



# 医学统计方法与实践

罗盈怡 杨万龄◎主编



press



上海交通大学出版社  
SHANGHAI JIAO TONG UNIVERSITY PRESS

# 医学统计方法与实践

罗盈怡 杨万龄◎主编



上海交通大学出版社  
SHANGHAI JIAO TONG UNIVERSITY PRESS

## 内容提要

医学统计方法是运用概率论与数理统计的原理及方法,结合医学实际,研究数字资料的搜集、整理分析与推断的一门方法性学科。而 SPSS 软件则是目前使用最广泛的优秀的统计分析软件。能使用该软件进行统计分析是掌握医学统计方法的重要内容之一。

本书内容主要包括绪论、SPSS 应用基础、不同类型资料的统计描述和统计推断、非参数检验、相关与回归、统计表、图,另外还有课堂实训和综合实训。

本书适用于医学类职业学校的各专业学生,也适合作为临床工作者自修或强化的读本。

## 图书在版编目(CIP)数据

医学统计方法与实践/罗盈怡,杨万龄主编. —上海:上海交通大学出版社,2014  
ISBN 978-7-313-10816-6

I. 医... II. ①罗... ②杨... III. 医学统计—医学院校—教材 IV. R195.1

中国版本图书馆 CIP 数据核字(2014)第 015837 号

## 医学统计方法与实践

主 编:罗盈怡 杨万龄

出版发行:上海交通大学出版社

地 址:上海市番禺路 951 号

邮政编码:200030

电 话:021-64071208

出 版 人:韩建民

印 制:上海交大印务有限公司

经 销:全国新华书店

开 本:787mm×960mm 1/16

印 张:11.75

字 数:206 千字

版 次:2014 年 5 月第 1 版

印 次:2014 年 5 月第 1 次印刷

书 号:ISBN 978-7-313-10816-6/R

定 价:25.00 元

版权所有 侵权必究

告读者:如发现本书有印装质量问题请与印刷厂质量科联系

联系电话:021-54742979

# 前 言

本书是我校营养与卫生教研室专任教师在以往教学实践积累的基础上编写的融理论和实践于一体的高职教材,其编写也是我校 2013 年度“飞跃计划”的子项目之一。

本书的适用对象涵盖医学类职业学校的各专业学生,也适合作为临床工作者自修或强化的读本。教学内容主要涉及各类基础统计方法和实践技能,并为学员日后深造作适当铺垫。为适应各专业的教学差异,课程所提供的教学单元可以让教师自由选配,以适应不同的教学时数之需。

对于本书的使用,我们建议采纳两种教学方式交替进行:一是在课堂上教学和实践,着重于统计学基础知识、各种统计应用技能、分析结论的专业应答能力等;二是在电脑房教学和实践,学习 SPSS 统计软件操作方法。先从熟悉数据检索并完成单项统计分析开始,逐渐过渡到利用仿真实训数据库实训。为了读者使用方便,数据库处理产生结果的表格直接在本书中引用。最终的教学目标是使学员能初步掌握论点-论据模式的综合统计分析技能。即根据所提供的专题数据库,自主择题或围绕某个论点,独立或合作完成能为证明论点提供合理论据的统计分析。

我们目前尚处在尝试用“实际职业任务引领”的教学理念编写教材,限于编者的水平和经验,本书存在疏漏和不足之处,恳切地希望师生们、读者们批评指正。

编 者

上海医药高等专科学校《营养与卫生教研室》

# 目 录

|                                   |    |
|-----------------------------------|----|
| 第 1 章 绪论 .....                    | 1  |
| 1.1 概述 .....                      | 1  |
| 1.2 统计资料的类型 .....                 | 3  |
| 1.3 统计学的基本概念 .....                | 5  |
| 第 2 章 SPSS 应用基础 .....             | 9  |
| 2.1 SPSS 简介 .....                 | 9  |
| 2.2 数据库的建立 .....                  | 12 |
| 2.3 数据与变量基本操作 .....               | 16 |
| 2.4 变量转换 .....                    | 19 |
| 第 3 章 计量资料的统计描述 .....             | 26 |
| 3.1 计量资料的频数分布 .....               | 26 |
| 3.2 计量资料统计描述的常用指标 .....           | 31 |
| 3.3 正态分布及其应用 .....                | 43 |
| 3.4 SPSS 软件在计量资料统计描述中的应用 .....    | 50 |
| 第 4 章 总体均数的估计和假设检验 .....          | 55 |
| 4.1 总体均数的估计 .....                 | 55 |
| 4.2 假设检验的基本思想和步骤 .....            | 58 |
| 4.3 $t$ 检验 .....                  | 61 |
| 4.4 SPSS 软件在总体均数估计和假设检验中的应用 ..... | 67 |
| 第 5 章 方差分析 .....                  | 75 |
| 5.1 方差分析的基本思想和基本步骤 .....          | 75 |
| 5.2 SPSS 软件在方差分析中的应用 .....        | 77 |



