

SHIYONG YIXUE TONGJIXUE

# 实用医学统计学

胡良平◎主 编

立足实用可操作性  
把握核心基础理论  
轻松解决实际问题



金盾出版社

# 实用医学统计学

主 编

胡良平

副主编

胡纯严 李子建

编著者

(以姓氏笔画为序)

王 琪 毛 玮 吕辰龙

刘一松 刘惠刚 关 雪

孙日扬 李长平 沈 宁

周诗国 柳伟伟 贾元杰

高 辉 郭 晋 郭辰仪

陶丽新 鲍晓蕾

金 盾 出 版 社

## 内 容 提 要

本书旨在全面介绍医学统计学的基础知识,内容涉及医学统计学基础和单组设计及其资料的统计处理,包括试验设计基础、统计分析基础与工具、试验设计关键技术、统计资料与基本概念、相对数与分位数、平均指标与变异指标、统计表与统计图,以及与单组设计有关的概念问题、区间估计问题、假设检验问题、相关与回归分析问题。在写作手法上,注重从实际出发,便于读者易学易懂易用。

本书适合于广大实际工作者,包括科技工作者、管理工作、临床医生、杂志编辑等,以及具有中专以上学历的各类学生、研究生和博士后人员学习和使用;也适合希望能正确运用中、高级统计学知识的所有学者和实际工作者参考。

### 图书在版编目(CIP)数据

实用医学统计学/胡良平主编. —北京:金盾出版社,2014.8  
ISBN 978-7-5082-9414-8

I. ①实… II. ①胡… III. ①医学统计—统计学 IV. ①  
R195.1

中国版本图书馆 CIP 数据核字(2014)第 093284 号

### 金盾出版社出版、总发行

北京太平路 5 号(地铁万寿路站往南)  
邮政编码:100036 电话:68214039 83219215  
传真:68276683 网址:www.jdcbs.cn  
封面印刷:北京精美彩色印刷有限公司  
正文印刷:北京万博诚印刷有限公司  
装订:北京万博诚印刷有限公司  
各地新华书店经销

开本:850×1168 1/32 印张:9 字数:226 千字  
2014 年 8 月第 1 版第 1 次印刷

印数:1~4 000 册 定价:22.00 元

---

(凡购买金盾出版社的图书,如有缺页、  
倒页、脱页者,本社发行部负责调换)

# 前言



奋战在科研、教学、医疗、管理等领域的广大实际工作者和管理人员,经常需要申请课题、开展课题研究、撰写学术论文、申报专利等,而这些工作都要求作者具备丰富的专业知识和统计学知识。事实上,实际工作者一般都具有一定的专业知识,然而,他们中的绝大多数却不具备丰富的统计学知识。尽管很多实际工作者至少具有大学以上文化程度,也学过一些统计学知识,但一到用时,还是十有八九会出错。这是什么原因呢?

笔者认为,其主要原因如下:第一,统计学本身比较复杂,尤其是生物统计学和医学统计学,它们要求使用者不仅要有较好的数学和计算机基础,要对所研究问题的专业背景十分熟悉,还要求使用者能灵活运用唯物辩证法。换句话说,统计学不是一门孤立的学科,而是多学科结晶。第二,现行的统计学教科书过于追求“科学和严谨”,严重脱离科研实际。第三,世界上的事物和现象错综复杂、千变万化,并非使用一两种统计分析技术就能解释或揭示它们的内在变化规律。第四,学习者往往低估了统计学的学习和运用难度,希望在很短的时间内学会诸多技术知识,投入的时间和精力却极其有限。更主要的是,在学习过程中,仅用眼,不动脑,不

