

# 临床医学研究中的 统计分析和图形表达 实例详解 | 第2版

Step-by-step Examples of Data Analysis and  
Statistical Graphs in Clinical Researches

统计分析

量表分析

Meta分析

EpiData + SPSS + GraphPad + LISREL + RevMan

主审 马骏

主编 周登远

 北京科学技术出版社

# 临床医学研究中的统计分析和 图形表达实例详解

(第2版)

主 审 马 骏

主 编 周登远

---

## 图书在版编目(CIP)数据

临床医学研究中的统计分析和图形表达实例详解/周登远主编. —2版. —北京:北京科学技术出版社,2017.7

ISBN 978-7-5304-9030-3

I. ①临… II. ①周… III. ①医学统计—统计分析—应用软件 IV. ①R195.1-39

中国版本图书馆CIP数据核字(2017)第110359号

---

## 临床医学研究中的统计分析和图形表达实例详解

---

主 编:周登远

策划编辑:于庆兰

责任编辑:杨 帆

责任印制:李 茗

封面设计:永诚天地

出 版 人:曾庆宇

出版发行:北京科学技术出版社

社 址:北京西直门南大街16号

邮政编码:100035

电话传真:0086-10-66135495(总编室)

0086-10-66113227(发行部) 0086-10-66161952(发行部传真)

电子信箱:bjkj@bjkjpress.com

网 址:www.bkydw.cn

经 销:新华书店

印 刷:三河市国新印装有限公司

开 本:700mm×1000mm 1/16

字 数:587千字

印 张:27

版 次:2017年7月第1版

印 次:2017年7月第1次印刷

ISBN 978-7-5304-9030-3/R·2301

---

定 价:85.00元



京科版图书,版权所有,侵权必究。  
京科版图书,印装差错,负责退换。

# 编委会名单

- 主 审 马 骏 (天津医科大学公共卫生学院)
- 主 编 周登远 (天津市中医药研究院)
- 副主编 王红梅 (天津市中医药研究院附属医院)
- 崔 壮 (天津医科大学公共卫生学院)
- 朱嘉文 (江苏省无锡市疾病预防控制中心)
- 程继旺 (天津市中医药研究院附属医院)
- 编 者 (以姓氏拼音排序)
- 陈 阳 (天津市红桥区疾病预防控制中心)
- 丁 军 (天津市胸科医院)
- 房慧媚 (上海市宝山区中西医结合医院)
- 郭海霞 (天津市中医药研究院附属医院)
- 韩凤娴 (天津市中医药研究院附属医院)
- 贾真琳 (天津医科大学公共卫生学院)
- 焦振山 (天津市中医药研究院)
- 兰 岚 (天津医科大学护理学院)
- 李 妍 (天津市中医药研究院附属医院)
- 李 志 (天津市中医药研究院)
- 李晓莉 (天津市中医药研究院附属医院)
- 刘 栋 (天津市中医药研究院附属医院)
- 刘 飞 (军事医学科学院卫生学环境医学研究所)
- 罗 欣 (天津市中医药研究院附属医院)
- 穆新亚 (天津财经大学理工学院统计学系)
- 于 洋 (天津市胸科医院)
- 赵琳娜 (天津中医药大学第一附属医院)
- 赵林胜 (天津市儿童医院)
- 朱 君 (天津市中医药研究院附属医院)

# 再版前言

《临床医学研究中的统计分析和图形表达实例详解》自 2011 年上市以来,由于其简单、实用的特点,受到临床医生和医学科研工作者的欢迎,时隔 6 年,现进行修订改版。但是笔者发现,很多书改版之后,反而不如第 1 版精炼了,因为在第 2 版中,作者往往会在原有的章节内增加了许多比如公式、理论探讨的内容,书变厚了,但实质性的内容还是那些。因此,在编写《临床医学研究中的统计分析和图形表达实例详解》第 2 版时,就本着保持其简单、实用风格,同时又符合科研进展的宗旨,对原有的核心章节未大幅增加,仅增加了量表分析和 Meta 分析两个篇章。由此,这还是那本临床医生看得懂、用得上的数据处理参考书籍。以下介绍一下本书的特点。

