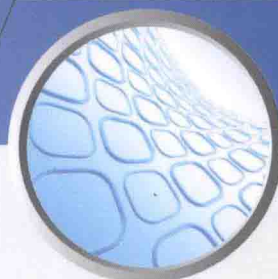


生物医学信息基础课系列教材

医学统计方法 及SPSS实现

主 编 陈 卉 李冬果

副主编 武文芳 华 琳



科学出版社

生物医学信息基础课系列教材

医学统计方法及 SPSS 实现

主 编 陈 卉 李冬果

副主编 武文芳 华 琳

编 者 王 宇 杨秋英

周 震 赵相坤

科 学 出 版 社

北 京

内 容 简 介

本书根据医学类高级专业人才培养目标的要求以及医学研究生学习和应用医学统计学的现状编写而成,主要介绍了医学统计方法及统计软件包在医学科研中的应用。本书从医学统计学的基本原理和方法入手,在详细介绍了各种医学科研中常用的统计学方法的基本思想、适用场合以及应用条件的基础上,结合实际医学科研案例,完整地介绍了统计方法的软件实现过程,包括数据文件的建立、统计方法选择、适用条件判断、统计方法实现、输出结果解释以及统计学结论和专业结论的获得,特别强调如何利用统计软件解决实际科研工作中遇到的统计学问题。本书以 IBM SPSS Statistics 21.0 版本为例介绍统计软件包的操作过程,其主要步骤也适用于 SPSS 的其他版本。

本书内容涉及医学科研中常用的统计分析方法,既可以作为高等医学院校硕士及博士研究生的医学统计学课程教材(建议 50~70 学时),也适用于医疗卫生系统科研工作者、临床医生进行统计分析的学习参考。

图书在版编目(CIP)数据

医学统计方法及 SPSS 实现/陈卉,李冬果主编. —北京:科学出版社, 2016.8

生物医学信息基础课系列教材

ISBN 978-7-03-049411-5

I. ①医… II. ①陈… ②李… 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(2016)第 166356 号

责任编辑:张中兴 王 刚 / 责任校对:钟 洋

责任印制:徐晓晨 / 封面设计:迷底书装

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>

北京京华虎彩印刷有限公司 印刷

科学出版社发行 各地新华书店经销

*

2016 年 8 月第 一 版 开本:720×1 000 B5

2016 年 8 月第一次印刷 印张:21 3/4

字数:422 000

定价:59.00 元

(如有印装质量问题,我社负责调换)

前 言

医学统计学对于医学博士研究生来说是一个十分重要、不可或缺的科研工具，它不仅在学生博士学习期间发挥着重要作用，也将一直伴随着学生今后的临床工作和基础研究。

本书从医学统计学的基本原理和方法入手，在详细介绍了各种医学科研中常用的统计学方法的基本思想、适用场合以及应用条件的基础上，结合实际医学科研案例，完整地介绍了统计方法的软件实现过程，包括数据文件建立、统计方法选择、适用条件判断、统计方法实现、输出结果解释以及统计学结论和专业结论的获得，特别强调如何利用统计软件解决实际科研工作中遇到的统计学问题。本书以 IBM SPSS Statistics 21.0 版本为例介绍统计软件包的操作过程，其主要步骤也适用于 SPSS 的其他版本。

全书共分 14 章，包括医学统计学基础、在 IBM SPSS Statistics 中管理数据、基本统计描述、抽样误差与 χ^2 -检验、定量资料的 t 检验、定量资料的方差分析、基于秩的非参数检验、定性资料的假设检验、直线相关与回归、二分类 Logistic 回归分析、生存分析、诊断试验评价和 Meta 分析简介。为了便于读者加深对统计学基本原理的理解，熟练软件的操作过程，本书各章均配有适量的思考题和上机操作题。

本书思路清晰，内容翔实，图文并茂，案例丰富，实用性强，适合作为基础医学、临床医学等医学专业硕士和博士研究生统计学相关课程的教材，也可作为临床医务工作者在科研过程中进行统计分析的参考书。

