

5

Modeling the Mean: Analyzing Response Profiles

5.1 INTRODUCTION

In this chapter we present a method for analyzing longitudinal data that imposes minimal structure or restrictions on the mean response over time and on the covariance among the repeated measures. The method focuses on analyzing response profiles and can be applied to longitudinal data when the design is balanced, with the timing of the repeated measures common to all individuals in the study. Although we focus on study designs where all subjects are measured at the same set of n occasions (i.e., *balanced* longitudinal designs), as we show, the analysis of response profiles can also handle incompleteness due to missing data (i.e., incomplete longitudinal studies with balanced designs).

Methods for analyzing response profiles are appealing when there is a single categorical covariate (perhaps denoting different treatment or exposure groups) and when no specific *a priori* pattern for the differences in the response profiles between groups can be specified. When repeated measures are obtained at the same sequence of occasions, the data can be summarized by the estimated mean response at each occasion, stratified by levels of the group factor. At any given level of the group factor, the sequence of means over time is referred to as the mean *response profile*.

For example, consider the blood lead level data from the TLC trial. The mean response profiles for the two groups randomized to succimer and placebo are presented in Figure 5.1. This plot is produced by simply calculating the arithmetic average of the responses at each occasion, within each treatment group, and joining adjacent means with a series of line segments. In settings where data on some subjects are missing, such a plot can still be made but it is obtained from the estimated means

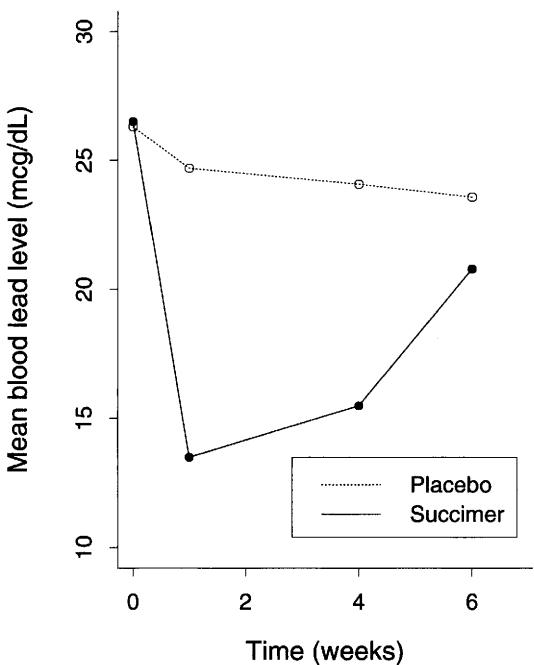


Fig. 5.1 Mean blood lead levels at baseline (week 0), week 1, week 4, and week 6 in the succimer and placebo groups.

for each occasion, stratified by group. We will show how to estimate these mean response profile curves in Section 5.3.

The main goal in the analysis of response profiles is to characterize the patterns of *change* in the mean response over time in the groups and to determine whether the shapes of the mean response profiles differ among the groups. For example, in the TLC trial, the major question of scientific interest is concerned with whether *changes* in the mean blood lead levels are the same for the succimer and placebo groups. In Sections 5.2 and 5.3 we show how questions about whether the patterns of change are the same in all groups translate into hypotheses about the interaction between the group factor and time.

Methods for analyzing response profiles can be extended in a straightforward way to handle the case where there is more than a single group factor and when there are baseline covariates that need to be adjusted for. However, for ease of exposition, we focus on the case where there is only a single group factor. For example, in an observational study the groups might be defined by characteristics of the study subjects, such as age, gender, or exposure level. Alternatively, groups might be defined by random assignment to different treatments or interventions. The distinction

between observational studies and randomized trials is important and, as we will see later, has ramifications for the analysis of response profiles.

A characteristic feature of longitudinal studies is the presence of a baseline measurement. In the TLC trial, the objective is to compare the patterns of change in blood lead levels from baseline over time across the treatment groups. The baseline measurement is an outcome like those measured subsequently, but is unique in that, being pre-randomization, it can be assumed not to depend on treatment group. Indeed, this is apparent in the plot of the mean response profiles in Figure 5.1. This is a common feature of longitudinal studies which involve randomization after baseline. The baseline response may play a special role in other settings as well. For example, sometimes the baseline response is range restricted, as when only subjects with values greater than or less than a threshold are included in the study. With observational studies of growth or decline, groups may be known to differ at baseline, or comparison groups may be selected by matching so that baseline means are comparable.

Thus the question naturally arises as to how to handle the baseline measurement in the assessment of change. This is important, since it will affect how we construct hypothesis tests, and how they should be interpreted. In addition, how we handle the baseline response in the analysis will have an impact on efficiency and the power of tests of hypotheses. In Section 5.6, we describe two ways of adjusting for the baseline value in a simple setting and discuss their relative merits under different longitudinal study designs. In Section 5.7, we compare and contrast a number of alternative strategies for handling the baseline response in more general settings and make recommendations about the preferred strategies in different situations. Many of our readers may find the level of detail in Section 5.7 somewhat daunting. We note that Section 5.7 can be omitted at first reading without loss of continuity. However, we encourage all of our readers to eventually tackle the material in Section 5.7 since appropriate adjustment for baseline is an important aspect of the analysis of longitudinal change.

5.2 HYPOTHESES CONCERNING RESPONSE PROFILES

In our discussion of the analysis of response profiles, we focus initially on the two-group design, but generalizations to more than two groups are straightforward. Given a sequence of n repeated measures on a number of distinct groups of individuals, three main questions concerning the response profiles can be posed:

1. Are the mean response profiles similar in the groups, in the sense that the mean response profiles are parallel?

This is a question that concerns the *group × time interaction effect*. A graphical representation of the null hypothesis of parallel mean response profiles is displayed in Figure 5.2(a).

2. Assuming that the population mean response profiles are parallel, are the means constant over time, in the sense that the mean response profiles are flat?

This is a question that concerns the *time effect*. A graphical representation of

the null hypothesis that the mean response profiles are flat is displayed in Figure 5.2(b).

3. Assuming that the population mean response profiles are parallel, are they also at the same level in the sense that the mean response profiles for the groups coincide?

This is a question that concerns the *group effect*. A graphical representation of the null hypothesis that the mean response profiles are at the same level is displayed in Figure 5.2(c).

In longitudinal studies, the first question that is of main scientific interest. The hypothesis of parallel response profiles corresponds to the hypothesis that the patterns of change in the mean response over time are the same across groups. This comparison of change in the response over time is the *raison d'être* of a longitudinal study. In contrast, as we will see later, the second and third questions may not have any scientific relevance. That is, even when the response profiles can be assumed to be parallel, any interest in the second and third questions is secondary and depends on the longitudinal study design.

Note that the second and third questions have implicitly made an assumption about the answer to the first. There is a very good reason for doing so. Except in very rare circumstances, it is not meaningful to ask the second and third questions if the mean response profiles are not parallel. Indeed, this is consistent with the general principle that *main effects* (e.g., group or time effects) are ordinarily not of interest when there is an interaction among them. That is, when there is a group \times time interaction, the mean response profiles in the groups are different (non-parallel profiles), consequently their shape can be described only with reference to a specific group, and their level can be described only with reference to a specific time.

The appropriate scientific hypotheses in any particular study must be derived from the relevant scientific issues in that investigation. Here it becomes important to distinguish between longitudinal data arising from a randomized trial and from an observational study. In the former case, when study participants have been randomized to treatment groups and the baseline value of the response has been obtained prior to any study interventions, the mean response at occasion 1 is independent of treatment assignment. That is, by design, the group means are equal at baseline (occasion 1). In contrast, in an observational study, there is no a priori reason to assume that the groups have the same mean response at baseline unless the groups were selected by matching on baseline response.

Consider a randomized longitudinal clinical trial comparing treatments where the measurement at the first occasion is a baseline response, obtained prior to any study interventions. For example, in the TLC trial, the blood lead levels at baseline were obtained prior to receiving placebo or succimer. In that case the only question of scientific interest is the first because it addresses whether the patterns of change in the mean response over time are the same in all groups. For example, in the TLC trial, the test of the group \times time interaction assesses whether changes in the mean blood levels are the same for the succimer and placebo groups. In a randomized trial, the second question is usually of less importance because it does not involve a direct

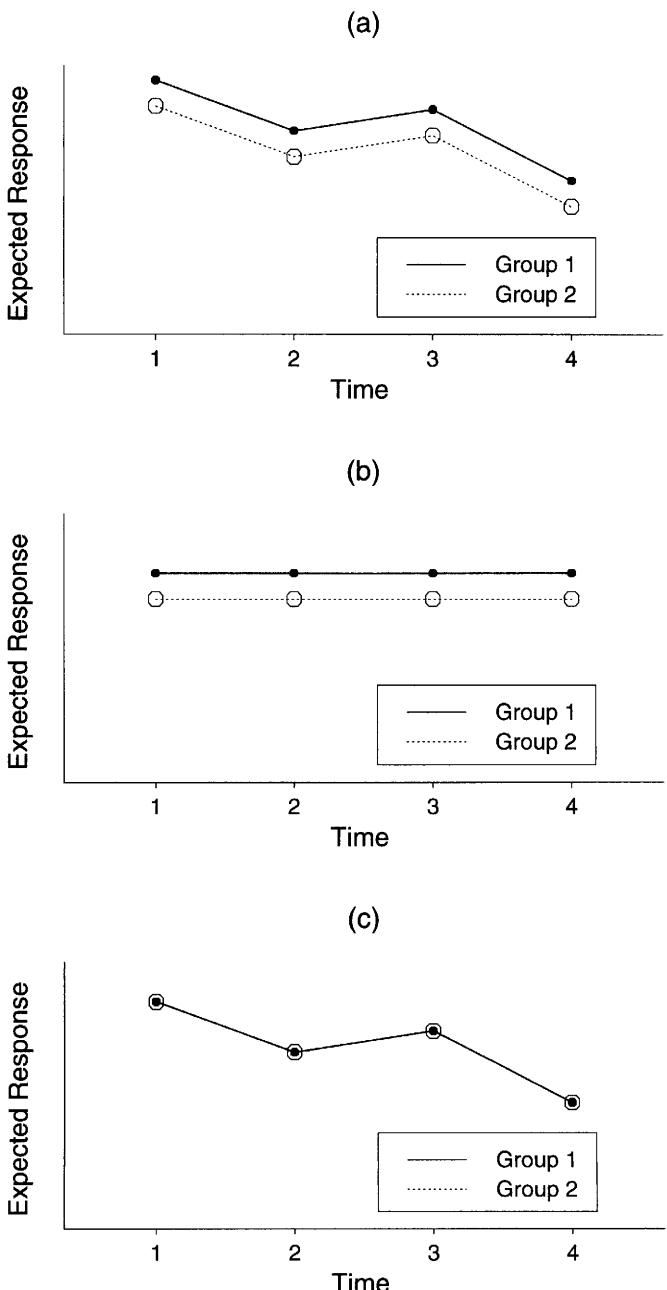


Fig. 5.2 Graphical representation of the null hypotheses of (a) no group \times time interaction effect, (b) no time effect, and (c) no group effect.

comparison of groups. The second question concerns the time effect, where the focus is on the comparison of the mean response at each occasion averaged over the groups. Hypotheses concerning the main effect of time translate into questions concerning whether the overall (i.e., averaged over groups) mean response has changed from baseline. Finally, in a randomized trial where the baseline response has been obtained before any study interventions, there is no interest in the third question. The third question concerns the group effect. However, in this setting the absence of a group \times time interaction implies that there is no group effect. That is, if the groups have the same pattern of change over time and, by design, do not differ at baseline, their mean response profiles must necessarily coincide. As a result the test of group effect is subsumed within the test of group \times time interaction.