动手,很少有人边学边用。由于上述这些主要原因,导致很多学过统计学的人没有学到统计学的真本领,尤其是统计学的基础知识打得很不牢固。

俗话说得好,“欲速则不达”,为了使广大实际工作者真正学会统计学并在实践中发挥作用,笔者认为有必要写一本着重于统计学基础的《实用医学统计学》。首先,应帮助实际工作者纠正一个错误的认识,即“统计学是一个计算工具”。很多人认为,统计学就是一门关于如何进行数值计算的学问。持有这样的认识,学生学不好统计学,教师教不好统计学,实际工作者也用不对统计学!道理显而易见,在进行数据分析之前,连资料是否值得分析都没搞清楚,岂能得出正确的统计分析结果?有鉴于此,本书从“试验设计基础”开始阐述,延续至单组设计的区间估计、假设检验和简单相关与回归分析。所有这些内容,虽然都是统计学中的基础知识,但只有正确地掌握了它们,才有可能学好更复杂的统计学知识,才有可能用对统计学知识。

由于笔者水平有限,书中难免会有不妥之处,恳请广大读者不吝赐教,以便再版时修正。笔者的电子邮箱:lphu812@sina.com。

**胡良平**



## 第一章 试验设计基础

|                        |      |
|------------------------|------|
| 一、与确定三要素有关的问题 .....    | (1)  |
| (一)三要素的概述 .....        | (1)  |
| (二)与影响因素的确定有关的问题 ..... | (1)  |
| (三)与受试对象的选取有关的问题 ..... | (6)  |
| (四)与试验效应的确定有关的问题 ..... | (9)  |
| 二、与落实四原则有关的问题 .....    | (12) |
| (一)四原则的概述 .....        | (12) |
| (二)与落实随机原则有关的问题 .....  | (12) |
| (三)与落实对照原则有关的问题 .....  | (18) |
| (四)与落实重复原则有关的问题 .....  | (28) |
| (五)与落实均衡原则有关的问题 .....  | (30) |
| 三、与设计类型有关的问题 .....     | (35) |
| (一)何为试验设计类型 .....      | (35) |
| (二)试验设计类型的种类 .....     | (36) |
| 四、与四种比较类型有关的问题 .....   | (37) |
| (一)四种比较类型的概念 .....     | (37) |

|                              |      |
|------------------------------|------|
| (二)四种比较类型下检验假设及结论的正确陈述 ..... | (38) |
| (三)如何合理选择临床试验的比较类型 .....     | (40) |
| 五、与质量控制有关的问题 .....           | (42) |
| (一)在设计方案中的具体体现 .....         | (42) |
| (二)应着重把握的重要内容 .....          | (43) |

## 第二章 统计分析基础与工具

|                                              |      |
|----------------------------------------------|------|
| 一、统计分析方法概述 .....                             | (45) |
| (一)统计分析方法的分类 .....                           | (45) |
| (二)统计分析方法的合理选择 .....                         | (47) |
| (三)实现统计计算的方法 .....                           | (48) |
| 二、SAS 软件简介 .....                             | (49) |
| (一)SAS 软件概述 .....                            | (49) |
| (二)SAS 界面简介 .....                            | (49) |
| (三)SAS 过程简介 .....                            | (51) |
| (四)SAS 程序简介 .....                            | (51) |
| 三、SAS 用法简介 .....                             | (52) |
| (一)用 SAS 非编程模块进行卡方检验 .....                   | (52) |
| (二)用 SAS 编程法并利用已有 SAS 过程进行卡方检验<br>.....      | (59) |
| (三)当没有相应的 SAS 过程时用 SAS 编程法实现某种<br>统计分析 ..... | (63) |
| (四)归纳与总结 .....                               | (65) |
| 四、SASPAL 软件简介 .....                          | (67) |
| (一)SASPAL 符合初学者的需求 .....                     | (67) |

|                      |      |
|----------------------|------|
| (二)SASPAL 使用方法 ..... | (67) |
| (三)SASPAL 界面简介 ..... | (68) |

