



普通高等教育“十三五”规划教材

# 医学统计学 与SPSS软件实现方法

Medical Statistics and SPSS Software Application

(第二版)

郭秀花 主编



科学出版社



普通高等教育“十三五”规划教材

# 医学统计学与 SPSS 软件实现方法

(第二版)

郭秀花 主编

科学出版社

北京

## 内 容 简 介

医学统计学是我国各医学院校所有专业本科生、研究生的必修课，也是从事医学科学研究工作者不可缺少的方法学。本教材将医学统计学理论方法与常用的 SPSS 软件操作相结合，分为十八章：绪论、数据管理与 SPSS 软件实现方法、定量资料的统计描述、定量资料的参数估计与假设检验基础、定量资料的  $t$  检验、定量资料的方差分析、定量资料的非参数检验、定性资料的统计描述、定性资料的参数估计与  $\chi^2$  检验、有序定性资料的假设检验方法、直线相关与回归、多重线性回归分析、Logistic 回归分析、生存分析基本统计方法、Cox 比例风险回归分析、统计表与统计图、观察性研究设计、实验性研究设计。本教材的附录部分增加了自测试题。此外，教材中所有例题数据、各章练习题答案、自测试题答案和样本含量估计方法，采用了二维码形式，方便使用。

本教材可供临床医学、基础医学、口腔医学、护理学、医学检验、中医学、中药学、卫生法学、卫生信息管理学、卫生事业管理学、生物医学工程学、药学等专业本科生开设医学统计学课程教学使用；也可供各专业研究生开设 50~60 课时以内的医学统计学教学使用，同时，对从事医学科学研究的工作者也是方便自学的有益方法学书籍。

### 图书在版编目(CIP)数据

医学统计学与 SPSS 软件实现方法/郭秀花主编. —2 版. —北京: 科学出版社, 2017.7  
(普通高等教育“十三五”规划教材)

ISBN 978-7-03-051893-4

I. ①医… II. ①郭… III. ①医学统计-统计分析-软件包-高等学校-教材  
IV. ①R195.1-39

中国版本图书馆 CIP 数据核字(2017)第 036796 号

责任编辑: 闵捷 / 责任校对: 贾娜娜 高明虎

责任印制: 谭宏宇 / 封面设计: 殷靓

科学出版社 出版

北京东黄城根北街 16 号

邮政编码: 100717

<http://www.sciencep.com>

天津市新科印刷有限公司印刷

科学出版社发行 各地新华书店经销

\*

2014 年 1 月第 一 版 开本: 889×1194 1/16

2017 年 7 月第 二 版 印张: 22 1/2

2018 年 7 月第七次印刷 字数: 716 000

定价: 48.00 元

(如有印装质量问题, 我社负责调换)

# 《医学统计学与 SPSS 软件实现方法》(第二版)编委会名单

---

主 编 郭秀花

副主编(以姓氏笔画为序) 方 亚 刘美娜 贾 红 黄水平 薛付忠

编 委 (按在书中出现的先后顺序)

郭秀花(首都医科大学)

钟晓妮(重庆医科大学)

罗艳侠(首都医科大学)

刘 芬(首都医科大学)

宇传华(武汉大学)

黄水平(徐州医科大学)

曾 平(徐州医科大学)

薛付忠(山东大学)

刘云霞(山东大学)

王素珍(潍坊医学院)

王 虹(首都医科大学)

刘美娜(哈尔滨医科大学)

张秋菊(哈尔滨医科大学)

郭淑霞(石河子大学)

李述刚(石河子大学)

钱 聪(中国医科大学)

刘红波(中国医科大学)

彭志行(南京医科大学)

王立芹(河北医科大学)

颜素容(北京中医药大学)

潘发明(安徽医科大学)

范引光(安徽医科大学)

尹素凤(华北理工大学)

贺 佳(第二军医大学)

孙 忠(天津医科大学)

王 媛(天津医科大学)

余红梅(山西医科大学)

杨兴华(首都医科大学)

曹明芹(新疆医科大学)

闫宇翔(首都医科大学)

吴立娟(首都医科大学)

方 亚(厦门大学)

王 玫(滨州医学院)

贾 红(西南医科大学)

陈 征(南方医科大学)

艾自胜(同济大学医学院)