In an observational study where the group factor might represent different exposures or inherent characteristics of the individuals, the first question is usually of primary interest. It addresses the fundamental question of whether patterns of change over time in the mean response vary by group. In contrast to a randomized trial, however, the second and third questions may also be of substantive interest. For example, in a longitudinal study of growth or aging, there may be interest in the pattern of change in the mean response over time, even when the pattern of change is the same in all groups. This concerns the main effect of time and is addressed by the second question. Ordinarily, when there is interest in the time effect, the preferred way to describe the trend in the mean response is with a relatively simple parametric curve; methods for fitting parametric or semiparametric curves to the mean response are described in Chapter 6. Finally, in an observational study, there may be interest in group comparisons of the mean response averaged over time. This concerns the group effect and is addressed by the third question. However, in the absence of any group \times time interaction, it must be recognized that the test for the main effect of group represents a comparison of the groups in terms of their baseline (occasion 1) response. That is, if the groups have the same pattern of change over time, any group differences in the overall (i.e., averaged over occasions) mean response must reflect existing baseline differences among the groups.

To highlight the main features of the analysis of response profiles, consider the following example from a two-group study comparing a novel *treatment* and a *control*. We assume that the two groups have repeated measurements at the same set of n occasions. The analysis of response profiles is based on comparing the mean response profiles in the two groups. In a somewhat relaxed notation, let $\mu(T) = \{\mu_1(T), \dots, \mu_n(T)\}'$ denote the mean response profile for the treatment group and $\mu(C) = \{\mu_1(C), \dots, \mu_n(C)\}'$ denote the mean response profile for the control group. The population means in the two groups at each occasion are given in Table 5.1.

In this hypothetical study we are primarily interested in testing scientific hypotheses that compare the novel treatment and the control in terms of *changes* in the mean response over time. This can be determined by considering the null hypothesis of no group \times time interaction; that is, the null hypothesis that the mean response profiles are parallel. If the mean response profiles are parallel, the difference in the means between the two groups is constant over time. As a result in terms of Δ_j , the null

Table 5.1 Mean response profile over time in the treatment and control groups.

Group	Measurement Occasion			
	1	2	...	n
Treatment	$\mu_1(T)$	$\mu_2(T)$...	$\mu_n(T)$
Control	$\mu_1(C)$	$\mu_2(C)$...	$\mu_n(C)$
Difference	Δ_1	Δ_2	...	Δ_n

Note: $\Delta_j = \mu_j(T) - \mu_j(C)$.

hypothesis is

$$H_{01}: \Delta_1 = \Delta_2 = \cdots = \Delta_n,$$

where $\Delta_j = \mu_j(T) - \mu_j(C)$. If the null hypothesis is rejected, the two groups have non-parallel mean response profiles and the patterns of change over time differ in the two groups. Note that the number of constraints on the mean responses under this null hypothesis is $n - 1$. As a result the test of this null hypothesis has $n - 1$ degrees of freedom. In Section 5.3 we describe how the constraints can be expressed in terms of specific contrasts of the means.

The previous illustration focused on the special case of $G = 2$ groups; however, the main ideas can be generalized in a straightforward way when there are more than two groups. When there are G groups with repeated measurements at the same set of n occasions, we let $\mu(g) = \{\mu_1(g), \dots, \mu_n(g)\}'$ denote the mean response profile for the g^{th} group ($g = 1, \dots, G$). The population means in the G groups at each occasion are given in Table 5.2. With $G > 2$, we can compare groups in a number of different ways. However, with G groups, there are only $G - 1$ non-redundant comparisons. We define $\Delta_j(g) = \mu_j(g) - \mu_j(G)$, (for $j = 1, \dots, n$; $g = 1, \dots, G - 1$). That is, $\Delta_j(g)$ is a contrast or comparison of the mean response at the j^{th} occasion for the g^{th} group (for $g = 1, \dots, G - 1$) with the mean response at the j^{th} occasion in group G . Then the null hypothesis that the mean response profiles are parallel is

$$H_{01}: \Delta_1(g) = \Delta_2(g) = \cdots = \Delta_n(g); \text{ for } g = 1, \dots, G - 1.$$

With $G \geq 2$, the test of the null hypothesis of no group \times time interaction effect has $(G - 1) \times (n - 1)$ degrees of freedom.

So far our discussion of the analysis of response profiles has focused on an omnibus test of the group \times time interaction. However, unless the test of the group \times time interaction has only a single degree of freedom, this test does not help in discerning in what manner the patterns of change over time differ across groups. For the latter we must consider estimates (and their standard errors) of relevant contrasts of the

Table 5.2 Mean response profile over time in G groups.

Group	Measurement Occasion			
	1	2	...	n
1	$\mu_1(1)$	$\mu_2(1)$...	$\mu_n(1)$
2	$\mu_1(2)$	$\mu_2(2)$...	$\mu_n(2)$
⋮	⋮	⋮	⋮	⋮
g	$\mu_1(g)$	$\mu_2(g)$...	$\mu_n(g)$
⋮	⋮	⋮	⋮	⋮
G	$\mu_1(G)$	$\mu_2(G)$...	$\mu_n(G)$

means. In the next section, we describe a general linear model formulation of the analysis of response profiles. As we will show, the analysis of response profiles can be formulated in such a way that certain regression parameters have interpretations that bear directly on the questions of main scientific interest.

5.3 GENERAL LINEAR MODEL FORMULATION

Before we illustrate the main ideas with a numerical example, we consider how the analysis of response profiles can be implemented in the general linear model

$$E(Y_i|X_i) = \mu_i = X_i\beta,$$

for appropriate choices of X_i . We also describe how the main hypothesis of no group \times time interaction effect can be expressed in terms of β . Let n be the number of repeated measures and N the number of subjects. To express the model for the longitudinal design with G groups and n occasions of measurement, we require $G \times n$ parameters for the G mean response profiles.

For example, with two groups measured at three occasions, there are $2 \times 3 = 6$ mean parameters (see Table 5.2). For the first group, let the design matrix X_i be the following 3×6 matrix:

$$X_i = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix},$$

while for the second group, let the design matrix be

$$X_i = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

Then, in terms of the model

$$E(Y_i|X_i) = \mu_i = X_i\beta,$$

where $\beta = (\beta_1, \dots, \beta_6)'$ is a 6×1 vector of regression coefficients,

$$\mu(1) = \begin{pmatrix} \mu_1(1) \\ \mu_2(1) \\ \mu_3(1) \end{pmatrix} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix};$$

similarly

$$\mu(2) = \begin{pmatrix} \mu_1(2) \\ \mu_2(2) \\ \mu_3(2) \end{pmatrix} = \begin{pmatrix} \beta_4 \\ \beta_5 \\ \beta_6 \end{pmatrix}.$$

As a result hypotheses about the mean response profiles in the two groups that were previously expressed in terms of $\mu(1) = \{\mu_1(1), \mu_2(1), \mu_3(1)\}'$ and $\mu(2) = \{\mu_1(2), \mu_2(2), \mu_3(2)\}'$ can easily be re-expressed in terms of hypotheses about the components of β . Specifically, the hypothesis of no group \times time interaction effect can be expressed as

$$H_{01}: (\beta_1 - \beta_4) = (\beta_2 - \beta_5) = (\beta_3 - \beta_6).$$

In this parameterization, hypotheses about the group \times time interaction cannot be expressed in terms of certain components of β being zero; instead, these hypotheses can be expressed in terms of $L\beta = 0$, for particular choices of vectors or matrices L . For example, the null hypothesis of no group \times time interaction effect,

$$H_{01}: (\beta_1 - \beta_4) = (\beta_2 - \beta_5) = (\beta_3 - \beta_6),$$

can be expressed as

$$H_{01}: L\beta = 0,$$

where

$$L = \begin{pmatrix} 1 & -1 & 0 & -1 & 1 & 0 \\ 1 & 0 & -1 & -1 & 0 & 1 \end{pmatrix}.$$

An attractive feature of the general linear model formulation,

$$E(Y_i|X_i) = \mu_i = X_i\beta,$$

is that it can handle settings where the data for some subjects are missing. For example, suppose that the i^{th} subject belongs to the first group and is missing the response at the third occasion. The appropriate design matrix for that subject is the following 2×6 matrix, obtained by removing the last row of the full data design matrix for subjects from the first group:

$$X_i = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

For more general patterns of missingness, the appropriate design matrix for the i^{th} subject is simply obtained by removing rows of the full data design matrix corresponding to the missing responses. This allows the analysis of response profiles to be based on all available observations of the subjects.

