

## Preface

This is the third edition of a textbook with the same title. The last two editions were originally published in Chinese. It has been used in teaching by many leading medical schools for 7 years with a lot of positive feedback. The readers have found that the content of this book is necessary for medical researchers and the computer experiments in each chapter are helpful. As a result, the Ministry of Education in China has officially recommended the second edition for the postgraduate program. Some colleagues in other countries have recognized this book as a typical reflection of the teaching materials and style that Chinese statisticians have been working with. They have encouraged us to publish this English version for international exchange.

This English edition is not a simple translation of the second Chinese edition, but a harmonized revision and improvement of the previous editions. According to the suggestions from the referees and publisher, the 3 parts have been re-ordered as: Part I — Basic concepts of statistics; Part II — Multivariate statistics and Part III — Statistics in medical sciences and public health.

In this revision, we have reorganized related topics in a systematic and logical order to improve the clarity and consistency. For instance, the closely related topics such as “simple correlation — simple regression — multiple regression — nonlinear regression”, “cross sectional study — case control study — prospective study — logistic regression — survival analysis — log linear model”, and “discriminant analysis — principle components analysis — factor analysis — canonical correlation analysis — correspondence analysis” are all sorted out towards a goal of eliminating the redundancies and amending the gaps.

Even within chapters, some materials have been carefully revised based on the readers’ responses to our previous Chinese versions. For instance,

on the statistical description for discrete variables, we used to classify the relative measures into “proportion, rate and ratio” as of most similar books; however, students and even instructors were often confused about proportion and rate, and found no room for the measures such as incidence rate and mortality rate. Therefore, in this edition, we decide to classify the relative measures into “frequency, intensity and ratio” to clarify the concepts more precisely. Again, in the chapter of logistic regression, the nature of the conditional logistic regression has been illustrated more clearly than in the previous versions. The basic model is still the logistic model for two categories, but limited by the 1:1 paired design, one can only consider the conditional probability within the couples, and only the “difference within couple” can be used. Hence the intercept cannot be estimated and the model cannot be used for prediction.

The 81 computer experiments can be categorized as two types: one is to perform the commonly used statistical designs and analyses, by which the reader can easily deal with their own data by changing a few parameters; another is to perform simulation experiments, for which we have provided the programs for generating and analyzing “data”. After replication, the reader can easily observe the statistical phenomena and rules. All necessary data and programs for the computer experiments have been revised, and are all printed in the book and stored in the attached CD-Rom.

Besides the editors, many friends have offered their helps to this book. Firstly, we would like to acknowledge the contributions of Dr. Tong Wang, Dr. Kai W. Ng, Dr. Lin Fei, Dr. Futian Luo and Dr. Binghua Su, who have helped us to revise and translate Chapters 10, 13, 20, 21 and Appendix III respectively. Also Dr. Sin Kai Luo, Dr. Min Yang, Dr. Boguang Zeng, Dr. Zongli Xu and Dr. Caixia Li, who have reviewed some chapters of the early draft and kindly provided valuable comments respectively. We sincerely thank Dr. Jie Yan and Dr. Xuemei Wang, who have checked and run through the programs for computer experiments, and picked out several potential errors. We would also like to thank Ms. Shaomin Wu who has revised and managed the whole package of the plots. Finally, a special thanks goes out to Dr. Jinxin Zhang who has patiently reviewed and revised the reference answers for the “Thinking, Practice and Experiments” in each chapter, which you will find in Appendix D.

Anyway, through a great effort of the authors as well as many colleagues and friends, this edition has its quality further promoted and becomes more mature than previous editions. We truly hope this book will be of great use to all students of Medical Statistics, Biostatistics and Applied Statistics.

Ji-Qian Fang, Ph.D.  
Chair Professor  
Department of Medical Statistics  
School of Public Health  
Sun Yat-Sen University  
Guangzhou, 510080, China

## Preface to the First Edition

Most of the current postgraduate programs in medical schools and institutions in China require about 100 teaching hours of medical statistics; in addition to this, some programs even set up various selective courses in medical statistics, about 50 teaching hours for each. From this, one can see the important role of medical statistics in the medical sciences. Many colleagues and students have encouraged us to write a textbook to serve the postgraduate programs of medical sciences in China. Our editorial team was established in 1993. We named this book “*Medical Statistics and Computer Experiments*”. The backbone of this book is formed by the basic concepts and commonly used methods on design and analysis in medical statistics, incorporating with the operation of statistical package SAS and the computer experiments for the important statistical phenomena related to each chapter. We all commit ourselves to create the following prominent features: to clearly explain the profound statistical concepts and logic in a simple way without too much mathematics; to cover both the traditional and advanced statistical methodology; to naturally merge both the knowledge on statistical design and analysis; to carefully arrange the computer experiments in each chapter in order to make the abstract theory more intuitive and interesting; to make the textbook exercises challenging and stimulating. The first draft was finished in June 1995, and after a run of cross reviewing, all the manuscripts were sent to me to be merged and polished for unity of content, format and style. Before sending to press, several chapters were test-taught by the authors in their own courses and the feedback was collected for the last-minute revision.

This book consists of 3 parts, with 10 chapters for each. The first part includes the basic concepts in statistics; the second includes the design

and analysis for medical researches; and the third includes the multivariate statistical methods. Each part needs 50 teaching hours (1:1 shared by the theory and practice sessions). The first two parts can be used together for the required course of the postgraduates in medical field; the third part can be used for the selective course. In fact, undergraduate program in clinical medicine may use the first part, and undergraduate program in public health can reference the first two parts.