### 第三章 试验设计关键技术

|                            |       |
|----------------------------|-------|
| 一、常用样本含量估计方法与 SAS 实现 ..... | (70)  |
| (一)样本含量估计的概述 .....         | (70)  |
| (二)估计样本含量时应具备的前提条件 .....   | (70)  |
| (三)调查研究中样本含量的估计 .....      | (71)  |
| (四)试验研究中样本含量的估计 .....      | (74)  |
| 二、常用随机化方法及 SAS 实现 .....    | (99)  |
| (一)各种常见随机化抽样的 SAS 实现 ..... | (99)  |
| (二)各种常见随机化分组的 SAS 实现 ..... | (104) |

### 第四章 统计资料与基本概念

|                      |       |
|----------------------|-------|
| 一、正确认识统计资料 .....     | (111) |
| (一)统计资料的现代划分方法 ..... | (111) |
| (二)四种统计资料的定义 .....   | (111) |
| (三)统计资料的三要素 .....    | (113) |
| (四)课题的统计资料 .....     | (113) |
| 二、需要弄清的基本概念 .....    | (113) |
| (一)因素与水平 .....       | (113) |
| (二)一般变量与随机变量 .....   | (121) |
| (三)离散与连续型随机变量 .....  | (122) |
| (四)总体与样本 .....       | (122) |
| (五)统计量与参数 .....      | (123) |

|                       |       |
|-----------------------|-------|
| (六)频率与概率·····         | (123) |
| (七)两种类型随机变量的概率分布····· | (124) |

## 第五章 相对数与分位数

|                                      |       |
|--------------------------------------|-------|
| 一、相对数·····                           | (127) |
| (一)概念与种类·····                        | (127) |
| (二)相对数的计算·····                       | (129) |
| 二、分位数·····                           | (135) |
| (一)分位数的概念、种类和作用·····                 | (135) |
| (二)四分位数与百分位数的计算·····                 | (137) |
| (三)百分位数在确定偏态频数分布资料参考值范围中的<br>应用····· | (137) |

## 第六章 平均指标与变异指标

|                             |       |
|-----------------------------|-------|
| 一、概述·····                   | (140) |
| 二、对称和非对称分布定量资料的统计描述·····    | (141) |
| (一)对称分布定量资料的平均指标——算术均数····· | (141) |
| (二)非对称分布定量资料的四个平均指标·····    | (144) |
| (三)对称分布定量资料的四个变异指标·····     | (149) |
| (四)非对称分布定量资料的三个变异指标·····    | (153) |

## 第七章 统计表与统计图

|                              |       |
|------------------------------|-------|
| 一、合理编制统计表·····               | (155) |
| (一)表达单组设计一元定量资料频数分布的频数表····· | (155) |
| (二)表达多组定量资料的统计表·····         | (156) |

|                        |       |
|------------------------|-------|
| (三)表达多组定性资料的统计表.....   | (157) |
| 二、正确绘制统计图 .....        | (158) |
| (一)常用统计图知识点概述.....     | (158) |
| (二)用 SAS 绘制常用统计图 ..... | (159) |

## 第八章 与单组设计有关的概念问题

|                             |       |
|-----------------------------|-------|
| 一、单组设计的例子及解说 .....          | (183) |
| (一)无标准值的单组设计定量资料的例子及解说..... | (183) |
| (二)有标准值的单组设计定量资料的例子及解说..... | (188) |
| (三)无标准值的单组设计定性资料的例子及解说..... | (190) |
| (四)有标准值的单组设计定性资料的例子及解说..... | (194) |
| 二、单组设计的概念 .....             | (194) |
| (一)单组设计的基本概念.....           | (194) |
| (二)与单组设计有关的重要概念.....        | (195) |

