

主编  
I V  
胡良平

# 医学统计 实用手册

人民卫生出版社



R195.1  
HLPc  
C.1

127083

# 医学统计 实用手册

主 编 胡良平

编 委 李子建 刘惠刚 童中彪  
李峰青 胡良平



人民卫生出版社

解放军医学图书馆[书]



\*C0249123\*

**图书在版编目(CIP)数据**

医学统计实用手册/胡良平主编. —北京:  
人民卫生出版社, 2004. 12  
ISBN 7-117-06528-1

I. 医… II. 胡… III. 医学统计-手册  
IV. R195.1

中国版本图书馆 CIP 数据核字(2004)第 124516 号

**医学统计实用手册**

主 编: 胡 良 平

出版发行: 人民卫生出版社(中继线 67616688)

地 址: (100078)北京市丰台区方庄芳群园 3 区 3 号楼

网 址: [http://www. pmph. com](http://www.pmph.com)

E - mail: [pmph @ pmph. com](mailto:pmph@pmph.com)

印 刷: 原创阳光印业有限公司

经 销: 新华书店

开 本: 787 × 1092 1/16 印张: 20. 25

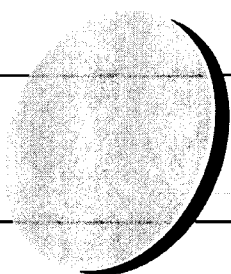
字 数: 485 千字

版 次: 2004 年 12 月第 1 版 2004 年 12 月第 1 版第 1 次印刷

标准书号: ISBN 7-117-06528-1/R · 6529

定 价: 30.00 元

著作权所有, 请勿擅自用本书制作各类出版物, 违者必究  
(凡属质量问题请与本社发行部联系退换)



# 前 言

市面上医学统计学教科书不计其数,笔者为何还要写这本书呢?因为在医学科研和医学期刊论文中有大量误用和滥用统计学的现象,笔者还经常被人问起:“为什么我们学了多遍统计学,一到用时就出错?”面对医学研究中涌现出来的大量误用统计学的案例和许多想学好、用对统计学的实际工作者,作为一名统计学工作者,笔者深感有必要帮助他们正确使用统计学,这也正是驱使笔者撰写本书的动因。

本书通过分析医学科研和医学期刊中存在大量误用和滥用统计学的现象的根源和实质,有针对性地提出了解决这些问题的对策。本书共分3篇,第1篇**医学统计典型错误辨析**,从实验设计、统计描述、定量与定性资料统计分析、相关与回归分析等方面,辨析误用统计学的案例;第2篇**医学统计学理论与方法**,从正面较详细地阐述了掌握实验设计和常用统计处理技术所必需的理论和方法;第3篇**用手工和 SAS 实现统计分析**,即用两种途径实现常用统计分析中的计算问题,为读者解决实际问题排除了计算方面的障碍,提供了理论、方法和技术指导。

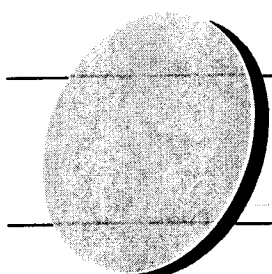
本书可作为医学科研工作者和临床医生辨析常见统计学应用错误和合理选用实验设计方法与统计分析方法的实用手册,也可作为生物医学领域内的本科生、研究生学习医学统计学、领会其精神实质的参考书。

在本书即将问世之际,笔者衷心地感谢首都医科大学继续教育学院刘惠刚副教授和医学统计教研室的三名硕士研究生为本书提供了大量的素材和部分编写工作,尤其是李子建同志,查阅了大量的文献资料,提供了很多生动的典型案例,并作了精辟地分析和释疑。为了方便广大读者学习和使用本书,笔者将书中的全部 SAS 引导程序放在生物医学统计咨询网站上供读者下载,网址:<http://www.stathlp.com>。

由于笔者水平有限,缺点和错误在所难免,恳请广大读者批评指正!

胡良平

2004 年 11 月



# 目 录

## 第1篇 医学统计典型错误辨析

|                       |    |
|-----------------------|----|
| 第1章 实验设计典型错误辨析        | 3  |
| 1.1 指导思想上的问题          | 3  |
| 1.1.1 只有专业设计,缺乏统计研究设计 | 3  |
| 1.1.2 临床科研流程倒置        | 3  |
| 1.1.3 缺乏完善的实验设计方案     | 3  |
| 1.2 实验设计三要素方面存在的问题    | 5  |
| 1.2.1 不能正确区分因素与水平     | 5  |
| 1.2.2 盲目选用受试对象        | 6  |
| 1.2.3 盲目选用观测指标        | 6  |
| 1.3 对照原则实施中存在的问题      | 7  |
| 1.3.1 对照过剩            | 7  |
| 1.3.2 假对照             | 7  |
| 1.3.3 对照不全            | 8  |
| 1.3.4 缺乏必要的对照组        | 9  |
| 1.3.5 盲目设立对照组         | 10 |
| 1.4 随机原则实施中存在的问题      | 11 |
| 1.4.1 按进入研究的时间顺序分组    | 11 |
| 1.4.2 用随意取代随机         | 11 |
| 1.5 重复原则实施中存在的问题      | 12 |
| 1.5.1 重复的三层含义         | 12 |
| 1.5.2 样本含量 $n=1$      | 12 |
| 1.6 均衡原则实施中存在的问题      | 13 |
| 1.6.1 均衡原则的含义         | 13 |
| 1.6.2 对照组与实验组基线不可比    | 14 |
| 1.6.3 提高均衡原则的策略       | 14 |
| 1.7 多因素实验研究中存在的问题     | 15 |

