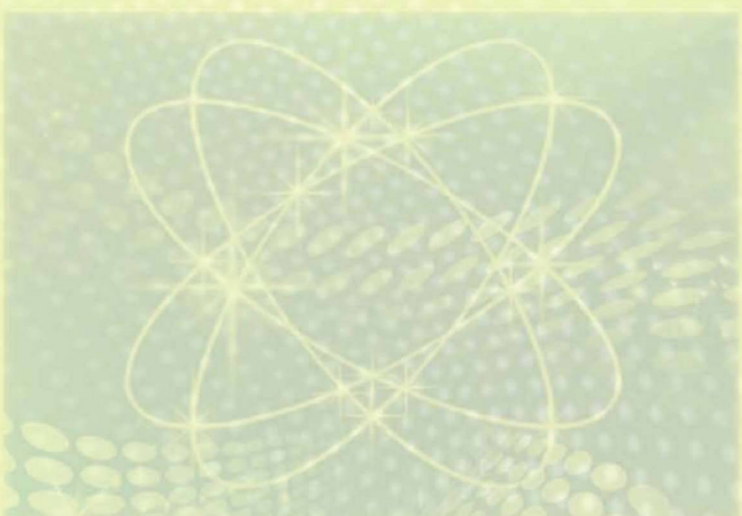
医学统计学 =Medical Statistics 英文

陆贤杰　主编



广西科学技术出版社

图书在版编目(CIP)数据

医学统计学＝Medical Statistics：英文／陆贤杰主编. —南宁：广西科学技术出版社，2013.10
ISBN 978-7-80763-902-2

Ⅰ.①医… Ⅱ.①陆… Ⅲ.①医学统计—统计学—双语教学—高等学校—教材—英文 Ⅳ.①R195.1

中国版本图书馆CIP数据核字(2013)第003506号

Medical Statistics

医学统计学

YIXUE TONGJIXUE

主　编　陆贤杰
副主编　黄高明　邓树嵩　马金香

出版发行　广西科学技术出版社
（社址/南宁市东葛路66号　邮政编码/530022）
网　　址　http://www.gxkjs.com
经　　销　广西新华书店
印　　刷　广西大华印刷有限公司
（地址/南宁市高新区科园路62号　邮政编码/530007）
开　　本　787 mm×1092 mm　1/16
印　　张　20.25
字　　数　414千字
版　　次　2013年10月第1版
印　　次　2013年10月第1次印刷
书　　号　ISBN 978-7-80763-902-2
定　　价　48.00元

高等医学院校双语教材

（供医药、预防、护理各专业研究生及本科生选用）

Medical Statistics

医学统计学

编委会

主　编　陆贤杰

副主编　黄高明　邓树嵩　马金香

编委（以姓氏笔画为序）

马迎教　右江民族医学院
王　玖　滨州医学院
石武祥　桂林医学院
陆贤杰　右江民族医学院
曹明芹　新疆医科大学
黄志碧　广西医科大学
谭盛葵　桂林医学院
马金香　广州医学院
邓树嵩　右江民族医学院
宇传华　武汉大学
郭　蕊　右江民族医学院
黄高明　广西医科大学
黄鲜桃　广西医科大学

秘书：漆光紫　右江民族医学院

前　言

当今世界正处在国际化科技、信息化环境、全球化经济的快速发展和重大变革时期，通用的语言、文化交流能力越来越受到重视。这种能力也成为我国各类人才参与竞争的一种有效工具。国家教育机构和各类大专院校一直在思考如何顺应这一发展潮流，推动各层次人员通过学习来获取这种能力，双语教学就是在这种背景下的一种尝试。

