

卫生部规划教材

全国中等卫生学校教材

供医学检验专业用

# 医用统计学

第一版

主编 严日树



浙江科学技术出版社

封面设计：王士忠

全国中等卫生学校教材

**医用统计学**

供医学检验专业用

第 一 版

主 编 严日树

编者 严日树 胡中杰 钱伟钧 高越明

\*

浙江科学技术出版社出版

淳安千岛湖环球印务有限公司印刷

浙江省新华书店发行

\*

开本 787×1092 1/16 印张 12 字数 270000

1997年12月第一版

1997年12月第一次印刷

**ISBN 7-5341-0979-5/R · 159**

**定 价：10.20 元**

### 第三轮中等医学教材出版说明

卫生部曾于1983年组织编写、陆续出版全国中等卫生学校11个专业使用的77种教材。1992年又组织小修订，出版第二轮教材。为我国的中等医学教育作出了积极贡献。

为适应中等医学教育改革形势的需要和医学模式的转变，1993年11月，卫生部审定、颁发了全国中等卫生学校新的教学计划及教学大纲。在卫生部科教司领导下，我们组织编写（修订）出版第三轮全国中等医学12个专业96种规划教材，供各地教学使用。

本轮教材以培养中级实用型卫技人才为目标，以新的教学计划及大纲为依据，体现“思想性、科学性、先进性、启发性、适用性”，强调“基本理论知识、基本实践技能、基本态度方法”。教材所用的医学名词、药物、检验项目、计量单位，注意规范化，符合国家要求。

编写教材仍实行主编负责制；编审委员会在教材编审及组织管理中，起参谋、助手、纽带作用；部分初版教材和新任主编，请主审协助质量把关。第三轮中等医学教材由人民卫生、河北教育、山东科技、江苏科技、浙江科技、安徽科技、广东科技、四川科技和陕西科技九家出版社出版。

希望各校师生在使用规划教材的过程中，提出宝贵意见，以便教材质量能不断提高。

卫生部教材办公室

1995年10月

### 全国中等医学教材编审委员会

主任委员：姜寿葆

副主任委员：陈咨夔 殷冬生

委员：（以姓氏笔画为序）

|     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|
| 马惠玲 | 王同明 | 方茵英 | 王德尚 | 延民  | 那功伟 |
| 朱国光 | 吕树森 | 李绍华 | 李振宗 | 李振林 | 陈心铭 |
| 吴忠礼 | 杨华章 | 洪启中 | 洪思勉 | 郭常安 | 张冠玉 |
| 张审恭 | 殷善莹 | 董品泸 | 谭筱芳 |     |     |

## 前 言

随着医学的发展,医用统计方法作为医学卫生工作者必须掌握的工具,已得到普遍承认,在医学检验工作中亦得到广泛应用。《医用统计学》是全国中等卫校医学检验专业新开设的一门必修课程。

本教材原则上根据1994年卫生部新颁医学检验专业《医用统计学》大纲编写,对大纲部分内容报卫生部教材办公室和教材编审委员会审批后,作了完善性的修改。编写的指导思想是:贯彻具有必要的理论知识、较强的实践技能、良好的职业素质的中等医学检验人材的培养目标。编写特点是:密切结合医学实际,深入浅出,通俗易懂,力求充分体现思想性、实用性、科学性、启发性、先进性。全书共计13章,包括绪论、各种资料类型的统计描述和统计推断,专章讲述统计表和统计图、医学正常值范围的制订方法、实验设计、统计方法在实验室质量控制中的应用;并编写了实习作业,包括函数型计算器的应用、复习思考题、实习题,以提供学生掌握知识情况的反馈信息。各校可根据具体情况,把实习作业安排在课内或课外完成,建议将第十三章的部分内容安排在实习前以讲座形式进行。

本书第一、三、五、八、九、十三章由严日树、高越明编写或合作编写;第六、十、十二章由胡中杰编写;第二、四、七、十一章由钱伟钧编写;实习作业由编者共同编写。严日树、高越明参加统稿工作,高越明兼学术秘书。本书插图由钱伟钧、高越明绘制,浙江省卫生学校严中浩、范春红帮助把初稿输入电脑,卫生医士9319班同学核对了计算数据。

本书的编写得到了浙江省卫生厅科教处汤菟菟副处长、邵祥珍主任,浙江省卫生学校朱金元校长和苏州卫生学校以及浙江省卫生统计学会领导的大力支持,受到了上海医科大学等高等医学院校教授们的热情指导,张照贵教授审阅了较简便的小样本正态性检验方法,并给予充分肯定。浙江省卫生厅科教处原处长王雪燕,浙江省卫生学校原主任严启之,浙江医科大学谢隆化教授,浙江医学科学院何渐生研究员,浙江省卫生防疫站莫世华副站长等,十分关心本书的编写。浙江省临床检验中心的领导和同志们以及浙江省人民医院检验科周永利副主任提供了大量实际工作数据,为本教材体现实用性、科学性、先进性提供了重要保证。此外,在本书编写过程中还得到全国部分医疗、卫生防疫单位的同志们们的热情帮助,他们有的提供现场工作的实验数据,有的提供许多有益的意见。在此,全体编者对支持、指导、帮助、关心我们编写工作的同志们表示衷心的感谢。