## 统计分析(第一至三篇)

统计分析的完整分析流程包括:统计方法选择—统计分析—统计图形绘制。

- 统计方法选择:传统的统计学教科书是以统计理论为框架,从分布开始谈起,不断深入;而统计软件教程则往往以软件界面为框架,逐个菜单栏进行介绍。这两种框架在实际运用中会产生一些问题,即要面临这些数据资料“能进行哪些统计分析”的问题,以及研究者提出的研究假设“需要采用何种统计分析”。所以“能做什么”和“想做什么”成为研究过程中研究者最关心的事情。本书以数据资料(能做什么)为框架,在各种资料中又以研究假设(想做什么)作为分类标准,详尽介绍了计量资料、计数资料、生存资料和诊断试验这四类临床科研中最常用的统计方法。本书还单独列出一章“统计方法的选择”,以帮助大家更好、更迅速地选用恰当的统计方法。

- 统计分析:每个统计方法单独成节,采用 SPSS 中文版,由“方法原理—分析示例—研究假设—数据录入—操作流程—结果解释—注意事项”7 个步骤组成。读者只要按书中提示一步一步完成即可掌握该统计方法。

- 统计图表绘制:以往的书介绍统计图形习惯就图论图,逐一介绍了直方图、条图、饼图和线图等的绘制,但在实际应用中,对数据采用何种统计图形表达却

是个难题。本书则以统计方法为框架,将统计分析和统计图形联系起来,让大家迅速选用合适的统计图形。所有章节均由“示例—该示例的图形表达—图形类型的选择—数据录入—坐标轴和图形调整—文字修饰”6个步骤组成。可以帮助读者迅速绘制出出版发表级的统计图。

## 量表(问卷)分析(第四篇)

随着公众对生活质量、健康状况等综合而抽象概念的日益重视,量表(问卷)的使用日渐广泛,但是鲜见这方面的系统介绍,该篇章系统介绍了量表(问卷)的制作和分析。

- 网络量表(问卷)制作:使用问卷星网站,以一个实例来说明一个网络量表(问卷)的制作、分发和回收过程,研究者可以通过微信、QQ、邮件等多种方式发放和回收问卷,相对传统的纸质问卷,大大扩大了问卷发放范围。

- 问卷分析篇:介绍了有特色的多选题分析。

- 量表分析篇:一份成熟的量表,不管是自制量表还是国外翻译过来的量表,在正式使用之前都需要进行区分度、信度和效度的检验。本书以一个简单的实例介绍了区分度、内部信度、外部信度、探索性因子分析(以上几种分析采用 SPSS 软件)和验证性因子分析(LISREL 8.7)分析流程,让大家迅速了解并掌握量表分析的基本技能。

## Meta 分析(第五篇)

Meta 分析已经被科研工作者大量使用,但是最基本的 Meta 分析主要针对试验性研究。本书采用 RevMan 软件,以实例分别介绍了分类资料(结局指标为 RR、OR、RD)和连续资料(结局指标为 SMD、WMD)的分析步骤和结果解释。

本书以“一切为了应用”为出发点,深刻把握临床医生在科研中的实际需求,帮助大家迅速有效地掌握统计知识,打造出科研核心竞争力。

## 本书数据文件

作者信箱 [zhoudengyuan@live.cn](mailto:zhoudengyuan@live.cn)

周登远  
2017年1月

# 目 录

## 第一篇 预 备

第一章 统计学的基本概念 .....	(3)
第二章 统计方法的选择 .....	(6)
第一节 计量资料的统计方法选择 .....	(6)
第二节 计数资料的统计方法选择 .....	(9)
第三节 生存资料的统计方法选择 .....	(11)
第四节 诊断试验资料的统计方法选择 .....	(12)
第三章 数据文件的建立 .....	(13)
第一节 用 SPSS 建立数据文件 .....	(13)
第二节 EpiData 数据录入 .....	(19)

## 第二篇 统计分析(SPSS 中文版)

第四章 $t$ 检验 .....	(27)
第一节 单样本 $t$ 检验 .....	(27)
第二节 配对 $t$ 检验 .....	(30)
第三节 成组 $t$ 检验 .....	(33)
第五章 方差分析 .....	(38)
第一节 完全随机设计资料的方差分析 .....	(39)
第二节 随机区组设计资料的方差分析 .....	(46)
第三节 析因设计资料的方差分析 .....	(55)
第四节 重复测量资料的方差分析 .....	(60)
第五节 协方差分析 .....	(70)
第六章 秩和检验 .....	(80)