国家教学部2001年提出加强大学生本科教学的12项措施，其中要求各高等学校在3年内开设5%～10%的双语课程。2003年，教育部在《高等学校教学质量与教学改革工程纲要》第十二条中明确指出："要继续推进双语教学。"在我国，双语教学较为常用的语言是汉语和英语。双语教学从其诞生的那一天起就被包围在赞成和反对的声浪中。如今，依然有人赞成有人反对，但无论怎样，双语教学的规模和影响都在原来的基础上不断扩大，且呈大发展之势。一些率先进行双语教学的高等学校在教学实践中积累了经验，不断加以改进和提高；一些待进入者也在模仿中学习，并静待时机成熟加入这一行列。由于我国内地长期缺乏讲第二语言的环境，开展双语教学面临特殊的困难，因此选用合适的教材是双语教学成功与否的关键。目前国内许多专业尚未有统一的专业双语教材，是阻碍各高等学校开展双语教学的最大问题。对于专业课双语教材的选择，根据国内有关调查，不同专业本科生对双语教材的选择有一定的差异。有的学生愿意选用国外原版英语专业教材，而多数学生愿意选择自编实用的英语教材。我们认为，双语教材用原版的各类学科教材，可以避免中国式英语问题，保证语言的原汁原味。但是，我国对学生的培养计划和要求与国外的不完全一致，照搬原版英语教材不符合我国国情，甚至会误导学生，影响教学效果。而且引进原版英语教材费用巨大，增加学生的经济负担。因

此，按照国家教学大纲和教学基本要求，组织编写专业双语教材，可以满足各高等学校双语教学的需要，且自编教材费用相对较低。

顺应我国目前双语教学发展趋势，我们组织编写了 *Medical Statistics* 双语教材。本教材共分十七章，内容包含了本科生、研究生需要掌握的医学统计学的基本原理和各类统计资料常用的统计分析方法，也增添了圆形分布描述和圆形分布资料的假设检验等新的内容。本教材的特点是编写条理清楚，讲究实用性，不强调统计学公式的数学演绎和推理。教材适用于医学院校各专业本科生和研究生双语教学使用，各医学院校可以根据本校不同专业的实际情况选择教学内容。对从事临床医学、预防医学、护理学、基础医学和管理学科的卫生工作者也是一本有用的参考书。

编写本教材的作者都是长期从事医学统计教学和科研工作的老师，他们多数经历在国外学习或工作 2 年以上，比较了解国外教育背景，总结多年积累的教学经验编写成一本实用性的双语教材，奉献给他们的学生，并用来与同行交流，这就是编者的初衷。

在编写教材过程中得到国内同行专家、老师的支持和指导，特别是得到右江民族医学院预防医学系各位同仁的关心、支持和协作，外语系程家惠教授对部分章节实例提出中肯的建议，并对语言文字进行编写、修改，借此机会表示衷心的感谢！

虽然我们全力投入编写并反复修改，但是限于时间和我们所处的文化背景，难免出现疏误，敬请读者批评指正。

为了使我们推出的双语教材更适于教学，我们真诚期待广大读者提出宝贵的意见和建议。

陆贤杰

2012 年 11 月 8 日

Preface to the first edition

To improve the college education and the quality of teaching, and to enhance the practical competence of foreign languages and students' social adaptation in learning their advanced professional knowledge, we have compiled this *Medical Statistics* (English version) as the bilingual teaching materials for students taking undergraduate and master courses, and an ideal reference book of medicine in English.

Statistical analysis is becoming increasingly important as a tool in scientific research, not only in industry, but also in the field of biological medicine. Most biologists are not so much concerned with acquiring detailed knowledge of the mathematical theory of statistics, as in learning how to select and how to use those statistical techniques which will help them interpret the results of their experiments.

We have emphasized the practical applications of statistical analysis, and have kept mathematics to a minimum. There is a limit to the number of techniques which can be dealt with in an introductory text of this size.