本教材亦可作为各类医学卫生工作者的参考书,并可作为举办医学统计方法培训班及其他专业医用统计学选修课的教材。本书的基本内容,曾在浙江省杭州市各类医务人员初级升中级医用统计学考前辅导班上作了讲授,受到学员好评。殷切希望使用本教材的师生和同志们,多提宝贵意见,以便在再版时修正。

严日树

1996年8月于浙江省卫生学校

# 目 录

|                                            |    |
|--------------------------------------------|----|
| <b>第一章 绪论</b> .....                        | 1  |
| 第一节 医用统计学的学科性质.....                        | 1  |
| 第二节 统计的基本步骤.....                           | 2  |
| 第三节 统计学的几个基本概念.....                        | 4  |
| 第四节 统计资料的类型.....                           | 6  |
| <b>第二章 统计表与统计图</b> .....                   | 8  |
| 第一节 统计表.....                               | 8  |
| 第二节 统计图.....                               | 10 |
| <b>第三章 计量资料的集中趋势和离散程度</b> .....            | 17 |
| 第一节 频数表.....                               | 17 |
| 第二节 集中趋势指标.....                            | 19 |
| 第三节 离散程度指标.....                            | 23 |
| <b>第四章 正常值范围的制订方法</b> .....                | 25 |
| 第一节 正态分布的概念.....                           | 25 |
| 第二节 制定正常值范围的原则和步骤.....                     | 28 |
| 第三节 正态分布资料制订正常值的统计方法.....                  | 29 |
| 第四节 偏态分布资料制订正常值的统计方法.....                  | 31 |
| <b>第五章 总体均数的估计和假设检验</b> .....              | 34 |
| 第一节 均数的抽样误差.....                           | 34 |
| 第二节 $t$ 分布.....                            | 34 |
| 第三节 总体均数的可信区间估计.....                       | 36 |
| 第四节 两均数比较的假设检验.....                        | 36 |
| 第五节 正态性检验.....                             | 42 |
| 第六节 假设检验应注意事项.....                         | 44 |
| <b>第六章 方差分析</b> .....                      | 47 |
| 第一节 方差分析的概念.....                           | 47 |
| 第二节 成组设计的多个样本均数比较（单因素方差分析）.....            | 47 |
| 第三节 随机区组设计的多个样本均数比较（两因素方差分析）.....          | 49 |
| 第四节 多个样本均数间的两两比较.....                      | 51 |
| <b>第七章 相对数</b> .....                       | 54 |
| 第一节 相对数的种类与计算.....                         | 54 |
| 第二节 应用相对数时应注意的事项.....                      | 56 |
| <b>第八章 <math>\chi^2</math>（卡方）检验</b> ..... | 59 |
| 第一节 四格表资料的 $\chi^2$ 检验.....                | 59 |

|                  |                              |     |
|------------------|------------------------------|-----|
| 第二节              | 配对计数资料的 $\chi^2$ 检验          | 61  |
| 第三节              | 行 $\times$ 列表资料的 $\chi^2$ 检验 | 63  |
| <b>第九章</b>       | <b>二项分布与泊松分布及其应用</b>         | 66  |
| 第一节              | 二项分布的概念及其应用条件                | 66  |
| 第二节              | 二项分布在医学上的应用                  | 69  |
| 第三节              | 泊松分布的概念及其应用条件                | 74  |
| 第四节              | 泊松分布在医学上的应用                  | 77  |
| <b>第十章</b>       | <b>秩和检验</b>                  | 82  |
| 第一节              | 配对比较的符号秩和检验 (Wilcoxon 配对法)   | 82  |
| 第二节              | 两样本比较的秩和检验 (Wilcoxon 两样本比较法) | 83  |
| 第三节              | 多个样本比较的秩和检验                  | 86  |
| <b>第十一章</b>      | <b>相关与回归</b>                 | 89  |
| 第一节              | 直线相关                         | 89  |
| 第二节              | 标准曲线的直线回归拟合法                 | 92  |
| 第三节              | 拟合和应用回归直线时应注意的事项             | 96  |
| 第四节              | 等级相关                         | 98  |
| <b>第十二章</b>      | <b>实验设计</b>                  | 102 |
| 第一节              | 实验设计的概念和基本要素                 | 102 |
| 第二节              | 实验设计的基本原则                    | 102 |
| 第三节              | 实验设计中常用的几种随机化方法              | 104 |
| 第四节              | 样本含量的估计                      | 108 |
| <b>第十三章</b>      | <b>统计方法在实验室质量控制中的应用</b>      | 111 |
| 第一节              | 实验室内部的质量控制                   | 111 |
| 第二节              | 实验室间的质量控制                    | 121 |
| 第三节              | 质量控制中应用统计方法应注意事项             | 126 |
| <b>医学统计学实习作业</b> |                              | 128 |
|                  | 函数型电子计算器使用方法                 | 128 |
|                  | 复习思考题                        | 133 |
|                  | 实习题                          | 136 |
| <b>统计附表</b>      |                              | 153 |
| 附表 1             | 标准正态曲线下的面积表                  | 153 |
| 附表 2             | $t$ 界值表                      | 155 |
| 附表 3             | $F$ 界值表 (方差齐性检验用)            | 156 |
| 附表 4             | $F$ 界值表 (方差分析用)              | 157 |
| 附表 5             | $q$ 界值表                      | 163 |
| 附表 6             | 百分率的可信限                      | 164 |
| 附表 7             | 泊松分布均数的可信限                   | 168 |
| 附表 8             | $\chi^2$ 界值表                 | 169 |
| 附表 9             | 相关系数 $r$ 界值表                 | 170 |