第一节	单样本秩和检验	(81)
第二节	配对样本秩和检验	(85)
第三节	两组独立样本秩和检验	(90)
第四节	多组独立样本秩和检验	(96)
<b>第七章</b>	<b>相关分析</b>	(102)
第一节	线性相关	(103)
第二节	秩相关	(108)
<b>第八章</b>	<b>线性回归分析</b>	(112)
第一节	简单线性回归分析	(113)
第二节	多重线性回归分析	(119)
<b>第九章</b>	<b>四格表卡方检验</b>	(126)
第一节	一般四格表卡方检验	(127)
第二节	配对四格表卡方检验	(132)
第三节	分层卡方检验	(137)
<b>第十章</b>	<b>列联表分析</b>	(144)
第一节	双向无序的列联表分析	(144)
第二节	单向有序的列联表分析	(151)
第三节	双向有序且属性不同的列联表分析	(156)
第四节	双向有序且属性相同的列联表分析	(163)
<b>第十一章</b>	<b>Logistic 回归</b>	(168)
第一节	非条件 Logistic 回归	(169)
第二节	条件 Logistic 回归	(177)
<b>第十二章</b>	<b>生存分析</b>	(187)
第一节	寿命表法	(189)
第二节	单因素生存曲线比较(Kaplan - Meier 法)	(194)
第三节	多因素生存分析(Cox 回归分析)	(201)
<b>第十三章</b>	<b>诊断实验的统计分析</b>	(211)
第一节	准确性检验(ROC 曲线)	(211)
第二节	一致性检验(Bland - Altman 图)	(218)

### 第三篇 统计绘图(GraphPad Prism 5)

<b>第十四章</b>	<b>统计图基本知识</b>	(231)
-------------	----------------	-------



第十五章	GraphPad Prism 5 绘图界面介绍 .....	(236)
第十六章	各种统计方法所对应的统计图绘制 .....	(243)
第一节	配对 $t$ 检验的图形绘制 .....	(243)
第二节	成组 $t$ 检验的图形绘制 .....	(247)
第三节	完全随机设计资料方差分析的图形绘制 .....	(251)
第四节	析因设计资料方差分析的图形绘制 .....	(255)
第五节	重复测量资料方差分析的图形绘制 .....	(259)
第六节	两组独立样本秩和检验的图形绘制 .....	(264)
第七节	多组独立样本秩和检验的图形绘制 .....	(268)
第八节	简单线性回归和线性相关的图形绘制 .....	(272)
第九节	列联表分析的图形绘制 .....	(278)
第十节	生存分析的图形绘制 .....	(284)
第十七章	统计图的排版与导出 .....	(292)

## 第四篇 量表(问卷)分析

第十八章	网络问卷(量表)的制作和回收 .....	(305)
第十九章	问卷分析 .....	(319)
第一节	多选题分析 .....	(319)
第二节	定向多选题分析 .....	(326)
第二十章	量表分析 .....	(331)
第一节	量表的基本概念 .....	(331)
第二节	区分度分析 .....	(335)
第三节	信度分析 .....	(340)
第四节	效度分析 .....	(345)
第五节	因子分析 .....	(347)

## 第五篇 Meta 分析

第二十一章	Meta 分析概述 .....	(367)
第二十二章	分类资料的 Meta 分析 .....	(371)
第一节	比值比 OR 的 Meta 分析 .....	(371)
第二节	相对危险度 RR 的 Meta 分析 .....	(379)

第三节 率差 RD 的 Meta 分析 .....	(387)
第二十三章 连续变量资料的 Meta 分析 .....	(395)
第一节 加权均数 WMD 的 Meta 分析 .....	(395)
第二节 标准均数差 SMD 的 Meta 分析 .....	(404)
附录 EpiData、SPSS、GraphPad Prism、LISREL、RevMan 介绍 .....	(414)
参考文献 .....	(420)

## 第一篇

# 预 备



## 第一章

# 统计学的基本概念

## 一、样本与总体

1. 总体(population) 是根据观察目的而确定的同质观察单位的全体,即同质的所有观察单位某种变量值的集合。

2. 样本(sample) 样本是总体中随机抽取部分观察单位实测值的集合。

科学研究一般是通过样本来推断总体特征,其做法是从研究总体中抽取少量有代表性的个体,称为**抽样**(sampling),对这些个体组成的**样本**(sample)进行深入的观察与测量,获取**数据**(data),利用统计知识,通过样本数据对研究总体的规律进行**推断**(inference)。

## 二、变异与同质

1. 变异(variation) 同质个体同指标之间的差异称为变异。

2. 同质(homogeneity) 指研究事物现象存在的共性。它是统计研究的基础,是资料整理和分析的前提。

## 三、变量的分类

**变量(variable)**:总体中的个体特征总是通过一个或多个变量来描述,变异性的存在决定了我们要处理的是变量。本书把变量分为定性(qualitative)和定量(quantitative)两种。

1. 定性变量 又分为**分类变量**和**有序变量**(等级变量)。

(1)**分类变量**(categorical variable):又称**名义变量**(nominative variable)。例如,职业是一个分类变量,其可能的“取值”不是数字,而是工、农、商、学、兵等,这些成为分类变量的水平(level),为便于输入计算机,一般采用代码(code)1、2、3、4、5等来表示各水平。最简单也是最常用的变量为**二分类变量**(binary variable),如性别(男女)、疾病(有无)和结局(生死)等。

