Part I:

Introduction to Longitudinal and Clustered Data
Longitudinal and Clustered Data

Research on statistical methods for the design and analysis of human investigations expanded exponentially in the second half of the twentieth century. Beginning in the early 1950s, the U.S. government channeled substantial financial research support from military to biomedical research. The legislative foundation for the modern National Institutes of Health (NIH), the Public Health Service Act, was passed in 1944 and NIH grew rapidly throughout the 1950s and 1960s. During these “golden years” of NIH expansion, NIH's biomedical grant program grew from $2 million in 1949 to more than $1 billion in 1966. This NIH investment in the biomedical epidemiological studies and clinical trials of that period, including the influential Framingham Heart Study (Chamberlain et al., 1951; Merskey, 1961).

The typical focus of these early studies was morbidity and, especially, mortality. Investigators sought to identify the causes of early death and to evaluate the effectiveness of treatments for reducing mortality. In the Framingham Heart Study, participants were seen at two-year intervals. Survival estimates during successive two-year periods were based on independent events and standard using multiplicative logistic regression. The success of this novel analytical approach in this study, and the recognition that it could be applied to the complex data, led to widespread use of this methodology beginning in the 1960s. The analysis of these two-year data was revolutionized by the seminal 1970 paper of D.R. Cox, describing the proportional hazards model (Cox, 1972). This paper was followed by a rich and important body of work that established the conceptual basis and the computational tools for modern survival analysis.
Although the design of the Honolulu Heart Study and other cohort studies called for periodic measurement of the patient characteristics thought to be determinants of chronic disease, interest in the levels and patterns of changes of these characteristics over time was initially limited. As the research advanced, however, investigations began to ask questions about the behavior of these risk factors. In the Honolulu Heart Study, for example, investigators began to ask whether blood pressure levels in childhood were predictive of hypertension in adult life. In the Coronary Artery Risk Development in Young Adults (CARDIA) Study, investigators sought to identify the determinants of the transition from normotension to normohypertension in early adulthood; in hypotension, and hypertension, and in middle age (Klatsky et al., 1988). In the treatment of asthma, anemia, and other diseases that are not typically life threatening, investigators began to study the effects of treatments on the levels and changes in measures of severity of disease. Similar questions were being posed in many disease settings. Investigators began to follow populations of all ages over time, both in observational studies and clinical trials, to understand the development and persistence of diseases and to identify factors that affect the course of disease development.

This interest in the temporal patterns of change in human characteristics came at a particular useful moment in computing power made ever more apparent and useful: computational methods of analyzing longitudinal data were being developed. Thus, in the early 1970s, Linus Pauling pioneered the use of the Kalman algorithm for fitting linear mixed-effects models appropriate for the analysis of repeated measurements (Pauling and White, 1982). Meanwhile and independently, in the late 1970s, statisticians developed the methodological framework for the analysis of longitudinal data. The past 30 years have seen substantial progress in the development of statistical methods for the analysis of longitudinal data. Despite these important advances, nonetheless, the analysis of longitudinal data has remained a major challenge to researchers, as the analysis of longitudinal data has been overshadowed by issues involving the study of the association between the individual disease and the trajectory of change over time.
prioritised study design. For example, in a case-to-control study, whether the exposure is measured as a single measurement, using only study subjects matched on a single variable of interest, or whether it is measured as a number of measurements taken at different times, or even different individuals, the differences in the responses. That is, a case-control study may allow the comparison of the disease risk between individuals exposed to a certain factor and those not exposed, but it does not provide any information about the differences among individuals exposed to the corresponding factor.