The general plan of this book is to show first, in Chapter 1 (Introduction to Statistics). The definition of medical statistics has been made, which is an applied branch of mathematical statistics in medical research area. Some basic concepts of statistics have been introduced such as population and samples, types of statistical data, variable and value of variable, and probability. Chapter 2 (Descriptive Statistics) introduces some ways that you can organize and describe the data, the tabular and graphical representations, provide the

distribution pattern of data and help determine the statistical methods to be used in data analysis. Measures of central tendency, introduce three commonly used measures for describing the central tendency of a population or sample, namely, mean, median and geometric mean, and the choice of the appropriate measures mainly depending on the distributional property of data. Measures of dispersion and variability, introduces the commonly used measures of dispersion: range, inter-quartile range, variance, standard deviation and coefficient of variation. In Chapter 3 (The normal distribution and its application), properties of a normal distribution and its application are introduced. In addition, Sampling distributions and the central limit theorem and distribution of mean and standard error have also been discussed. In Chapter 4, t-distribution and confidence interval estimation, and basic process of hypothesis testing are introduced. In Chapter 5 (Statistical inference of measurement data), testing the difference between means (small independent samples), testing the difference between means (dependent samples), and testing the difference between means (large independent samples) are all covered. Chapter 6 (Analysis of variances) introduces two kinds of ANOVA accordingly, used for analysis of data from completely randomized design and randomized block design. Besides, the commonly methods of multiple comparison between each two means are also introduced. Relative number indexes used for categorical data including rate, proportion (constituent ratio) and relative ratio, applicability as well as the formulas of both direct and indirect standardized rate are discussed in Chapter 7.

Chapter 8 (Analysis of enumeration data) is about using a z-test testing the difference between two population proportions p_1 and p_2 using a sample proportion from each population. Chapter 9 (Chi-square (x^2) testing) covers

the Chi-square goodness of fit, and the 2×2 contingency table. The frequency table for categorical data includes $R\times C$ table which is called contingency table. $R\times C$ Table includes two or more sample rates (or proportions). Chi-square test can be used to infer whether there is a difference among two or more rates (or proportions), and an association between two categorical variables. Binominal distribution and Poisson distribution, and two important probability distributions of discrete variables are introduced in Chapter 10 and Chapter 11. A variety of rank sum tests including Wilcoxon signed rank test, Wilcoxon test for independent two samples, Kruskal-Wallis and Friedman test, and their applications in the area of public health and biomedicine are introduced in Chapter 12. Simple Linear Correlation and Linear Regression, and the Spearman rank correlation coefficient are introduced in Chapter 13. Two commonly used statistical models of multivariate analysis in the area of biomedicine and public health are briefly introduced in Chapter 14, together with multiple linear regression model and logistic regression model. Statistical tables and graphs are introduced in Chapter 15. In Chapter 16, we introduce a special type of interval scale, namely, a circular scale. This chapter also introduces some basic considerations useful in calculating descriptive statistics, including graphical presentation of data, sines and cosines of circular data, the mean angle, angular dispersion and confidence limits for the mean and median angles, etc. And the final Chapter 17 discusses the hypothesis testing of circular data.

An important feature of this book is the summary in the end of each chapter. This is intended for a quick-reference guide for readers with some knowledge of statistics. Many workers who have already acquired a basic training in statistics, either from this book or elsewhere, are frequently

required to have their memories refreshed.

The eighteen Appendix Tables provided enable readers to carry out the commonest statistical tests without special reference to more extensive compilations. There is, however, some advantages in possessing a good set of tables for more general application.

Finally, a word should be said about further reading. This book attempts to provide the groundwork basic to most statistical methods. However, those workers who are closely concerned with special fields need to know something about methods more advanced. It is possible in the subject of experimental design, for example, to learn to use relatively sophisticated patterns of experimentation without becoming involved in higher mathematics. To some extent the choice of the textbook is a personal one, depending on the reader's own interests and way of looking at things. Specific recommendations are thus liable to be difficult. The section on Suggestions for more Advanced Reading, therefore, includes a variety of statistical books, some of them rather specialized, which the readers may find useful to consult.