|            |                          |           |
|------------|--------------------------|-----------|
| 1.7.1      | 用单因素设计取代多因素设计的含义及其不良后果   | 15        |
| 1.7.2      | 用单因素设计取代多因素设计实例          | 15        |
| 1.7.3      | 看似多因素设计实际仍为单因素设计         | 16        |
| 1.7.4      | 实验因素各水平组合不完全             | 17        |
| <b>第2章</b> | <b>统计描述典型错误辨析</b>        | <b>19</b> |
| 2.1        | 指导思想上的问题                 | 19        |
| 2.1.1      | 轻敌思想极为严重                 | 19        |
| 2.1.2      | 缺乏起码的常识                  | 19        |
| 2.2        | 统计资料类型识别方面的错误            | 19        |
| 2.2.1      | 误将定性资料视为定量资料             | 19        |
| 2.2.2      | 误将定量资料视为定性资料             | 21        |
| 2.3        | 相对数应用中存在的错误              | 23        |
| 2.3.1      | 误将“百分比”当作“百分率”           | 23        |
| 2.3.2      | 误将“百分率”当作“百分比”           | 24        |
| 2.3.3      | 分母很小( $n < 20$ )时仍计算相对数  | 24        |
| 2.4        | 平均指标与变异指标应用中存在的错误        | 25        |
| 2.5        | 编制统计表中存在的错误              | 26        |
| 2.6        | 绘制统计图中存在的错误              | 27        |
| <b>第3章</b> | <b>定量资料统计分析典型错误辨析</b>    | <b>30</b> |
| 3.1        | 定量资料统计分析常见错误及其不良后果       | 30        |
| 3.2        | 多因素实验设计类型的辨析法            | 30        |
| 3.2.1      | 平衡组合多因素实验设计资料“对号入座”辨析法   | 30        |
| 3.2.2      | 平衡组合多因素实验设计资料“结构变形”辨析法   | 33        |
| 3.2.3      | 非平衡组合多因素实验设计资料的“拆分”辨析法   | 40        |
| 3.3        | 多因素实验设计多指标资料一元与多元“兼顾”辨析法 | 47        |
| 3.4        | 正交实验设计资料处理中常见错误辨析法       | 50        |
| 3.5        | 不完善的多因素实验设计问题的辨析法        | 52        |
| 3.6        | 用单因素分析取代多因素分析问题的辨析法      | 54        |
| 3.7        | 存在组间均衡性较差错误的辨析法          | 56        |
| <b>第4章</b> | <b>定性资料统计分析典型错误辨析</b>    | <b>59</b> |
| 4.1        | 定性资料统计分析常见错误及其不良后果       | 59        |
| 4.2        | 定性资料分析中常见错误辨析            | 59        |
| 4.2.1      | 二维列联表资料分析中常见错误辨析         | 59        |
| 4.2.2      | 高维列联表资料分析中常见错误辨析         | 63        |
| <b>第5章</b> | <b>相关与回归分析典型错误辨析</b>     | <b>69</b> |
| 5.1        | 相关与回归分析中存在问题的概述          | 69        |
| 5.2        | 相关与回归分析中常见错误举例           | 69        |