学术秘书 闫宇翔

## 第二版前言

医学统计学是全国各医学院校所有专业本科生、研究生的必修课，也是从事医学科学研究不可缺少的一门方法学课程。虽然有各种版本的医学统计学或统计软件操作手册方面的书籍，然而，医学统计学方面的教材偏重于理论，详细介绍原理、公式和计算过程，缺少直观而详细的统计软件操作方法；而医学统计学软件操作手册又偏重软件介绍，缺乏统计学基本概念与理论。因此，缺少将医学统计学理论方法与常用的 SPSS 软件操作结合起来的书籍，学生需要购买两本教材。为方便教学、弥补缺憾，我们在 2012 年出版了《医学统计学与 SPSS 软件实现方法》，受到了许多高校学生的欢迎。本次应出版社要求，我们编写了第二版。

编写本教材秉承了第一版的四大原则：第一，以内容的科学性为主，兼顾理论的前瞻性。正确阐述医学统计学学科的科学理论和概念定义，在理论联系实际、以实例解释理论、对实践起到指导作用的基础上，注意将本领域的最新发展成果以及新技术、新方法纳入教材；第二，把握好写作条理性。注重教材的层次分明、条理清楚，教材体系能反映内容的内在联系及统计学的思维方式；第三，以本科学生或同等水平阅读能力群体为主要对象，并兼顾研究生医学统计学教学。从认知规律出发，富有启发性，便于学生学习，所选教学内容在满足学生未来职业活动所需的最基本、最常用的理论知识和方法基础上，增加样本含量估计方法等扩展内容；第四，突出实践技能，强化应用。注重科研实际案例引入，统计方法计算以 SPSS 20.0 软件操作和结果解释为主，使学生真正掌握实践操作技能。

本教材分为十八章：绪论、数据管理与 SPSS 软件实现方法、定量资料的统计描述、定量资料的参数估计与假设检验基础、定量资料的  $t$  检验、定量资料的方差分析、定量资料的非参数检验、定性资料的统计描述、定性资料的参数估计与  $\chi^2$  检验、有序定性资料的假设检验方法、直线相关与回归、多重线性回归分析、Logistic 回归分析、生存分析基本统计方法、Cox 比例风险回归分析、统计表与统计图、观察性研究设计、实验性研究设计，其中“生存分析基本统计方法”“Cox 比例风险回归分析”两章，是新增加的内容。此外本教材采用二维码形式，给出了教材中所有例题数据、各章练习题答案、自测试题答案和样本含量估计方法。

作者在写作过程中，参考了大量的资料，涉及医学统计学、医学统计学习题集、SPSS 软件操作方法等方面的书籍，与这些书籍相比，本教材具有以下特点：第一，在内容安排上注意与医学科研实际相结合，注意统计知识的整体性与前后连贯性，将科研统计设计(重点是统计设计)、数据管理与质量控制、数据统计分析几个步骤进行有机结合，强调数据管理与数据质量的必要性。第二，教材系统介绍医学统计学的基本概念、基本原理与基本方法，实用性强。重点在于，什么样的问题采用怎样的统计设计；什么样的实际数据，采用怎样的统计分析方法；以及如何对统计分析结果进行合理的解释。第三，结合 SPSS 20.0 统计软件窗口式操作简单、方便的特点，为学习者节省大量统计计算工作量和时间，从而将学习重点转移到对统计的三基的理解，而非统计公式的具体使用与计算。第四，注重统计学方法的适用性与通用性，并将之与现代统计学的理论相结合，如介绍有序列联表的假设检验方法、多因素统计分析模型、非参数的多重比较、样本含量估计方法等内容。第五，本教材后面的附录部分给出了各章练习题答案以及自测试题答案，为课堂教学和自学提供了方便。第六，教材中所有例题数据、各章练习题答案、自测试题答案和样本含量估计方法，采用了二维码形式，方便使用。

在本教材即将问世之际,我首先感谢首都医科大学有关领导对本教材编写工作的关心与指导;感谢各高校同仁参加本教材的编写;感谢学术秘书闫宇翔副教授为本教材做的大量而繁杂的具体工作;同时,感谢我的研究生刘梦阳、刘相佟、黄芳芳、胥芹、陈斯鹏、杨昆、李海彬、孙扬、田思佳、马圆等同学对本教材的所有例题、SPSS 20.0 统计软件操作进行的复核,并认真校对和排版书稿。最后,还要感谢我的丈夫和我可爱的女儿对我的理解和支持!

限于我们的学识和精力,本教材的缺点在所难免,恳请广大读者批评指正,给我们提出宝贵意见 (Email:guoxiuh@ccmu.edu.cn),以便再版时及时改正,谢谢!