|       |                           |     |
|-------|---------------------------|-----|
| 附表 10 | 等级相关系数 $r_s$ 界值表          | 172 |
| 附表 11 | 配对比较的秩和检验 $T$ 界值表         | 173 |
| 附表 12 | 两组比较的秩和检验 $T$ 界值表         | 174 |
| 附表 13 | 二组比较的秩和检验 $U$ 界值表         | 176 |
| 附表 14 | 两样本率比较时所需样本含量             | 177 |
| 附表 15 | 配对比较 ( $t$ 检验) 时所需样本含量    | 179 |
| 附表 16 | 两样本均数比较 ( $t$ 检验) 时所需样本含量 | 180 |
| 附表 17 | 随机数字表                     | 181 |
| 附表 18 | 随机排列表 ( $n=20$ )          | 182 |

# 第一章 绪 论

## 第一节 医用统计学的学科性质

### 一、随 机 事 件

在观察自然和社会现象中,有些是必然发生的,也就是确定发生的现象,例如把石块投在水中必然下沉现象;异性电荷必然相吸现象;人到一定年龄必然要死亡的现象;用变质的试剂去测定某种物质,其数据必然是不准确的。但是更多遇到的是一种可能发生亦可能不发生的现象,也就是不确定的现象,例如吸烟的人生肺癌的现象;与传染性肝炎病人密切接触发生传染性肝炎的现象;用某种检验方法去检验疟疾病人的血样,能找到疟原虫的现象;在一个装有红、白球数相等的袋中,摸取若干个球,其红、白球比例相等的现象。我们通常把不确定发生的现象称为随机现象,还对应的把随机现象是否发生称为随机事件(random matter)是否发生。上述吸烟的人生肺癌;与传染性肝炎病人密切接触发生传染性肝炎;用某种检验方法去检验疟疾病人的血样,能找到疟原虫;在一个装有红白球数相等的袋中,随机抽取若干个球,其红、白球比例相等的事件都是随机事件。也就是可能发生,也可能不发生的事件称为随机事件。

### 二、统计推断

随机事件的发生虽然是不确定的,但应用统计方法,通过对部分事件的观察可发现其有统计规律性,称为统计推断(statistical inference)。例如,在其他条件基本控制均衡的前提下,对吸烟的人和不吸烟的人作对比观察,可作出吸烟者生肺癌的可能性远超过不吸烟者;对密切接触传染性肝炎病人和不密切接触者作对比观察,可作出密切接触者生肝炎的可能性超过不密切接触者;对某种检验方法检查疟疾病人的血样中是否找到疟原虫,可推断其阳性符合率一个范围;从一个装有红、白球数相等的袋中作随机抽取每次球数相等的 $n$ 次观察,可发现袋中红、白球比例数相等或相近出现的频数较多。

要特别指出的是,通过局部情况的观察对整体进行统计推断,必须应用正确的统计方法。例如,用两种检验方法各检验100个阳性标本,甲法检出阳性70个,乙法检出阳性60个,甲法阳性符合率为70%,乙法阳性符合率为60%。如果不对局部的信息进行统计处理,就作出甲法阳性符合率高于乙法的统计推断,是不科学的。

这一点从下面的例子也同样可以说明:如果从两个装有红、白球比例相等且球数很多的袋中,各摸取100个球,则每个袋中摸出的100个球中,红、白球比例不一定相等。这就意味着从局部观察红、白球比例是不等的,并不意味着整体中红、白球比例是不等的。