(2)**有序变量**(ordinal variable):分类变量在种种可能的“取值”中自然地存在

着次序。例如,问卷调查中常问对某件事情的满意程度给出5个答案:极不满意、有点满意、中度满意、很满意、极满意。有些临床体检或实验室检验常用“-、±、+、++”和“+++”来表示测量结果。

2. 定量变量 又分为离散变量和连续变量。

(1)离散变量(discrete variable):离散型变量只能取整数值,如1个月中的手术患者数、一年里的新生儿数。

(2)连续变量(continuous variable):连续型变量可以取实数轴上的任何数值。有些变量的数值由测量而得到,它们大多属于连续变量,如血压、身高、体重等。而有一些测量值如红细胞数,虽然以“个”为单位时只能取整数值,但当数值很大而以“千”或“万”为单位时,又可以取小数值,所以通常把这些变量也称为连续变量。

有时为了数据分析的方便,人们将一种类型的变量转化为另一种类型,但变量只能由“高级”向“低级”转化:定量变量—有序变量—分类变量—二分类变量。

## 四、频率与概率

1. 频率(frequency) 指在相同条件下,进行有限 $n$ 次重复试验,某随机事件 $A$ 发生的次数 $x$ 与 $n$ 次试验的比值。频率是个变数,随样本变化而改变。

2. 概率(probability) 是描述随机事件 $A$ 发生的可能性大小的度量。概率是一个定值。假设在相同的条件下,独立进行 $n$ 次重复试验,随着 $n$ 充分增加,频率摆动的幅度越来越小,则该事件 $A$ 为随机事件,其频率可作为概率的估计值。

## 五、误差的分类

误差(error):可以分为随机误差和非随机误差。

1. 随机误差 又分为抽样误差和随机测量误差。

(1)抽样误差:由于产生的根本原因是生物个体的变异性,故抽样误差的分布具有规律性。

(2)随机测量误差:对同一观察单位的某项指标在同一条件下进行反复测量所产生的误差。

2. 非随机误差 又分为系统误差和过失误差。

(1)系统误差:可产生于设计人员、调查者或调查对象,也可由于考虑不当、汇总计算有误等造成,一般带有倾向性,其产生原因复杂,贯穿于研究的全过程并对研究结果有影响,很难用统计方法评价影响。

(2)过失误差:是错误,一般应杜绝。

## 六、统计分析的流程

1. 根据临床实践,提出研究问题,进行科研设计 在医学实践过程中提出科

研问题,然后围绕提出的研究问题制订研究方案,统计分析人员应当从设计阶段就参与研究项目,而不是临床医生获得数据之后才想到统计分析。医学研究一般有干预性研究(intervention study)和观察性研究(observational study)两种。医学干预研究是人们通过规定对象的准入条件(entry criteria)、随机化、重复、匹配(match)以及盲法(blinding)等措施来控制主要的混杂因素。公共卫生方面的研究大多属于观察性研究,这类研究不可能人为地控制许多混杂因素,人们能做的主要是观察已经或将要发生的事情。对于混杂因素的处理办法是精心设计抽样方法、无误地记录可能有用的信息。

2. 进行科学研究,分析清楚资料的性质,并分解出其观察与变量 研究方案出来后,需要严格按照研究方案进行,资料大体上可以分为计量资料和计数资料。计量资料指测定每个观察单位的某项指标量的大小所得的资料;而将观察单位按照某种属性或类别分组计数,所得的各组观察单位数称为计数资料。分清楚资料类型后,需要将资料分解成观察与变量。变量在临床上称为指标,是指具有相同属性的测量值的集合;而观察是指对同一观察对象的不同属性的集合就构成了一条观察;观察与变量结合起来就能准确地描述二维空间的物体特征。区分资料为计量资料还是计数资料,然后将资料分解为观察与变量,这是资料分析的基础。

3. 结合以上两点,罗列出能够回答该问题的可选统计方法 根据研究目的和资料性质,选用相应的统计学方法,如计量资料中的 $t$ 检验、方差分析、线性回归分析等,计数资料中的卡方检验、Logistic回归分析等。不同的资料和研究目的有多种可供选择的统计方法;而多种不同的统计方法可以对应多种资料类型、回答多种问题,如秩和检验能处理不符合方差分析条件的计量资料,也可以分析等级资料。

