters for the mean total pain relief for headache, by sequence and period.

Sequence	Period 1	Period 2
AB	$\beta_1 + \beta_2$	$\beta_1 + \beta_3 + \beta_4 + \beta_5$
BA	$\beta_1 + \beta_3$	$\beta_1 + \beta_2 + \beta_4 + \beta_6$
AP	$\beta_1 + \beta_2$	$\beta_1 + \beta_4 + \beta_5$
PA	β_1	$\beta_1 + \beta_2 + \beta_4$
BP	$\beta_1 + \beta_3$	$\beta_1 + \beta_4 + \beta_6$
PB	β_1	$\beta_1 + \beta_3 + \beta_4$

between-subject information be ignored, and the analysis use only the within-subject contrasts. The rationale behind this approach is that the within-subject contrast yields a simple structure with no need to estimate variance and covariance components for Y_{i1} and Y_{i2}, while using all the information requires estimating the between- and within-subject error variance. Of note, $\widehat{\rho} \approx 0.4$ for the pain relief data in our example (see Table 21.1).

The within-subject analysis is especially easy to do with just two periods by subtracting Y_{i1} from Y_{i2}, and analyzing the difference $d_i = Y_{i2} - Y_{i1}$:

$$E(d_i) = \beta_2(X_{2i2} - X_{2i1}) + \beta_3(X_{3i2} - X_{3i1}) + \beta_4 + \beta_5 X_{5i2} + \beta_6 X_{6i2}. \quad (21.2)$$

Note that β_1 has vanished from the model, and β_4 acts as the constant (or intercept) in the model, since $X_{4i1} = 0$ and $X_{4i2} = 1$ for all i. The parameters β_2 and β_3 remain the main effects of treatment and β_5 and β_6 remain the carryover effects. In our analysis the main focus is on the active drug comparison, $\beta_2 - \beta_3$. It is a special feature of this design that carryover effects can be estimated from within-subject contrasts (Koch et al., 1989), but as we will show, most of the information about the carryover effects comes from the between-subject information. This issue is similar to that raised in Chapter 9 (see Section 9.5) in the discussion of cross-sectional and longitudinal information. In the context of a balanced incomplete randomized trial, or a complete randomized trial with carryover effects, the issue of bias does not arise, only that of efficiency.

We will illustrate the analysis using both the simple regression on the differences (d_i) and a mixed effects model analysis on the full data (Y_{i1} and Y_{i2}), with subject as a random effect (b_i). Table 21.3 illustrates the conventional wisdom that the within-subject (differences) analysis is highly efficient for treatment effects when carryover is absent. (Compare the two rows of Table 21.3 for the treatment effect estimates assuming no carryover effects.) However, if carryover effects are present the within-subject analysis is very inefficient for both the main effect of treatment and for carryover effect. (Compare the two rows of Table 21.3 for the treatment and carryover effects estimates.)

Table 21.3 Results of comparison of the two active treatments (estimate ± SE) based on analysis of differences and mixed effects model analysis, with and without carryover effects.

	No Carryover	With Carryover	
	Treatment	Treatment	Carryover
Method of Analysis	$(\beta_2 - \beta_3)$	$(\beta_2 - \beta_3)$	$(\beta_5 - \beta_6)$
Differences	1.06 ± 0.24	0.46 ± 0.67	-1.19 ± 1.26
Mixed Effects Model[a]	1.02 ± 0.23	0.55 ± 0.32	-1.06 ± 0.52

[a]Linear mixed effects model with random subject effect.