To highlight that case-control differences between case-control and longitudinal follow-up designs are, in many cases, less precise than those that are observed in cross-sectional comparisons. However, in longitudinal follow-up designs, different comparisons of the average responses are made in the study groups of individuals with the same characteristics (measured) at follow-up. While comparisons of differences in outcomes in cross-sectional studies for longitudinal associations in groups of subjects from 10 to 30 years old. This effect of growth on appetite can be considered as a longitudinal study that shows no similarity between outcomes of the same individuals changing with time. In a cross-sectional study, the effect of aging is presented with cross-sectional methods with the same characteristics. That is, the difference in way, age, and many characteristics that individual factors in these same different age groups that can affect the relationship between age and body fatness. On the other hand, a longitudinal study that measures a single cohort of individuals at different ages 10 and 15 years can provide a valid estimate of the changes in body fatness at different ages. It is noteworthy that the associations in longitudinal studies in general are not determined on the basis of the control of the study and the control of differences in body fatness.

A longitudinal study in a longitudinal study in that they are measured. The longitudinal analyses the changes associated with the repeated measurements obtained from a single individual at different occasions. Comparisons within a subject will typically be subject to random measurement error. Comparisons between subjects can be measured for these associations. Longitudinal studies also have a temporal aspect; that is, the repeated measurements provide a clearer measure of the extent to which the variables change. The within-individual comparison provides a valid estimate of the changes in body fatness at different ages between 10 and 15 years. Furthermore, subjects in which the body mass index at different ages, changes in the presence of obesity but throughout the duration of the study and the control of differences in body fatness.

A longitudinal study with the same characteristics in that they are measured. The longitudinal analyses the changes associated with the repeated measurements obtained from a single individual at different occasions. Comparisons within a subject will typically be subject to random measurement error. Comparisons between subjects can be measured for these associations. Longitudinal studies also have a temporal aspect; that is, the repeated measurements provide a clearer measure of the extent to which the variables change. The within-individual comparison provides a valid estimate of the changes in body fatness at different ages between 10 and 15 years. Furthermore, subjects in which the body mass index at different ages, changes in the presence of obesity but throughout the duration of the study and the control of differences in body fatness.
pounds. Therefore, household-level, country-level, and national practices, all substantial and all instances of understated data differ in the amount of time needed for the replication that might be the prevalence mapping units in a study. Finally, household-level and country-level data are the baseline examination of household and country-level data. Each study multiple regression of these different measurement base cases.

In all these cumulative household data, we might reasonably expect that measurements on units within a household and units within the same country are data in different countries. When elements of longitudinal data are compared or because of the country-level data from units within the same country. This conclusion is based on the empirical evidence of the longitudinal and cross-sectional analyses. However, statistical techniques that longitudinal data would explicitly be addressed are not accounted for in these conclusions. Because longitudinal data are a special case of cross-sectional data for which a regular rotation of longitudinal data are used, both trends to provide a description of longitudinal and cross-sectional data, whereas longitudinal data are used to provide a description of cross-sectional data. Therefore, one of the goals of this branch in the cross-sectional and longitudinal data for the analysis of longitudinal data may have been to develop a special regression models that longitudinal data. As a result, a comprehensive understanding of methods for the analysis of longitudinal data is provided. This leads to a thorough understanding of methods for analyzing these special types of longitudinal data, which commonly occurs in studies involving longitudinal and cross-sectional data.

The longitudinal data should be considered only as samples from the longitudinal, longitudinal, cross-sectional, and longitudinal data are similar. A longitudinal data set is a description of the longitudinal data, whereas longitudinal data are used to describe the longitudinal data. Therefore, the current longitudinal data are not the same as the longitudinal data. For example, the data may consist of longitudinal and cross-sectional data, which are described by ethics. Alternatively, the data may consist of longitudinal and cross-sectional data, which are described by ethics. Although the analysis of longitudinal data is not the primary focus of this book, longitudinal data are discussed in Chapter 22.

Interest in the analysis of longitudinal and cross-sectional data continues to grow. New and more flexible models have been developed and inferences in computation, such as Markov dependence (MCMC) models, have allowed greater flexibility in modern regression analysis. Moreover, improvements in statistical software packages, especially R, Stata, SPSS, and S-Plus, have made these models much more accessible for use in routine data analysis. Despite these advances, however, methods for the analysis of longitudinal data are not strictly used and are seen to be accessible to many nonstatisticians with some statistical experience.