## 第2篇 医学统计学理论与方法

|                                               |     |
|-----------------------------------------------|-----|
| 第6章 实验设计的理论与方法 .....                          | 79  |
| 6.1 正确识别实验设计类型的意义和方法 .....                    | 79  |
| 6.2 常用实验设计类型介绍 .....                          | 80  |
| 6.2.1 掌握实验设计类型的必要性 .....                      | 80  |
| 6.2.2 单因素 $k$ 水平设计( $k=1, 2, \geq 3$ ) .....  | 80  |
| 6.2.3 多因素析因设计 .....                           | 82  |
| 6.2.4 多因素正交设计 .....                           | 87  |
| 6.2.5 重复测量设计 .....                            | 87  |
| 6.2.6 配伍组设计 .....                             | 91  |
| 6.2.7 交叉设计 .....                              | 92  |
| 6.2.8 拉丁方设计 .....                             | 93  |
| 6.3 实验设计类型的合理选用 .....                         | 94  |
| 6.3.1 选用实验设计类型的指导思想 .....                     | 94  |
| 6.3.2 “配对+成组”设计等价于“单因素 $k$ 水平”设计吗 .....       | 94  |
| 6.3.3 “配对、成组、交叉”设计中应选择哪一个 .....               | 94  |
| 6.3.4 “单因素 $k$ 水平设计、配伍组设计、拉丁方设计”中应选择哪一个 ..... | 96  |
| 6.3.5 “析因设计、正交设计”中应选择哪一个 .....                | 97  |
| 6.4 实验设计中三要素的考虑 .....                         | 100 |
| 6.4.1 受试对象的考虑 .....                           | 100 |
| 6.4.2 实验因素的考虑 .....                           | 100 |
| 6.4.3 实验效应的考虑 .....                           | 100 |
| 6.5 实验设计中四个基本原则的实施 .....                      | 100 |
| 6.5.1 对照原则的实施 .....                           | 100 |
| 6.5.2 随机原则的实施 .....                           | 101 |
| 6.5.3 重复原则的实施 .....                           | 109 |
| 6.5.4 均衡原则的实施 .....                           | 109 |
| 6.6 与实验设计有关的概念及设计举例 .....                     | 109 |
| 6.6.1 实验设计基本概念举例 .....                        | 109 |
| 6.6.2 实验设计类型用法举例 .....                        | 113 |
| 第7章 定量资料统计分析的理论与方法 .....                      | 117 |
| 7.1 定量资料统计分析的基本概念 .....                       | 117 |
| 7.1.1 何为定量资料的统计分析 .....                       | 117 |
| 7.1.2 何为参数检验与非参数检验 .....                      | 117 |
| 7.1.3 参数检验要求资料应满足的前提条件 .....                  | 117 |
| 7.1.4 当资料不满足前提条件时怎么办 .....                    | 118 |
| 7.1.5 统计计算的实现策略 .....                         | 118 |

|                                                  |     |
|--------------------------------------------------|-----|
| 7.2 参数检验的基本思想与计算步骤 .....                         | 118 |
| 7.2.1 常用参数检验方法概述 .....                           | 118 |
| 7.2.2 参数检验的基本思想 .....                            | 118 |
| 7.2.3 参数检验的前提条件和理论依据 .....                       | 119 |
| 7.2.4 参数检验的计算步骤 .....                            | 119 |
| 7.3 非参数检验的基本思想与计算步骤 .....                        | 120 |
| 7.3.1 常用非参数检验方法概述 .....                          | 120 |
| 7.3.2 非参数检验的基本思想 .....                           | 120 |
| 7.3.3 符号检验的计算步骤 .....                            | 121 |
| 7.3.4 符号秩和检验的计算步骤 .....                          | 122 |
| 7.3.5 秩和检验的计算步骤 .....                            | 122 |
| <b>第8章 定性资料统计分析的理论与方法</b> .....                  | 124 |
| 8.1 定性资料统计分析方法概述 .....                           | 124 |
| 8.1.1 何为定性资料的统计分析 .....                          | 124 |
| 8.1.2 正确处理定性资料的关键技术 .....                        | 124 |
| 8.2 定性资料统计分析的基本思想和原理 .....                       | 125 |
| 8.2.1 $\chi^2$ 检验的基本思想和原理 .....                  | 125 |
| 8.2.2 秩和检验的基本思想和原理 .....                         | 125 |
| 8.2.3 Ridit 分析的基本思想和原理 .....                     | 126 |
| 8.3 $2 \times 2$ 表(或四格表)资料及对应的统计分析方法 .....       | 127 |
| 8.3.1 来自横断面研究设计的 $2 \times 2$ 表资料及统计分析方法 .....   | 127 |
| 8.3.2 来自队列研究设计的 $2 \times 2$ 表资料及统计分析方法 .....    | 128 |
| 8.3.3 来自病例-对照研究设计的 $2 \times 2$ 表资料及统计分析方法 ..... | 128 |
| 8.3.4 来自配对研究设计的 $2 \times 2$ 表资料及统计分析方法 .....    | 129 |
| 8.4 $R \times C$ 表资料及对应的统计分析方法 .....             | 130 |
| 8.4.1 双向无序的 $R \times C$ 表资料及统计分析方法 .....        | 130 |
| 8.4.2 单向有序的 $R \times C$ 表资料及统计分析方法 .....        | 131 |
| 8.4.3 双向有序且属性不同的 $R \times C$ 表资料及统计分析方法 .....   | 131 |
| 8.4.4 双向有序且属性相同的 $R \times C$ 表资料及统计分析方法 .....   | 132 |
| 8.5 高维列联表资料及对应的统计分析方法 .....                      | 133 |
| 8.5.1 结果变量为二值变量的高维列联表及统计分析方法 .....               | 133 |
| 8.5.2 结果变量为多值有序变量的高维列联表及统计分析方法 .....             | 133 |
| 8.5.3 结果变量为多值名义变量的高维列联表及统计分析方法 .....             | 133 |
| <b>第9章 直线相关与回归分析的理论与方法</b> .....                 | 135 |
| 9.1 直线相关与回归分析方法概述 .....                          | 135 |
| 9.1.1 何为直线相关与回归分析 .....                          | 135 |
| 9.1.2 进行直线相关与回归分析的正确步骤 .....                     | 135 |
| 9.1.3 相关与回归分析的种类 .....                           | 136 |
| 9.1.4 直线相关与回归分析的异同点 .....                        | 137 |