郭秀花

2017年2月于北京

# 第一版前言

医学统计学是统计学原理和方法与医学相结合的一门应用科学，是高校医学专业的必修课，是进行医学科学研究不可缺少的一门方法学课程。目前虽然有各种版本的医学统计学或统计软件操作手册方面的书籍，然而，医学统计学方面的教材偏重于理论，详细介绍原理、公式和计算过程，缺少直观而详细的统计软件操作方法；大多数医学统计学软件操作手册又偏重软件介绍，缺乏统计学基本概念与理论。我们编写的教材，是将医学统计学方法与常用的 SPSS 18.0 统计软件操作结合起来，以弥补目前高校医学统计学教材的缺憾。

编写本教材是本着以下原则：第一，内容的科学性为主，兼顾理论的前瞻性。正确阐述医学统计学学科的科学理论和概念定义，在理论联系实际，以实例解释理论，对实践起到指导作用的基础上，注意将本领域的最新发展成果，把新技术、新方法纳入教材；第二，把握好写作条理性。注重教材的层次分明、条理清楚，教材体系能反映内容的内在联系及统计学的思维方式；第三，以本科学生或同等水平阅读能力群体为主要对象，并兼顾了课时数少的研究生医学统计学教学。从认知规律出发，富有启发性，便于学生学习，所选教学内容可以满足学生未来职业活动所需的最基本、最常用的理论知识和方法。第四，突出实践技能，强化应用。注重科研实际案例引入，统计方法计算以 SPSS 18.0 软件操作和结果解释为主，使学生真正掌握实践操作技能。

本教材共分为十六章：绪论、数据管理与 SPSS 统计软件简介、定量资料的统计描述、定量资料的参数估计与假设检验基础、定量资料的  $t$  检验、定量资料的方差分析、定量资料的非参数检验、定性资料的统计描述、定性资料的参数估计与  $\chi^2$  检验、有序定性资料的统计分析方法、直线相关与回归、多重线性回归分析、Logistic 回归分析、统计表与统计图、观察性研究设计、实验性研究设计。本教材后的附录部分有统计用表、各章练习题答案要点与综合测试题、英汉统计名词对照。本教材的作者在写作过程中，参考了大量的资料，涉及医学统计学、医学统计学习题集、SPSS 软件操作方法等方面的书籍，与这些书籍相比，本教材具有以下特点：第一，在内容安排上注重与医学科研实际相结合，注意统计知识的整体性与前后连贯性，将科研统计设计（重点是统计设计）、数据管理与质量控制、数据统计分析几个步骤进行有机结合，强调数据管理与数据质量的必要性。第二，全书系统介绍医学统计学的基本概念、基本原理与基本方法，实用性强。重点在于，什么样的问题采用怎样的统计设计；什么样的实际数据采用怎样的统计分析方法；如何对统计分析结果进行合理的解释；等等。第三，结合 SPSS 18.0 统计软件窗口式操作简单、方便的特点，为学习者节省了大量的统计计算工作量和时间，从而将学习重点转移到对统计的三基的理解，而非数据公式的具体使用与计算。第四，注重统计学方法的适用性与通用性，并将之与现代统计学的理论相结合，如介绍有序列联表的统计学分析、多因素统计分析模型、非参数的多重比较等内容。第五，本书后面的附录部分除了一般医学统计学书中给出的统计用表、关键词语的中英文对照外，还给出了各章练习题答案以及综合测试题，为课堂教学和自学提供了方便。

在本教材即将问世之际，我们首先感谢首都医科大学有关校领导、教务处各位领导对本教材编写工作的关心与指导；感谢全国高校素质教育教材研究编审委员会刘思祺主任、中国教师发展基金会教师出版专项基金办公室邱巍主任，对本教材出版工作给予的大力支持；感谢各高校同仁参加本书的编

写;感谢首都医科大学公共卫生与家庭医学学院王崑院长、郭爱民书记对出版本书提供的诸多指导;感谢学术秘书闫宇翔副教授为本书做了大量而繁杂的具体工作;同时,感谢我的研究生陶丽新、霍达、孙涛、潘蕾、周涛等同学对本书的所有例题进行了复核,并认真校对和排版书稿。最后,还要感谢我的丈夫和我可爱的女儿对我的理解和支持!