Note that the general linear model for two groups measured at three occasions,

$$E(Y_i|X_i) = \mu_i = X_i\beta,$$

could also have been expressed in terms of the following two design matrices:

$$X_i = \begin{pmatrix} 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 \end{pmatrix},$$

for the first group and

$$X_i = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \end{pmatrix},$$

for the second group. In that case

$$\mu(2) = \begin{pmatrix} \mu_1(2) \\ \mu_2(2) \\ \mu_3(2) \end{pmatrix} = \begin{pmatrix} \beta_1 \\ \beta_1 + \beta_2 \\ \beta_1 + \beta_3 \end{pmatrix};$$

and

$$\mu(1) = \begin{pmatrix} \mu_1(1) \\ \mu_2(1) \\ \mu_3(1) \end{pmatrix} = \begin{pmatrix} \beta_1 + \beta_4 \\ (\beta_1 + \beta_4) + (\beta_2 + \beta_5) \\ (\beta_1 + \beta_4) + (\beta_3 + \beta_6) \end{pmatrix}.$$

The choice of “reference group” (here the second group) is arbitrary, and we have used the convention adopted by many of the procedures in SAS; a more detailed

discussion of the “reference group” parameterization is given at the end of this section. With this choice of design matrices for the two groups, the interpretation of the regression coefficients β has changed. Re-expressing hypotheses about the mean response profiles for the two groups in terms of hypotheses about the components of β , we write the hypothesis of no group \times time interaction as

$$H_{01}: \beta_5 = \beta_6 = 0.$$

Although both alternative parameterizations considered thus far allow for testing of hypotheses about the response profiles, the second parameterization is more convenient since the hypothesis of no group \times time interaction is represented by the vanishing (or setting to zero) of certain components of β . Also the second parameterization, often called the “reference group” parameterization, is the one that is commonly adopted by many statistical software packages (e.g., PROC MIXED in SAS).

As indicated earlier, when the hypothesis of parallel profiles cannot be rejected, hypotheses concerning the main effects of time and/or group may be of secondary interest, although their relevance depends on the design of the study. Hypotheses concerning the main effects of time and group can similarly be represented by the vanishing (or setting to zero) of certain components of β . For example, with two groups measured at three occasions and assuming parallel profiles ($\beta_5 = \beta_6 = 0$), the hypothesis of no time effect is

$$H_{02}: \beta_2 = \beta_3 = 0;$$

the hypothesis of no group effect is

$$H_{03}: \beta_4 = 0.$$

For the more general case with G groups measured at n occasions, the number of constraints under H_{02} is $n - 1$ and is the same regardless of the number of groups; the test of H_{02} has $n - 1$ degrees of freedom. Similarly the number of constraints under H_{03} is $G - 1$ and is the same regardless of the number of occasions; the test of H_{03} has $G - 1$ degrees of freedom. Both of these hypotheses can be tested by considering a model with only main effects of group and time. That is, valid tests of the main effects of group and time are obtained from the reduced model that excludes the group \times time interaction.

Finally, given that the analysis of response profiles can be expressed in terms of the linear regression model

$$E(Y_i|X_i) = \mu_i = X_i\beta,$$

where $\beta = (\beta_1, \dots, \beta_p)'$ is a $p \times 1$ vector of regression coefficients (with $p = G \times n$), maximum likelihood estimation of β , and the construction of tests of the group \times time interaction (and the main effects of time and group), are possible once the covariance of Y_i has been specified. In the analysis of response profiles, the covariance of Y_i

is usually assumed to be unstructured with no constraints on the $\frac{n(n+1)}{2}$ covariance parameters other than the requirement that they yield a symmetric matrix that is positive-definite (the condition that the covariance matrix is positive-definite ensures that while the repeated measures can be highly correlated, there must be no redundancy in the sense that one of the repeated measures can be expressed as a linear combination of the others; the condition also ensures that no linear combination of the responses can have a negative variance). Given REML (or ML) estimates of β , and their standard errors (and the estimated covariance of $\hat{\beta}$), tests of the group \times time interaction (and the main effects of time and group), can be constructed using multivariate Wald tests. Alternatively, likelihood ratio tests can be constructed but require that the model be fit to the data with and without the constraints under the null hypothesis (i.e., fitting the “reduced” and “full” models, respectively). Before illustrating the analysis of response profiles, we present a brief review of the “reference group” parameterization.

Review: Reference Group Parameterization

Consider a group factor with G levels. To represent this factor in a linear model, we can define a set of “dummy” or indicator variables:

$$Z_{ig} = \begin{cases} 1 & \text{if } i^{\text{th}} \text{ subject belongs to group } g, \\ 0 & \text{otherwise.} \end{cases}$$

Letting $X_i = (Z_{i1}, \dots, Z_{iG})$, the mean response in the G groups, denoted by $\mu_i(1), \dots, \mu_i(G)$, can be expressed in terms of the following linear model:

$$E(Y_i|X_i) = \mu_i = X_i\beta.$$

In this parameterization,

$$\begin{pmatrix} \mu_i(1) \\ \mu_i(2) \\ \vdots \\ \mu_i(G-1) \\ \mu_i(G) \end{pmatrix} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_{G-1} \\ \beta_G \end{pmatrix}.$$

If we wish to include an “intercept”, say β_1 , by setting the first column of X_i to 1 (for all $i = 1, \dots, N$), then there is redundancy in X_i if all G indicator variables, Z_{i1}, \dots, Z_{iG} , are also included in the design vector, X_i . To avoid this over-specification, one of the indicator variables must be excluded from X_i . Arbitrarily, we can drop Z_{iG} . Then, with $X_i = (1, Z_{i1}, \dots, Z_{iG-1})$, the mean response in the G groups can be expressed in terms of the following linear model:

$$E(Y_i|X_i) = \mu_i = X_i\beta,$$

where $\beta = (\beta_1, \dots, \beta_G)'$. In this parameterization,

$$\begin{pmatrix} \mu_i(1) \\ \mu_i(2) \\ \vdots \\ \mu_i(G-1) \\ \mu_i(G) \end{pmatrix} = \begin{pmatrix} \beta_1 + \beta_2 \\ \beta_1 + \beta_3 \\ \vdots \\ \beta_1 + \beta_G \\ \beta_1 \end{pmatrix}.$$

Because the “intercept” term, β_1 , is also the mean of group G , and all of the remaining components of β represent deviations from the mean of group G , this parameterization is often referred to as the “reference group” parameterization. Here the last level of the group factor (i.e., group G) is the reference group, and it is no coincidence that this is the same group whose indicator variable was excluded from X_i . Other choices of reference group can be obtained by excluding the relevant indicator variable for the group in question from X_i .

5.4 CASE STUDY

Next we illustrate the main ideas by conducting an analysis of response profiles of the blood lead data of the 100 children from the succimer and placebo groups of the Treatment of Lead-Exposed Children (TLC) Trial.

Treatment of Lead-Exposed Children Trial

Recall that the TLC trial was a placebo-controlled, randomized trial of an orally administered chelating agent, succimer, in children with confirmed blood lead levels of 20 to 44 $\mu\text{g}/\text{dL}$. The children in the trial were aged 12 to 33 months and lived in deteriorating inner city housing. The following analysis is based on data on blood lead levels at baseline (or week 0), week 1, week 4, and week 6 during the first treatment period. The mean response profiles for the two groups were displayed in Figure 5.1.

In Table 5.3 the REML estimates of the components of the unstructured covariance matrix are displayed. Note the discernible increase in the variability in blood lead levels from pre- to post-randomization. This increase in variability from baseline is probably due to two factors. First, within each treatment group there may be natural heterogeneity in the individual response trajectories over time. Second, the trial had an inclusion criterion that blood lead levels at baseline were in the range of 20 to 44 $\mu\text{g}/\text{dL}$; this may partially account for the smaller variance at baseline.

In Table 5.4 the results of the analysis of response profiles are presented. The main interest is in the test of the group \times time interaction. The test of the group \times time interaction is based on a multivariate Wald test. The test provides a simultaneous test of $H_0: L\beta = 0$ versus $H_A: L\beta \neq 0$, for a suitable choice of L , and the test statistic can be constructed as

$$W^2 = (\widehat{L\beta})' \{ L \widehat{\text{Cov}}(\widehat{\beta}) L' \}^{-1} (\widehat{L\beta}),$$

Table 5.3 Estimated covariance matrix for the blood lead levels at baseline, week 1, week 4, and week 6 for the children from the TLC trial.

Covariance Matrix			
25.2	19.1	19.7	22.2
19.1	44.3	35.5	29.7
19.7	35.5	47.4	30.6
22.2	29.7	30.6	58.7

Table 5.4 Wald tests of fixed effects based on an analysis of response profiles of the blood lead level data at baseline, week 1, week 4, and week 6 for the children from the TLC trial.

Variable	DF	Chi-Squared	P-Value
Group	1	25.43	<0.0001
Week	3	184.48	<0.0001
Group × Week	3	107.79	<0.0001

and compared to a χ^2 distribution with degrees of freedom equal to the number of rows of L . The corresponding likelihood ratio test could be constructed and would require the comparison of the maximized ML log-likelihood for two models, one model that incorporates the constraint that $L\beta = 0$ (i.e., the model without group \times time interaction), the other model unconstrained (i.e., the model with group \times time interaction).

In the TLC trial the question of main scientific interest concerns the comparison of the two treatment groups in terms of their patterns of change from baseline in the mean blood lead levels. This question translates directly into a test of the group \times time interaction. From Table 5.4 the test of the group \times time interaction yields a Wald statistic of 107.79 with 3 degrees of freedom (the corresponding likelihood ratio test yields $G^2 = 74.2$). When compared with the reference chi-squared distribution with 3 degrees of freedom, there is strong evidence to reject the null hypothesis and conclude that the patterns of change from baseline are not the same in the two groups. Given the pattern of observed responses (see Figure 5.1), this result is expected.

The tests of main effects of group and time in Table 5.4 are not meaningful, for two quite different reasons. First, the TLC data are from a randomized trial and the test of the main effect of time is not of subject-matter interest while the test of the

Table 5.5 Estimated regression coefficients and standard errors based on an analysis of response profiles of the blood lead level data at baseline, week 1, week 4, and week 6 for the children from the TLC trial.