### 三、统计学的学科性质

统计学 (statistics) 是研究资料 and 信息的搜集、整理、分析和推断的一门科学, 它以概率论为理论基础, 从观察实际资料出发, 由局部推断整体, 从而揭示随机现象的统计规律性。我们学了统计方法以后就会知道, 仅仅根据上面所述的两种检验方法的局部信息, 不经过数据处理, 还不能认为甲法的阳性符合率高于乙法。

把统计学的理论和方法应用于医学中, 就形成统计学的一个分支, 亦就是医用统计学 (medical statistics)。由于在医学中会遇到大量的随机事件, 因此医用统计学是探索其规律的一门必不可少的工具, 每一个医学工作者都应掌握这门学科的基本理论方法, 以便更好地做好医疗卫生工作。

## 第二节 统计的基本步骤

### 一、搜集资料

按搜集资料 (collection of data) 的性质, 可分为专题调查研究资料和实验研究资料。专题调查研究是对研究对象不进行任何干预, 客观反映实际情况, 如研究某地高级知识分子的血清胆固醇水平; 研究某地小学生的贫血患病率等。实验研究是在研究者根据研究目的对研究对象人为的进行一些干预措施, 例如为研究喝牛奶是否能增加儿童身高, 把儿童分为若干对, 每一对儿童的其他情况, 如性别、年龄、原来身高、父母身高等都基本相近, 随机地把每对儿童的一个分入实验组, 另一个分入对照组, 实验组除食用和对照组相同的饮食外, 每天增加 500ml 牛奶, 过若干时间测定其身高增加情况, 以判断喝牛奶有否增加身高的作用。

例 1.1 为研究某地高级知识分子血清胆固醇水平, 随机抽查某地参加体检的 132 名高级知识分子血清胆固醇 (mmol/L) 的原始数据如下:

|      |      |      |      |      |      |
|------|------|------|------|------|------|
| 4.63 | 4.57 | 4.19 | 6.15 | 6.90 | 5.09 |
| 3.51 | 4.16 | 5.56 | 3.39 | 6.10 | 6.00 |
| 6.31 | 5.56 | 4.21 | 5.92 | 4.63 | 5.30 |
| 4.73 | 5.30 | 4.88 | 5.06 | 5.69 | 5.92 |
| 5.48 | 5.95 | 4.57 | 4.78 | 5.51 | 4.81 |
| 5.43 | 5.82 | 4.31 | 5.27 | 4.24 | 3.93 |
| 4.50 | 4.83 | 3.75 | 5.45 | 4.55 | 4.24 |
| 5.35 | 5.35 | 5.20 | 5.18 | 6.41 | 6.59 |
| 6.00 | 5.74 | 4.70 | 5.89 | 5.30 | 4.45 |
| 5.43 | 5.58 | 5.27 | 5.87 | 4.86 | 6.10 |
| 5.76 | 4.55 | 4.26 | 4.55 | 6.96 | 5.95 |
| 4.34 | 5.22 | 5.84 | 5.40 | 4.21 | 4.11 |
| 3.13 | 4.55 | 6.10 | 5.12 | 4.39 | 4.45 |
| 5.43 | 5.71 | 5.97 | 5.38 | 5.27 | 5.04 |
| 4.26 | 4.86 | 4.03 | 5.40 | 4.52 | 3.95 |

|      |      |      |      |      |      |
|------|------|------|------|------|------|
| 5.87 | 5.22 | 5.95 | 3.57 | 4.47 | 4.14 |
| 3.10 | 5.95 | 5.71 | 5.53 | 5.82 | 5.74 |
| 5.43 | 4.78 | 4.14 | 4.37 | 5.12 | 4.70 |
| 4.14 | 4.94 | 5.10 | 7.16 | 4.68 | 5.01 |
| 5.09 | 5.32 | 5.09 | 5.82 | 3.67 | 5.56 |
| 6.21 | 6.23 | 6.12 | 6.15 | 6.08 | 6.62 |
| 6.29 | 6.27 | 6.38 | 6.35 | 6.42 | 6.88 |

## 二、整 理 资 料

从搜集资料得到的原始资料往往是零乱的、分散的,需要进行整理,使其系统化、条理化,以便于统计分析。整理资料(sorting data)一般可分为以下几个步骤。

### (一) 检查核对

观察资料项目填写是否完整正确、有否遗漏,如一般项目漏填了姓名、性别;胆固醇记录数据中出现 716,就需检查是否漏了小数点,在核对时发现问题应及时补正。

### (二) 设计分组

1. 数量分组 如按年龄、身高值、血色素值、血压值、胆固醇值分组。
2. 质量分组 如按职业、性别、疾病种类等分组。

### (三) 拟订整理表

根据研究目的拟定合适的整理表,如需了解参加体检的高级知识分子的血清胆固醇值的分布特征,可按血清胆固醇值分组,拟订整理表。如表 1·1 和表 1·2。

表 1·1 某医院 132 名参加体检的高级知识分子血清胆固醇值分布

| 血清胆固醇值 (mmol/L) | 划记 | 频数 |
|-----------------|----|----|
| 3.0~            |    |    |
| 3.5~            |    |    |
| 4.0~            |    |    |
| .....           |    |    |
| .....           |    |    |
| 7.0~            |    |    |