本书的编写得到了首都医科大学和校生物医学工程学院领导的大力支持。参加本书编写的教师均多年从事研究生统计课教学，积累了十余年博士研究生统计学课程的教学经验与体会。本书由陈卉、李冬果负责策划，具体编写工作如下：第 1 章由武文芳编写，第 2 章由王宇、杨秋英编写，第 3 章由周震、赵相坤编写，第 4~12 章由陈卉编写，第 13 章由华琳编写，第 14 章由李冬果编写，全书由陈卉统稿。朱碧云同学、陈婕卿同学对全书进行了认真的排版和校对，并验证了所有例题。感谢科学出版社的编辑，他们的认真负责和热情帮助使本书得以顺利出版。

由于编者水平有限，书中难免有一些不妥之处，欢迎读者批评指正。

编 者

2016 年 3 月

于首都医科大学

目 录

第 1 章	医学统计学基础	1
1.1	医学统计工作的基本步骤	1
1.1.1	统计设计	1
1.1.2	收集资料	1
1.1.3	整理资料	1
1.1.4	分析资料	2
1.2	医学统计学的几个基本概念	2
1.2.1	同质与变异	2
1.2.2	总体与样本	2
1.2.3	误差	3
1.2.4	频率与概率	3
1.2.5	参数与统计量	4
1.3	医学实验设计	4
1.3.1	实验设计的原则	4
1.3.2	常用的实验设计方案	5
1.4	统计资料的类型	6
1.5	练习题	6
第 2 章	在 IBM SPSS Statistics 中管理数据	8
2.1	IBM SPSS Statistics 简介	8
2.1.1	SPSS 的主要特点	8
2.1.2	SPSS 窗口介绍	8
2.2	建立数据文件	12
2.2.1	设计变量	13
2.2.2	定义变量	14
2.2.3	录入数据及保存	17
2.2.4	导入 Excel 文件数据	17
2.3	数据整理	19
2.3.1	记录排序	19
2.3.2	重复记录识别	20
2.3.3	记录筛选	21

2.3.4	数据文件合并	24
2.4	数据转换	25
2.4.1	计算变量	26
2.4.2	重新编码	27
2.4.3	可视化分箱	29
2.5	练习题	31
第 3 章	基本统计描述	32
3.1	定量资料的统计描述	32
3.1.1	集中趋势	32
3.1.2	离散趋势	33
3.1.3	正态分布及其应用	35
3.2	定性资料的统计描述	38
3.2.1	常用的相对数指标	39
3.2.2	应用相对数的注意事项	39
3.3	SPSS 的统计描述模块	40
3.3.1	Frequencies 过程	40
3.3.2	Descriptives 过程	42
3.3.3	Explore 过程	42
3.3.4	Crosstabs 过程	45
3.4	练习题	48
第 4 章	抽样误差与假设检验	49
4.1	抽样与抽样误差	49
4.2	总体参数估计	51
4.2.1	t 分布	51
4.2.2	总体均数的置信区间	52
4.2.3	总体率的置信区间	54
4.3	假设检验	55
4.3.1	假设检验的意义	55
4.3.2	假设检验的基本原理和步骤	56
4.3.3	应用假设检验的注意事项	57
4.4	样本量的估计	59
4.4.1	影响样本量大小的因素	59
4.4.2	样本量的计算公式	60
4.5	练习题	62

第 5 章	定量资料的 t 检验	64
5.1	单样本 t 检验	64
5.1.1	单样本 t 检验的基本思想	64
5.1.2	单样本 t 检验的 SPSS 实现	65
5.2	配对 t 检验	68
5.2.1	配对 t 检验的基本思想	68
5.2.2	配对 t 检验的 SPSS 实现	70
5.3	独立样本 t 检验	73
5.3.1	独立样本 t 检验的基本思想	73
5.3.2	独立样本 t 检验的 SPSS 实现	75
5.4	练习题	78
第 6 章	定量资料的方差分析	80
6.1	方差分析的基本思想和应用条件	80
6.1.1	方差分析的基本思想	80
6.1.2	方差分析的应用条件	82
6.2	完全随机设计资料的方差分析	83
6.2.1	单因素方差分析	83
6.2.2	组间均数的多重比较	84
6.2.3	单因素方差分析的 SPSS 实现	85
6.3	随机区组设计资料的方差分析	89
6.3.1	随机区组方差分析	89
6.3.2	随机区组方差分析的 SPSS 实现	90
6.4	析因设计资料的方差分析	96
6.4.1	析因设计方差分析	96
6.4.2	析因设计方差分析的 SPSS 实现	98
6.5	重复测量资料的方差分析	104
6.5.1	重复测量方差分析	105
6.5.2	重复测量方差分析的 SPSS 实现	106
6.6	练习题	113
第 7 章	基于秩的非参数检验	116
7.1	非参数检验概述	116
7.2	配对样本的秩和检验	117
7.2.1	Wilcoxon 符号秩检验	117
7.2.2	配对秩和检验的 SPSS 实现	118
7.3	两个独立样本的秩和检验	121