In this case investigators basing their conclusion on the within-subject (differences) analysis would be led to conclude erroneously that the active drug comparison is not confounded with carryover ($Z = -1.19/1.26 = -0.94$, $p > 0.30$) and to use the results of the model without carryover to obtain a significant difference between the active treatments ($Z = 1.06/0.24 = 4.42$, $p < 0.0001$). That is, the inefficient analysis of carryover effects based on differences leads to the erroneous conclusion that the treatment comparison does not require any adjustment for carryover effects. Using the mixed effects model analysis, however, we come to quite a different conclusion; there is a substantial carryover effect ($Z = -1.06/0.52 = -2.05$, $p < 0.05$), but no evidence of a statistically significant treatment effect ($Z = 0.55/0.32 = 1.71$, $p > 0.05$) when the carryover effects are included in the model and adjusted for in the analysis. The widespread availability of software to implement a mixed effects model analysis makes it relatively easy to capture the "between-subject information" even in complex repeated measures designs.

21.5 CASE STUDY 2: MULTIPLE SOURCE DATA

Data for this example come from two surveys of children's mental health (Zahner et al., 1992, 1993). A standardized measure of childhood psychopathology was used both by parents (Child Behavior Checklist, CBC) and teachers (Teacher Report Forms, TRF) to assess children in the study. We use here the externalizing scale, which assesses delinquent and aggressive behavior. The scale has been dichotomized at the cut point for borderline/clinical psychopathology. The cut points are normed separately for males and females; thus we expect to see small gender effects in these data. Because of the multiple levels of permissions and reporting, a substantial number of children were missing the TRF. Our analysis is based on 1428 children who had both parent and teacher responses, and an additional 1073 children with only a parent response; a total of 2501 children participated in the study. In this example the two sources or

respondents are the children's parents and teachers; in the psychiatric literature these sources are often referred to as "informants."

The objective of the analysis is to study the influence of several explanatory variables on the prevalence of externalizing behavior in these children. For simplicity we limit our analysis to single-parent status (coded 1: single, 0: otherwise) and child's physical health problems (coded 1: fair to bad health, 0: good health). In addition we are interested in describing the level of association between the two respondents, and determining whether the effects of the covariates depend on informant. We will use standard regression models to address these issues, but because we have two different measures of the outcome, we will use correlated data models, one for each outcome. Since externalizing behavior is dichotomous, we use logistic models for the regressions.

The basic approach uses two separate regression models, one with the CBC as an outcome and one with the TRF as an outcome. Both models have the same set of covariates (here single-parent status and physical health problems), but the coefficients may differ for the different sources. In addition we have an "informant" indicator variable that identifies the source of the response. To motivate the approach, we first fit two completely separate logistic regressions each with the full complement of covariates, one for each informant outcome. Let μ^P and μ^T denote the probability of a positive response on externalizing behavior as measured by parents and teachers, respectively. Then the two regression models are

$$\text{logit}(\mu_i^P) = X_i'\beta^P$$

and

$$\text{logit}(\mu_i^T) = X_i'\beta^T,$$

where X_i is a $p \times 1$ vector of covariates for the i^{th} subject and

$$E(Y_i^P|X_i) = \mu_i^P \text{ and } E(Y_i^T|X_i) = \mu_i^T.$$

The first logistic regression model was fit with the parent response as outcome using all 2501 children, and the second was fit with the teacher response using only the 1428 children with a teacher response. The results are displayed in Table 21.4.

The estimated coefficients for single parent status are similar and statistically significant (at the 0.05 level) for each informant report; the estimated coefficients for child health problems are rather different, and significant only for the parent report. In both cases standard errors for the parent response are smaller, reflecting the larger sample size. Fitting these two logistic regression models separately does not allow us to formally quantify the differences in $\widehat{\beta}^P$ and $\widehat{\beta}^T$ because the estimated regression parameters are correlated.

We now show how to fit these two regression models simultaneously using multivariate methods that take the association between the responses into account. To begin, we rewrite the two separate models as a set of bivariate models with a common regression coefficient β, which will have dimension six (or $2p$ in general). First, we simply change notation as follows. Let $Y_i^P = Y_{i1}, Y_i^T = Y_{i2}, \mu_i^P = \mu_{i1}, \mu_i^T =$

Table 21.4 Results of fitting separate logistic regressions to data on externalizing behavior from each source.