|       |                                 |     |
|-------|---------------------------------|-----|
| 9.1.5 | 直线相关与回归分析的关键点                   | 138 |
| 9.2   | 直线相关与回归分析的计算与检验                 | 140 |
| 9.2.1 | 相关系数 $r$ 的计算                    | 140 |
| 9.2.2 | 相关系数 $r$ 的假设检验                  | 141 |
| 9.2.3 | 总体相关系数 $\rho$ 的置信区间             | 141 |
| 9.2.4 | 截距 $a$ 和斜率 $b$ 的计算              | 141 |
| 9.2.5 | 截距 $a$ 和斜率 $b$ 的假设检验            | 142 |
| 9.2.6 | 总体截距 $\alpha$ 和斜率 $\beta$ 的置信区间 | 143 |
| 9.2.7 | 直线回归分析中其他有关的区间估计问题              | 143 |

### 第3篇 用手工与 SAS 实现统计分析

|        |                           |     |
|--------|---------------------------|-----|
| 第10章   | 定量资料参数检验                  | 147 |
| 10.1   | 单组设计定量资料 $t$ 检验的应用        | 147 |
| 10.1.1 | 单组设计定量资料 $t$ 检验的预备工作      | 147 |
| 10.1.2 | 单组设计定量资料 $t$ 检验的实例        | 147 |
| 10.1.3 | 用手工实现单组设计定量资料 $t$ 检验      | 147 |
| 10.1.4 | 用 SAS 实现单组设计定量资料 $t$ 检验   | 148 |
| 10.2   | 配对设计定量资料 $t$ 检验的应用        | 149 |
| 10.2.1 | 配对设计定量资料 $t$ 检验的预备工作      | 149 |
| 10.2.2 | 配对设计定量资料 $t$ 检验的实例        | 150 |
| 10.2.3 | 用手工实现配对设计定量资料 $t$ 检验      | 150 |
| 10.2.4 | 用 SAS 实现配对设计定量资料 $t$ 检验   | 151 |
| 10.3   | 成组设计定量资料 $t$ 检验的应用        | 151 |
| 10.3.1 | 成组设计定量资料 $t$ 检验的预备工作      | 151 |
| 10.3.2 | 成组设计定量资料 $t$ 检验的实例        | 152 |
| 10.3.3 | 用手工实现成组设计定量资料 $t$ 检验      | 152 |
| 10.3.4 | 用 SAS 实现成组设计定量资料 $t$ 检验   | 153 |
| 10.4   | 单因素多水平设计定量资料方差分析的应用       | 155 |
| 10.4.1 | 单因素多水平设计定量资料方差分析的预备工作     | 155 |
| 10.4.2 | 单因素多水平设计定量资料方差分析的实例       | 155 |
| 10.4.3 | 用手工实现单因素多水平设计定量资料的方差分析    | 155 |
| 10.4.4 | 用 SAS 实现单因素多水平设计定量资料的方差分析 | 156 |
| 10.5   | 配伍组设计定量资料方差分析的应用          | 159 |
| 10.5.1 | 配伍组设计定量资料方差分析的预备工作        | 159 |
| 10.5.2 | 配伍组设计定量资料方差分析的实例          | 159 |
| 10.5.3 | 用手工实现配伍组设计定量资料方差分析        | 159 |
| 10.5.4 | 用 SAS 实现配伍组设计定量资料方差分析     | 160 |
| 10.6   | 拉丁方设计定量资料方差分析的应用          | 162 |
| 10.6.1 | 拉丁方设计定量资料方差分析的预备工作        | 162 |