7.3.1	Mann-Whitney U 检验	121
7.3.2	两个独立样本秩和检验的 SPSS 实现	122
7.4	多个独立样本的秩和检验	126
7.4.1	Kruskal-Wallis 检验	127
7.4.2	多个样本的组间多重比较	128
7.4.3	多个独立样本秩和检验的 SPSS 实现	128
7.5	随机区组资料的秩和检验	136
7.5.1	Friedman 检验	136
7.5.2	随机区组秩和检验的 SPSS 实现	138
7.6	练习题	143
第 8 章	无序定性资料的 χ^2 检验	145
8.1	χ^2 检验的思想	145
8.1.1	χ^2 分布	145
8.1.2	χ^2 检验的基本思想	145
8.2	单样本 χ^2 检验	146
8.2.1	单样本 χ^2 检验的基本思想	146
8.2.2	单样本 χ^2 检验的 SPSS 实现	147
8.3	独立样本四格表资料的 χ^2 检验	150
8.3.1	独立样本四格表 χ^2 检验的基本思想	150
8.3.2	独立样本四格表 χ^2 检验的 SPSS 实现	152
8.4	行×列表资料的 χ^2 检验	155
8.4.1	行×列表资料 χ^2 检验的基本思想	155
8.4.2	行×列表资料 χ^2 检验的多重比较	156
8.4.3	行×列表资料 χ^2 检验的 SPSS 实现	157
8.5	配对四格表资料的 χ^2 检验	160
8.5.1	配对四格表资料 χ^2 检验的基本思想	160
8.5.2	配对四格表资料 χ^2 检验的 SPSS 实现	161
8.6	练习题	163
第 9 章	直线相关与回归	165
9.1	线性相关分析	165
9.1.1	线性相关的概念	165
9.1.2	相关系数	166
9.1.3	等级相关	169
9.1.4	线性相关分析的注意事项	170
9.1.5	相关分析的 SPSS 实现	170

9.2	线性回归分析	173
9.2.1	线性回归方程	173
9.2.2	回归系数的统计推断	174
9.2.3	线性回归分析的应用条件	176
9.2.4	应用线性回归分析的注意事项	177
9.2.5	线性回归分析的 SPSS 实现	177
9.3	多元线性回归分析	181
9.3.1	多元线性回归模型及其假设检验	181
9.3.2	自变量筛选	183
9.3.3	应用多元线性回归的相关问题	185
9.3.4	多元线性回归分析的 SPSS 实现	185
9.4	练习题	189
第 10 章	二分类 Logistic 回归分析	192
10.1	Logistic 回归模型简介	192
10.2	Logistic 回归的参数估计与假设检验	194
10.2.1	Logistic 回归的参数估计	195
10.2.2	Logistic 回归方程的假设检验	195
10.2.3	Logistic 回归系数的假设检验	196
10.3	Logistic 回归分析的实际应用	197
10.4	Logistic 回归分析中的若干问题	198
10.4.1	基本应用条件	198
10.4.2	样本含量估计	198
10.4.3	对自变量的预处理	199
10.4.4	自变量筛选	200
10.4.5	回归模型的评价	201
10.5	Logistic 回归分析的 SPSS 实现	202
10.5.1	示例一	202
10.5.2	示例二	210
10.6	练习题	216
第 11 章	生存分析	218
11.1	生存分析概述	218
11.1.1	生存分析中的基本概念	218
11.1.2	生存分析的主要内容及方法	220
11.2	描述生存过程	220
11.2.1	Kaplan-Meier 法	220