We believe that this methodology for the analysis of longitudinal data can be much more readily understood and applied. It in many cases that this branch will help readers to that possibilities. It provides a comprehensive introduction to nonalgebraic methods for analysis of longitudinal data, written for readers with a basic knowledge of mathematics and a strong background in regression analysis techniques. The book does not require a high level of mathematical preparation and does assume a willingness to read and understand mathematical formulas.
1.23  EXAMPLES

It is difficult to determine the number of times that children have been exposed to chemotherapy in clinical settings. A recent study found that children exposed to high levels of chemotherapy had significantly lower IQ scores compared to children who had received lower doses. This finding suggests the need for further research on the long-term effects of chemotherapy on children's cognitive development.

1.23.1  Treatment of Leukemia-Exposed Children (UCE) Trial

The UCE trial was conducted to evaluate the effectiveness of a new chemotherapy regimen in treating childhood leukemia. The results showed that the new regimen was more effective than the standard treatment in terms of reducing the risk of relapse and improving survival rates. The trial also highlighted the importance of early detection and intervention in improving outcomes for children with leukemia.
Table 4.8: Salivary lead levels (geometric mean) at baseline, week 1, week 4, and week 6 for 10 randomly selected children from the 

<table>
<thead>
<tr>
<th>Group</th>
<th>Baseline</th>
<th>Week 1</th>
<th>Week 4</th>
<th>Week 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>30.8</td>
<td>23.9</td>
<td>25.3</td>
<td>23.3</td>
</tr>
<tr>
<td>3</td>
<td>26.5</td>
<td>14.2</td>
<td>19.5</td>
<td>21.0</td>
</tr>
<tr>
<td>4</td>
<td>35.3</td>
<td>7.0</td>
<td>12.1</td>
<td>23.2</td>
</tr>
<tr>
<td>11</td>
<td>24.7</td>
<td>24.5</td>
<td>22.0</td>
<td>23.5</td>
</tr>
<tr>
<td>60</td>
<td>20.4</td>
<td>8.8</td>
<td>3.2</td>
<td>9.4</td>
</tr>
<tr>
<td>29</td>
<td>21.4</td>
<td>7.4</td>
<td>4.5</td>
<td>11.9</td>
</tr>
<tr>
<td>46</td>
<td>25.6</td>
<td>20.8</td>
<td>10.2</td>
<td>18.4</td>
</tr>
<tr>
<td>43</td>
<td>31.7</td>
<td>14.6</td>
<td>28.5</td>
<td>23.1</td>
</tr>
<tr>
<td>34</td>
<td>19.7</td>
<td>14.9</td>
<td>15.3</td>
<td>14.7</td>
</tr>
<tr>
<td>78</td>
<td>31.1</td>
<td>31.3</td>
<td>29.2</td>
<td>30.1</td>
</tr>
</tbody>
</table>

* 2: placebo; 3: micronized.

Table 5.2: Hemoglobin levels (mean standard deviation) at baseline, week 1, week 4, and week 6 for children from the MCT trial.

<table>
<thead>
<tr>
<th>Group</th>
<th>Baseline</th>
<th>Week 1</th>
<th>Week 4</th>
<th>Week 6</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Hemoglobin levels at baseline, week 1, week 4, and week 6 for children from the MCT trial.

<table>
<thead>
<tr>
<th>Group</th>
<th>Baseline</th>
<th>Week 1</th>
<th>Week 4</th>
<th>Week 6</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Hemoglobin levels at baseline, week 1, week 4, and week 6 for children from the MCT trial.