There is a specific section in each chapter for “Computer Experiments” which can be run by statistical package SAS. An introduction of SAS is given in the appendix. If a computer lab is available, it is encouraged to extend the current practice session to include computer experiments, where the students can actively observe the statistical regulation of the random phenomena using the SAS programs given in the book; otherwise, the computer experiments can be demonstrated by the instructors in the classroom. Since the computer experiments section in each chapter is relatively independent, if no computer is available, this section can be skipped.

Another section set at the end of each chapter is “Thinking, Practice and Experiments”, of which the majority come from practice or literatures, and for some even do not exist the best answers only but better ones. There is plenty of space for thinking and discussion. Part can be used in the practice session, and the rest can be used for homework.

In Appendix II, several medical examples and related data are provided, with which the readers can more experience the process of resolving real life problems. Appendix III is the glossary of terms with English-Chinese contrast, which are collected from the 30 chapters.

We sincerely appreciate the advice from the two advisors to this book, Professor Zuchao Guo (Fourth Military Medical University of China) and Professor Mengxuan Hu (Sun Yat-Sen University of Medical Sciences). We owe this book’s publishing to the support of the office of postgraduate education, Sun Yat-Sen University of Medical Sciences and to the assistance of Mr. Zhonghua Tang, senior editor, Shanghai Science and Technology Publishing House. Many thanks to my colleagues and postgraduate students of all the authors, who generously contributed in collecting data, checking program, typing, plotting and proofreading for this book.

Limited by the authors' knowledge and energy, there must be lots of deficiency in the book. We would warmly welcome comments and suggestions from colleagues and readers, we will carefully remedy this book's shortcomings in the next version.

Ji-Qian Fang

Guangzhou, September 1995

## Preface to the Second Edition

After a series of period review, the document No. [2000] 12 from the Office of Education Research, Ministry of Education, People's Republic of China lists our textbook "*Medical Statistics and Computer Experiments*" (Shanghai Science and Technology Publishing House, 1997) as a recommended textbook for postgraduate programs. We were suggested to further enrich its materials and include more universities into the editorial committee. A meeting for the second edition was held at Sun Yat-Sen University of Medical Sciences, Guangzhou, on 24–26 June 2000. The editorial committee was expended up to 20 outstanding statisticians who came from 14 well-known medical schools in China, such as Shanghai Medical University, Shanghai Second Medical University, Western China Medical University, Fourth Military Medical University, Tongji Medical University and Sun Yat-Sen University of Medical Sciences.

After the three-day discussion, the revision goals were determined as: keeping the original structure; supplying a reasonable amount of new materials; strengthening the concepts, focusing on the applications and relaxing the calculation requirements. With joint efforts in the next 4 months, the manuscript of the second edition was worked out.

Compared to the first edition, this second edition has made great progress in readability. First, each chapter starts with a challenge of the medical problems; and then discusses the ideas and strategies, introduces the concepts and methods focusing on the real problems; finally ends in resolving problems and interpreting the results. Besides introducing the statistical concepts, designs and methods commonly used in advanced medical researches, this book presents some of the caveats of the methods introduced by showing potential mistakes by commenting on real medical research projects. The performance of the computer experiments becomes much easier than the

first edition: besides the initial SAS programs, the parallel programs in SPSS and EXCEL are also provided in the CD-Rom.

The material used is more readily applicable than that of the previous version. First, two brand new chapters have been added: the analysis of repeat measures of continuous variable (Chapter 22) and the statistical analysis for genetic data (Chapter 31). Chapter 22 covers materials that are most often missed in the current medical literatures; Chapter 31 covers materials that have already been widely used in modern Genetics. Second, some important statistical issues in advanced medical research are supplied to most chapters, such as biostatistics principles in clinical trials (Chapter 11), ROC curve (Chapter 18), path analysis (Chapter 21 — the former Chapter 22), the validity, reliability and responsiveness of questionnaires (Chapter 13 — the former Chapter 20), confirmatory factor analysis (Chapter 26), correspondence analysis (Chapter 27), multinomial logistic regression (Chapter 28) and Poisson regression (Chapter 30). Third, the Appendix is expanded to two parts: paper print and computer disks; the former includes brief introduction of SAS, statistical tables, short answers, references, Chinese-English and English-Chinese glossary of statistical terms; the latter includes the SAS and SPSS programs for all computer experiments and the data sets of several medical research projects.

The second edition still has three parts: Part I includes 10 chapters on basic statistical concepts; Part II includes 9 chapters on design and analysis of medical researches; Part III includes 12 chapters on advanced statistical methods. The first two parts can be used for required courses, 50 teaching hours for theory session and 50 hours for practice session; the third part can be used for elective courses, 30 hours for the theory session and 20 hours for the practice sessions. As for the whole book, 170–180 teaching hours are needed. The instructors may select and drop according to their curriculum and the number of teaching hours. The chapters or sessions with \* can be used for the students' supplemental reading.

This revision benefits from the excellent young professors who teach low division classes at many famous universities. They have generously contributed their great wisdom and discernment. The faculty staffs of the Department of Medical Statistics, Sun Yat-Sen University of Medical Sciences provided their 5-year successful teaching experience of the first edition, with which they actively impregnated into the second edition. The

secretaries Ms. Shaomin Wu and Miss Fangfang Zeng and the postgraduate students Miss Meihua Wang, Miss Qiling Wang, Miss Yan Chen and Mr. Guohui Liu offered their great support to the authors. On behalf of the editorial committee, I would express our sincere thanks to all of the above-mentioned colleagues and friends who have been involved in and devoted to this second edition.

Without doubt there must be some unintentional omissions and mistakes. Perhaps the old mistakes are just rectified and the new ones come forth immediately. I earnestly invite all readers and colleagues to point out all the book's shortcomings so that we can continuously revise it to make the future editions better.

Ji-Qian Fang

Guangzhou, October 2000