本教材供临床、护理、检验、药学等专业本科生,开设 30~60 课时的医学统计学教学使用;也可供各专业研究生开设 60 课时以内的医学统计学教学使用。虽然我们力求在编写内容、体例、实用等方面有新的创新与突破,但限于我们的学识和精力,本书的缺点在所难免,恳请广大读者批评指正(E-mail:guoxiuh@ccmu.edu.cn),以便再版时改正。

郭秀花

2012 年 1 月于北京



# 目 录

第二版前言

第一版前言

## 第1章 绪 论

1

- 1.1 医学统计学概述 / 1
- 1.2 医学统计工作的基本步骤 / 2
- 1.3 医学统计学中的几组基本概念 / 4
- 1.4 实验设计基本概念 / 6
- 1.5 统计软件简介 / 7
- 小结 / 10
- 练习题 / 10

## 第2章 数据管理与SPSS软件实现方法

12

- 2.1 数据管理 / 12
- 2.2 质量控制 / 13
- 2.3 数据库和数据管理软件 / 14
- 2.4 SPSS 数据库与数据管理 / 16
- 小结 / 30
- 练习题 / 30

## 第3章 定量资料的统计描述

32

- 3.1 频数分布表与分布图 / 32
- 3.2 平均数 / 34
- 3.3 变异指标 / 38
- 3.4 正态分布及其应用 / 41
- 3.5 SPSS 软件实现定量资料的统计描述方法 / 48
- 小结 / 55
- 练习题 / 55

## 第4章 定量资料的参数估计与假设检验基础

58

- 4.1 抽样与抽样误差 / 58

- 4.2  $t$  分布 / 60
- 4.3 均数的参数估计 / 61
- 4.4 假设检验基础 / 64
- 4.5 区间估计的 SPSS 软件实现方法 / 66
- 小结 / 70
- 练习题 / 70

## 第 5 章 定量资料的 $t$ 检验

72

- 5.1 单样本定量资料的  $t$  检验 / 72
- 5.2 配对设计定量资料的  $t$  检验 / 73
- 5.3 两独立样本均数比较的  $t$  检验 / 75
- 5.4  $t$  检验注意事项 / 77
- 5.5  $t$  检验的 SPSS 软件实现方法 / 79
- 小结 / 90
- 练习题 / 90

## 第 6 章 定量资料的方差分析

94

- 6.1 方差分析的基本思想和应用条件 / 94
- 6.2 完全随机设计资料的方差分析 / 97
- 6.3 随机区组设计资料的方差分析 / 98
- 6.4 多个样本均数的两两比较 / 100
- 6.5 析因设计资料的方差分析 / 102
- 6.6 重复测量设计资料的方差分析 / 105
- 6.7 方差分析的 SPSS 软件实现方法 / 109
- 小结 / 130
- 练习题 / 131

## 第 7 章 定量资料的非参数检验

135

- 7.1 配对设计的符号秩和检验 / 135
- 7.2 成组设计两样本比较的秩和检验 / 137
- 7.3 成组设计多个样本比较的秩和检验 / 139
- 7.4 多个组间的多重比较 / 141
- 7.5 SPSS 软件实现定量资料非参数检验方法 / 142
- 小结 / 148
- 练习题 / 148

## 第 8 章 定性资料的统计描述

151

- 8.1 相对数的概念与计算 / 151
- 8.2 动态数列 / 152

- 8.3 率的标准法 / 154
- 8.4 应用相对数注意事项 / 157
- 小结 / 158
- 练习题 / 158

## 第 9 章 定性资料的参数估计与 $\chi^2$ 检验

162

- 9.1 总体率的估计 / 162
- 9.2 四格表资料的  $\chi^2$  检验 / 163
- 9.3 行×列表资料的  $\chi^2$  检验 / 168
- 9.4  $\chi^2$  检验方法的 SPSS 软件实现 / 171
- 小结 / 180
- 练习题 / 181

## 第 10 章 有序定性资料的假设检验方法

184

- 10.1 单向有序行×列表数据的假设检验 / 184
- 10.2 双向有序属性相同行×列表数据的假设检验 / 188
- 10.3 双向有序属性不同行×列表数据的假设检验 / 189
- 10.4 有序定性资料假设检验的 SPSS 软件实现 / 191
- 小结 / 197
- 练习题 / 197

## 第 11 章 直线相关与回归

200

- 11.1 直线相关 / 200
- 11.2 直线回归 / 203
- 11.3 直线相关与回归的区别与联系 / 208
- 11.4 直线相关与回归分析的 SPSS 软件实现方法 / 209
- 小结 / 213
- 练习题 / 214

## 第 12 章 多重线性回归分析

216

- 12.1 多重线性回归的数据结构和前提条件 / 216
- 12.2 多重线性回归的参数估计及假设检验 / 217
- 12.3 SPSS 软件实现多重线性回归方法 / 219
- 小结 / 223
- 练习题 / 224