**Note:** Presented for the first time in the study. The table shows the mean and standard deviation of hemoglobin levels at various time points for different groups.
1.3.2 Measuring Coronary Risk Factor Study

In 1998 the American Heart Association (AHA) announced that obesity had been added to its AHA's list of major preventable risk factors for coronary heart disease. These major preventable risk factors include smoking, high blood cholesterol, high blood pressure, and sedentary lifestyle. Unlike risk factors that cannot be altered, such as heredity, increasing age, and being male, obesity is a risk factor that many individuals can alter and control. The medical definition of obesity is quite simple: an excess of body fat. Obesity is primarily caused by consuming too many calories and not getting enough physical exercise. Obesity can lead to higher blood cholesterol and triglyceride levels, lower HDL cholesterol, high levels of triglycerides, the "bad" cholesterol, and higher blood pressure. Thus, obesity can contribute to higher coronary risk in a variety of different ways.

Public health scientists now accept that obesity is a chronic disease, just like high blood pressure or high blood cholesterol. Its causes are a complex, individualized combination of genetics, behavior, and lifestyle. There is also increased awareness that obese children are at increased risk for obesity as adults.
<table>
<thead>
<tr>
<th>Diagnosis</th>
<th>1987</th>
<th>1979</th>
<th>1981</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rhizomelic</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>20</td>
</tr>
<tr>
<td>Refusals</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>Refusals</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>9</td>
</tr>
<tr>
<td>Refusals</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td>Refusals</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td>Refusals</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>13</td>
</tr>
<tr>
<td>Refusals</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Refusals</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Refusals</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>152</td>
</tr>
<tr>
<td>Refusals</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>Refusals</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Refusals</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>Refusals</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>16</td>
</tr>
<tr>
<td>Refusals</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>Refusals</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>32</td>
</tr>
<tr>
<td>Refusals</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>Refusals</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>32</td>
</tr>
<tr>
<td>Refusals</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>Refusals</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>Refusals</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>48</td>
</tr>
</tbody>
</table>

**Notes:**

- 1 = Diagnosis
- 0 = Refusal
- = Indeterminate
In 1972, researchers from the University of Nevada began to examine the links between obesity and other community factors. Of particular interest were the associations between certain environmental factors and obesity in children. The researchers focused on obesity in children aged 5 to 13 years. A study was conducted to examine the development and prevalence of obesity in children. The study included 518 children of normal weight and 498 children of overweight weight. The study aimed to determine whether the prevalence of obesity increased with age and whether patterns of change in obesity were the same for boys and girls.

A summary of the obesity data for children in one of the study locations, aged 5 to 13 years, in 1977, is presented in Table 1.3. Because of the variability in obesity across the study areas, the data were analyzed using chi-square tests for each age group. For example, the test results are shown in Table 1.3. The chi-square test statistic for the age groups is 8 (out of 24) for 2 age groups. The prevalence of obesity increases with age, and the test statistic is significant. A similar analysis was conducted for each of the remaining age groups of children. The data are summarized in Table 1.3. The prevalence of obesity is related to the third category of the obesity status variable.

1.3.3 Clinical Trials of a New Anticonvulsant Therapy

Anticonvulsant therapy is a pharmacological treatment that may reduce the risk of brain injury, developmental delay in children with epilepsy. Anticonvulsant therapy is important for children with epilepsy. The anticonvulsant drugs used in this trial were divided into two groups: placebo and anticonvulsant. The first group received a placebo, and the second group received the anticonvulsant. In the second group, children were randomized to receive either a new anticonvulsant or placebo. The results of the study are presented in Table 1.4. The anticonvulsant therapy was more effective in reducing the number of seizures than placebo. CANADA is the primary institution involved in the trial.
Week 1-6: Mean score of adherence per week (mean ± standard deviation) at baseline, week 2, week 4, week 6, and week 8 in the pegylated and placebo groups.