## 第九章 与单组设计有关的区间估计问题

|                                            |       |
|--------------------------------------------|-------|
| 一、与单组设计一元定量资料区间估计有关的问题 .....               | (197) |
| (一)区间的概念.....                              | (197) |
| (二)几种区间的异同点.....                           | (199) |
| (三)单组设计一元定量资料总体参数置信区间的估计<br>.....          | (200) |
| (四)单组设计一元定量资料“SUB”容许区间和个体值<br>预测区间的估计..... | (206) |
| 二、与单组设计一元定性资料总体率区间估计有关的问题<br>.....         | (209) |

- (一)如何用 SAS 实现频数资料总体率区间估计 ..... (209)
- (二)如何用 SAS 实现数据库结构资料总体率区间估计  
..... (211)
- (三)如何用 SAS 实现比例的不同置信水平的区间估计  
..... (213)

## 第十章 与单组设计有关的假设检验问题

- 一、假设检验的概念 ..... (217)
  - (一)假设检验基本思想..... (217)
  - (二)假设检验的基本步骤..... (218)
- 二、假设检验的实现 ..... (219)
  - (一)单组设计一元定量资料假设检验的内容..... (219)
  - (二)单组设计一元定量资料的正常性检验..... (219)
  - (三)单组设计一元定量资料均值的参数检验( $t$  检验)或  
中位数的非参数检验..... (220)
  - (四)单组设计一元定性资料的假设检验..... (224)
  - (五)单组设计定量资料差异性分析方面的遗留问题..... (227)

## 第十一章 与单组设计有关的相关与回归分析问题

- 一、如何进行三种简单相关分析 ..... (228)
  - (一)单组设计二元定量资料简单相关分析..... (228)
  - (二)如何用 SAS 实现三种简单相关分析 ..... (231)
- 二、如何用直线回归方程描述两定量变量间依赖关系 ..... (250)
  - (一)直线回归分析概述..... (250)
  - (二)问题与数据结构..... (250)

|                                            |       |
|--------------------------------------------|-------|
| (三)如何分析数据结构.....                           | (251) |
| (四)如何合理选择统计分析方法.....                       | (252) |
| (五)如何用 SAS 实现直线回归分析 .....                  | (253) |
| (六)如何巧妙地将直线回归分析用于配对设计定量资料<br>以实现一致性评价..... | (259) |
| (七)实施直线回归分析的正确步骤.....                      | (262) |
| (八)简单线性回归分析的计算原理.....                      | (263) |
| 三、相关与回归分析的遗留问题 .....                       | (269) |
| (一)相关分析方面存在的遗留问题.....                      | (269) |
| (二)回归分析方面存在的遗留问题.....                      | (270) |

### 附录 胡良平统计学专著及配套软件

|               |       |
|---------------|-------|
| (一)统计学专著..... | (273) |
| (二)配套软件.....  | (276) |

## 第一章 试验设计基础

### 一、与确定三要素有关的问题

#### (一)三要素的概述

无论是以动物或样品为受试对象的试验研究,还是以人为受试对象的临床试验研究,都离不开“影响因素、受试对象和试验效应”这三个“物质基础”,它们被称为试验研究的三要素。既然是试验研究,就要涉及不同的影响因素(如药物种类、药物剂量、作用时间等,它们被称为试验因素)作用于具有不同状态(如患病时间长短、病情轻重、性别、血型等,它们被称为非试验因素)的人或不同状态的动物或样品(它们都被称为受试对象)后产生的效应(被称为试验效应,常通过各种指标及其取值来体现)。因此,一项科研课题的结果和结论是否可信,第一道关卡就是看三要素把握得是否到位。

#### (二)与影响因素的确定有关的问题

所有对试验结果有影响的因素统称为影响因素,包括试验因素和非试验因素。试验因素通常包括以下两类:第一类,研究者希望通过试验考察其对试验结果是否有影响的性质相同的不同试验条件的总称;第二类,试验因素的两个水平是“有或无”两种状态,如A药用否、结扎与否、去势与否。试验因素包括物理的(如针