表 1·2 某小学 100 名小学生蛔虫感染情况

| 性别 | 阳性划记 | 频数 |
|----|------|----|
| 男  |      |    |
| 女  |      |    |

### (四) 归组汇总

把各变量值划记归属在各个组,如将上述血清胆固醇值原始数据分别划记在各个组段,可得各组的频数(人数)。

## 三、统计资料的分析

搜集得到的原始资料经整理后就显得比较条理化、系统化,可根据研究目的,应用

统计方法揭示分布特征,进行比较分析,如根据上述资料可计算其集中趋势指标,如算术平均数;离散程度指标,如标准差;并可推断该地高级知识分子血清胆固醇总体平均水平在那一个范围。还可比较不同消毒方法的消毒效果,两种检验方法的优劣,送检时间不同的标本对检出率的影响等。

必须指出,如何进行搜集资料、整理资料、分析资料,事先应该有一个良好的统计全过程设计。

### 第三节 统计学的几个基本概念

#### 一、总 体

根据研究目的确定的相同性质的所有观察单位某种指标的变量值的集合称为总体(population)。如要研究某地高级知识分子血清胆固醇情况,该地每一个高级知识分子就是一个同质的观察单位,所有高级知识分子的血清胆固醇值就形成一个总体。

#### 二、样 本

从总体中随机抽取的一部分称为样本(sample)。所谓随机抽取亦就是总体中每一个观察单位都有同等可能被抽到,随机抽样的目的是为了减少偏性,使样本对总体有代表性,以便应用统计方法由样本推断总体。常用的抽样方法有四种:

##### (一) 单纯随机抽样

所谓单纯随机抽样(simple random sampling),即把总体中的观察单位编号,然后用抽签的方法或应用随机数字表抽取。例如,要在100个观察单位中随机抽取10个观察单位,可先把100个观察单位从1号编到100号,然后在随机数字表中任一行、任一列开始,例如从第21行第11列开始,连续取10个数字,遇到有重复的数字就删去,00即为100,这样就可得10个数字:68, 47, 22, 00, 20, 35, 55, 31, 51, 83。把原编号的相应的观察单位抽出,就是一个样本。

##### (二) 机械抽样

所谓机械抽样(systematic sampling),即把观察单位编号后,每隔若干个观察单位抽取一个观察单位。例如,从1000个观察单位需抽1:10样本,可先把观察单位编为1至1000号,然后从0至9编10个号,随机抽取一个号,例如为3,则每逢个位数为3就作为被抽到的观察单位,这样由3, 13, …… 993号就组成观察单位数100的样本。

##### (三) 分层抽样

所谓分层抽样(stratified sampling),即在各层中按原观察单位数按比例抽取组成一个样本。例如某地区有平原、山区两个地形,平原有700000人,山区有300000人,若为研究某病的发病率与地形的关系,想抽取观察单位数的百分之一组成一个样本,就应按比例在平原抽取7000人,在山区抽取3000人,组成一个观察单位数为1万的样本。

##### (四) 整群抽样

所谓整群抽样(cluster sampling),即在群体中随机抽取若干个群体作为样本。在抽到的群体中,其每一个观察单位都进行调查,在未抽到的群体中,一个观察单位也不调查,例如为研究某地中学生沙眼感染情况,如在各所中学沙眼感染分布情况基本上是均

匀的,可把每一所中学作为一个群体,对抽到的中学中的每一个学生都调查有无感染沙眼,未抽到的中学,一个亦不进行调查。

### 三、概 率

概率(probability)是说明一个事件发生的可能性,如在 $n$ 次试验中某事件发生了 $m$ 次,则 $m/n$ 称为某事件发生的频率,如当 $n$ 趋向无穷大时, $m/n$ 无限趋近于某常数 $P$ ,则把 $P$ 称为该事件发生的概率, $0 \leq P \leq 1$ 。在一般情况下,当一个事件的概率小于0.05时,称为小概率事件,而小概率事件在一次试验中,基本上是不会发生的。例如,在装有999个红球和1个白球的袋子中,随机摸取一个球,摸得白球的事件基本上是不会发生的。这个准则很重要,根据样本信息去推断总体情况,都需要应用这个准则。