|               |                                        |            |
|---------------|----------------------------------------|------------|
| 10.6.2        | 拉丁方设计定量资料方差分析的实例                       | 163        |
| 10.6.3        | 用手工实现拉丁方设计定量资料方差分析                     | 163        |
| 10.6.4        | 用 SAS 实现拉丁方设计定量资料方差分析                  | 164        |
| 10.7          | 交叉设计定量资料方差分析的应用                        | 166        |
| 10.7.1        | 交叉设计定量资料方差分析的预备工作                      | 166        |
| 10.7.2        | 交叉设计定量资料方差分析的实例                        | 166        |
| 10.7.3        | 用手工实现交叉设计定量资料方差分析                      | 167        |
| 10.7.4        | 用 SAS 实现交叉设计定量资料方差分析                   | 169        |
| 10.8          | 析因设计定量资料方差分析的应用                        | 170        |
| 10.8.1        | 析因设计定量资料方差分析的预备工作                      | 170        |
| 10.8.2        | 析因设计定量资料方差分析的实例                        | 170        |
| 10.8.3        | 用手工实现析因设计定量资料方差分析                      | 171        |
| 10.8.4        | 用 SAS 实现析因设计定量资料方差分析                   | 172        |
| 10.9          | 正交设计定量资料方差分析的应用                        | 173        |
| 10.9.1        | 正交设计定量资料方差分析的预备工作                      | 173        |
| 10.9.2        | 正交设计定量资料方差分析的实例                        | 174        |
| 10.9.3        | 用手工实现正交设计定量资料方差分析                      | 174        |
| 10.9.4        | 用 SAS 实现正交设计定量资料方差分析                   | 175        |
| <b>第 11 章</b> | <b>定量资料非参数检验</b>                       | <b>178</b> |
| 11.1          | 单组设计定量资料非参数检验的应用                       | 178        |
| 11.1.1        | 单组设计定量资料非参数检验的实例                       | 178        |
| 11.1.2        | 用手工实现单组设计定量资料符号检验                      | 178        |
| 11.1.3        | 用 SAS 实现单组设计定量资料符号检验                   | 178        |
| 11.2          | 配对设计定量资料非参数检验的应用                       | 179        |
| 11.2.1        | 配对设计定量资料非参数检验的实例                       | 179        |
| 11.2.2        | 用手工实现配对设计定量资料符号秩和检验                    | 179        |
| 11.2.3        | 用 SAS 实现配对设计定量资料符号秩和检验                 | 180        |
| 11.3          | 成组设计定量资料非参数检验的应用                       | 181        |
| 11.3.1        | 成组设计定量资料非参数检验的实例                       | 181        |
| 11.3.2        | 用手工实现成组设计定量资料 Wilcoxon 秩和检验            | 181        |
| 11.3.3        | 用 SAS 实现成组设计定量资料 Wilcoxon 秩和检验         | 182        |
| 11.4          | 单因素多水平设计定量资料非参数检验的应用                   | 184        |
| 11.4.1        | 单因素多水平设计定量资料非参数检验的实例                   | 184        |
| 11.4.2        | 用手工实现单因素多水平设计定量资料 Kruskal-Wallis 检验    | 184        |
| 11.4.3        | 用 SAS 实现单因素多水平设计定量资料 Kruskal-Wallis 检验 | 185        |
| 11.5          | 配伍组设计定量资料非参数检验的应用                      | 186        |
| 11.5.1        | 配伍组设计定量资料非参数检验的实例                      | 186        |
| 11.5.2        | 用手工实现配伍组设计定量资料 Friedman 检验             | 186        |
| 11.5.3        | 用 SAS 实现配伍组设计定量资料 Friedman 检验          | 187        |