## 第 13 章 Logistic 回归分析

228

- 13.1 Logistic 回归的数据结构 / 228

- 13.2 Logistic 回归模型的参数估计及假设检验 / 231
- 13.3 Logistic 回归的 SPSS 软件实现方法 / 235
- 小结 / 240
- 练习题 / 240

## 第 14 章 生存分析基本统计方法

243

- 14.1 基本概念 / 243
- 14.2 生存率估计 / 244
- 14.3 Log-rank 检验 / 247
- 14.4 生存分析基本统计的 SPSS 软件实现方法 / 250
- 小结 / 255
- 练习题 / 255

## 第 15 章 Cox 比例风险回归分析

257

- 15.1 Cox 比例风险回归的数据结构 / 257
- 15.2 Cox 回归的参数估计及假设检验 / 259
- 15.3 Cox 回归分析的 SPSS 软件实现方法 / 261
- 小结 / 262
- 练习题 / 263

## 第 16 章 统计表与统计图

266

- 16.1 统计表 / 266
- 16.2 统计图 / 267
- 16.3 绘制统计表与统计图的 SPSS 软件实现方法 / 271
- 小结 / 280
- 练习题 / 280

## 第 17 章 观察性研究设计

283

- 17.1 概论 / 283
- 17.2 问卷的设计技巧 / 286
- 17.3 抽样方法 / 291
- 17.4 观察性研究的质量控制 / 292
- 小结 / 294
- 练习题 / 295

## 第 18 章 实验性研究设计

297

- 18.1 实验性研究概论 / 297
- 18.2 实验性研究设计种类 / 300

- 18.3 随机化实现方法 / 304  
18.4 临床试验简介 / 306  
小结 / 309  
练习题 / 309

### 附录一 统计用表

..... 311 .....

### 附录二 英汉统计名词对照

..... 328 .....

### 附录三 《医学统计学与 SPSS 软件实现方法》自测试题

..... 334 .....

- 《医学统计学与 SPSS 软件实现方法》自测试题(一) / 334  
《医学统计学与 SPSS 软件实现方法》自测试题(二) / 339

### 参 考 文 献

..... 344 .....

# 第1章 绪论

## 1.1 医学统计学概述

统计对我们每个人来说并不陌生,报纸杂志、电视广播、网络媒体等每时每刻都传递着很多统计数据和信息,我们也常听到很多关于“统计”方面的结果。例如,据统计,去年国民生产总值增长率为8.2%;某地人均寿命为78.6岁;2月份房屋销售量环比下降15.6%。还有许多问题需要应用统计学知识才能给出正确答案,例如,治疗艾滋病的新药有效吗?明年中国股市涨跌走势如何?体育彩票中奖的概率是多少?子女像父母的程度有多大?目前居民对医疗改革政策的满意度有多高?可以说统计学知识和思维已经渗透到自然科学、社会科学以及人类生活的各个领域。在现代社会中,大到国家重大政策的制定,小到人们的日常生活,几乎都离不开统计学。

### 1.1.1 定义

**1. 统计学的定义** 在西方,统计学(statistics)一词,源于state(国家、情况),专指有关“国情”的学问,最初多用于文字记叙,后发展为数量比较,随着概率论思想和方法的引入,逐渐形成今天在理论与应用方面都已相当完备的独立学科。我国教育部1998年在《普通高等学校本科专业目录和专业介绍》中将统计学列为理学类一级学科。按照《教育部关于进行普通高等学校本科专业目录修订工作的通知》(教高〔2010〕11号)要求,2011年对《普通高等学校本科专业目录》的修订工作进入了公开征求意见阶段,2012年9月,正式出台的《普通高等学校本科专业目录(2012年)》,将统计学(编号0712)列为理学类一级学科。

那么,什么是统计?广义上是人们通常所遇见的任何以数字、表格与图形所表达的一个事实,狭义上是将统计作为一门学科。什么是统计学?统计学有其自身独有的知识体系和方法论。著名的《韦氏国际大辞典》中定义,统计学是“a science dealing with the collection, analysis, interpretation, and presentation of masses of numerical data”,即统计学是一门关于收集、分析、解释和表达数据的科学。

**2. 医学统计学的定义** 统计学与各个专业结合就形成了数十个学科分支,如社会统计学、经济统计学、人口统计学、心理统计学、遗传统计学等。统计学理论是随着人类社会生产需要而产生,同时也随着人类社会生产的发展而更新的,特别是近二十年来,统计学的理论方法和应用得到迅速发展,新的领域与统计学结合形成的新的分支如同雨后春笋般不断出现。如果把统计学应用在医学领域,即将统计学与医学相结合而形成的一个交叉科学,就形成了医学统计学,因此,医学统计学的定义就是用统计学的原理和方法来研究医学领域中不确定性现象规律性的一门学科。医学统计学是现在及未来一个世纪中最活跃、最有生命力的学科之一。