### 四、误 差

误差(error)就是测得值和真值的差,可分为抽样误差(sampling error)、随机测量误差(random measuring error)和系统误差(systematic error)。

#### (一) 抽样误差

由于抽样原因引起的样本指标和总体指标的差别称为抽样误差。例如,经普查得某地儿童蛔虫感染率为30%,如果从中抽取100个儿童的样本,其蛔虫感染率不一定就等于30%。

#### (二) 随机测量误差

同一个样品重复检验 $n$ 次,所得的测量值不一定和样品的真值相等,彼此测量结果亦不一定相等,这种误差称为随机测量误差。

上述两种误差是不能避免的,但其大小可以由样本含量及检测仪器精度等控制。

#### (三) 系统误差

由于确定原因引起的误差称为系统误差。例如天平在检验前如未校正好,每次称重都会在一边出现偏大或偏小的情况,检验中常因试剂变质,工作方法失误等原因引起系统误差,克服引起系统误差的原因,一般就可以消除系统误差。

### 五、有效数字

#### (一) 有效数字的意义

把一个真值经舍入法后得到的近似数,它与真值的误差的绝对值,不会超过最末一个数字的半个单位,这个数从第一个非零数字起,所有的数字都是有效数字。例如,把真值2.4153经舍入法到小数第3位(千分之一位)得2.415,就有4个有效数字,精确到千分之一;把真值0.00756经舍入法到小数第4位(万分之一位)得0.0076,就有2个有效数字,精确到万分之一。由此可见,有效数字多的并不意味着精确度高。必须注意近似数2.5和2.50有不同的含意,前者有2个有效数字,精确到小数第1位,它与真值的误差的绝对值不超过0.05,后者有3个有效数字,它与真值的误差绝对值不超过0.005。对于准确值来说,可以写成任意多个有效数字,可以精确到任意位。对于一个最小刻度单位为1ml的量筒,量得的数值,其有效数字到个位数,例如,写成12ml,有效数字为2个,而不能写成12.0ml。

## (二) 有效数字的取舍原则和运算规则

1. 取舍原则 要遵照“四舍六入，五前奇入，偶舍”的取舍原则。如果把一个近似数的最后一个有效数字定在小数第  $n$  位，则在小数第  $(n+1)$  位的数字为 4，就舍；数字为 6，就入。若在小数第  $(n+1)$  位的数字为 5，再以后不全为零则入，若再以后全为零，则小数第  $n$  位的数字是奇数就入，是偶数就舍。例如把一个近似数的最后一个有效数字定在小数第 2 位，则

$$61.044 \rightarrow 61.04$$

$$61.046 \rightarrow 61.05$$

$$61.0651 \rightarrow 61.07$$

$$61.035 \rightarrow 61.04$$

$$61.065 \rightarrow 61.06$$

2. 加减法运算规则 以小数位数最少（精确度最低）的作为标准，其他比作标准的数多保留一位小数，最后结果和小数位数最少的位数相同。

$$\text{例如 } 74.0 + 0.056 + 0.006 = 74.0 + 0.06 + 0.01 = 74.07 = 74.1$$

3. 乘除法运算规则 以有效数字最少的作为标准，其他的比作标准的数字多保留一个有效数字，最后结果和有效数字最少的相同。

$$\text{例如，当半径 } r = 0.0060 \text{ 时，圆周长 } C = 2\pi r = 2.00 \times 3.14 \times 0.0060 = 0.03768 = 0.038$$

注意，这里 2 是准确数，有效数字可取任意多个，作标准的数  $r = 0.0060$ ，有 2 个有效数字，其结果保留 2 个有效数字。

4. 乘方、开方运算法则 乘方和开方的运算结果和原被乘方数和开方数字的有效数字相同。

$$\text{例如 } (2.310)^2 = 5.3361 = 5.336$$

$$\sqrt{2.310} = 1.519868415 = 1.520$$

如用计算器进行连续运算时，中间步骤可不作取舍，但结果保留的数字要按照运算法则。

## 第四节 统计资料的类型

我们把研究事物的指标称为变量，可以分为数值变量和定性变量，按其所取变量值可得到统计资料的三种类型。

### 一、计量资料

所谓数值变量，即其变量值可以取数值，例如我们欲研究某地健康成年男子血色素的情况，则血色素就是变量，它所取的变量值是数值的。该地每一个健康成年男子的血色素值都是取数值的，这种资料称为计量资料。身高、体重、体温、血压、红细胞总数等资料都为计量资料。上述 132 名高级知识分子血清胆固醇值，亦是计量资料。