A book of this nature requires and benefits from the assistance of many people. I wish to acknowledge the invaluable help and advice given by Professor Cheng Jiahui, English department of Youjiang Medical College of Nationalities. During the preparation of the manuscript he has kept a watchful eye for errors of English translation. Nevertheless, responsibility for mistakes of any kind remains my own.

Lu Xianjie

Contents

Chapter 1 Introduction to Statistics

Many of the investigations in medical sciences have become quantitative, in that a great many types of medical observations consist of numerical facts called data. As medical entities are counted or measured, it becomes apparent that some objective methods are necessary to aid the investigator in presenting and analyzing research data.

The word "statistics" is derived from the Latin for "state", indicating the historical importance of governmental data gathering, which related principally to census taking and tax collecting. The layman uses this term as a synonym for "data"; One hears of college enrollment statistics (how many freshman students, how many students from each province, etc.). Statistics of diseases (how many people are suffering from hypertension in a certain area.), and so on. In this book, "statistics" will be used to refer to the analysis and interpretation of data with a view toward objective evaluation of the reliability of the conclusions based on the data.

Before data can be analyzed, they must be collected, and here statistical considerations can aid in the design of experiments and in the setting up of hypothsis to be tested. Many are the researcher who will attempt the analysis of their research data only to find that too few data were collected to enable reliable conclusions to be drawn, or that much extra effort was expended in collecting data that cannot be of ready aid in the analysis of the experiment. Thus, the knowledge of basic statistical principles and procedures is important even before an experiment is begun.

Once the data have been obtained, we may organize and summarize them in such a way as to arrive at their orderly presentation. Such procedures are often termed descriptive statistics. For example, a tabulation might be made of the heights of all members of freshman from Clinical Medicine class, indicating an average height for each sex, or for each age. However, it might be desired to make some generalizations from these data. We might, for example, we wish to make a reasonable estimate of the heights of all freshmen in the university. Or we might wish to conclude whether the males in the university are on the average taller than the females. The ability to make such generalized

conclusions, inferring characteristics of the whole from characteristics of its parts, lies within the realm of inferential statistics.

1.1 Definition of Medical Statistics

Statistics is a helpful science for inherent law deduction by means of fortuitous phenomena analysis, and a research for collection, collation, analysis and deduction of the data about uncertain phenomena, whose theory is based on the theory of probability and the mathematical statistics with the application of the principle and method in medical research. Statistics applied to medical problems is called medical statistics or health statistics.

1. Significance of medical statistics

The service objects of the medical and health care are all the inhabitants. To carry it out effectively, a survey and analysis of the health of the inhabitants must first be done for a scientific appraisal usually with the help of medical statistics.

Medical statistics is an essential tool for learning about the quantitative characteristics of the medical phenomena, a basic principle and method for the application of the probability theory and the mathematical statistics, and a process for medical research design and collection, collation, analysis and deduction of the data. Statistics is an important premise and approach for scientific research.

2. The use of medical statistics

(1) Study of influence of social condition and environment factors on the human health;

(2) Appraisal of quality and effect of treatment and prevention;

(3) Survey, experimental design and data processing of medical and health research.

3. Main contents of medical statistics

(1) Principle and method of health statistics

(2) Health statistics including demographic statistics, morbidity statistics, and growth and development statistics

(3) Health service statistics, etc.

1.2 Basic Procedure of Statistical Work

Plan and design, data collection, collation and analysis are the four steps for medical statistics, which are closely related. A scientific, deliberate design ensures the collection for precise and reliable data. Therefore, plan and design is the key for medical statistics, while deliberate, integral and timely data

collection and proper collation is the basic for statistical analysis. On that basis, only with the right method can a scientific conclusion be made. This chapter mainly covers analysis of statistical data.

1. Plan and design

The research plan should include research objective, method, manpower, resource, organization, etc. Design means further completion of survey and experiment design after a deliberate medical research plan is made. A good design should be scientific, deliberate and concise so as to attain more precise research data with less manpower, resource, and time.

