Reviewer: Kelcey Jasen, Luigi Salmaso
University of Central Florida, University of Padova

Applied Medical Statistics Using **SAS**

Applied Medical Statistics Using **SAS** is a thorough documentation of statistical methods, inclusive of medical data sets and **SAS** code. The book would make an excellent reference guide for medical data analysts with access to base **SAS** 9.3 or a textbook for an introductory and intermediate graduate biostatistics course.

The book starts with the necessary introduction to base **SAS** and its interface. The definition of procs and the language of **SAS** is explained in simplistic terms for a new user to pick up easily. The authors include an interesting background to the emergence of statistics in the medical field. The chapters follow a logical organization of statistical topics, each including a medical data set, **SAS** code, corresponding output (sometimes condensed), and mathematical models. This helps a reader first understand the topic, see an appropriate example, follow the **SAS** program, recognize the output, learn how to interpret and match it all to the theory. The length seems appropriate and justified by the abundance of valuable content it contains; there may be topics missing that are included in biostatistics courses but where the book lacks in procedures, it includes references to additional texts.

The necessary assumptions for the statistical procedures are included, emphasized where needed and also robust models are expressed. This allows the reader to make the final judgment call but illustrates how important it is to perform analyses and make proper inferences and conclusions only when the necessary assumptions are satisfied. In some cases, the corresponding nonparametric technique is provided.

The organization of the book is as follows: an introduction to **SAS**, statistics and measurement in medicine, clinical trials, epidemiology, meta-analysis, ANOVA and ANCOVA, scatter plots, correlation, simple regression, smoothing, multiple linear regression, logistic regression, generalized linear model, generalized additive models, analysis of longitudinal data, survival analysis, Cox's proportional hazards models for survival data, Bayesian methods, and missing values.

Chapter 1 introduces the reader to base **SAS** 9.3's user interface and the structure of **SAS** program steps using statements. Examples are provided to show the procedure outputs and

SAS logs when conducting the basic steps of a program. In addition, the chapter covers some of the techniques to read, input, and modify data into SAS which is usually a large proportion of the overall time and effort of a user's project, as mentioned by the authors. Final topics in this chapter include the proc step, global statements, SAS graphics, the Output Delivery System (ODS), and SAS macros. The chapter closes with a brief list of common user errors in SAS.

Once the reader is familiar with the basics of SAS, the book progresses to the topic of statistics in medicine in Chapter 2. A brief and interesting background to the history of statistics in the medical field is given before moving on to definitions of data and types of measurements. The first example in the text is the use of proc means and calculating Cochran's Q test on a data set of the presence or absence of cervical cancer from independent pathologists. The mathematical formula for Cochran's Q test is also supplied, allowing the reader to understand what SAS is calculating. The results of the test are analyzed and the inferences are explained. This organization is repeated throughout the book making it an easy read and helpful to navigate to the appropriate section if referencing the book during a project. The remaining topics in Chapter 2 are McNemar's test, Cohen's kappa, intraclass correlation coefficient, ANOVA for reliability data, Pearson's correlation coefficient, and diagnostic testing. Chapter 2 ends, as do all the chapters, with a summary of the methods discussed and their application.

In Chapter 3 "Clinical Trials", the two-sample t test is discussed. The authors state the necessary assumptions for a valid two-sample t test and include assessing these assumptions on an example. When the assumption of normality appears to be violated for one of the treatment groups, the corresponding nonparametric Wilcoxon Mann-Whitney rank sum test is introduced. The authors' inclusion of the assumptions, checking the assumptions, and alternative approaches is a very important inclusion if the reader is a non-statistician. Often, practitioners who lack a statistical background will perform the wrong analysis or will construct inferences when it is inappropriate to do so. The inclusion of these examples will help reduce mistakes.

A medical researcher can use this text as a resource in the beginning stages of a medical study referring to the sampling techniques in Chapter 3 "Clinical Trials", sample size estimation in Chapter 4 "Epidemiology", and synthesizing published results from the large number of annual clinical trials using a systematic approach in Chapter 5 "Meta-Analysis". The common method for exploring the relationship between variables using regression is covered in Chapter 7 "Scatter Plots, Correlation, Simple Regression, and Smoothing", Chapter 8 "Multiple Linear Regression", and Chapter 9 "Logistic Regression". In these chapters, the authors include visual representations of the data emphasizing how to visually deduce meaning in data and validate output produced by SAS. The multiple regression model, ANOVA, ANCOVA models (Chapter 6 "Analysis of Variance and Covariance"), and the logistic regression model are members of the family of generalised linear models (Chapter 10). Generalised linear models and generalized additive models (Chapter 11) are powerful tools for more realistic models for medical data. Longitudinal data, often collected from clinical trials, is spread over three chapters (Chapter 12−14) which are a particularly useful application of statistical techniques to medical data. Survival analysis (Chapter 15) and the specific and notably popular method of Cox's regression (Chapter 16) are statistical techniques appropriate in oncology research which included references to other topics discussed in the text like sample size estimation and random-effects models. The authors acknowledge the connecting ideas across chapters allowing the book to be a reference to itself. All of the data sets, SAS code, and output

are completely accessible for readers online at http://support.sas.com/amsus allowing the practice of the methods thoroughly covered in the text.

More recent topics in the general field of statistics are also included; such as, Bayesian analysis and missing values. Chapter 17 "Bayesian Methods" includes prior distributions in SAS and Markov chain Monte Carlo processes. The summary of Chapter 17 includes caveats and objections against the Bayesian approach but defends the practice in a medical context by the approval of the US Food and Drug Administration. Chapter 18 "Missing Values" includes data mining topics for example detecting missing data patterns and missing value imputation. The chapter is brief, and refers readers to an issue of the journal *Statistical Methods in Medical Research* for more information but is a good introduction to the topic and perhaps will perk the interest of readers to include or research missing values that they would have previously disregarded.

Applied Medical Statistics Using SAS comes to the market at an appropriate time in the extension of statistical applications to the medical industry thanks to the abundant examples, medical datasets, SAS code, and proper analysis of output in this book. There is a growing need for a simplistic explanation of statistical procedures to medical researchers who lack a statistical background and the use of software is evidently necessary for the modern analysis of medical data. The thoroughness of procedures and the consideration the authors included in the selection of graphs, SAS code, and theory allow this book to be a resourceful companion for medical analysts. If looking for a broad selection of medical analyses using base SAS 9.3 this is the book for you; in addition, if a particular topic is required for further analyses the book references additional sources.

Reviewer:

Kelcey Jasen
University of Central Florida
Department of Statistics
Orlando, Florida 32816-2370, United States of America
E-mail: Kelcey.Jasen@ucf.edu

Luigi Salmaso
University of Padova
Department of Management and Engineering
Vicenza, Italy 36100
E-mail: luigi.salmaso@unipd.it
URL: www.gest.unipd.it/~salmaso/