Informant	N	Intercept[a]	Single Parent[a]	Child Health[a]
Parent	2501	-2.156 ± 0.092	0.620 ± 0.124	0.600 ± 0.113
Teacher	1428	-1.694 ± 0.105	0.655 ± 0.157	0.175 ± 0.135

[a]Estimated coefficient $\pm$ standard error.

$\mu_{i2}, \beta' = (\beta^{P\prime}, \beta^{T\prime})$, and let Z_i be an $n_i \times 6$ matrix where n_i is the number of informants available for the i^{th} child. For a child with both informants ($n_i = 2$), the first row of Z_i is given by

$$Z'_{1i} = (X'_i, 0, 0, 0)$$

and the second row of Z_i simply interchanges X'_i with $(0, 0, 0)$,

$$Z'_{2i} = (0, 0, 0, X'_i).$$

If an informant is missing ($n_i = 1$), we delete the row corresponding to that informant (i.e., delete Z_{2i} if the teacher does not give a report for the i^{th} child). We now write

$$E(Y_{ij}) = \mu_{ij},\ i = 1, ..., N \text{ and } j = 1, ..., n_i,$$

and specify the following bivariate model,

$$\text{logit}(\mu_{ij}) = Z'_{ij}\beta,\ j = 1, 2. \tag{21.3}$$

This is exactly the mean model for a marginal model (see Chapters 12 and 13), with a special structure for the design matrix, here labeled Z_i instead of the usual X_i. To complete the marginal model, all we need is a specification of $\text{Cov}(Y_{i1}, Y_{i2})$, using one of the approaches discussed in Chapters 12 and 13. For the analysis here we use the odds ratio to measure the association. Given the model for the mean and $\text{Cov}(Y_{i1}, Y_{i2})$, GEE methods can be used to estimate β. If we specify the entire joint distribution for the Y_{ij}'s, we can use maximum likelihood estimation (Fitzmaurice et al., 1995). We choose to present a GEE analysis since it is easier to implement. Note that, because of the way we defined Z_i and β, the first three components of β correspond to β^P and the second three correspond to β^T. The variance-covariance matrix of $\widehat{\beta}$ is now 6×6; the 3×3 diagonal blocks are the variance-covariance matrices of $\widehat{\beta}^P$ and $\widehat{\beta}^T$, while the off-diagonal 3×3 block gives the covariance between the two.

Table 21.5 shows the results of fitting the regression model, using the log odds ratio to model association between parent and teacher response. The estimates and standard errors for β^P are nearly unchanged, as we might expect, but those for β^T

Table 21.5 Results of fitting two regression models simultaneously to externalizing behavior data on 2501 children using GEE method.

Informant	Intercept[a]	Single Parent[a]	Child Health[a]
Parent	-2.154 ± 0.091	0.616 ± 0.124	0.598 ± 0.113
Teacher	-1.683 ± 0.104	0.602 ± 0.155	0.146 ± 0.135

[a] Estimated coefficient $\pm$ empirical standard error.

are different, reflecting the fact that many children were missing the teacher response. The parent response provides some information about teacher response because of the relatively high association between parent and teacher response (estimated odds ratio is 4.75, with a 95% confidence interval of 3.52 to 6.39). For both $\widehat{\beta}^T$ and $\widehat{\beta}^P$ the standard errors are slightly smaller than with separate logistic regressions, but only very slightly so for the $\widehat{\beta}^P$.