|                                                  |            |
|--------------------------------------------------|------------|
| 11.6 多个样本间两两比较秩和检验的应用                            | 190        |
| 11.6.1 多个样本间两两比较秩和检验的概述                          | 190        |
| 11.6.2 多个样本间定量资料两两比较秩和检验的实例及 SAS 实现              | 190        |
| 11.6.3 单向有序列联表资料两两比较秩和检验的实例及 SAS 实现              | 193        |
| <b>第 12 章 二维列联表资料的统计分析</b>                       | <b>197</b> |
| 12.1 横断面研究设计 $2 \times 2$ 表资料的统计分析               | 197        |
| 12.1.1 横断面研究设计 $2 \times 2$ 表资料的实例               | 197        |
| 12.1.2 用手工实现横断面研究设计 $2 \times 2$ 表资料的统计分析        | 197        |
| 12.1.3 用 SAS 实现横断面研究设计 $2 \times 2$ 表资料的统计分析     | 198        |
| 12.2 队列研究设计 $2 \times 2$ 表资料的统计分析                | 199        |
| 12.2.1 队列研究设计 $2 \times 2$ 表资料的实例                | 199        |
| 12.2.2 用手工实现队列研究设计 $2 \times 2$ 表资料的统计分析         | 199        |
| 12.2.3 用 SAS 实现队列研究设计 $2 \times 2$ 表资料的统计分析      | 200        |
| 12.3 病例-对照研究设计 $2 \times 2$ 表资料的统计分析             | 201        |
| 12.3.1 病例-对照研究设计 $2 \times 2$ 表资料的实例             | 201        |
| 12.3.2 用手工实现病例-对照研究设计 $2 \times 2$ 表资料的统计分析      | 202        |
| 12.3.3 用 SAS 实现病例-对照研究设计 $2 \times 2$ 表资料的统计分析   | 202        |
| 12.4 配对研究设计 $2 \times 2$ 表资料的统计分析                | 202        |
| 12.4.1 配对研究设计 $2 \times 2$ 表资料的实例                | 202        |
| 12.4.2 用手工实现配对研究设计 $2 \times 2$ 表资料的统计分析         | 203        |
| 12.4.3 用 SAS 实现配对研究设计 $2 \times 2$ 表资料的统计分析      | 203        |
| 12.5 双向无序 $R \times C$ 列联表资料的统计分析                | 203        |
| 12.5.1 双向无序 $R \times C$ 列联表资料的实例                | 203        |
| 12.5.2 用手工实现双向无序 $R \times C$ 列联表资料的统计分析         | 204        |
| 12.5.3 用 SAS 实现双向无序 $R \times C$ 列联表资料的统计分析      | 204        |
| 12.6 单向有序 $R \times C$ 列联表资料的统计分析                | 206        |
| 12.6.1 单向有序 $R \times C$ 列联表资料的实例                | 206        |
| 12.6.2 用手工实现单向有序 $R \times C$ 列联表资料的统计分析         | 207        |
| 12.6.3 用 SAS 实现单向有序 $R \times C$ 列联表资料的统计分析      | 208        |
| 12.7 双向有序且属性不同 $R \times C$ 列联表资料的统计分析           | 211        |
| 12.7.1 双向有序且属性不同 $R \times C$ 列联表资料的实例           | 211        |
| 12.7.2 用 SAS 实现双向有序且属性不同 $R \times C$ 列联表资料的统计分析 | 211        |
| 12.8 双向有序且属性相同 $R \times C$ 列联表资料的统计分析           | 213        |
| 12.8.1 双向有序且属性相同 $R \times C$ 列联表资料的实例           | 213        |
| 12.8.2 用手工实现双向有序且属性相同 $R \times C$ 列联表资料的统计分析    | 213        |
| 12.8.3 用 SAS 实现双向有序且属性相同 $R \times C$ 列联表资料的统计分析 | 214        |
| <b>第 13 章 直线相关与回归分析</b>                          | <b>215</b> |
| 13.1 直线相关与回归分析的应用                                | 215        |
| 13.1.1 直线相关与回归分析的实例                              | 215        |

|               |                                         |            |
|---------------|-----------------------------------------|------------|
| 13.1.2        | 用 SAS 实现直线相关与回归分析 .....                 | 215        |
| 13.1.3        | 直线相关与回归分析输出结果及其解释 .....                 | 217        |
| 13.2          | 可直线化回归分析的应用 .....                       | 219        |
| 13.2.1        | 简单曲线拟合的实例 .....                         | 219        |
| 13.2.2        | 用 SAS 实现简单曲线拟合 .....                    | 220        |
| 13.2.3        | 曲线拟合输出结果及其解释 .....                      | 220        |
| <b>第 14 章</b> | <b>多元回归分析 .....</b>                     | <b>224</b> |
| 14.1          | 多元回归分析概述 .....                          | 224        |
| 14.1.1        | 多元回归分析的种类 .....                         | 224        |
| 14.1.2        | 将定性变量转换成哑变量 .....                       | 224        |
| 14.1.3        | 筛选自变量的必要性及常用方法 .....                    | 225        |
| 14.1.4        | 回归诊断的概念及意义 .....                        | 225        |
| 14.1.5        | 回归方程优劣的评价 .....                         | 226        |
| 14.1.6        | 各自变量作用大小的评价 .....                       | 226        |
| 14.2          | 回归模型与回归方程 .....                         | 226        |
| 14.2.1        | 多元线性回归模型与多元线性回归方程 .....                 | 226        |
| 14.2.2        | 多元 Logistic 回归模型与多元 Logistic 回归方程 ..... | 226        |
| 14.2.3        | 多元回归分析的任务 .....                         | 227        |
| 14.3          | 多元线性回归分析的应用 .....                       | 227        |
| 14.3.1        | 多元线性回归分析的实例 .....                       | 227        |
| 14.3.2        | 用 SAS 实现多元线性回归分析 .....                  | 228        |
| 14.3.3        | 用 SAS 实现多元 Logistic 回归分析 .....          | 234        |
| 14.3.4        | 可用对数线性模型分析高维列联表资料 .....                 | 237        |
| <b>第 15 章</b> | <b>生存分析简介 .....</b>                     | <b>238</b> |
| 15.1          | 生存资料的特点 .....                           | 238        |
| 15.2          | 生存分析方法简介 .....                          | 238        |
| 15.2.1        | 统计描述 .....                              | 238        |
| 15.2.2        | 非参数检验 .....                             | 239        |
| 15.2.3        | 半参数检验 .....                             | 239        |
| 15.2.4        | 参数检验 .....                              | 240        |
| 15.3          | 用 Kaplan-Meier 法分析单因素多水平设计生存资料 .....    | 240        |
| 15.3.1        | 计算方法 .....                              | 240        |
| 15.3.2        | 应用举例 .....                              | 240        |
| 15.4          | 用 COX 模型分析多因素生存资料 .....                 | 243        |
| 15.4.1        | COX 模型回归分析概述 .....                      | 243        |
| 15.4.2        | COX 模型回归分析的应用 .....                     | 244        |
| <b>第 16 章</b> | <b>样本含量的估计 .....</b>                    | <b>247</b> |
| 16.1          | 估计样本含量所需要的前提条件 .....                    | 247        |