11.2.2	寿命表法	223
11.2.3	描述生存过程的 SPSS 实现	224
11.3	比较生存曲线	229
11.3.1	Log-rank 检验的基本思想	230
11.3.2	Log-rank 检验的 SPSS 实现	230
11.4	分析生存时间的影响因素	232
11.4.1	Cox 回归模型	232
11.4.2	Cox 回归分析的 SPSS 实现	233
11.5	练习题	236
第 12 章	诊断试验评价	237
12.1	诊断试验的准确性评价	237
12.1.1	诊断试验准确性评价的常用指标	237
12.1.2	诊断试验准确性评价应注意的问题	241
12.2	ROC 曲线及其应用	242
12.2.1	ROC 曲线的含义	242
12.2.2	绘制 ROC 曲线	243
12.2.3	计算 ROC 曲线下面积	244
12.2.4	比较 ROC 曲线下面积	245
12.2.5	ROC 曲线分析的 SPSS 实现	246
12.3	诊断试验的一致性评价	254
12.3.1	定性指标的一致性评价	254
12.3.2	定量指标的一致性评价	258
12.4	练习题	265
第 13 章	Meta 分析简介	269
13.1	系统评价与 Meta 分析	269
13.2	Meta 分析的过程	269
13.2.1	确定效应量	269
13.2.2	异质性分析	271
13.2.3	合并效应量	272
13.2.4	识别发表性偏倚	274
13.3	Meta 分析软件及实例	275
13.3.1	Meta 分析常用软件	275
13.3.2	二分类资料的 Meta 分析	277
13.3.3	连续型资料的 Meta 分析	281

第 14 章 绘制统计图	288
14.1 统计图概述	288
14.1.1 统计图的构成	288
14.1.2 常用的统计图	289
14.2 生成统计图	292
14.2.1 绘制条图	293
14.2.2 绘制线图	296
14.2.3 绘制散点图	298
14.2.4 绘制直方图和箱式图	301
14.2.5 绘制饼图和百分条图	304
14.3 编辑统计图	306
14.3.1 编辑条图	306
14.3.2 编辑线图	310
14.3.3 编辑散点图	313
14.4 应用 GraphPad Prism 绘图	317
14.4.1 GraphPad Prism 简介	317
14.4.2 绘制和编辑统计图形	319
14.4.3 导出图形	331
14.5 练习题	332
附录 部分练习题参考答案	335

第 1 章 医学统计学基础

医学统计学 (medical statistics) 是应用概率论和数理统计的基本原理和方法, 结合医学研究领域的实际问题, 研究资料的搜集、整理、分析与推断的方法学, 是认识医学现象数量特征的重要工具。本章就医学统计工作中的一些基本概念进行介绍。

1.1 医学统计工作的基本步骤

医学统计工作的全过程可分为统计设计、搜集资料、整理资料和分析资料 4 个步骤。它们紧密相连、缺一不可, 任何一步的缺陷都将影响整个研究结果。

1.1.1 统计设计

统计设计 (design) 是指根据明确的研究目的, 从统计学的角度对资料的搜集、整理和分析过程制定一个周密、具体的要求和计划。因此, 统计设计是整个统计工作过程中最关键的一环, 也是后面 3 个步骤实施的依据。

在整个统计设计中, 有关收集资料的计划是核心部分。它主要包括拟定研究方案、确定观察对象和观察单位、确定样本含量和抽样方法、制定实验过程中的质量控制措施、初步确定拟使用的统计方法等。对于医学实验性研究, 实验设计是研究过程中极其重要的一个环节, 它应遵循随机、重复、对照、均衡的原则, 具体介绍见 1.3 节。

1.1.2 收集资料

收集资料 (data collection) 是统计工作的基础, 是根据设计取得准确可靠的原始数据的过程。资料必须完整、准确和及时。医学研究中资料的来源有以下三方面。

- 统计报表: 如法定传染病报表、职业病报表、医院工作报表、卫生统计年鉴等。
- 经常性工作记录: 如卫生监测记录、医院住院患者记录、化验结果报告等。
- 开展专题研究获得的实验数据和调查资料。