### 二、计数资料

所谓定性变量（分类变量），即其变量值是定性的。如果其变量值取无等级顺序的二

项分类，或多项分类，例如我们如要研究某地成年男子钩虫感染情况，则钩虫卵检查结果就可以看成变量，其变量值是定性的，可能取阳性或阴性两类；又如我们要研究某地人群的血型情况，则血型是变量，其变量值可以取A、B、O、AB四类。如检查1000人中，经清点钩虫卵阳性50人，阴性950人；又如94个人的血型中A型30人，B型29人，O型31人，AB型1人，这种资料称为计数资料。

### 三、等 级 资 料

如果定性变量的变量值取有等级顺序的几类，例如临床医学中的疗效是定性变量，如其变量值取治愈，显效，好转，无效有等级顺序的四类，在治疗65人中，经清点，治愈30人，显效20人，好转10人，无效5人；又如，在检验中可把肾炎病人尿中的红细胞数作为变量，它可以取一，二，三，四，五，有等级顺序的几类，经检验，在70名肾炎病人中，红细胞数为十有20人，二十有20人，三十有25人，四十有5人。这种资料称为等级资料。

不同类型的统计资料，应采用不同的统计方法进行数据处理。

## 第二章 统计表与统计图

统计资料的表达形式通常采用文字、表格和图表三种，而统计表和统计图是表达数据关系和特征的重要工具。统计表可以代替冗长繁琐的文字叙述，便于分析比较。统计图是用点、线、面积等图形反映事物间的数量关系和变化情况，具有生动、形象和一目了然的特点，仅在图中反映的数字显得粗略一些。

### 第一节 统计表

统计表 (statistical table) 是表达统计分析的事物及其指标的一种表格。正规合理的统计表，可以代替长篇的文字叙述，直接将众多的有关联的统计数据和结果简明扼要地表达出来，使数据系统化、条理化，便于理解、计算、分析和对比。这里指的统计表，是指分析阶段的统计表。

#### 一、统计表的结构与绘制要求

一张统计表的基本结构应包括：标题、标目、线条和数字。  
对统计表的各结构部分的要求如下：

##### (一) 总体要求

统计表简明，重点突出，便于分析比较。

##### (二) 标题 (title)

标题要简明确切，用一句话概括说明表的主要内容，必要时应注明资料的时间、地点，标题应写在表的上端中央，文章中的附表应有表号，表号的位置与标题同行。

##### (三) 标目

标目是表格内的项目，可分横标目和纵标目。横标目表示表中被研究事物的主要标志，并列在表的左侧；纵标目用来说明横标目的各个统计指标的内容，列在表的上端。简言之，横标目与纵标目要主谓分明，层次清楚，主语在左，谓语在右。

标目的文字应简明，指标的计算单位应当用括号注明单位，如年龄组 (岁)、身高 (cm)、阳性率 (%)、出生率 (‰)。

标目内容一般应按顺序，从小到大排列，小的放在上面，时间数列的动态资料可按年份、月份先后排列，以揭示事物的规律性。

##### (四) 线条

一般除顶线和底线外，还有纵标目下面和合计上面的横线，即表内主要横线最多四条，其余线条均可省略。在表中如只需个别列要合计数字，可在相应列的合计栏上方画一短线即可。

##### (五) 数字

表内数字必须准确，一律用阿拉伯数字表示，同一列的小数点位数保持一致，位次对齐。表内不留空格，不出现文字，如不应出现“同左”、“同上”等字样。数字若为

统计

“0”，则填写“0”，若为暂缺或未记录可用“……”表示。

#### (六) 备注

备注不是表的必要结构，一般不列入表内，特殊情况须用备注说明时，可用“\*”等符号标出，在表格的下面用文字说明。

## 二、统计表的种类

常用的统计表可分为简单表、复合表和相关表。

### (一) 简单表

只按一种主要标志分组的统计表称简单表。即由一组横标目和一组纵标目组成，以表示它们之间的关系。如表 2·1。

表 2·1 中枢神经系统埃博病毒感染干扰素检测情况

| 病 原      | 病例数 | 急性期干扰素阳性数 | 阳性率(%) |
|----------|-----|-----------|--------|
| ECHO 病毒  | 15  | 29        | 61.4   |
| 风疹病毒     | 17  | 12        | 70.6   |
| 流行性腮腺炎病毒 | 67  | 60        | 89.6   |
| 水痘带状疱疹病毒 | 11  | 11        | 100.0  |
| 病因不明     | 6   | 4         | 66.7   |
| 合 计      | 116 | 116       | 79.5   |