|               |                       |            |
|---------------|-----------------------|------------|
| 16.2          | 参数估计中样本含量的估计          | 247        |
| 16.2.1        | 估计总体均数时所需的样本含量        | 247        |
| 16.2.2        | 估计总体率时所需的样本含量         | 248        |
| 16.3          | 定量资料假设检验中样本含量的估计      | 248        |
| 16.3.1        | 采用单组设计或配对设计时所需样本含量的估计 | 248        |
| 16.3.2        | 采用成组设计时所需样本含量的估计      | 248        |
| 16.3.3        | 采用单因素多水平设计时所需样本含量的估计  | 248        |
| 16.4          | 定性资料假设检验中样本含量的估计      | 249        |
| 16.4.1        | 单组设计下检验总体率时样本含量的估计    | 249        |
| 16.4.2        | 成组设计下两总体率比较时样本含量的估计   | 250        |
| 16.4.3        | 配对设计定性资料假设检验时样本含量的估计  | 250        |
| 16.4.4        | 两总体生存率比较时样本含量的估计      | 251        |
| 16.4.5        | 多个总体生存率比较时样本含量的估计     | 251        |
| 16.5          | 直线相关分析中样本含量的估计        | 251        |
| 16.5.1        | 直线相关分析时样本含量的估计        | 251        |
| 16.5.2        | 两总体相关系数比较时样本含量的估计     | 252        |
| <b>第 17 章</b> | <b>正确应用统计学的策略</b>     | <b>253</b> |
| 17.1          | 人们误用和滥用统计学的根本原因是什么    | 253        |
| 17.1.1        | 对统计学的复杂性和严谨性认识不足      | 253        |
| 17.1.2        | 业务主管部门对统计学的重要性缺乏应有的重视 | 253        |
| 17.2          | 正确应用统计学的策略            | 253        |
| 17.2.1        | 加强正面宣传、提高认识水平         | 253        |
| 17.2.2        | 建立健全科研质量监督检查管理机制      | 254        |
| 17.2.3        | 改进教学方法、真正提高统计学教学质量    | 254        |
| 17.2.4        | 建立完善的继续教育体系           | 254        |
| 17.2.5        | 提高统计学水平的举措            | 255        |
| 17.2.6        | 正确应用统计学的思想            | 255        |
| 17.2.7        | 正确应用统计学的要领            | 255        |
| <b>附录</b>     | <b>统计用表</b>           | <b>258</b> |
|               | <b>参考文献</b>           | <b>309</b> |

1

医学统计典型  
错误辨析

第1篇



【本章要点】 了解人们在实验设计中常犯的错误,围绕实验设计的三要素(即实验因素、受试对象、实验效应)和四个基本原则(即对照、随机、重复、均衡)去查找问题,同时,不要忽视人们在实验设计中常用“单因素设计取代多因素设计”和“仅依据专业知识去设计对照组和处理组”的不良习惯。

## 1.1 指导思想上的问题

### 1.1.1 只有专业设计,缺乏统计研究设计

很多实际工作者,在开展实验研究之前,只从专业角度考虑问题,很少从统计学角度考虑问题,以至于仅根据“1、2个”动物或患者的观测结果,就作出带有规律性的结论来;从一些中高档科研成果的档案中我们不难看到一卷又一卷的科研流水账,但很难寻觅到完善的实验设计方案;有些研究工作从表面上看似乎也遵守了实验设计的四个基本原则,但每一个原则都遵守得不严,以至于实验设计漏洞百出。显然,在这样的实验方案的指导下,得到的结果和结论恰似“墙上芦苇,头重脚轻根底浅”,有时甚至是“荒唐透顶,颠倒黑白”!