1.1.3 整理资料

整理资料 (data sorting) 是对收集到的原始资料有目的地进行审核、校正、整理并使之系统化、条理化, 以便于进一步计算统计指标和分析。整理资料的过程包括:

- 对原始数据进行检查、核对、纠错和改正, 消除收集和录入数据的过程中可能出现的差错;

- 根据常识、专业知识、逻辑关系对资料的合理性和一致性进行检查与核对，如研究某种妇科疾病发生的影响因素，排查资料中是否有男性病例的记录；
- 根据统计分析的要求，将观察单位按类别（如性别、职业、疾病分类、婚姻状况等）或数值大小（如年龄、体重指数等）归类分组，并分组汇总资料。

1.1.4 分析资料

分析资料（data analysis）是按设计的要求，根据研究目的和资料类型，对整理出的基础数据作进一步的计算分析和统计处理，最后结合专业知识作出结论。经过统计分析，可以反映数据的综合特征，阐明事物内在联系和规律，得出科学结论。统计分析一般包括：

- 统计描述（descriptive statistics）：用统计指标与统计图（表）等方法对样本资料的数量特征及其分布规律进行描述；
- 统计推断（inferential statistics）：用样本信息推断总体特征，如推断总体均数，推断两个或多个总体某指标值是否存在差异等。

进行统计分析时，需根据研究目的、设计类型和资料类型选择恰当的描述性指标和统计推断方法。

1.2 医学统计学的几个基本概念

医学统计学涉及许多概念和原理，其中的几组基本概念是掌握统计学方法的基础。

1.2.1 同质与变异

同质（homogeneity）是指个体的观察指标所受的影响因素相同。这里所说的“相同”有时是相对的，也就是说观察单位所受的影响因素有时不可能达到绝对的一致。特别是在医学研究中，有些影响因素是难以控制、甚至是未知的，如遗传、心理等因素，这时只要保证主要影响因素相同或基本相同就可以认为是同质的。

变异（variation）是指在同质的基础上个体间的差异性，如相同年龄、性别、身高的正常儿童的体重不是都相同的，这种情况就称为体重的变异。对于人和其他生物体，变异是广泛存在的。尽管变异具有随机性和偶然性，但变异是有规律的。统计学正是探讨并利用变异规律性的科学，若无变异则无需统计。

1.2.2 总体与样本

总体（population）就是根据研究目的确定的、同质个体（研究对象）的全体，实际上是某一指标值的集合。样本（sample）从总体中随机抽取的部分个体，样本中个体的数目称为样本含量（sample size）。

由于直接研究总体通常是不可能或不必要的，因此实际工作中总是从总体中抽取

个体组成样本, 根据样本所提供的信息推断总体的特征。从总体中抽取的样本必须具有代表性、随机性和可靠性。

- 代表性 (representation): 样本要满足总体对个体同质性的规定, 属于所规定的总体的范围。
- 随机性 (randomization): 总体中任何一个个体都有同等机会被抽取进入样本, 样本中任何一个个体都有同等机会被分配到任何一个处理组中。
- 可靠性 (reliability): 组成样本的个体的数量即样本含量要足够大, 这时根据样本推断总体的结论才是可靠的。在实际工作中, 确定样本含量大小的原则是既要保证研究结果可靠, 又要避免不必要的浪费。

1.2.3 误差

统计学所说的误差 (error) 泛指测量值与真实值之差, 以及样本指标与总体指标之差。根据误差产生的原因和性质, 可以将误差分为 3 类。

1) 系统误差

系统误差 (systematic error) 是指由确定原因, 如测量仪器未校准、医生掌握判断标准偏高或偏低等, 所引起的观测值与真实值呈倾向性的偏差。由于系统误差影响了原始资料的准确性, 其产生原因是明确的, 因此在收集资料的过程必须消除系统误差。

2) 随机测量误差

随机测量误差 (random measurement error) 是由于非人为的、偶然因素造成的同一个体多次观测结果之间存在的没有固定倾向的偏差。对于这类误差应尽可能加以控制, 至少控制在一个允许的范围内。在实际工作中, 可以通过多次测量后计算平均值的方法减小甚至消除随机测量误差。