If we use a "working independence" model for $\text{Cov}(Y_{i1}, Y_{i2})$, setting the log odds ratio for the association between parent and teacher response to zero, then the results (not shown) of a GEE analysis yield exactly the same result as separate regressions for $\widehat{\beta}$ (i.e., GEE estimates of β are identical to those reported in Table 21.4). The model-based standard errors are also identical to those reported in Table 21.4 and the 3×3 off-diagonal block of the covariance matrix for $\widehat{\beta}$ is zero. The empirical (or "sandwich") standard errors are very similar to the model-based standard errors. However, they differ slightly because the 3×3 off-diagonal block of the covariance matrix for $\widehat{\beta}$ is estimated as zero by the model-based variance estimator, whereas the empirical variance estimator correctly estimates the covariance between $\widehat{\beta}^P$ and $\widehat{\beta}^T$. That is, the empirical standard errors account for the correlations (ranging from approximately 0.15 to 0.25) between components of $\widehat{\beta}^P$ and $\widehat{\beta}^T$.

The model given by (21.3) is a very general model; its advantages over the separate regressions are that:

1. we can test whether $\beta_k^P = \beta_k^T$ for the k^{th} covariate (or for the whole vector, test $\beta^P = \beta^T$, using $\text{Cov}(\widehat{\beta})$ provided by the GEE analysis);

2. we can use all available data; and

3. it provides a measure of association between the two informants.

With a large number of covariates, we will usually want to fit simpler models. The way we have defined β and Z_i in model (21.3) means that the first p components of β correspond to β^P and the second p correspond to β^T. To formulate simpler models, we need to create a dichotomous indicator variable of informant.

Table 21.6 Results of fitting regression models, with common or shared parameters, simultaneously to data on externalizing behavior using GEE method.

Variable	Estimate	SE[a]	Z
Intercept	−1.685	0.100	−16.85
Informant	−0.467	0.118	−3.96
Single Parent	0.611	0.108	5.68
Child Health	0.146	0.135	1.08
Informant × Child Health	0.452	0.157	2.87

[a]Empirical standard error.

To illustrate, we introduce a dichotomous variable (X_2) which is 1 if the informant is the parent, and 0 if the informant is the teacher. Denoting single-parent status and child health problems by X_3 and X_4, consider a model of the form

$$\text{logit}(\mu_{ij}) = \beta_1 + \beta_2 X_{2ij} + \beta_3 X_{3ij} + \beta_4 X_{4ij}, \tag{21.4}$$

where $X_{1ij} = 1$ for all i and j. This model specifies that the effects of single-parent status (measured by β_3) and child's physical health problems (measured by β_4) are the same regardless of the source of the information, but the mean level may be higher or lower (measured by β_2) depending on informant. A positive β_2 suggests that a positive rating ($Y_{ij} = 1$), here denoting externalizing behaviors in the borderline/clinical range, is more likely from a parent's report than from a teacher's, holding single parent status and physical health problems constant. Notice that only informant is a within-subject variable; that is, $X_{2i1} \neq X_{2i2}$ while X_3 and X_4 are both between-subject variables. Forcing the coefficients of single-parent status to be equal seems reasonable in view of the results presented in Tables 21.4 and 21.5, but not for child health problems.

We construct a model which allows the effect of physical health problems to depend on informant by simply adding the interaction:

$$\text{logit}(\mu_{ij}) = \beta_1 + \beta_2 X_{2ij} + \beta_3 X_{3ij} + \beta_4 X_{4ij} + \beta_5 X_{5ij}, \tag{21.5}$$

where $X_{5ij} = X_{2ij} X_{4ij}$. Fitting a model that includes the interaction allows the probability of a positive rating to depend on informant, single-parent status, and child health problems, and allows the effect of child health problems to differ by informant. The results of this model are displayed in Table 21.6. The estimated odds ratios change very little from the model given by (21.3). The interpretation of the results in Table 21.6 becomes more transparent if the model given by (21.5) is written separately for

the parent and teacher informant:

$$\begin{aligned}\text{logit}(\mu_{i1}) &= \text{logit}(\mu_i^P) &&= (\beta_1+\beta_2)+\beta_3 X_{3ij}+(\beta_4+\beta_5)X_{4ij};\\ \text{logit}(\mu_{i2}) &= \text{logit}(\mu_i^T) &&= \beta_1+\beta_3 X_{3ij}+\beta_4 X_{4ij}.\end{aligned}$$

