_{Chapter} **1** Introduction

1.1 About This Book

The fourth edition of *SAS® for Linear Models*, like earlier editions, plays a role somewhere between a textbook on applied linear models and a manual for using certain procedures in SAS. It can serve as a companion to various courses offered at colleges and universities. Most universities offer a course on statistical methods directed at graduate students in scientific fields that utilize analysis of variance and regression analysis. These topics are covered at an appropriate level for such a course. In addition, this book makes a useful companion to a graduate course in linear models for statistics students to bridge the gap between theory and application. It also is useful to persons engaged in data analysis as a reference for statistical topics and SAS programming to implement a multitude of methods that fall under the general classification of "linear models."

The list of topics covered in *SAS® for Linear Models* is intentionally very broad to at least touch the bases mentioned in the previous paragraph. Other books offer more detailed information of a similar nature on regression analysis, mixed models, and other topics. These are *SAS® System for Regression*, *Third Edition*, *SAS® System for Mixed Models,* and *Categorical Data Analysis Using the SAS® System, Second Edition*, all published by SAS.

1.2 Statistical Topics and SAS Procedures

The broad range of topics are itemized and detailed below, according to chapter. Previous editions of *SAS® System for Linear Models* used the REG, MEANS, TTEST, ANOVA, and GLM procedures in SAS. In the decade since publication of the third edition, there have been enormous advances in computer technology and statistical methodology, as well as enhancements to SAS. These have resulted in changes in the most efficient and appropriate ways of analyzing data, and standards for data analysis have been raised by scientific journals and regulatory agencies. Accordingly, we have shifted emphasis of the procedures and topics in the fourth edition. Most of the applications of PROC ANOVA in the first three editions now use PROC GLM, and PROC MIXED is used for most mixed-model applications. A brief treatment of a new topic, *generalized* linear models (as opposed to *general* linear models), is introduced, illustrated primarily with PROC GENMOD.

Chapter 2 describes linear regression analysis for single and multiple independent variables. However, the main purpose is not to provide a comprehensive illustration of regressions methods. Rather, the intent is to lay the groundwork for linear models and the inferential techniques for estimating and testing hypotheses regarding linear combinations of parameters by using the REG and GLM procedures.

Chapter 3 introduces the basic methods for estimating and comparing means by using *t*-tests and analysis of variance. The MEANS, TTEST, and GLM procedures are demonstrated. The ANOVA procedure is only briefly mentioned because PROC GLM has all the capabilities of PROC ANOVA.

Chapter 4 describes ANOVA methods for data with random effects. Analyses for nested sampling classifications, randomized-blocks designs, and two-way mixed-models are illustrated using the GLM and MIXED procedures. Results from GLM and MIXED are compared for the two-way mixed model.

Chapter 5 presents analysis of unbalanced data. The fixed-effects situation is discussed first, covering the four types of sums of squares in PROC GLM. Least-squares means and the concept of estimable functions are covered next, and then analysis of unbalanced mixed-model data is discussed, using both PROC GLM and PROC MIXED.

Chapter 6 discusses details of the linear model. This chapter presents the principles and theory needed to understand the methods used by PROC GLM and PROC MIXED.

Chapter 7 describes analysis of covariance—that is, methods for comparing means in the presence of a continuous concomitant variable. Also, general issues are described when both qualitative and quantitative variables are included in a model.

Chapter 8 focuses on repeated-measures analysis. Analysis methods using PROC GLM are presented, similar to the third edition. We also present an introduction to the topic using PROC MIXED, which models the covariance structure of the repeated measures.

Chapter 9 discusses multivariate analysis, and presents essentially the same material as in the third edition.

Chapter 10 introduces the topic of *generalized* linear models. These are models with a non-normally distributed response variable. A common subtopic is *logistic regression* for the case of a binary response variable. This is a new topic for the fourth edition, which introduces PROC GENMOD. Since the third edition, this methodology has become accepted in many subject matter areas, and expected in some.

Finally, Chapter 11 presents examples of special analyses. These are examples that extend and combine methods that have been described in previous chapters.