<table>
<thead>
<tr>
<th>Group</th>
<th>Baseline</th>
<th>Week 2</th>
<th>Week 4</th>
<th>Week 6</th>
<th>Week 8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pegylated</td>
<td>3.5±1</td>
<td>4.2±1</td>
<td>4.2±1</td>
<td>4.0±1</td>
<td>3.3±1</td>
</tr>
<tr>
<td>Placebo</td>
<td>3.2±1</td>
<td>4.5±1</td>
<td>4.1±1</td>
<td>4.3±1</td>
<td>4.0±1</td>
</tr>
</tbody>
</table>

*Fig. 1.1: Difference score of performance (over weeks) and baseline, week 2, week 4, week 6, and week 8 in the pegylated and placebo groups.*
1.3.4) Connecticut Child-Teen Surveys:

These new remarkable findings, which involves all psychiatric disorders in childhood and adolescence, with a particular focus on psychopathology, are drawn from several psychiatric epidemiological studies. Children's psychiatric status has been found to be related to their early development and environment. This is thought to occur in the development of the child's psychopathology.

In addition, the present study surveyed multiple environmental factors on the child's psychopathology in order to explore the relationship between the child's mental health status and multiple environmental factors. These factors include the child's demographic characteristics, such as age, sex, and socioeconomic status, and the child's health and educational background.

Data from these surveys have shown a significant association between environmental factors and the child's mental health status. In particular, the child's sociodemographic characteristics, such as age, sex, and socioeconomic status, are found to be related to the child's mental health status. Additionally, the child's health and educational background are found to be related to the child's mental health status.
The four examples considered in this section differ in terms of outcome variables, study design, and goals on objectives of the analyses. In the first example from the NHNP trial, the outcome variable, length head level, is continuous. In the second example from the NHNP study, the outcome variables, obesity status, in infancy. In the third example from the clinical trial of pre-school, the outcome variable is change. These three examples illustrate the diverse types of longitudinal data that arise in the health and medical sciences. A notable feature of the second example is the amount of missing data. Missing data are a common problem in longitudinal studies in the health sciences. As we will discuss in later chapters, one will need to examine the reasons for any missingness to determine the validity of inferences about changes in the response over time. Next, consider the design of these studies. The first and third examples are experiments, whereas the treatments have been chosen by the investigators and randomly assigned to the study participants. The second example in an observational study where the study participants are followed between...
in time to observe the outcome variable at future time points; however, unlike the
same longitudinal cohort study, the investigator cannot directly control the comparability of
participants (home, schools, and baseline). While these factors may confound the group
inferences, the final outcomes in the case-control observational study. In the
Commensalism Child Survey, variables are measured at a single time point in a sample
of children. Because information on the outcome variable of interest is obtained from
two sources (the parent and teacher), these data are also observational. Similarly, we
must test the general effect of all factors to see whether the final three examples characterize
the changes in the outcome variables over time; small changes that influence outcomes.
In the final example, however, the objective of the analysis is to understand changes in the
continuous variable in question. Instead, the goal is to examine the effects of subject-
specific covariates on the outcomes. In later chapters, we describe model
methods for analyzing diverse types of longitudinal data, starting from both experimental
and observational studies. Because longitudinal data are a special case of clustered data,
we also describe methods of analysis for clustered data, namely cluster-variant.

4.1 READING RESIDUALS FROM CONTROLLED ANALYSES

In the last 30 years we have seen remarkable advances in methods for analyzing
longitudinal and clustered data. In particular, the use of fixed and flexible classes of models for correlated data based on a regression paradigm. Indeed, all the
methods that we described in later chapters can be thought of as regression models
for correlated responses. In this section, we provide conditions for the regression
paradigm for correlated responses.

Regression models are widely used and provide a very general and versatile
approach for analyzing data. One way of thinking about regression models is to
think of the dependent variable as being a function of a linear predictor, which
is a function of the independent variables. In this section, we describe a
number of different types of regression models and discuss how they can be
applied to different types of data. We begin by describing the general form
of a regression model. In this chapter, we consider several examples of
regression models and discuss how they can be used to analyze different types
of data. In later chapters, we consider other types of data and regression
models for those data. In this chapter, we consider the general form
of a regression model.
\[
M(Y | X) : \beta_1 + \beta_2 X,
\]

\[
N(Y | X) : \beta_1 + \beta_2 \log(X),
\]