3) 抽样误差

抽样误差 (sampling error) 是指由于研究的对象是样本而非总体以及总体中个体存在变异而造成的样本指标与总体指标之间的差别。抽样误差是不可避免的, 但抽样误差是有规律的因而是可控制的。统计推断过程正是对抽样误差进行计算后由样本估计总体的。

1.2.4 频率与概率

频率 (frequency) 是指某事件出现的次数相对于总试验次数的百分比。随着试验次数无限增加, 某事件出现的频率在某一确定值附近摆动并无限趋近于这个值, 这个值就确定为该事件出现的概率 (probability)。在医学统计中常将频率简称为率, 如患病率、有效率等, 这时频率也称为统计概率。

概率是描述随机事件发生可能性大小的量度, 用符号 “ P ” 表示。概率的范围在 0 到 1 之间, 必然发生的事件其概率为 1, 不可能发生的事件概率为 0。某事件的发生概

率 P 越大, 该事件发生的可能性越大。在医学研究中, 通常将 $P < 0.05$ 或 $P < 0.01$ 的事件称为小概率事件。一般认为在一次随机抽样中小概率事件基本上不会发生, 这就是小概率事件原理。小概率事件原理是统计推断的一个重要原理。

1.2.5 参数与统计量

参数 (parameter) 是根据总体中个体值计算出来的描述总体特征的指标。参数一般用希腊字母表示, 如总体均数 μ 、总体方差 σ^2 、总体率 π 等。

统计量 (statistic) 是根据样本中个体值计算出来的描述样本特征的指标。统计量一般用英文字母表示, 如样本均数 $\bar{X}$ 、样本方差 S^2 、样本率 p 等。

1.3 医学实验设计

医学实验根据研究对象的不同分为动物实验 (animal experiment) 和临床试验 (clinical trial) 两类。它们分别以动物或生物材料和人为研究对象, 在研究过程中对研究对象施加干预或处理 (临床试验干预或处理必须遵从人道主义和医学伦理学原则), 观察不同干预的结果是否存在差异。

1.3.1 实验设计的原则

为了有效排除非处理因素影响、提高实验效率、节省人力物力, 在实验设计时必须遵守 3 个基本原则。

1) 随机化原则

随机化 (randomization) 是指在抽样时总体中任何一个个体都有同等的机会被抽取进入样本, 在分组时样本中任何一个个体都有同等的机会被分配到任何一个组中。随机化是保证非处理因素均衡一致的重要手段。实现实验设计随机化的方法有很多, 如抽签法、随机数字表法、机械抽样法等。

2) 重复原则

重复 (replication) 是指在相同的实验条件下进行多次观察, 即要求各处理组的观察单位都要有一定的数量。重复的原则实际上就是确定样本含量的原则, 根据研究目的和统计学知识估计样本含量, 做到既保证实验结果可靠, 又避免不必要的浪费。样本含量的估算方法见 4.4 节。

3) 对照原则

对照 (control) 是指在实验设计中设立与实验组进行比较的对照组, 通过比较反映出研究因素对实验结果的作用。常用的对照方法有空白对照、安慰剂对照、实验对照、标准对照和自身对照。

在设立对照时必须满足均衡性要求, 即除研究因素或施加的处理不同外, 对照组

与实验组中的非研究因素都应相同或相近,以消除非研究因素对实验效应的影响,从而分离出研究因素的效应。例如,在临床试验中,各治疗组患者的年龄、性别、病情、病程、病型等应基本一致。

1.3.2 常用的实验设计方案

根据研究的目的、期望获得的结果和现有的条件,可采用不同的实验设计方案。明确实验设计方案也是正确选择统计方法的前提。

1) 完全随机设计

完全随机设计(completely randomized design)是指将观察对象随机分配到各个处理组或对照组中,每个组接受一种处理,形成两个或多个样本。在实际设计时,将研究对象随机分配到各个实验组进行观察,或从不同人群中随机抽样进行对比观察都属于完全随机设计。例如,为研究和比较3种降转氨酶药物的效果,将60名肝功能异常患者随机分配到3个药物组中,服药一段时间后观察转氨酶指标的变化。