4. 选用统计软件,尝试着进行相关的统计分析 大家经常认为,只要形成了数据表格,选对了统计方法,用软件一操作,就万事大吉了。其实,事情没有这么简单,统计分析是一个反复的过程,是一个系统工程,需要进行预分析、正式分析等,如针对一份计量资料我们首先考虑进行方差分析,但是分析过程中发现方差不齐,我们可以改做秩和检验。如在Logistic回归过程中,可以选用全部进入法和逐步回归法,我们可以两者均尝试一下,然后比较两种方法结果的差异,再根据专业知识和分析目的做出判断,可见统计分析不是一锤定音、一成不变的过程,而是不断尝试、不断思考、不断判断的过程。

5. 评估统计结果,结合专业来回答提出的研究问题 从统计结论到专业结论,大家都需要特别慎重,不可妄加推断、任意发挥。

## 第二章

## 统计方法的选择

统计方法的选择是个非常具有艺术性的问题。同样的数据可以采用不同的统计方法,而不同的数据也可以采用同一种统计方法,因此我们需要在统计方法的选择中把握其根本:数据资料的性质决定“能做什么”,而研究设计或研究目的则决定了“想做什么”。本书将数据按性质分为计量资料和计数资料,对于两种资料又根据不同的研究设计或研究目的讨论其具体的统计方法。而生存分析和诊断试验两章则属于特定的分析方法,其数据特征和分析目的有其固定的特征,所以单独列出进行讨论。

需要特别说明的是,计量资料和计数资料只是对资料的一个通俗叫法,并不意味着计量资料中没有定性变量,或者计数资料中没有定量变量,只是我们测量和感兴趣的指标为定量变量(如每组病例的血压值)或定性变量(如每组中治愈的人数),我们就称之为计量资料或计数资料。下面就计量资料、计数资料、生存资料和诊断试验的统计学方法选择进行解释和讲解。

### 第一节 计量资料的统计方法选择

按照研究设计或研究目的的不同,计量资料可以分为三类。

- **成组设计**:其目的在于比较各组所代表的总体之间均数或中位数的差别,包括  $t$  检验、方差分析和秩和检验三种类型。
- **相关分析**:其目的在于研究两个变量之间联系的密切程度,又可分为线性相关和秩相关。
- **因果联系**:其目的在于探讨自变量和因变量之间的因果关系,因变量又称为结果变量,通常为身高、血压等连续变量;自变量又称解释变量,可以有多个自变量,自变量可以为连续变量、等级变量和分类变量,其分析方法称为线性回归分析。



## 一、成组设计

下表这组统计方法是成组设计中最简单的统计方法,也是应用非常广泛的统计方法。 $t$  检验中要求数据来自于正态总体,如果这一前提条件被违反,则采用对应的非参数检验(秩和检验)。后面介绍的方差分析和秩和检验可以视为该表的扩展。

单组或两组计量资料的统计方法

设计名称	前提条件是否满足及假设检验方法的选择	
	满足	不满足
单组设计	单样本 $t$ 检验	单样本秩和检验
配对设计	配对 $t$ 检验	配对样本秩和检验
成组设计	成组 $t$ 检验	两组独立样本秩和检验

在谈到复杂成组设计之前,有三个概念必须弄清楚。

- **因素 (factor)**: 因素指对测量结果可能有影响的变量。一般来说,因素会有不止一个水平,而分析的目的就是比较同一因素内各水平之间的测量结果是否相同。在方差分析中,一般有一个或多个因素。

- **水平 (level)**: 因素的不同取值称为水平,例如因素“性别”中有男、女两个水平。需要注意的是有时水平是人为划分的,比如身高被分为高、中、低三个水平。

- **交互作用 (interaction)**: 如果一个因素的效应大小在另一个因素的不同水平下明显不同,则称为两因素间存在交互作用。当存在交互作用时,单纯研究某个因素的作用是没有意义的,必须在另一个因素的不同水平下研究该因素的作用大小。有时两因素之间的交互作用无法测量,如在随机区组设计的方差分析中。

根据因素、水平和交互作用的不同,可以将方差分析分为如下几类。

方差分析的类型	因素	水平	交互作用
完全随机设计	1	>2	无
随机区组设计	2	>2	无
析因设计	2	$\geq 2$	考察

完全随机设计可以视为成组  $t$  检验的扩展,两者均为一个因素,但是成组  $t$  检验中的水平数为 2,而完全随机设计方差分析中的水平数大于 2。由于只有一个因