and

\[
M(Y | X) : \beta_1 + \beta_2 X + \beta_3 X^2,
\]

are all cases where the mean is linear in the regression parameters (where \(M(Y | X)\) denotes the conditional mean or expectations of \(Y\) given \(X\)). All these models are linear in the regression parameters, even if the latter term are non-linear in the covariate. In this latter case we usually consider models where the covariate, or some suitable transformations of the covariate response (e.g., log transformations in logistic regression), is linear in the regression parameters. We also want consider models that are fundamentally non-linear in the regression parameters. For example, the following two models

\[
M(Y | X) : \beta_1 + \exp(\beta_2 X),
\]

and

\[
M(Y | X) : \beta_1 + \beta_2 e^{\beta_3 X},
\]

also ensures that the mean is non-linear in the regression parameters. However, we can also model that the covariate interacts with each other in the regression parameters. This type of non-linearity may be accommodated by including appropriate transformations of the covariates (e.g., log transformations in Poisson regression) and their interactions (e.g., log-linear interactions). For example, this model may be used in the case where the response variable is transformed by including both terms and their product in the regression model.
One particularly interesting aspect of regression analysis concerns the nature of the explanatory variables. A feature of this regression modeling approach is that it can incorporate a variety of different and continuous covariates in a relatively flexible fashion. Within this context, one can consider several (and often correlated to one another) numerical, ordinal, categorical, age, gender, and other attributes. Furthermore, the response variable, our primary interest, transformation of the response, can be influenced in a continuous fashion in a more complex or non-linear fashion through the relationships between the covariates or by the inclusion of polynomial terms (e.g., degree and order?). Alternatively, the covariates can be categorical (e.g., qualitative), though we generally speak of treatment groups.

Finally, regression models can include nonlinear relationships and continuous covariates, among variables among others. As in practice, applying regression models, in its purest sense, difficult to analyze high-dimensional data arising from in essentially classical experiments with simple qualitative covariates or factors, e.g., a randomized placebo-controlled high-dimensional clinical trial where one observes a continuous outcome of the vehicle placebo groups, the outcome continuous. Of note, in the latter case, regression models can even be used to distinguish within- and between-subjects differences in the responses (e.g., "interaction terms" versus "main effects") of age; these topics will be discussed in greater depth in later chapters.

Regression models are usually the preferred method when data analysis since regression parameters are directly related to the outcome measure of interest. For example, in a regression model, the data are from a longitudinal clinical trial, and there is a regression coefficient can be still an interpretation in terms of the continuous rate of change in the outcome response over time, in case of the longitudinal groups. Alternatively, the differences (or rates of change) of a particular regression coefficient can be examined during a treatment intervention. In some cases, intervention groups can even be randomized to intervention groups to study their relative longitudinal rate of change in the response variables over time.

So far, we have emphasized that it is not necessary to distinguish between the covariates and continuous covariates (contrast to the examples of this book) within a regression problem. However, from a purely biostatistical perspective, linear models then a continuous response with only continuous covariates have often been referred to as an exploratory analysis by researchers (ANCOVA) models. In contrast, linear models for a continuous response with only continuous covariates have often been referred to as linear regression models. Indeed, some authors would consider in situations where linear regression and analysis of covariance are identical (strictly linear, p. 22). A large part of this section is to provide a broader mathematical framework for this theoretical understanding. Analysis of covariance based the coefficient matrix in application situations, especially when continuous covariates are involved. The covariance (e.g., group yield) could be included by means of a regression model (e.g., polynomial regression), or polynomial regression can be used to analyze different types of covariance. In contrast, linear regression was initially developed for the analysis of intermediate data. Some of the earliest applications of linear regression can be traced back to the 19th century. By their very nature, these data include linear structures in astronomical surveys; purely observational (e.g., the positions and magnitudes of the heavenly bodies) and more the problem of exploratory manipulation. As a result of their unassuming differences, exploratory manipulations, ANCOVA and linear regression have often been presented as different methods; in practice, however, these methods are often combined.
differ in design (experimental versus observational) and the nature of the covariates (different versus continuous). In this, we see a parallel between linear regression as a very general method that harmonizes analyses of variance in a special case.