刺、射线、理疗等),化学的(如药物、毒物等),生物的(如细菌、病毒等),以及社会的因素等;而那些与试验因素同时存在,能使受试对象产生效应的其他因素被称为非试验因素,一般是受试对象自身具有的、可影响其发展过程的某些因素(如性别、年龄、血型、病情轻重、患病时间长短、是否有某病的家族史等)。因考察的效应指标不同,有些重要的非试验因素对试验结果的影响是不容忽视的,必须在试验设计阶段或在统计分析阶段加以控制,而有些非试验因素是可以忽略不计的。

值得注意的是,试验因素取决于研究者的研究目的。例如,探讨高血压性脑出血微创术对清除血肿后脑水肿患者血浆细胞间黏附分子-1的影响。试验因素为是否采用高血压性脑出血微创术。治疗前的血浆细胞间黏附分子-1含量就是一个定量的重要非试验因素。研究者在试验设计中应注意以下几点。

**1. 确定试验因素和非试验因素** 试验研究的目的不同,对试验的要求也不同。若在整个试验过程中影响观察结果的因素很多,就必须结合专业知识,对众多的因素做全面分析,必要时做一些预试验,区分哪些是重要的试验因素,哪些是重要的非试验因素,以便选用合适的试验设计方法妥善安排这些因素。任何试验效应都是多种因素作用的结果,我们不可能也没有必要在一次或几次研究中,研究所想到的一切有关因素,只能在设法严格控制重要非试验因素的前提下,抓住对试验结果影响较大的主要试验因素加以研究。

**2. 合理安排试验因素与非试验因素** 一个因素在试验中所处的不同状态称为水平。不同因素或同一因素的水平不同,就造成试验条件的多样性。

(1)单因素单水平(本质上应属于单因素两水平)设计:这是科研中较常见的一种试验设计类型,如一种新的治疗方案对某种疾病的疗效,比较时常以“常规疗法”的特定结果(如理论平均值或

率)为基准,就属于这类试验。

(2)单因素多水平设计:比较某药物不同剂量对某病的治疗效果,属于这类试验,特别是珍贵药物、毒性较大的药物或最佳新药剂量的选择,往往需要采用这类试验。

(3)多因素多水平设计:例如,探索联合用药方案、研究中药复方等。由于生物体的复杂性,决定了多因素多水平研究成为最常用的试验设计类型。

值得注意的是,多因素多水平设计并非是一种标准的设计名称,在实际工作中,应将其具体化,如无重复试验的两因素随机区组设计、无重复试验且为单个体型的三因素拉丁方设计、四因素析因设计、具有一个重复测量的三因素设计、五个3水平因素的正交设计、十个7水平因素的均匀设计等。

若试验中含有重要的非试验因素,设计时应明确这些非试验因素,有意识地控制和消除其干扰作用。若非试验因素是定性的,如性别、窝别等,可以采用分层随机化的方法对这些非试验因素加以控制,常见的区组随机化和配对随机化可看做是分层随机化的实际应用。此类试验设计包括配对设计、随机区组设计、平衡不完全随机区组设计、含区组因素的析因设计等。有些非试验因素虽然是定量的,但可以通过分组将它看做定性的因素,如年龄是定量的,若按照年龄大小分为儿童、少年、青年、中年、老年,就转变成定性的因素了,也可以按上述方法安排这种定量的非试验因素。若重要的非试验因素是定量的又不便将其定性化,则应从试验开始之时,就记录每个受试对象的定量影响因素的数值大小,将来进行数据处理时采用“协方差分析”等技术将其影响降至最低程度。例如,研究不同饲料对动物体重的影响,饲料种类是试验因素,而动物进食量是一个很重要的非试验因素,在试验设计阶段难以严格控制,可以在试验中记录动物的每餐进食量,最后可求出每只动物的平均进食量,将来进行数据处理时采用相应设计定量资料的协

方差分析方法,从而排除不同饲料组中动物平均进食量不同对试验结果造成的影响。像“进食量”这样的重要的定量非试验因素,只有在试验之前就考虑到并在试验过程中认真记录下来;否则,等试验结束了再来弥补,就是绝对不可能的了!