Variable	Group	Week	Estimate	SE	Z
Intercept			26.272	0.710	36.99
Group	S		0.268	1.005	0.27
Week		1	-1.612	0.792	-2.04
Week		4	-2.202	0.815	-2.70
Week		6	-2.626	0.889	-2.96
Group × Week	S	1	-11.406	1.120	-10.18
Group × Week	S	4	-8.824	1.153	-7.66
Group × Week	S	6	-3.152	1.257	-2.51

main effect of group is subsumed within the test of group × time interaction. Second, in general, the tests of main effects of time and group in Table 5.4 are not meaningful in the presence of a significant group × time interaction. This underscores our earlier advice that tests of main effects should only be considered when the assumption of parallel response profiles is tenable. When the profiles are parallel, and there is scientific interest in the main effects of time and/or group, the tests of the main effects of time and group require that the model be re-fit to the data, excluding the group × time interaction. The resulting Wald tests for the main effects from this reduced model have the desired interpretations.

So far our analysis of response profiles has provided an omnibus test of the group × time interaction. However, unless the test of the group × time interaction has only a single degree of freedom, this test does not indicate how the two groups differ. For the latter, we must consider the REML estimates of β and their standard errors presented in Table 5.5; alternative single degree of freedom tests for group × time interaction will be discussed in Section 5.5. For ease of interpretation, the baseline (week 0) is chosen as the reference level for time and the placebo group is chosen as the reference level for treatment group. From an examination of the three single-degree-of-freedom contrasts for the group × time interaction, the results indicate that children treated with succimer have a discernibly greater decrease in mean blood lead levels from baseline at all occasions when compared to the children treated with placebo. For example, when compared to the placebo group, the succimer group has an additional 3.152 $\mu\text{g}/\text{dL}$ (with SE = 1.257) decrease in mean blood lead levels from baseline to week 6. Of note, there are even larger differences between the two treatment groups earlier in the trial. For example, when compared to the placebo group, the succimer group has an additional 11.406 $\mu\text{g}/\text{dL}$ decrease in mean blood

lead levels from baseline to week 1. The apparent rebound in blood lead levels after week 1 in the succimer group is thought to be due to lead that is stored in the bones being mobilized, resulting in a new equilibrium in blood lead levels in the children treated with succimer.

We remind the reader that our earlier warning about testing main effects in the presence of interactions applies also to the results in Table 5.5. For example, the test of the main effect of group in Table 5.5 ($Z = 0.27$) does not compare the average (over occasions) response in the succimer and placebo groups; instead, it compares the mean response at baseline (here the reference level for time) in the two treatment groups. The lack of equivalence between the tests for the main effect of group in Tables 5.4 and 5.5 is a direct consequence of the reference group parameterization adopted here (and commonly used by many statistical software packages; e.g., PROC MIXED in SAS).

Finally, because the TLC data are from a randomized trial, the mean response at baseline is independent of treatment assignment (as was confirmed by the non-significant test of the main effect of group in Table 5.5). Because of the random assignment to treatment groups, this suggests that the analysis of response profiles could be simplified by fitting a model that omits the main effect of group, thereby forcing the two groups to have the same mean response at baseline. In Sections 5.6 and 5.7 we consider the merits of such an adjustment and compare and contrast alternative strategies for handling the baseline response in different settings. In these sections we highlight how the analysis of response profiles is a flexible method that can easily be adapted to account for the design of the study and to address questions that are scientifically relevant to any particular study.

5.5 ONE-DEGREE-OF-FREEDOM TESTS FOR GROUP BY TIME INTERACTION

As we saw in the previous section, the test for group \times time interaction is quite general. It posits no specific pattern for the difference in the response profiles between groups. This lack of specificity becomes a problem in studies with a large number of occasions of measurement because the general test for group \times time interaction, with $(G - 1) \times (n - 1)$ degrees of freedom, becomes less sensitive to an interaction with a specific pattern as n increases. Even with as few as three or four occasions of measurement, the general test for interaction will not be as sensitive to specific departures from parallelism as the more focused tests we discuss in this section.

In the typical randomized trial of interventions, subjects are randomized to the intervention groups at baseline and the investigator seeks to determine whether the pattern of response after randomization differs between groups. Randomization implies that the mean at baseline is independent of treatment group; that is, by design, the groups have the same mean response at baseline. In that setting, analysts frequently specify a single contrast believed to best represent the direction in which the pattern of response will differ most markedly. For example, if we assume the

first parameterization described in Section 5.3 with two groups and wish to test for equality of the difference between the average response at occasions 2 through n and the baseline value in the two groups, we can choose the contrast

$$L = (-L_1, L_1),$$

where

$$L_1 = \left(-1, \frac{1}{n-1}, \frac{1}{n-1}, \dots, \frac{1}{n-1} \right).$$

Here L_1 computes the mean response from occasions 2 through n and subtracts the mean response at baseline for a single group. The latter can be thought of as the average change over the interval for a single group. Thus L is a group contrast of this average change in the two groups.

A variant of this approach, known as *area under the curve minus baseline*, or sometimes simply AUC, corresponds to a calculation of the area under the trapezoidal curve created by connecting the responses plotted at the respective time points and subtracting $y_1 \times (t_n - t_1)$, the area of the rectangle of height y_1 and width $t_n - t_1$. For illustrative purposes, the AUC of the profile of blood lead levels for a single subject in the TLC trial is shown in Figure 5.3. For this participant, as for most participants in the TLC trial, the AUC is negative because the responses after intervention begins are smaller than the baseline value. The AUC (minus baseline) can be constructed by subtracting the baseline mean, μ_1 , from each of the means, μ_1 through μ_n , and calculating the area under the trapezoid constructed by connecting these differences. To test for the equality of the AUC in two groups, we would employ the contrast

$$L = (-L_2, L_2),$$

where

$$L_2 = \frac{1}{2} \times (t_1 + t_2 - 2t_n, t_3 - t_1, \dots, t_{j+1} - t_{j-1}, \dots, t_n - t_{n-1})$$

and $\frac{1}{2} \times (t_{j+1} - t_{j-1})$ is the value of the contrast vector for time points other than 1 (baseline) or n (the last occasion). These contrast weights are not intuitively obvious, but can be derived from the formula for the area of a trapezoid. Although the curve presented in Figure 5.3 suggests that L is applied to the individual observations, we must emphasize that the contrast weights are applied to the estimated means, not the individual observations. As a result the AUC can be estimated in setting where some subjects have missing response data.

A third popular method for constructing a single-degree-of-freedom test corresponds to a test of the hypothesis that the trend over time is the same in the several treatment groups. Because this method is a special case of growth curve analysis, to be discussed in depth in the next chapter, and because the expected pattern of response to chelation therapy in the TLC trial would not predict a linear trend in blood lead levels during the treatment period in the group receiving succimer therapy, we defer a discussion of this approach to Chapter 6.

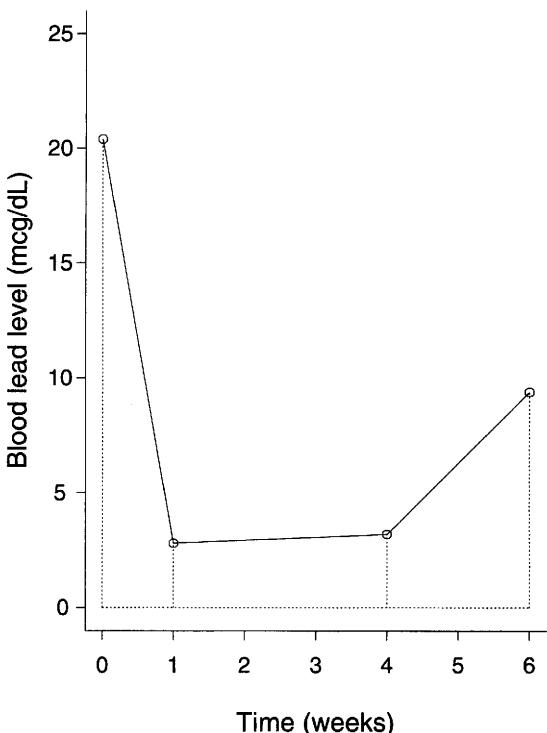


Fig. 5.3 Area under the curve, calculated using the trapezoidal rule, for the profile of blood lead levels for a single subject in the TLC trial.

Application to the Treatment of Lead-Exposed Children Trial

Since the TLC trial measured blood lead levels at four time points during the first treatment period, the vector representing the contrast based on the mean response at times 2 through n minus baseline is given by

$$L = (-L_1, L_1) = \left(1, -\frac{1}{3}, -\frac{1}{3}, -\frac{1}{3}, -1, \frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right).$$

For the data displayed in Figure 5.3, the change in mean response relative to baseline is -15.27 . Because blood lead levels declined for this subject, as for most participants in the TLC trial, the mean response relative to baseline is negative. From the descriptive statistics in Table 5.6 we can easily determine that the average value of the mean response minus baseline is -9.90 in the succimer group and -2.17 in the placebo group. Thus, if we assume the first parameterization described in Section 5.3, then $\widehat{L\beta} = 7.73$ and the value of the Wald test statistic is $Z = 8.21$ (or $W^2 = 67.4$, with

Table 5.6 Mean blood lead levels (and standard deviation) at baseline, week 1, week 4, and week 6 for the children from the TLC trial.

Group	Baseline	Week 1	Week 4	Week 6
Succimer	26.5 (5.0)	13.5 (7.7)	15.5 (7.8)	20.8 (9.2)
Placebo	26.3 (5.0)	24.7 (5.5)	24.1 (5.8)	23.6 (5.6)

one degree of freedom), indicating a highly significant difference in the response pattern between treatment groups.

Similarly, because the time points in the TLC trial were 0, 1, 4, and 6 weeks, the contrast for comparing the AUC (minus baseline) in the two treatment groups is given by

$$L = (-L_2, L_2) = (5.5, -2, -2.5, -1, -5.5, 2, 2.5, 1).$$

The area under the curve for the single subject shown in Figure 5.3 is -89.2 . From Table 5.6, the estimated mean AUC is -59.20 in the succimer group and -11.40 in the placebo group. Thus, if we assume the first parameterization described in Section 5.3, then $L\hat{\beta} = 47.8$, yielding a Wald statistic of $Z = 8.97$ (or $W^2 = 80.5$, with one degree of freedom), again highly statistically significant. Thus both methods of analysis provide a clear signal that the response profile differs in the two treatment groups.