**3. 医学统计学应用现况** 现在,生物医学实验、临床试验、流行病学调查和公共卫生管理都要寻求统计学家的合作。医学科研基金申请要求有统计学家参与合作,申请书必须包含详尽的统计设计与分析;新药开发和报批必须依法执行统计学准则,递交统计分析报告;公共卫生项目的确立和验收,必须基于抽样调查的数据和完善的评价体系;医学杂志发布统计学指南,邀请统计学家审稿,严控论文的统计学缺陷。美国国立卫生研究院(National Institutes of Health, NIH)的基金申请体系中明确要求基金申请团队中有统计学家,并且在所立项中有统计学方面的内容。美国国家食品药品监督管理局(Food and Drug Administration, FDA)要求新药的研发试验中,必须有统计学家来指导研究的设计、数据的分析、报告的呈递等。总之,统计学思维和方法学已经渗透到医学研究和卫生决策之中。

但是,很多一线医学科研工作者对统计学的作用重视不够,突出表现为忽视医学科研设计、在统



计分析时盲目套用统计分析方法、对统计分析结果解释时轻描淡写,一笔带过。把统计学当成无关紧要的“修饰物”,严重影响了医学科研工作的科学性与严谨性。由于轻视或误用统计学而导致得出错误结论的例子并不鲜见。2002年11月9日《科学时报》登载了军事医学科学院情报研究所胡良平教授公布的一个令人触目惊心的数据结果:全国各类医学期刊中,有统计学错误的论著竟占到80%。2001年西班牙赫罗约大学的 Emili Garcia-Berthou 和 Carles Alcaraz 查阅了 *Nature* 上发表的 181 篇论文,发现 38% 的文章至少有一处有统计学错误。2005 年 *Nature Medicine* 发表过一篇社论,题目为 *Statistically significant*,一开头就指出“*Nature* 和 *Nature Medicine* 因为登载的某些文章统计分析欠佳而遭到公众批评”。

### 1.1.2 怎样学好医学统计学

许多学生习惯于传统的医学统计教学模式,往往是“填鸭灌输式”或“知识继承型”的教学方法,教师在上面讲,学生在下面听,忙于记笔记,死记硬背应付考试,以后科研中遇到统计学问题还是束手无策或误用滥用。因此,在如何学好医学统计学上我们提出如下建议。

**1. 培养严谨、科学的态度** 在医学科研中应用统计学的目的是要探究客观事物的规律性,提出检验或验证科学问题。当应用统计学处理实际资料得不到理想的结果(或阳性结果)时,有的人就拼凑数据甚至修改数据,这是严重违背统计学主旨的,也是严重的学术造假行为。我们要遵从客观事实、认真分析原因,例如,各种因素是否考虑全面了?研究对象的选取是否合理?样本量是否足够大?指标选取的如何?收集资料的方法是否可靠?统计方法应用是否有误?统计计算是否正确?如果各个环节都没有问题,也许是我们最开始从专业上提出的科学问题就应该是这样的结论。学习医学统计学,就是要培养严谨、科学的态度。

**2. 抓住三基,即基本概念,基本原理,基本方法** 在学习过程中,对复杂公式本身及其推导过程不做硬性要求,学生只需要了解其作用,而不必死记硬背其具体的形式,也不必深究其数学原理。医学科学研究所应用的统计学知识中约 70% 是最基本概念和经典的统计分析方法,其余则是较为复杂的、近代发展起来的统计理论和技术,而出现错误最多的却偏偏是前一部分。

**3. 重视统计应用,把实际问题转化为统计问题** 学统计时一定要结合实例,最好从问题的原型入手,将其转化成统计问题,这是正确使用统计学的关键一步。根据设计类型、资料性质和分析目的,选择合适的统计分析方法进行资料处理。要经过从理论到实践、再从实践到理论的反复过程,循序渐进,才能逐渐掌握统计学,进而运用统计学解决实际问题时,才能得心应手。能否把各种实际问题转化为统计问题,能否合理选用统计学方法、正确运用统计学的理论和方法解决实际问题,是学好医学统计学的难点所在,也是衡量医学统计学教学质量的“金标准”。