**3. 注意因素之间的交互作用** 多种因素之间往往存在交互作用。交互作用是在原效应的基础上产生增强效应或减弱效应的作用,最易于理解的是两药共用可发生药效的协同作用或拮抗作用。事实上,不仅两种药物同时使用可能会产生交互作用,非药物的试验因素之间也常可产生交互作用。所谓交互作用,就是一个因素不同水平对观测结果的影响情况会随着另一个因素水平的改变而改变。因此,在试验设计时除了要考虑各因素单独施加于受试对象的效应,还应特别关注因素配合施加于受试对象后产生的效应。

**4. 试验因素要标准化** 试验因素标准化就是要保证试验因素在整个试验的过程中始终如一,保持不变,包括处理因素的施加方法、强度、频率和持续时间等。例如,试验因素是药品,除应确定药品的名称、性质、成分、作用及用法外,还应明确生产厂家、药品批号、出厂日期及保存方法等;若试验因素是某种手术方法,操作者的熟练程度应当自始至终保持恒定,否则将会影响结果的稳定性。同样,其他因素也必须遵循标准化原则,否则那些非试验因素或多或少会影响试验结果,导致结论出现较大偏差。因此,在设计时应制订或摸索出标准化的具体措施和方法,使试验因素真正达到标准化,并尽可能使非试验因素始终处于可控状态。

### 5. 实际工作者在确定试验因素方面常犯的错误

(1)喜欢用“组别”这个词作为全部试验分组的总称,无论试验中涉及多少个组,他们总把“组别”理解为“一个试验因素”,很自然地将其下的各组理解为该因素的各个水平。

(2)毫无根据地选择很多试验因素,使试验研究变得十分复

杂,往往因为试验条件、人力和时间都达不到起码的要求,使试验研究半途而废;或因安排不当,组间夹杂着某些混杂因素的影响,导致结论可信度低。

(3)由于研究者试验设计水平较低,不会科学地利用多因素试验设计技术,只能将其他试验因素视而不见或固定在各自的特定水平上,每次只让自己关心的某个试验因素取不同的水平。当试验因素之间不独立时,对基于这样的试验方法获得的结果进行分析,常得出错误结论。

例如,为观察吞咽康复训练联合抗抑郁药治疗脑卒中后吞咽障碍的临床疗效,将脑卒中后并发吞咽障碍的 133 例住院患者随机分为治疗组(68 例)和对照组(65 例),两组患者均接受神经内科常规治疗,1 周内给予吞咽康复训练,治疗组在此基础上给予抗抑郁药帕罗西丁,初始剂量 10 mg/d,早餐后服用,第 2 周增加至 20 mg/d,4 周后进行疗效评定。治疗组有效率 92.65%,对照组有效率 76.93%,两组比较差异有统计学意义。研究者认为,抗抑郁药帕罗西丁能明显提高脑卒中后吞咽障碍的疗效。

**辨析与释疑:**研究者忽视了重要的非试验因素。本试验中,试验因素为是否使用抗抑郁药帕罗西丁,受试者是脑卒中后并发吞咽障碍的 133 例住院患者,试验效应是对吞咽困难的治疗效果。根据医学知识,吞咽障碍由轻到重共分为 7 级,研究者在试验设计阶段应将两组患者吞咽障碍的严重程度作为一个重要的非试验因素,即在分组时,确保两组患者病情之间的差异无统计学意义。若两组患者吞咽障碍的程度不一致,得到的结论是不可信的,甚至是错误的。此外,抗抑郁药的疗效与患者在初始阶段的心理状态及患者的性格都有关系,建议研究者考虑将患者的心理状态及性格作为另两个重要非试验因素。