2) 配对设计

配对设计(paired design)是指将研究对象按照某些重要的非处理因素配成对子,每对中的两个研究对象随机分配到两个组中。配对设计分为自身配对设计,即对每个研究对象某种处理前后的观察、身体对称或不同部位的观察以及接受两种不同处理后的观察和异体配对设计,即将条件相同的两个研究对象配成对子并将它们随机分配到不同组。例如,为研究某药物是否有降压作用,有25名高血压患者服用该药,测量每名患者服药前及服药一段时间后的舒张压。

3) 随机区组设计

随机区组设计(randomized block design)又称随机配伍组设计,是指将研究对象按照可能影响实验结果的非处理因素配成区组(block),再将每个区组中的研究对象随机分配到处理组或对照组中。如果每个区组中只有两个研究对象则为配对设计。例如,比较3种抗癌药对小白鼠肉瘤的抑瘤作用,将45只染有肉瘤的小白鼠按体重大小配成15个区组,每个区组的3只小白鼠随机分配到3个药物组之一,观察服药后肉瘤重量的改变。

4) 重复测量设计

重复测量设计(repeated measurement design)是指对一组或多组受试对象,在多个不同的时间点上,对每个个体进行某一观察指标的多次重复观测,从而分析观察指标在不同时间点的变化趋势和特点。例如,观察某种药物对儿童过敏性鼻炎的治疗效果,观测50名病情相同的2~5岁患儿在服药前及服药后7天、14天、28天的鼻炎症状评分。

5) 析因设计

析因设计(factorial design)是指对两个或多个处理因素的各水平的所有组合进行实验,从而探讨各处理因素的主效应(main effect)、简单效应(simple effect)以及各

因素间的交互作用 (interaction effect)。例如, 观察甲、乙两种镇痛药物在产妇分娩时的镇痛效果, 每种药物各分为大、小两种剂量, 形成 4 个处理组, 将 60 名产妇随机等分到这 4 个组中, 记录分娩时的镇痛时间。

1.4 统计资料的类型

在对观察对象进行某项指标或特征的测量和观察时, 观察对象的这些指标或特征具有变异性, 因此我们将这样的指标或特征称为变量 (variable), 观察对象对变量的测量值称为变量值, 全部或部分变量值构成了研究资料 (data)。例如, 为比较糖尿病患者与当地非糖尿病患者的肥胖程度是否存在差异, 观察对象的年龄、性别、体重、身高、是否为糖尿病即是变量, 40 个观察对象的这些指标或特征的变量值构成研究资料。

根据变量的观察结果是定量的还是定性的, 可以将资料分为两大类型。

1) 定量资料

定量资料 (quantitative data) 是指对每个观察对象观测某个指标 (变量) 的具体数值而获得的资料, 如观察对象的体重、身高、白细胞计数等。

2) 定性资料

定性资料 (qualitative data) 又称分类资料, 是指对每个观察对象观测某个指标的属性或类别而获得的资料, 如观察对象的性别、是否患糖尿病等。定性资料又根据属性或类别之间是否有程度上的差别, 分为两种:

- 无序分类资料, 如血型分为 A、B、AB 和 O 型, 地区分为东北、华北、西北、华中、西南等。在进行统计分析时, 对于二分类资料, 如性别分为男和女、诊断结果分为良性和恶性、血压分为正常和异常等, 均按照无序分类资料对待。
- 有序分类资料, 也称等级资料, 如化验结果分为-、±、+、++和+++, 治疗效果分为退步、无效、好转和治愈等。

上述 3 种资料可以根据需要进行转换。如测定 50 个人的身高 (m) 和体重 (kg) 并计算体重指数 (kg/m^2), 这些都是定量资料; 将体重指数按 $18.5 \text{ kg}/\text{m}^2$ 、 $24 \text{ kg}/\text{m}^2$ 、 $28 \text{ kg}/\text{m}^2$ 划分, 可分为偏瘦、理想体重、超重和肥胖 4 类, 构成等级资料; 若将体重指数按 $24 \text{ kg}/\text{m}^2$ 划分, 可分为不胖和胖两类, 构成无序分类资料。通常情况下, 定量资料可以转换为定性资料, 而定性资料不能转换为定量资料。

1.5 练习题

一、思考题

1. 参数和统计量有什么区别和联系?
2. 从总体中进行抽样有哪几条基本原则?