### (二) 复合表

按两个或两个以上主要标志结合起来分组的统计表，称复合表。通常由两组及两组以上的横标目和纵标目相结合，以表示它们之间的关系。如表 2·2。

表 2·2 血清抗-HCV 在各型肝炎中分布及与输血的关系

| 临床分型    | 病例数 | 抗-HCV 阳性数 | 阳性率(%) | 有输血史者抗-HCV 阳性数 | 输血占阳性(%) |
|---------|-----|-----------|--------|----------------|----------|
| 急性肝炎    | 127 | 27        | 21.26  | 23             | 85.19    |
| 慢性迁延性肝炎 | 12  | 1         | 8.33   | 0              | 0.00     |
| 慢性活动性肝炎 | 30  | 4         | 13.33  | 1              | 25.00    |
| 亚急性重型肝炎 | 5   | 2         | 40.00  | 1              | 50.00    |
| 慢性重型肝炎  | 22  | 11        | 50.00  | 0              | 0.00     |
| 肝炎后肝硬化  | 13  | 6         | 46.15  | 2              | 33.33    |
| 肝细胞性肝癌  | 11  | 0         | 0.00   | 0              | 0.00     |
| 合 计     | 220 | 51        | 23.18  | 27             | 52.94    |

### (三) 相关表

是复合表的一种形式，表示两种标志间的相互关系。如表 2·3。

表 2·3 男性 7 岁男孩身高与体重的关系

| 体重<br>(kg) | 身 高 (cm) |      |      |      |      |      |      |      |      |      |      |  | 合计 |
|------------|----------|------|------|------|------|------|------|------|------|------|------|--|----|
|            | 106~     | 108~ | 110~ | 112~ | 114~ | 116~ | 118~ | 120~ | 122~ | 124~ | 126~ |  |    |
| 15~        | 1        | 2    | 1    |      |      | 1    |      |      |      |      |      |  | 5  |
| 16~        | 3        | 4    | 4    | 2    |      |      | 1    | 1    |      |      |      |  | 16 |



续表

| 体重<br>(kg) | 身 高 (cm) |      |      |      |      |      |      |      |      |      | 合计 |      |
|------------|----------|------|------|------|------|------|------|------|------|------|----|------|
|            | 106~     | 108~ | 110~ | 112~ | 114~ | 116~ | 118~ | 120~ | 122~ | 124~ |    | 126~ |
| 17~        |          | 4    | 5    | 7    | 6    | 3    |      | 1    |      |      |    | 26   |
| 18~        | 1        | 3    | 6    | 9    | 13   | 6    | 7    |      | 1    |      |    | 46   |
| 19~        |          |      | 3    | 6    | 8    | 6    | 12   | 3    |      |      |    | 38   |
| 20~        |          |      | 1    | 4    | 1    | 1    | 10   | 7    |      | 2    |    | 32   |
| 21~        |          |      |      |      | 3    | 5    | 1    | 5    | 4    | 2    |    | 23   |
| 22~        |          |      |      |      |      | 3    | 2    | 5    | 3    | 4    | 1  | 18   |
| 23~        |          |      |      |      |      |      | 1    | 1    |      |      | 2  | 4    |
| 24~        |          |      |      |      |      | 1    |      | 2    |      | 1    | 1  | 5    |
| 合计         | 5        | 13   | 20   | 28   | 35   | 29   | 37   | 25   | 8    | 9    | 4  | 213  |

## 第二节 统 计 图

统计图是用几何图形表达统计分析结果的一种形式。它可以使统计资料形象化,通俗易懂,使人们印象清晰、一目了然。统计图通常采用面积大小、线条升降和点的位置,将统计资料的动态趋势、相互关系、内部结构、对比分布和数量多少鲜明地表达出来,因此应用非常广泛。

### 一、绘制统计图的基本要求

统计图有多种类型。如直条图、构成图、线图、直方图等。各种图形适用范围不同,一般应根据资料性质和分析目的加以选择。

1. 首要的是根据资料的性质绘制合适的图形。
2. 必须有图题,以简要地表明图的时间、地点和内容,图题一般都写在图的下端中央。
3. 图中如有纵轴和横轴,两轴均应有标目和尺度,标目应注明单位。
4. 图中用不同线条或颜色表示不同事物时,需用图例说明。
5. 图形应作整体布局设计,要注意纵横、高低、大小、间隔的尺寸和比例。如图的长宽比例一般以 7:5 左右较为美观,但为了说明问题,也可灵活掌握。

### 二、常用几种类型的统计图

#### (一) 直条图

以同样宽度的直条长短来表示指标数量的大小,它适用于相互独立、性质相似的时间性资料的指标数值大小的比较。

直条图 (bar chart) 可分单式和复式两种。绘制时一般以横轴为基线,表示各个不同类别;纵轴表示数值。纵轴尺度必须从 0 开始,中间不宜折断。各直条宽度应相等,各条之间的间隙也应相等,间隙的宽度应在直条宽度的 1/2 至等于直条宽度范围。直条排列应按高低顺序或习惯顺序排列,以便比较。复式直条图是一个组内包括两个及以上直条,直条所表示的类别应用图例说明,同一组的直条间不留空隙,各组内直条排列次序