Because the TLC trial data provide unequivocal evidence of an effect of succimer on blood lead level, the added sensitivity to treatment effects achieved by the greater specificity of a one-degree-of-freedom test is not important in this application. In many applications, however, the one-degree-of-freedom test will be statistically significant when the overall test for group \times time interaction is not. For valid application of conventional significance levels, the form of the contrast must be specified prior to data analysis. Otherwise, one would be at risk of seeking the best contrast and testing its significance as if it had been chosen in advance. To guard against this criticism, the protocols for randomized trials usually specify the form of the contrast. This requirement highlights a hazard of one-degree-of-freedom tests. The added sensitivity comes at the price of reduced generality. If the difference between treatment groups takes a form quite different from the pattern anticipated by the contrast, one can fail to obtain a statistically significant result for a one-degree-of-freedom test even when the overall test for group \times time interaction is statistically significant. Thus one-degree-of-freedom tests should be employed only when there is sufficient prior information to specify the contrast with confidence.

5.6 ADJUSTMENT FOR BASELINE RESPONSE

When the data are complete, each of the one-degree-of-freedom tests described in the previous section can be constructed by calculating a univariate summary statistic for each study participant and performing a test for equality of means of these summary statistics in the G groups. With complete data, group comparisons of these summary statistics are equivalent to applying the corresponding contrast weights to the mean responses. This is because the difference in the means is the mean of the differences when each subject is measured at every occasion. Moreover, for each of the two tests described in detail, mean response minus baseline and AUC minus baseline, the summary statistic corresponds to subtracting the baseline value from a summary of the responses on occasions 2 through n . For example, for the test for equality of mean response minus baseline, the summary statistic for the i^{th} participant is given by

$$\frac{(Y_{i2} + Y_{i3} + \cdots + Y_{in})}{n - 1} - Y_{i1}. \quad (5.1)$$

With this representation in mind, some analysts have suggested an alternative approach analogous to analysis of covariance (ANCOVA), in which a summary of the response at times 2 through n becomes the dependent variable and the baseline value enters the analysis as a covariate. When the response variable is the mean at occasions 2 through n and we wish to test for the equality of the mean in two treatment groups, we can write the corresponding univariate model as

$$Y_i^* = \beta_1 + \beta_2 Y_{i1} + \beta_3 \text{trt}_i + e_i^*, \quad (5.2)$$

where

$$Y_i^* = \frac{(Y_{i2} + Y_{i3} + \cdots + Y_{in})}{n - 1}$$

is the mean response at occasions 2 through n for the i^{th} subject, trt_i is an indicator variable distinguishing the two treatment groups, and e_i^* is the error term in the univariate model. This model assumes that the data are complete, and it cannot be fit with missing data; we defer a discussion of more general approaches for handling baseline response to Section 5.7.

An analysis based on either (5.1) or (5.2) will be especially appealing in settings where initial changes from baseline are expected to persist throughout the duration of follow-up. For example, in a trial where the impact of the intervention on changes in the mean response at the start of follow-up is expected to be similar to that toward the end of follow-up; this pattern for the mean response profiles is illustrated in Figure 5.4. Tests based on (5.1) or (5.2) correspond to a comparison between groups of the mean responses on occasions 2 through n , with adjustment for baseline, and have $G - 1$ degrees of freedom irrespective of the number of occasions of measurement.

This raises a question about whether one should incorporate the baseline value through the contrast given by (5.1) or through the analysis of covariance model given by (5.2) in a specific application. The answer depends critically on whether the data arose from an observational study or a randomized trial. If the study is an

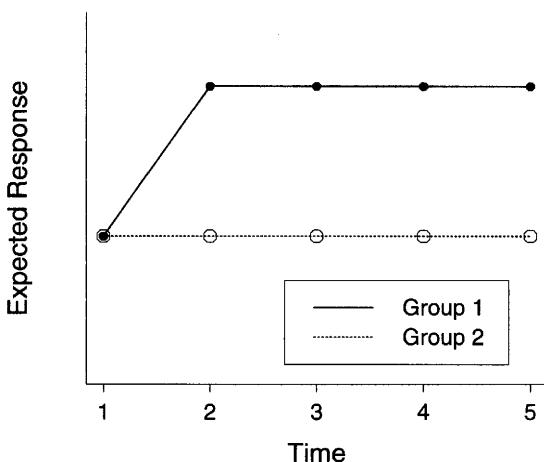


Fig. 5.4 Graphical representation of changes in the mean response from baseline (in Group 1) that persist throughout the duration of follow-up.

observational one, for example, a longitudinal study of the determinants of rate of decline of pulmonary function in adults, it is usually not advisable to employ the analysis of covariance approach because the baseline value may be associated with other variables whose effects are to be studied, raising problems of confounding in an analysis intended to describe how the pattern of response over time is influenced by the characteristics of study participants. For example, individuals who are smokers as adults might have smoked during adolescence. If smoking affected the attained pulmonary function level for young adults, then smoking will likely be associated with pulmonary function level later in adult life, even if cigarette smoking does not influence the rate of decline of pulmonary function with age. Thus adjustment for baseline pulmonary function level using (5.2) could introduce an association between smoking status and rate of decline of pulmonary function, even if the unadjusted rates of decline are nearly equivalent in the various smoking groups.

When participants have been randomized to the several treatment groups and the baseline value has been obtained before any study interventions, adjustment for baseline through analysis of covariance is of interest. In that setting the mean response at time 1 is independent of treatment assignment. One can then show the one-degree-of-freedom test for equality of response profiles based on a contrast and the corresponding test based on analysis of covariance represent alternative tests of the same null hypothesis and that the test based on the analysis of covariance approach will always be more efficient. That is, the analysis of covariance approach yields estimates of treatment effects with smaller standard errors than those obtained by calculating contrasts.

For example, the greater efficiency of the analysis of covariance can be highlighted by examining the relative efficiency of (5.1) to (5.2) in simple settings. The relative efficiency is defined as the ratio of the variance of the estimator based on (5.2) to the variance of the estimator based on (5.1); a relative efficiency less than one implies that the estimator based on (5.2) has smaller variance. When the covariance among the repeated measures is assumed to have a compound symmetry pattern, with common variance σ^2 and common correlation ρ , the relative efficiency is given by

$$\frac{1}{n} \{1 + (n - 1)\rho\}. \quad (5.3)$$

The derivation of (5.3) is not important. What this simple expression indicates is that the two methods of adjustment for baseline response are equally efficient only when $\rho = 1$. When $\rho = 0$, the analysis based on (5.1) is only $\frac{1}{n}$ times as efficient as the analysis of covariance. The greater efficiency of the analysis of covariance depends on both the number of repeated measures and the strength of the correlation among them. For example, when $n = 5$ and $\rho = 0.4$, the analysis of covariance is approximately twice as efficient as subtracting the baseline response.

In general, the analysis of longitudinal data from a randomized trial is the only setting where we recommend adjustment for baseline through analysis of covariance. In that setting, in contrast to observational studies, adjustment leads to meaningful tests of hypothesis of scientific interest. Moreover the tests based on the analysis of covariance approach will be more powerful. The notion of adjustment for baseline can also be applied more generally in the analysis of response profiles; in Section 5.7 we compare and contrast a number of alternative strategies for handling the baseline response in more general settings and make recommendations about the preferred strategies in different situations.

We conclude this section by noting that adjustment for baseline in the analysis of longitudinal change is a topic that has generated heated debate among analysts. When longitudinal data arise from an observational study, the two methods of adjusting for baseline described in this section can yield discernibly different and, apparently conflicting, results. This conundrum is also known as *Lord's paradox* (named after Frederic Lord, who eloquently brought the issue to light) and has led many researchers astray over the years. The paradox lies in the interpretation of the two types of analyses and is resolved by noting that these two alternative methods of adjusting for baseline answer qualitatively different scientific questions when the data arise from an observational study. This can be illustrated in the simplest setting where there are two groups or sub-populations (e.g., males and females) measured at two occasions. The overall goal of such a study is to compare the changes in response for the two groups. The analysis that subtracts baseline response, thereby creating a simple change score, addresses the question of whether the two groups differ in terms of their mean change over time. In contrast, adjustment for baseline using analysis of covariance addresses the question of whether an individual belonging to one group is expected to change more (or less) than an individual belonging to the other group, *given that they have the same baseline response*. The latter question is a conditional one and, depending on the study design, may address a different scientific question than the former.

For example, in an observational study examining gender difference in weight gain of infants between the ages of 12 and 24 months, a measure of body weight might be obtained at 12 months (baseline) and at 24 months. The analysis of the simple change score addresses the question of whether boys and girls differ in terms of their changes in mean body weight over the 12 months of follow up. At baseline, boys are on average $1\frac{1}{2}$ pounds heavier than girls, but there is no evidence of a gender effect on the 12 month changes in body weight, with boys and girls both gaining approximately $5\frac{1}{4}$ pounds. In contrast, the analysis of covariance of the same data reveals a discernible gender effect, with boys showing more weight gain than girls. Thus, even though the unadjusted (or *unconditional*) increases in body weight are approximately the same for this age cohort of boys and girls, the analysis of covariance is directed at the *conditional* question of whether boys are expected to gain more weight than girls, *given that they have the same initial weight at 12 months*. That is, if we compare boys and girls within sub-populations with the same initial weight at 12 months, are their average weights at 24 months the same? When the conditional question is posed this way, we would expect boys to gain more weight than girls. The reasoning is that if a boy and girl have the same initial weight at 12 months, then there are two possibilities: (1) the girl is initially overweight and is expected to gain less weight over the 12 months, or (2) the boy is initially underweight and is expected to gain more.

A more thorough discussion of this issue is beyond the scope of this book, but we advise readers to employ the analysis of covariance approach in longitudinal settings only if the approach and its implications are fully understood.

In summary, the choice between the two methods of adjusting for baseline discussed in this section should be made on substantive grounds. That is, the design of the longitudinal study and the research question of interest should guide the choice of analytic method. The analysis that subtracts baseline response is appropriate when the primary goal of the study is to compare distinct populations in terms of their average change over time. The analysis addresses the question: Do the populations differ in terms of their average change? and this is appropriate when the data have arisen from either an observational study or a randomized trial. On the other hand, analysis of covariance will, in general, be appropriate in cases where individuals have been assigned to groups at random (e.g., a randomized trial) or where the population distributions of the baseline responses can reasonably be assumed to be equal (even though the sample means of the baseline responses may differ across groups). In cases where the population distributions of the baseline responses are equal, it is then meaningful to ask the question: Is the expected change the same in all groups, when we compare individuals having the same baseline response? Furthermore the analysis of covariance will provide a more powerful test of group differences. The latter has often been touted as the main reason why analysis of covariance should be the preferred method of adjusting for baseline. This faulty rationale, however, has blinded many researchers to the potential difficulties in interpreting the results of analysis of covariance when the assumption of equal population distributions of baseline response is not tenable. In conclusion, it is the study design and the scientific question of interest, and not issues of statistical precision and power, that should primarily determine the choice of analytic methods for adjusting for baseline response.