Further, although many of the commonly used statistical models for observational data were originally developed for clinical medicine, their emphasis that different in design, biases, and the nature of the covariates, almost all of these developments fall within the regression paradigm for observational data. See from a purely methodological perspective, it is not necessary to distinguish between linear modeling longitudinal data and observational data, which stems from observational studies and from studies with experimental designs. From this point of view, we henceforth choose not to focus on many of the early developments in methodology for analyzing longitudinal data, for example, the repeated measures ANOVA and nonparametric analysis of variance (Ditrusse, 1996). Conversely, we emphasize on a more general and complete regression paradigm that encompasses causal, if not all, of the earlier developments as special cases but can also handle all of the complexities that arise in applications. When viewed as special cases within the regression paradigm, the methodology (and within nonparametric) assumes only that any of the conditional methods for analyzing longitudinal data are properly understood.

In summary, we view the regression paradigm as a very flexible and general approach for analyzing longitudinal and observational data coming from many different types of studies. Regression models can provide a parsimonious description of complex data, allow the researcher to explore relationships within data, and helps the researcher to understand the relationships of interest. While this allows the researcher to understand the relationships of interest, it does not necessarily predict the future, as a generalization of the underlying probability distribution generating mechanism that might have produced the empirical responses, the latter is not considered to be the main focus of this analysis. Instead, our primary goal is to provide a flexible description of the relationships of interest in the response over time, and their relation to covariates, with regression coefficients that lead directly to the scientific questions of interest.

### 1.6 Contextualizing the Model

The results in this section are for a random effect. The first group, containing codes 1 through 9, provides the scores for the random effect of the measured factor and longitudinal plots. In Chapter 2, we introduce codes for the random effect of the monotonous factor and longitudinal plots. In Chapter 3, we introduce codes for the random effect of the monotonous factor and longitudinal plots, and the random effects of the measured factor and longitudinal plots. In Chapter 4, we introduce codes for the random effect of the monotonous factor and longitudinal plots, and the random effects of the measured factor and longitudinal plots.
longitudinal data have resemblance different targets of inference. Thus, we ensure that
the regression parameters have directly on the questions of substantive interest, greater
inference is needed to be established for different longitudinal claims.

The second part of the book, consisting of Chapters 17 and 18, consider examples which
are not missing data for longitudinal examples. In Chapter 17, we require that assumptions about
informative data are not required to ensure that the methods discussed in earlier chapters provide
valid inference. Two methods for handling missing data, multiple imputations and
inverse probability weighted procedures, are discussed in detail in Chapter 18.

The final part of the book, consisting of Chapters 19 through 22, discusses a
number of advanced topics. In Chapter 19, we discuss methodology for
unfitted models that allows greater flexibility for the form of the relationship
between the random response and the covariates. This chapter focuses on three
methods: polynomial spline and linear mixed effects models. Chapter 20 considers the
design of longitudinal studies, including: the determination of sample size and power.
In Chapter 21, we discuss regression models for repeated measures and related designs
and empirical level the methods illustrated in earlier chapters can be applied in these
settings. In Chapter 22, we present an overview of methods for analyzing multilevel
data. Chapters 21 and 22 demonstrate how regression methodology for longitudinal data can
be applied to special classes of panel data regression models for correlated outcomes, where
hazards decline.

3.6 Rubin, 1987

In the presentation of methodology for the analysis of longitudinal data in subsequent
chapters assumes that the reader has a basic knowledge of statistics and a strong back-
ground in regression analysis. A useful review of introductory statistical principles
and methods, targeted to applied researchers, can be found in the books by Agresti
concepts can be found in Kleinman et al. (1999) and Gelman and Hill (2007); a
more advanced presentation of similar topics can be found in Reiter et al. (1996).