**4. 熟练掌握统计软件的使用** 目前可以用来进行数据分析的统计软件很多,如: SAS, SPSS, Stata, R 语言等。在解决实际问题时,要重视各种检验方法适用的前提条件及应用场合,可以忽略其具体的计算推导过程;要熟练地掌握一种统计软件(如最简单、直观的操作软件 SPSS),学会正确使用统计软件和正确选择统计方法,对软件输出结果及统计学结果作出正确解释。随着现代统计学和计算机技术的迅猛发展,一些新的统计学方法和技术逐渐成熟并得到广泛应用,统计软件的功能也日益强大,并促使医学研究向更深和更广的方向发展。

## 1.2 医学统计工作的基本步骤

医学科学研究的全过程应有统计方法与统计工作相伴随,统计方法已广泛渗透到医学科研的各个环节,统计工作就是统计方法在医学科研中恰当和正确地使用。从统计学角度来说,统计工作的基本步骤是对科研项目进行设计、收集资料、整理资料和分析资料。

### 1.2.1 设计

设计(design)就是根据研究的问题与目的,从统计学的角度对各步提前作出的周密计划和安排是



科学研究工作的纲领和完成研究工作的关键环节。

设计包括专业设计和统计设计。专业设计是从专业角度考虑实验的科学安排,是科学研究的基础,包括选题、建立假说、确定研究对象和技术方法等;统计设计是在明确研究目的的前提下,从统计学角度对资料进行收集、整理和分析并提出全面具体的计划和要求,作为统计工作实施的依据,用尽可能少的人力、物力和时间获得准确可靠的结论。对于实验性研究的统计设计,根据研究目的制定研究方案,包括研究对象的纳入标准和排除标准、样本获取方法、实验与对照的分组、确定观察指标、实验过程中的质量控制和拟使用的统计方法等。对于观察性研究的统计设计,采用调查问卷或访问的方法,直接从某社会群体中收集资料,通过对资料的统计分析回答科研问题。

无论实验性研究还是调查性研究的统计设计,都强调如何获得符合研究目的的可靠研究资料,正确的整理资料过程和分析方法,使结果能很好地回答所研究的问题。具体内容应体现为:明确同质的研究对象;明确取得原始资料的方法;如何整理资料的过程;计算哪些指标;用何种统计推断方法及对结果的预测。例如,研究补钙对绝经期妇女骨密度的影响:研究对象为绝经1年以上、年龄为50~65岁,排除影响骨密度的相关疾病、手术史、服用过激素类药物等因素的绝经期妇女;通过调查表和干预实验获得原始资料;利用计算机建立数据库进行资料的整理;计算指标的均数、标准差和率等;统计推断方法主要是方差分析、 $\chi^2$ 检验和多因素回归分析等;结果预测为排除干扰因素后给予不同剂量钙绝经期妇女的骨密度有差别。

## 1.2.2 收集资料

收集资料(data collection)是获得研究所需原始数据的过程,要根据研究目的与设计确定。实验性研究收集资料主要是通过专项实验,如动物实验、临床观察实验;调查性研究收集资料主要是通过专题调查。无论何种途径收集到的资料,都应强调它的准确性、完整性。医学科学研究原始资料的来源如下。

**1. 报表资料** 医疗卫生领域里的各种报表,如传染病报表、疾病监测报表、医院年度统计报表、卫生统计年鉴等。例如,研究中国传染病的疾病负担,要收集几年内中国疾病预防控制中心或卫生部传染病的报告数据;对十年后中国卫生技术人员中医生和护士人数进行预测,要对近二十年的卫生统计年鉴里中国卫生技术人员中医生和护士人数进行收集,建立数据集进行预测。

**2. 医疗、预防机构的日常工作记录** 如住院病例、经常性工作记录和数据库等。疾病治疗质量评价的研究中,确定了评价指标后,要对医院住院患者的病例数据进行收集,利用统计分析方法进行影响因素调整和治疗质量评价。

**3. 专题研究的实验数据和调查资料** 如补钙对280名绝经期妇女骨密度影响的数据收集有两个部分:一是通过调查表调查绝经期妇女的一般情况、饮食情况、体育锻炼情况、生育史、心理健康与应对等调查资料;二是实验研究数据,实验分为4组,每组70人,信息干预组(只透露本人检查结果,不给干预措施),其余三组分别给A, B, C三种含不同剂量钙的奶粉,补钙一年后和两年后分别测其骨密度值、血液、尿液中的实验室指标值。

## 1.2.3 整理资料

整理资料(data sorting)是指对收集到的原始资料进行归类整理汇总的过程,即有目的地对收集到的原始资料进行科学加工,使资料系统化、条理化,以便进行统计分析。包括三方面的内容。