5.7 ALTERNATIVE METHODS OF ADJUSTING FOR BASELINE RESPONSE*

One feature of longitudinal studies that sets them apart from repeated measures and related designs is the presence of a baseline measurement. In a randomized longitudinal trial comparing treatments, the measurement at the first occasion is usually a baseline response obtained prior to any study interventions. For example, in the TLC trial, the blood lead levels at baseline were obtained prior to receiving placebo or succimer. In that case, due to randomization, we can assume that the treatment group means are equal at baseline. Thus the question naturally arises as to how to handle the baseline measurement in the assessment of whether patterns of change in the mean response over time are the same in the groups.

In the previous section we considered two methods of adjustment for baseline in a relatively simple setting. The notion of adjustment for baseline can also be applied more generally in the analysis of response profiles. In this section[†] we compare and contrast a number of alternative strategies for handling the baseline response and make recommendations about the preferred strategies in different situations.

We consider four ways of handling the baseline value:

1. We can retain it as part of the outcome vector and make no assumptions about group differences in the mean response at baseline.
2. We can retain it as part of the outcome vector and assume the group means are equal at baseline, as might be appropriate in a randomized trial.
3. We can subtract the baseline response from all of the remaining post-baseline responses, and analyze the differences from baseline.
4. We can use the baseline value as a covariate in the analysis of the post-baseline responses.

We now consider the appropriateness and merits of each of these four strategies and illustrate their application to the blood lead level data from the TLC trial.

The first two strategies retain the baseline measurement as part of the outcome vector, but differ in terms of assumptions about the mean response at baseline. The first strategy corresponds to a standard analysis of response profiles without incorporating any constraints on the group means at baseline. This was the method of analysis highlighted in Sections 5.2 through 5.4. The second strategy corresponds to an analysis of response profiles where the group means at baseline are constrained to be equal. In a randomized trial, where treatment assignment is random, both strategies yield valid estimates of treatment group comparisons, but the second strategy is, in general, more powerful. Thus, in randomized trials, or in observational studies

[†]Readers may find the level of detail in this section challenging; this section can be omitted at first reading without loss of continuity. However, the reader who returns to it will find that it yields important insights about baseline adjustment.

Table 5.7 Estimated regression coefficients and standard errors based on an analysis of response profiles of the blood lead level data assuming equal mean blood lead levels at baseline in the succimer and placebo groups.

Variable	Group	Week	Estimate	SE	Z
Intercept			26.406	0.500	52.83
Week		1	-1.645	0.782	-2.10
Week		4	-2.231	0.807	-2.76
Week		6	-2.642	0.887	-2.98
Group × Week	S	1	-11.341	1.093	-10.38
Group × Week	S	4	-8.765	1.131	-7.75
Group × Week	S	6	-3.120	1.251	-2.49

where there is good reason to assume that the groups have the same mean response at baseline (e.g., due to matching on baseline response), the second strategy for handling baseline is preferred and should be routinely used. In contrast, in observational studies where there is no a priori reason to assume the groups have the same mean response at baseline, the second strategy is not appropriate and only the first strategy should be used.

In the analyses of the blood lead level data from the TLC trial presented in Section 5.4, the first strategy was employed in the results presented in Tables 5.4 and 5.5. The second strategy can be implemented by excluding the treatment group main effect from the model for the response profiles. This model is unusual in that it contains an interaction between group and time but no main effect of group. This model appears to contradict the conventional wisdom that interactions should not be included in a regression model without their main effects. However, this is an important exception to the rule. Because baseline (week 0) was chosen as the reference level for time, the exclusion of the group main effect forces the two groups to have the same mean response at baseline. The results of such an analysis are presented in Table 5.7 and are qualitatively similar to those in Table 5.5. Note the absence of the main effect of group, which has been set to zero. Also the omnibus test of the group × time interaction from this model yields a Wald statistic of 111.96 with 3 degrees of freedom. In contrast, the analysis of response profiles without any adjustment for baseline (strategy 1) yielded a Wald statistic of 107.79, with 3 degrees of freedom (see Table 5.4). The difference between these two statistics reflects the increased power of the second method for handling baseline response.

The third and fourth strategies do not retain the baseline response as part of the outcome vector. Instead, they focus on raw and adjusted changes from baseline and restrict the outcome vector to measurements obtained post-baseline. The third strategy

is to subtract the baseline response from the remaining post-baseline responses, and analyze the differences from baseline. We refer to these differences from baseline as “raw change scores.” With responses at n occasions, we can define the $(n - 1) \times 1$ vector of raw change scores,

$$D_i = (Y_{i2} - Y_{i1}, Y_{i3} - Y_{i1}, \dots, Y_{in} - Y_{i1})'$$

and conduct an analysis of response profiles with D_i as the outcome vector. Because the outcome is a change score (the change from baseline), this approach alters the interpretation of the tests for all three effects in the analysis of response profiles. The test for group \times time interaction becomes a test for parallel profiles for the changes from baseline in the mean response on occasions 2 through n ; the test for group effect becomes a test that the changes from baseline at occasion 2 are the same across groups (assuming that occasion 2 is chosen as the reference level for time). Thus, to address the question of whether patterns of change over time are the same in all groups, the test of interest under the third strategy must be modified. It is now a joint test that combines the main effect of group and the group \times time interaction. The test has the same $(G - 1) \times (n - 1)$ degrees of freedom because it combines a $(G - 1)$ degrees of freedom test for group with a $(G - 1) \times (n - 2)$ degrees of freedom test for group \times time interaction. This is in contrast to the conventional analysis of response profiles (with baseline response included as part of the response vector) where only the group \times time interaction addresses important questions concerning group comparisons of the patterns of changes in the mean response and the group main effect is not of scientific interest.

Of note, this joint test of the main effect of group and the group \times time interaction is formally equivalent to the test of the group \times time interaction (with $(G - 1) \times (n - 1)$ degrees of freedom) under the first strategy. Moreover, for the purposes of analyzing changes in the mean response and how these changes differ among groups, the first and third strategies are completely equivalent; that is, the first and third strategies for handling baseline produce identical tests and estimates of effects. Thus it is clear that the third strategy offers no efficiency gain.

Using the blood lead level data from the TLC trial, the results of the analysis of change scores are presented in Table 5.8. At first glance the regression parameter estimates in Table 5.8 do not appear to agree with those in Table 5.5. However, it can easily be shown that the six parameter estimates in Table 5.8 are simple linear combinations of the estimates for the time and group \times time interaction effects reported in Table 5.5. For example, the group effect, -11.406 , in Table 5.8 is identical to the estimate of the group \times time interaction in Table 5.5 that represents the group contrast of the changes from baseline (week 0) to week 1. Similarly the estimated group contrast of the changes from baseline to week 4, -8.824 , from Table 5.5 can also be obtained from the estimates in Table 5.8 ($-11.406 + 2.582 = -8.824$). It can be shown that all estimates of change from baseline, and the group comparisons of these changes, reported in Tables 5.5 and 5.8 are identical.

The analysis of change scores reported in Table 5.8 produced a Wald statistic of 107.79, with 3 degrees of freedom, for jointly testing the effect of group and the group \times time interaction. As expected, this agrees with the omnibus test of the group \times

Table 5.8 Estimated regression coefficients and standard errors based on an analysis of response profiles of the changes from baseline in blood lead levels at week 1, week 4, and week 6 for the children from the TLC trial.

Variable	Group	Week	Estimate	SE	Z
Intercept			-1.612	0.792	-2.04
Group	S		-11.406	1.120	-10.18
Week		4	-0.590	0.643	-0.92
Week		6	-1.014	0.934	-1.09
Group × Week	S	4	2.582	0.909	2.84
Group × Week	S	6	8.254	1.321	6.25

time interaction reported in Table 5.4 from the analysis of response profiles without any adjustment for baseline (strategy 1).

Because the first and third strategies yield identical analyses of changes in the mean response, and how these changes differ among groups, in principle either method can be used. However, from a practical standpoint, we recommend the first strategy over the third for two main reasons. First, the analysis of change scores has implications for the interpretation of the hypothesis tests that are more consequential. For example, while the test of primary interest in the conventional analysis of response profiles is usually the test for group \times time interaction (with $(G - 1) \times (n - 1)$ degrees of freedom), the test of interest in the analysis of change score is a $(G - 1) \times (n - 1)$ degrees of freedom test that incorporates the main effect of group (with $(G - 1)$ degrees of freedom) and the $(G - 1) \times (n - 2)$ degrees of freedom group \times time interaction in this model. The analysis of change scores requires the construction of joint tests of main effects and interactions; these tests are not routinely produced as standard output from statistical software for analyzing response profiles. Second, when there are subjects with missing baseline response, all their data are excluded from the analysis of change scores; in contrast, the first strategy incorporates all available data in the analysis.

The fourth strategy is to analyze the post-baseline responses and make an adjustment for the baseline response by including it as a covariate. If we have responses at n occasions, the analysis of response profiles is based on the $(n - 1) \times 1$ vector,

$$Y_i = (Y_{i2}, Y_{i3}, \dots, Y_{in})',$$

and the baseline response, Y_{i1} , is regarded as a covariate. This type of analysis corresponds to an analysis of covariance (ANCOVA), albeit one where the outcome is a $(n - 1) \times 1$ vector of responses. This strategy for handling baseline is appropriate when analyzing data from randomized trials. Note that because of randomization, hypotheses of equality of the *conditional* means of the response at occasions 2 through

n , given the baseline response, imply hypotheses of equality of the *unconditional* means of the response at occasions 2 through n . This strategy for handling baseline is appropriate also for observational studies where there is good reason to assume the groups have the same mean response at baseline. It should not be used, however, in observational studies where there is no a priori reason to assume that the groups have the same mean response at baseline.