|        |                            |     |
|--------|----------------------------|-----|
| 第 6 章  | 计数资料的统计描述 .....            | 81  |
| 6.1    | 计数资料统计描述的常用指标 .....        | 81  |
| 6.2    | 常见相对数的应用 .....             | 83  |
| 6.3    | SPSS 软件在计数资料统计描述中的应用 ..... | 87  |
| 第 7 章  | 卡方检验 .....                 | 92  |
| 7.1    | 四格表卡方检验 .....              | 93  |
| 7.2    | 行 $\times$ 列表卡方检验 .....    | 97  |
| 7.3    | 配对资料的卡方检验 .....            | 100 |
| 7.4    | SPSS 软件在卡方检验中的应用 .....     | 103 |
| 第 8 章  | 非参数检验 .....                | 109 |
| 8.1    | 两个独立样本的秩和检验 .....          | 109 |
| 8.2    | 配对样本的秩和检验 .....            | 112 |
| 第 9 章  | 相关与回归 .....                | 115 |
| 9.1    | 直线相关分析 .....               | 115 |
| 9.2    | 直线回归分析 .....               | 120 |
| 第 10 章 | 统计表和统计图 .....              | 124 |
| 10.1   | 统计表 .....                  | 124 |
| 10.2   | 统计图 .....                  | 127 |
| 10.3   | SPSS 在统计图表中的应用 .....       | 133 |
| 第 11 章 | 课堂实训 .....                 | 141 |
| 11.1   | SPSS 应用基础 .....            | 141 |
| 11.2   | 计量资料统计描述 .....             | 144 |
| 11.3   | $t$ 检验 .....               | 145 |
| 11.4   | 方差分析 .....                 | 146 |
| 11.5   | 计数资料的统计描述 .....            | 147 |
| 11.6   | 卡方检验 .....                 | 148 |

|               |                     |            |
|---------------|---------------------|------------|
| 11.7          | 非参数检验 .....         | 149        |
| 11.8          | 相关与回归 .....         | 150        |
| 11.9          | 统计图表 .....          | 151        |
| <b>第 12 章</b> | <b>综合实训 .....</b>   | <b>153</b> |
| 12.1          | 老年人高血压现况调查 .....    | 153        |
| 12.2          | 艾滋病认知度调查 .....      | 156        |
| 12.3          | 校园课余活动满意度调查 .....   | 159        |
| 12.4          | 内地西藏班学生适应情况调查 ..... | 161        |
| 12.5          | 腹型肥胖现况调查 .....      | 167        |
| 12.6          | 焦虑抑郁问卷调查 .....      | 169        |
| 12.7          | 儿童肥胖、超体重现况调查 .....  | 174        |
| 12.8          | 自杀态度调查 .....        | 176        |

# 第1章 绪 论

## 1.1 概述

### 问题导向

1. 期末考试结束了,老师想了解某年级3个班级学生成绩情况如何,他该做些什么?

2. 某社区有2万人口,社区医生想要了解该社区人群的高血压、糖尿病、恶性肿瘤等慢性病的流行情况,以及这类疾病与年龄、性别、职业等因素的关系,他该做些什么?