### 1.1.2 临床科研流程倒置

笔者在审阅大量临床研究论文时,产生的印象是很多科研工作的“流程”倒了。按理说,进行一项科研工作,首先应当有科研假设,对科研假设深思熟虑之后,提出研究目标和实现目标的思路。在此基础上,结合专业知识和统计学知识,制定出完善的统计研究设计方案,在方案的指导下,有的放矢、循序渐进地开展科研工作。然而,很多临床科研工作采取的是“回顾性研究策略”。也就是说,临床研究工作已经结束了,才撰写所谓的“临床科研论文”。这种回顾性的研究工作,往往因事先缺乏完善的统计研究设计,在“随机、对照、重复、均衡”四个实验设计原则方面存在严重问题,以至于推理证据不充分,说理底气不足,其结论难以令人置信。有些科研工作,由于在实验设计方面、资料的统计分析方面存在严重错误,导致其得出了错误的结论。

### 1.1.3 缺乏完善的实验设计方案

原文题目:静息细胞法研究影响西索米星生物合成因素,原作者需要摸索不同的条件以筛选出较为理想的静息细胞培养基,静息细胞培养基的组成应采用尽可能简单的已知成分

的原料,能维持菌种最低限度的生命活动和继续合成西索米星(SISO)的能力。通过表 1-1-1 列出的不同成分与配比的培养基进行实验,从中可以看出 1#、2#配方发酵效价较高但成分较复杂,菌体总核酸( $\Delta C_{NA}$ )较大,菌丝生长不符合静息的要求,说明天然成分虽能提高效价,但菌体也增加,一般不采用;3#配方发酵效价较 5#低,在促进 SISO 合成能力方面不理想;4#、6#由于缺乏  $Mg^{2+}$ ,合成产物能力较差,说明  $Mg^{2+}$  与合成产物的酶类的活性有关,并与细胞的通透性有关, $Mg^{2+}$  的存在有利于产物分泌,5#为成分完全已知的较为理想的静息细胞培养基。然后以 5#培养基为基础,选择不同种龄(26~42 小时),均在 34℃ 培养 20 小时后测其生物效价,结果发现种龄不足 26 小时生物效价很低,38~40 小时达高峰,故确定 40 小时种龄的菌丝为静息培养系统的种子。

表 1-1-1 静息细胞培养基配方的筛选

| 成分(%)                  | 1#    | 2#   | 3#   | 4#    | 5#   | 6#    |
|------------------------|-------|------|------|-------|------|-------|
| 淀粉                     | 1.0   | X    | X    | X     | X    | X     |
| 酵母膏                    | 1.5   | X    | X    | X     | X    | X     |
| 玉米浆                    | 1.0   | 0.5  | X    | X     | X    | X     |
| 葡萄糖                    | 1.0   | 1.0  | 1.0  | 1.0   | 1.0  | X     |
| 硫酸镁                    | 0.1   | 0.1  | 0.1  | X     | 0.2  | X     |
| 氯化钠                    | 1.0   | 0.5  | 0.5  | 0.5   | 0.5  | 0.5   |
| 硝酸钾                    | 0.5   | 0.5  | 0.1  | X     | 0.5  | 0.1   |
| 碳酸钙                    | 0.5   | 0.5  | 1.0  | 0.2   | 0.5  | 0.3   |
| 生物效价(u/ml)             | 139.7 | 68.9 | 43.6 | 32.4  | 64.3 | 20.1  |
| $\Delta C_{NA}$ (g/5L) | 2.01  | 0.73 | 0.37 | -0.15 | 0.07 | -0.18 |

“X”表示各号配方中未用的左边相应的成分。

为了确定静息培养基的最佳 pH,又用 5#培养基,40h 种龄的菌种固定其他条件,在不同 pH 下进行发酵,得到结果认为以 pH7.0~7.2 时对 SISO 合成最有利。又固定其他条件,用不同培养时间进行实验,确定最佳培养时间为 20 小时。在确定了建立静息细胞培养系统的条件后,原作者又采取类似的做法,来摸索影响 SISO 合成的条件。最佳条件的摸索都是先固定其他条件,然后比较某一因素的不同水平的取值,选择取值最大的所对应的水平为最适的水平。这样依次确定的最适条件为:甲硫氨酸浓度为 0.1%,磷酸盐浓度为 0.03%,开始合成 SISO 的时间是培养后 30 小时。

**【辨析与释疑】** 从文章中可以看出原作者的目的主要有两个,一是筛选出建立静息细胞培养系统的最优实验条件,另一个是在此培养系统的基础上进一步找出合成 SISO 的最适条件。原文涉及的因素众多,而原作者仅做了 6 批实验,就认为 5#方案是最优方案,从某种意义上讲,这只能算是预试验,通过这 6 批试验,认为不宜采用淀粉、酵母膏等天然成分,这样可以在实验设计中不加天然成分,并可根据预实验的结果,确定各因素不同水平的取值。原作者在确定每个因素最佳水平时,总是先固定其他因素在各自的某个水平上,每次只让该因素的水平变化,等找到该因素的所谓“最佳水平”时就固定不变了,用同样的思路再依次去寻找其他因素的所谓“最佳水平”,最后将全部因素的所谓“最佳水平”组合在一起,就得出所谓“最优实验条件”。这就是人们进行多因素实验研究时习惯的做法,即用单因素设计取代多因素设计。一般说来,当因素之间存在一定程度的交互作用时,各因素不同水平的组合对结果的影响较大,用多次单因素实验所得出的条件组合起来,其结果往往并不是很理想