Interestingly, the fourth strategy for handling baseline can be implemented by conducting the analysis of response profiles (with Y_{i1} included as a covariate) on either the post-baseline responses or the post-baseline change scores. That is, the estimates of all effects of interest are identical whether the analysis is based on the $(n - 1) \times 1$ vector of post-baseline response, $Y_i = (Y_{i2}, Y_{i3}, \dots, Y_{in})'$, or the $(n - 1) \times 1$ vector of post-baseline differences, $D_i = (Y_{i2} - Y_{i1}, Y_{i3} - Y_{i1}, \dots, Y_{in} - Y_{i1})'$. The intuition for why these two analyses are identical is as follows: Because the two outcomes differ by Y_{i1} , and both analyses estimate effects that are adjusted for Y_{i1} by holding the baseline value fixed, they produce the same regression coefficients for all effects of interest. The two analyses differ only in terms of the estimated slope for Y_{i1} . (The estimated slope from the analysis based on Y_i is simply one unit larger than the estimated slope from the analysis based on D_i .) Because of this equivalence, we can regard the fourth strategy as an analysis of the “adjusted change scores” (i.e., D_i adjusted for Y_{i1}) in contrast to an analysis of the raw or unadjusted change scores (strategy 3).

Because the outcome is an adjusted change score, the fourth strategy for handling baseline also alters the interpretation of the tests for all three effects in the analysis of response profiles. The test for group \times time interaction becomes a test for parallel profiles for the adjusted changes from baseline in the mean response on occasions 2 through n ; the test for group effect becomes a test that the adjusted changes from baseline in the mean response at occasion 2 are the same across groups (assuming that occasion 2 is chosen as the reference level for time). Thus, similar to the third strategy, the test of interest is a joint test of the main effect of group and the group \times time interaction.

Using the blood lead level data from the TLC trial, the results of the analysis of adjusted change scores are presented in Table 5.9. Although the results of this analysis are qualitatively similar to the results from the analysis of raw change scores (strategy 3), note that the estimated main effect of group (and the intercept) in Table 5.9 is slightly different from the corresponding estimate in Table 5.8.

The analysis of adjusted change scores reported in Table 5.9 produced a Wald statistic of 111.13, with 3 degrees of freedom, for jointly testing the effect of group and the group \times time interaction. This is larger than the corresponding statistic under the third strategy, reflecting the greater efficiency of the analysis of adjusted change scores. The greater efficiency of analysis of covariance (adjusted change score analysis) over simple contrasts (raw change score analysis) was highlighted in Section 5.6.

Observe that the Wald statistic of 111.13 produced by the analysis of adjusted change scores is quite similar to that obtained under the second strategy for adjusting for baseline. Thus the question naturally arises as to which approach is preferred:

Table 5.9 Estimated regression coefficients and standard errors based on an analysis of response profiles of the adjusted changes from baseline in blood lead levels at week 1, week 4, and week 6 for the children from the TLC trial.

Variable	Group	Week	Estimate	SE	Z
Intercept			-1.638	0.777	-2.11
Baseline ^a ($Y_{i1} - 26.406$)			-0.196	0.094	-2.08
Group	S		-11.354	1.099	-10.34
Week		4	-0.590	0.643	-0.92
Week		6	-1.014	0.934	-1.09
Group \times Week	S	4	2.582	0.909	2.84
Group \times Week	S	6	8.254	1.321	6.25

^aCentering baseline response on its overall mean (26.406) gives the intercept a meaningful interpretation.

strategy 2 or strategy 4. One could argue that strategy 2 is preferred over strategy 4 for exactly the same reasons given for preferring strategy 1 over strategy 3. That is, the analysis of adjusted change scores requires the construction of joint tests of main effects and interactions, and these tests are not routinely produced as standard output from statistical software for analyzing response profiles. Second, when there are subjects with missing baseline response, all their data are excluded from the analysis of adjusted change scores; in contrast, the second strategy incorporates all available data in the analysis. Finally, there is a third reason why strategy 2 might be preferred over strategy 4. There is an implicit assumption in the adjusted change score analysis that the regression slope relating Y_{ij} to Y_{i1} (for $j = 2, \dots, n$) is the same at all $n - 1$ post-baseline occasions. This implies a strong assumption about the covariances among Y_{i1}, \dots, Y_{in} . Specifically, it constrains

$$\text{Cov}(Y_{i1}, Y_{i2}) = \text{Cov}(Y_{i1}, Y_{i3}) = \dots = \text{Cov}(Y_{i1}, Y_{in}).$$

As a result of these constraints, there is the potential for misspecification of the model for the covariance and misleading inferences about change over time. In contrast, the second strategy imposes no such structure on the covariance matrix. Finally, we note that if the assumption of homogeneous regression slopes (relating Y_{ij} to Y_{i1} , for $j = 2, \dots, n$) is relaxed and $n - 1$ separate regression slopes are estimated, then strategy 4 and strategy 2 are completely equivalent and produce identical tests and estimates of effects. Thus the second strategy for handling baseline can be seen to enjoy all of the efficiency gains that have been highlighted in Section 5.6 for ANCOVA (adjusted change score analysis). However, on practical grounds, for the reasons outlined above, it can be argued that strategy 2 is preferred over strategy 4.

To summarize, there are many ways to handle baseline response in the analysis of longitudinal data. In this section we have reviewed four strategies. We have seen that the two methods that retain the baseline value as part of the outcome are completely equivalent to corresponding strategies that restrict the outcome vector to measurements obtained post-baseline. However, on practical grounds, it can be argued that the first and second strategies are preferable. The first and second strategies differ in terms of efficiency. The second strategy enjoys all the efficiency gains that have been highlighted in Section 5.6 for ANCOVA. But the choice between the first and second strategies should be guided by the study design. In randomized trials, or in observational studies where there is good reason to assume the groups have the same mean response at baseline, the second strategy is, in general, more powerful and should be routinely used. In contrast, in observational studies where there is no a priori reason to assume that the groups have the same mean response at baseline, the second strategy is not appropriate and the first should be used.

5.8 STRENGTHS AND WEAKNESSES OF ANALYZING RESPONSE PROFILES

The analysis of response profiles is a conceptually straightforward way to analyze data from a longitudinal study when the design is balanced, with the timing of the repeated measures common to all individuals in the study, and when all the covariates are discrete (e.g., representing different treatments, interventions, or characteristics of the study subjects). The main feature of the analysis of response profiles is that it allows arbitrary patterns in the mean response over time and arbitrary patterns in the covariance of the responses. As a result this method for longitudinal analysis has a certain robustness, since the potential risks of bias due to misspecification of the models for the mean and covariance are minimal. Although the analysis of response profiles requires that the data arise from a balanced design, it can be applied when the data are incomplete due to missing response data.

The method for analyzing response profiles described in this chapter is related to a more traditional approach known in the statistical literature as “profile analysis”. However, we make a distinction between the method presented in this chapter and a traditional profile analysis. In a traditional profile analysis the three hypotheses concerning response profiles described at the beginning of Section 5.2 are placed on an equal footing, and there is an overwhelming emphasis on hypothesis testing rather than estimation of effects. As we have seen, however, tests of hypotheses concerning main effects of time and/or group often have no direct bearing on questions of scientific interest in a longitudinal study. This is especially the case for longitudinal data arising from randomized trials. Consequently the routine use of traditional profile analysis for longitudinal data coerces the analyst to test certain hypotheses that do not necessarily translate into meaningful scientific questions about longitudinal change in the response. In addition, because profile analysis is often implemented within a multivariate analysis of variance (MANOVA) that requires a complete response

vector on each individual, it does not permit subjects with missing responses. The resulting analysis is very inefficient because it is based only on data from the so-called complete cases; it can also produce biased estimates of change in the mean response over time when such “completers” are not a random sample from the target population. Finally, traditional profile analysis lacks flexibility in handling the baseline response and requires that it be part of the response vector. In contrast, the method for analyzing response profiles presented in this chapter can be readily adapted to address specific questions that are well grounded in the science, can be applied when the data are incomplete due to missing response data, and permit alternative approaches for making adjustments for the baseline response.

Although it was not considered here, the analysis of response profiles can be extended in a straightforward way to handle the case where individuals can be grouped according to more than a single factor. For example, if there are two covariates that are discrete (e.g., treatment group and gender), the analysis will include tests of the 3-way and 2-way interactions among these two factors and time (in addition to their main effects). The general linear model can also be used to provide estimated means for summary measure analyses that are based on linear combinations of the mean response vector, for instance, “area under the curve” analysis, when the data are incomplete.

The analysis of response profiles does have a number of potential drawbacks that make it either unappealing or unsuitable for analyzing data from many longitudinal studies. First, the requirement that the longitudinal design be balanced implies that the method cannot be applied when the vectors of repeated measures are obtained at different sequences of time, except by “moving” an observation to the nearest planned measurement time. As a result the method is not well suited to handle mistimed measurements, a common problem in many longitudinal studies. Note, however, that the general method for analyzing response profiles can handle unbalanced patterns of observations due to missing response data. Second, the analysis of response profiles ignores the time ordering of the repeated measures in a longitudinal study. Indeed, the analysis of response profiles could be applied when each individual has a vector of multivariate outcomes that are distinct and non-commensurate (i.e., measures of more than one outcome) rather than repeated measures of a single outcome. Because the analysis of longitudinal response profiles allows for an arbitrary pattern in the mean responses, and does not impose any time trends, the results of the analysis provide only a very broad or general statement about group differences in patterns of change over time. Ordinarily a significant group \times time interaction effect that has more than a single degree of freedom will require additional analysis to provide a more informative description of how the groups differ in their patterns of change in the mean response. Third, because the analysis of response profiles produces an overall or omnibus test of effects, it may have low power to detect group differences in specific trends in the mean response over time (e.g., linear trends in the mean response). Single-degree-of-freedom tests of specific time trends are more powerful. Finally, in the analysis of response profiles, the number of estimated parameters ($G \times n$ mean parameters and $\frac{n(n+1)}{2}$ covariance parameters) grows rapidly with the number of measurement occasions. For example, with two groups measured at three

occasions, the number of parameters is 12. However, with two groups measured at 10 occasions, the number of parameters is 75. Consequently this method is more appealing when the total number of subjects, N , is relatively large in comparison to the number of measurement occasions, n .