**1. 数据净化(data cleaning)** 对数据进行去伪存真的过程,即对原始数据进行检查、核对、纠错和改正。

**2. 逻辑检查(logical check)** 通过计算机对数据进行检查与核对的过程。根据逻辑关系、常识和专业背景知识,对所研究的资料进行检查与核对。对产生怀疑的数据,要进行深入核查并予以纠正。

**3. 统计核查(statistical check)** 为了进行统计分析,需要对原始数据进行加工,将其转化为频数分布表(frequency distribution)数据,可以根据数据间的关联性和频数分布图或表等进行核查。

整理资料主要步骤是审核资料、拟整理表和归纳汇总。在补钙对绝经期妇女骨密度的影响研究中,





根据补钙前和补钙一年及两年后的一般情况、饮食情况、体育锻炼情况、骨密度值及实验室等指标,建立 EpiData 数据库,采用双向比对的形式对数据进行录入、检查与核对。将数据导入 SPSS 分析软件内,再进一步对数据进行逻辑学检查。对数据进行粗加工、拟整理表、计算中间指标(如 BMI 指数)、进行归纳汇总等。

### 1.2.4 分析资料

分析资料(data analysis)就是对整理的资料进行统计分析,获取资料中有关信息的过程,包括统计描述(statistical description)和统计推断(statistical inference)两个方面。统计描述是通过计算有关的统计指标,对资料进行全面概括的描述,即统计指标的计算和统计图表的绘制。统计推断是从样本中的信息推断总体特征,包括两部分:一是参数估计(estimation of parameter),用样本统计量估计总体参数;二是假设检验(hypothesis test),用样本信息检验关于总体之间的差别。

例如,补钙对绝经期妇女骨密度影响的分析资料中:统计描述为计算骨密度的均值和标准差,根据腰椎骨密度  $T < -2.5$  为骨质疏松来计算患病率;绘制统计表和统计图。对调查数据的统计推断是以骨密度值为主要指标,用多重线性回归分析一般情况、饮食因素及其他因素与骨密度之间的关系;还可将人群分为骨质疏松症和非骨质疏松症两组,用 Logistic 回归分析骨质疏松症的相关危险因素。对实验干预数据的统计推断,按照重复测量设计,用方差分析研究三次骨密度值在不同干预分组之间及与时间变化的关系。

## 1.3 医学统计学中的几组基本概念

### 1.3.1 总体和样本

总体(population)是根据研究目的确定的,所有同质研究对象某一(组)指标值的集合。总体分为有限总体(finite population)和无限总体(infinite population),其中,同质研究对象个体数有限者,称为有限总体;在没有时间、空间限定情况下,同质研究对象个体数无限者称为无限总体。如研究某年某地 12 岁男孩身高状况,该年该地全部 12 岁男孩为研究对象,其身高值所构成的总体为有限总体。如研究某条江水中甲基汞含量,则全部江水为研究对象,这些江水中的甲基汞含量值所构成的总体则为无限总体。在医学科学研究的过程中,对于无限总体,不可能做到将总体中的个体都进行研究,对于有限总体,理论上即使能做到,但实际上没有必要对每一个体都进行研究。因此,医学科学研究常采用的是抽样研究。

样本(sample)是从总体中随机抽取的、数量足够的、能代表总体特征的部分研究对象某一(组)指标值的集合。样本一定要具备代表性和可靠性。上述实例中,如要了解该江水中甲基汞含量,可从该江的上中下游分别抽取一定数量的水样作为研究对象,所测水样的甲基汞含量构成样本,然后根据样本研究结果对江水中甲基汞含量进行推断。

统计学把描述总体特征的指标称为参数(parameter),描述样本特征的指标称为统计量(statistic)。医学科学研究的目的是由样本推断总体,依据统计量的特征或性质对总体参数作出推论,进而阐明总体的特征与规律。

代表性(representation)就是要求样本能够充分反映总体的特征。根据研究目的对总体有一个明确的规定后,样本必须是从总体中随机抽取出来,抽取样本的过程称为抽样(sampling)。随机(random)即需要保证总体中的每个个体有相同的可能性被抽出作为样本,要避免抽样的主观偏性。必须指出的是,随机化抽样绝不等于随意抽样。为了保证抽样的随机性,可用抽签法、机械抽样法、分层抽样法、随机数字表及计算机随机化抽样等方法。

可靠性(reliability)就是在总体范围内,样本数量要足够大。样本包含的个体数目称为样本含量(sample size)。由于个体之间存在差异,只有观察的样本含量达到一定数量才能体现出其客观规律性,样本量不足时,根据样本所下结论是不可靠的。样本含量越大,结论可靠性越大,但随着例数增加,