2. Data collection

(1) Data source

Regularly data: ① statistical report forms, ② reporting card, ③ daily health care record.

Temporary data: Recording of particular survey or experiment.

Plan for data collection includes: ①options for personnel, place and time ②draw-up for the training of the data collection personnel ③draw-up for pre-survey and pre-experiment ④ways of data record ⑤draw-up for survey list and printing ⑥ survey or preparation for experiment instruments and reagent ⑦ sampling method for survey data ⑧provision for outlay of the data collection, etc.

(2) Types of statistical data

There are three types of data: ①Measurement data (include ratio data and interval data), ② Enumerate data (nominal data), ③ Rank data (ordinal data).

Data on a ratio scale. The data values in ratio data do have meaningful ratios.

First, there is a constant size interval between any adjacent units on the measurement scale. That is, the difference in body height between a 50 cm and 51 cm is the same as the difference between a 164 cm and 165, and someone who is 40 is twice as old as someone who is 20. Temperature is ratio data, weight, body height, and the number of heart attacks also is ratio data.

Second, it is important that there exists a zero point on the measurement scale and that there is a physical significance to this zero. This enables us to say something meaningful about the ratio of measurements. We can say that a 90 cm tall body height is half as tall as a 180 cm.

Measurement scales having a constant interval size and a true zero point are said to be ratio scales of measurement. Besides lengths and numbers of items, ratio scales include weights (mg, g, kg, etc), volumes (cc, cu etc), capacities (ml, qt, etc), age (year), and the number of heart attacks also is

ratio data.

Data on an interval scale. Some measurement scales possess a constant interval size but not a true zero; they are called interval scales. An outstanding example is that of the two common temperature scales: Celsius (C) and Fahrenheit (F). We can see that the same difference exists between 5 ℃ (40 ℉) and 10 ℃ (50 ℉) as between 20 ℃ (68 ℉) and 25 ℃ (77 ℉); i. e., the measurement scale is composed of equal-sized intervals, but it cannot be said that a temperature of 40 ℃ (104 ℉) is twice as hot as a temperature of 20 ℃ (68 ℉); i. e., the zero point is arbitrary.

Some interval scales encountered in biological data collection are circular scales. Time of day and time of the year are examples of such scales. Circular biological data are occasionally compass points, as if one records the compass direction in which an animal or plant is oriented. Since the designation of north as 0° is asbitrary, this circular scale is a form of interval scale of measurement. Some special statistical procedures are available for circular data; these are discussed in Chapters 16 and 17.

Data on a nominal scale. Sometimes the variable under study is classified by some quality it possesses rather than by a numerical measurement. In such cases the variable is called an attribute, and we are said to be using a nominal scale of measurement. i. e., nominal data refers to data that represent categories or names, and there is no implied order to the categories of nominal data. Some examples of nominal data are: (1)Eye color, an animal's eye color may be blue or brown, and if human hair color were the attribute of interest, we might record black, brown, or red; (2) Race; (3) Sex; and (4) Marital status. etc. Sometimes data from an ordinal, interval, or ratio scale of measurement may be recorded in nominal scale categories. For example, heights may be recorded as tall or short, or performances on an examination as pass or fail.

Data on an ordinal scale. The preceding paragraphs on ratio and interval scales of measurement discuss data between which we know numerical differences. For example, if man A body weights 50 kg and man B body weights 60 kg, then man A is known to weigh 10 kg more than B. But our data may, instead, be a record only of the fact that man A weights more than man B (with no indication of how much more). Thus, we may be dealing with relative differences rather than with quantitative differences. Such data consist of an ordering or ranking of measurements and are said to be on an ordinal scale of measurement ("ordinal" being from the Latin word for "order"). Ordinal data refers to data that are ordered, but the spaces or intervals between the