5.9 COMPUTING: ANALYZING RESPONSE PROFILES USING PROC MIXED IN SAS

The MIXED procedure in SAS is a very general and versatile procedure for fitting linear models to longitudinal and clustered data. No attempt is made here to give a comprehensive review of the main features of PROC MIXED. Instead, we present illustrative source code for an analysis of response profiles in general terms and then describe the most salient parts of the command syntax. Many of the later chapters will include a description of additional commands and features of PROC MIXED as they are needed. Although these concluding sections in each chapter will not provide a training manual for the use of PROC MIXED, they should provide a firm basis for understanding the command syntax required for analyzing longitudinal data using PROC MIXED in SAS.

Before discussing the command syntax for PROC MIXED, we note that the procedure requires each repeated measurement in a longitudinal data set to be a separate “record.” For example, in the TLC trial, the data are recorded as follows:

ID	Group	Baseline	Week 1	Week 4	Week 6
001	P	30.8	26.9	25.8	23.8
002	S	26.5	14.8	19.5	21.0
003	S	25.8	23.0	19.1	23.2
004	P	24.7	24.5	22.0	22.5
005	S	20.4	2.8	3.2	9.4
006	S	20.4	5.4	4.5	11.9
:	:	:	:	:	:
100	P	31.1	31.2	29.2	30.1

with a single “record” of the four repeated measurements for each child in the study. When the data set is in this form, it is said to be in a *multivariate* mode or *wide* format. Prior to analysis these data must be converted to a data set with four records for each child, one for each measurement occasion. In the latter form the data set is said to be

Table 5.10 Illustrative commands in SAS for transforming a data set with a single record for each individual to a data set with multiple records corresponding to each measurement occasion.

DATA lead;

```
INFILE 'tlc.dat';
INPUT id group $ y1 y2 y3 y4;
y=y1; time=0; OUTPUT;
y=y2; time=1; OUTPUT;
y=y3; time=4; OUTPUT;
y=y4; time=6; OUTPUT;
DROP y1-y4;
```

in a *univariate* mode or *long* format. This can be accomplished using the illustrative SAS commands in Table 5.10, which produce the following data set:

ID	Group	Time	Y
001	P	0	30.8
001	P	1	26.9
001	P	4	25.8
001	P	6	23.8
002	S	0	26.5
002	S	1	14.8
002	S	4	19.5
002	S	6	21.0
⋮	⋮	⋮	⋮
100	P	0	31.1
100	P	1	31.2
100	P	4	29.2
100	P	6	30.1

To conduct an analysis of response profiles with data from two or more treatment groups measured repeatedly over time, we can use the illustrative SAS commands

Table 5.11 Illustrative commands for an analysis of response profiles using PROC MIXED in SAS.

```
PROC MIXED;
  CLASS id group time;
  MODEL y=group time group*time / S CHISQ;
  REPEATED time / TYPE=UN SUBJECT=id R RCORR;
```

given in Table 5.11. This model assumes that the covariance matrix is unstructured. Alternative assumption about the covariance can be considered, and this may be advantageous when the number of measurement occasions is relatively large in comparison to the number of subjects. Choosing a model for the covariance matrix is a topic that will be discussed in Chapter 7. Next we present a brief description of each of the command statements in Table 5.11.

PROC MIXED <options>;

The PROC MIXED statement calls the procedure MIXED in SAS. It can also include an option for the choice of method of estimation. By default, PROC MIXED uses REML estimation; ML estimation can be invoked by including the option METHOD=ML.

CLASS variables;

The CLASS statement is used to define all variables that are to be regarded as categorical or factors. By default, this statement will create indicator variables for each factor using a reference group coding, with the last level (where “last” here refers to the level with the largest alpha-numeric value) regarded as the reference group. Of note, this default indicator variable coding is not the most natural for a categorical variable denoting the occasions of measurement; for the latter, the “first” level of the factor (e.g., the baseline measurement occasions) is usually the natural reference group.

The default coding can be changed with the inclusion of the ORDER= option in the PROC MIXED statement. For example, the ORDER=DATA option forces the levels of all variables included in the CLASS statement to be sorted by their order of appearance in the input data set. Therefore, by previously sorting the data set in descending order of a categorical variable denoting the occasions of measurement, a more natural reference group coding is obtained for that variable (with the lowest, rather than the highest, level of the categorical variable for time used as the reference). However, one unappealing consequence of circumventing the default coding of time in this way is that the estimates of the covariance matrix are printed in reverse (or descending) order of time. To avoid

potential confusion when extracting the estimates of the covariance matrix, it is advisable to re-run the analysis without the ORDER=DATA option.

MODEL dependent = <fixed effects> / <options>;

The MODEL statement specifies the response variable and the fixed effects. The fixed effects can include both discrete (defined in the CLASS statement) and quantitative (excluded from the CLASS statement) covariates.

The covariates included in the MODEL statement determine the design matrix X_i . Of note, by default, PROC MIXED includes a column of 1's in X_i for the intercept. The option NOINT requests that no intercept be included in the model.

Various options that can be included on the MODEL statement modify how test statistics are computed and the type of output produced. The option DDFM=SATTERTH requests Satterthwaite's approximation for the denominator degrees of freedom for tests of the fixed effects. Alternatively, the option CHISQ requests that multivariate Wald tests be computed and compared to the reference chi-squared distribution. The option S (or SOLUTION) requests that the estimates of the fixed effects, and their standard errors, be displayed.

REPEATED <repeated effect> / SUBJECT = effect <options>;

The REPEATED statement is primarily used to distinguish which observations are correlated and which can be regarded as independent of one another. This is achieved with the SUBJECT option, which is used to denote a variable that distinguishes clusters of correlated responses. By including a variable in the SUBJECT option (e.g., a subject identifier), pairs of observations with the same value of that variable are regarded as correlated while pairs of observations with distinct values are regarded as independent.

The REPEATED statement also includes options for specifying assumptions about the nature of the covariance among the errors. This is achieved with the TYPE=<pattern> option (e.g., TYPE=UN specifies an unstructured covariance matrix). A full listing and description of all the possible covariance patterns can be found in the SAS documentation. There are also various options that modify the type of output that is produced. The option R and RCORR print the covariance and correlation matrices, respectively.

Finally, a variable denoted the "repeated effect" can also be included on the REPEATED statement, and this identifies "units within a cluster." In the context of longitudinal data, the "repeated effect" identifies the measurement occasions. While it is not always necessary to include this variable, failure to do so may have unforeseen consequences when there are vectors of repeated measures of different length and/or when the vector of responses are not in the same order for all subjects. In Table 5.11 the REPEATED statement identifies "time" as the repeated effect. To avoid any potential problems, it is recommended that this variable be included in the REPEATED statement to ensure that the covariance is structured and estimated appropriately.

5.10 FURTHER READING

A useful review of traditional profile analysis, targeted at applied researchers, can be found in Chapter 3 (Section 3.4, pp. 48–52) of Hand and Taylor (1987).

A more detailed discussion of the subtle issues surrounding adjustment for baseline response in the analysis of change can be found in the articles by Lord (1967), Laird (1983), Fitzmaurice (2001), and Glymour et al. (2005), and in Chapter 7 (Section 7.3, pp. 489–496) of Bock (1975).

Bibliographic Notes

One of the earliest descriptions of traditional profile analysis appeared in an article by Greenhouse and Geisser (1959). Profile analysis is also discussed in detail in Chapter 6 of Johnson and Wichern (2002), Chapters 4 and 5 of Morrison (1990), and Chapters 5 and 6 of Rencher (2002).

Problems

5.1 In the National Cooperative Gallstone Study (NCGS), one of the major interests was to study the safety of the drug chenodiol for the treatment of cholesterol gallstones (Schoenfeld et al., 1981; Wei and Lachin, 1984). In this study, patients were randomly assigned to high-dose (750 mg per day), low-dose (375 mg per day), or placebo. We focus on a subset of data on patients who had floating gallstones and who were assigned to the high-dose and placebo groups.

In the NCGS it was suggested that chenodiol would dissolve gallstones but, in doing so, might increase levels of serum cholesterol. As a result serum cholesterol (mg/dL) was measured at baseline and at 6, 12, 20, and 24 months of follow-up. Many cholesterol measurements are missing because of missed visits, laboratory specimens were lost or inadequate, or patient follow-up was terminated.

The NCGS serum cholesterol data are stored in an external file: `cholesterol.dat`

Each row of the data set contains the following seven variables:

Group ID Y₁ Y₂ Y₃ Y₄ Y₅

Note: The categorical variable Group is coded 1 = High-Dose, 2 = Placebo.

5.1.1 Read the data from the external file and keep it in a “multivariate” or “wide” format.

5.1.2 Calculate the sample means, standard deviations, and variances of the serum cholesterol levels at each occasion for each treatment group.

- 5.1.3** On a single graph, construct a time plot that displays the mean serum cholesterol versus time (in months) for the two treatment group. Describe the general characteristics of the time trends for the two groups.
- 5.1.4** Next read the data from the external file and put the data in a “univariate” or “long” format, with five “records” per subject.
- 5.1.5** Assuming an unstructured covariance matrix, conduct an analysis of response profiles. Determine whether the patterns of change over time differ in the two treatment groups.
- 5.1.6** Display the estimated 5×5 covariance and correlation matrices for the five repeated measurements of serum cholesterol.
- 5.1.7** With baseline (month 0) and the placebo group (group 2) as the *reference group*, write out the regression model for mean serum cholesterol that corresponds to the analysis of response profiles in Problem 5.1.5.
- 5.1.8** Let L denote a matrix of known weights and β the vector of linear regression parameters from the model assumed in Problem 5.1.7. The null hypothesis that the patterns of change over time do not differ in the two treatment groups can be expressed as $H_0: L\beta = 0$. Describe an appropriate weight matrix L for this null hypothesis.
- 5.1.9** Show how the *estimated* regression coefficients from an analysis of response profiles can be used to construct the time-specific means in the two groups. Compare these estimated means with the sample means obtained in Problem 5.1.2.
- 5.1.10** With baseline (month 0) and the placebo group (group 2) as the *reference group*, provide an interpretation for each of the estimated regression coefficients in terms of the effect of the treatments on the patterns of change in mean serum cholesterol.