Because we code informant as 1 for parent response, $\widehat{\beta}_1$ and $\widehat{\beta}_4$ can be regarded as estimated teacher parameters for the intercept and child physical health problems. For single-parent status, $\widehat{\beta}_3$ is the common coefficient for both informants; notice that its standard error is considerably smaller than the two corresponding standard errors for model (21.3) reported in Table 21.5. Finally, from $\widehat{\beta}_5$ we have the difference in effects of child health problems as estimated by parent and teacher evaluations. Comparison of $\widehat{\beta}_5$ to its standard error provides a formal test of the null hypothesis that informant does not matter in evaluating the effect of child health problems. The rejection of this hypothesis may reflect the fact that the rating of child health problems was given by the parent, and the teacher may lack information about child's physical health.

As a rule, if informant interactions are included for all the covariates, then the model is basically equivalent to fitting separate regressions. However, the estimated coefficients obtained from using GEE with a non-zero correlation will differ from those obtained by fitting separate regressions because of the non-linearity of the logistic regression model, even with complete data on each informant. If there are missing data, the coefficients may differ substantially. In our example the estimated effects for the parent respondents are very similar to those obtained from a separate regression on Y_{i1}, with all 2501 observations. The differences for the teacher respondents were more pronounced because of the substantial missing data, and because the two informant responses are highly associated. When we use GEE with a non-zero correlation, the coefficients for the teacher responses use the information in the parent response to provide some information for the missing teacher respondents.

When simpler models are fit (i.e., not all interactions with informant are present) we can expect to gain efficiency in the analysis for the common coefficients. This point is illustrated by comparing the standard errors of the coefficient for single-parent status in Table 21.4 or 21.5 with Table 21.6.

21.6 SUMMARY

This chapter has illustrated how two other types of studies, repeated measures and multiple sources, can be analyzed using methods for correlated data that are comparable to those used for longitudinal data analysis. This is certainly not an exhaustive list, as many study designs produce repeated measures and multiple source data, and their proper analysis requires linear or generalized linear models for correlated data.

Our repeated measures example is also an example of a crossover design, but there are many examples of simpler repeated measures designs with no period or carryover effects; in these cases there are often between-subject variables and the interaction of those with the within-subject variables will be of interest. For our example the response variable is continuous and a linear mixed model is appropriate, but in other

settings one may wish to use a more general covariance structure or, when the response variable is discrete, a generalized linear model for correlated data.

Our multiple informant example used dichotomous reports and logistic regression models, but often multiple source data will be continuous responses. The general approach to constructing multivariate correlated regression models remains the same, but now maximum likelihood and GEE approaches are viable alternatives for the analysis. Likewise there may be more than two sources ($n > 2$). The general approach to analyzing multiple source data can handle any number of informants or sources using $n - 1$ (one less than the number of informants) dichotomous indicators of informant and their interactions with the covariates of interest.

21.7 FURTHER READING

Repeated measures designs are frequently encountered in applications and there is a very large literature on their design and analysis. Accessible descriptions of methods for analyzing repeated measures data can be found in the review articles by Keselman and Keselman (1984), Everitt (1995), and Omar *et al.* (1999).

Methods for analyzing crossover trials are discussed in the review articles by Matthews (1994) and Jones and Donev (1996). Finally, Fitzmaurice et al. (1995) and Goldwasser and Fitzmaurice (2001) discuss the use of regression models for analyzing multiple source data and present a substantive analysis of the multiple informant data from the Connecticut Child Surveys.

Bibliographic Notes

A comprehensive description of the multitude of techniques available for analyzing repeated measures data can be found in the books by Crowder and Hand (1990), Lindsey (1999), and Davis (2002), and the references therein.

A discussion of split-plot designs can be found in Chapter 7 of Cochran and Cox (1957). A comprehensive description of the design and analysis of crossover trials can be found in the books by Jones and Kenward (1989) and Senn (2002), and the references therein.