### 1.1.1 统计学的定义

统计学是应用概率论和数理统计的基本原理和方法,研究数据的搜集、整理、分析、表达和解释的一门应用数学。通过对数据的搜集、整理和综合分析,去伪存真,透过表面现象揭示事物内部的发展规律。

统计学的应用范围非常广泛,只要存在数据的领域,都有统计学存在的必要。它是研究事物发展规律的工具,可以协助经济学家研究国民经济的发展走势,可以协助市场分析师了解产品的销售情况及影响因素,也可以协助证券分析师预测股票的涨跌等等。

当统计学与医学结合起来时就形成了医学统计学。医学现象和结果也是由大量的数据表现出来的。临床医学工作者有了统计学这门工具,就可以通过相应指标的计算分析来了解疾病发生的严重程度和影响因素,也可以通过对比分析来回答几种医护手段中哪种更好;药物分析师可以应用统计学分析各种药物的疗效差别;医学检验工作者可以利用统计学做检验结果的质量控制;护理人员可以用于评价哪种预防措施能更有效地减少院内感染的发生等等。因此,医学统计学



是医学工作者不可缺少的工具,借此可以科学地揭示大量医学现象背后的本质规律。

### 1.1.2 统计工作的基本步骤

医学统计工作的基本步骤分为研究设计、搜集资料、整理资料和分析资料。

#### 1. 研究设计

医学统计工作最重要的一个环节就是研究设计,是科学研究的第一步。设计就是要根据研究的目的而制定具体的统计策略,包括如何选择研究对象和数量,设计调查表获取原始资料,如何控制研究误差,采用何种指标评价统计结果,确定资料进行整理和分析的方案等。设计的好坏直接影响着统计结果的真实性,是统计过程中最关键的环节。

#### 2. 搜集资料

按照设计的方法进行资料的搜集是统计工作的第二步,资料的搜集要力求完整、准确、及时、可靠,在搜集的过程中要尽量减少漏填和误填。常见的资料搜集来源有以下三类:①常规工作记录:如病史记录、体检记录、病理报告等;②统计报表:如传染病报告卡、恶性肿瘤报告卡、死亡和出生登记等;③专题调查和实验:如在社区进行慢性病患病情况调查;在病房观察某种医学措施(药物、手术、护理等)的效果;在实验室完成的小白鼠药物实验数据的搜集等。

#### 3. 整理资料

包括对搜集的数据进行检查和整理的过程。目前一般采用计算机软件建立数据库,将相关数据录入数据库并做好相关的数据核查与清理工作,包括对漏填的数据进行补填,对误填的数据进行纠错等。系统的资料整理是进行进一步统计分析的基础。

#### 4. 分析资料

资料的分析是给出统计结果的阶段,包括相应统计指标的计算,如计算患病率以描述某疾病的严重程度,还可以进行事物之间的比较,如对比分析分别经手术和化疗的两组肿瘤患者的生存率来回答哪种医疗手段更好等。

## 1.2 统计资料的类型

### 问题导向

某研究者对 10 名同学的三种特征进行了调查,整理得到以下数据:

| 学生    | 赵×  | 钱×  | 孙×  | 李×  | 周×  | 吴×  | 郑×  | 王×  | 冯×  | 陈×  |
|-------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 身高/cm | 160 | 165 | 168 | 170 | 175 | 166 | 173 | 180 | 178 | 159 |
| 血型    | B   | AB  | A   | B   | O   | B   | O   | A   | A   | B   |
| 体型    | 正常  | 正常  | 超重  | 消瘦  | 正常  | 正常  | 超重  | 正常  | 正常  | 肥胖  |

请说出三种资料的资料类型?

### 1.2.1 变量和变量值

在医学研究中,研究对象的某个需要观察的特征称为变量。换言之,对研究对象的视角不同可观察到的特征也不同。根据调查需要设计调查表,调查表中的每个项目都是变量。变量值则是变量的实际测量结果,有定量和定性之分。

以“问题导向”中的“身高”变量和“血型”变量为例。某对象身高 160 cm,则身高为定量变量,变量值为 160 cm;某对象血型为 B 型,则血型为定性变量,其变量值为“B”等。

### 1.2.2 统计资料的类型

通常按变量的类型将统计资料分为以下三种类型。

#### 1. 数值变量资料

又称计量资料,是用定量的方法对一组观察对象的某项指标进行测量获得的资料。其变量值是定量且通常有度量衡单位。例如临床上常见的数值变量包括身高、体重、血压、血糖、血红蛋白浓度等。

#### 2. 分类变量资料

又称计数资料,是把一组观察对象按照某个属性定性分组,然后清点各组例数得到的资料。例如“问题导向”中的血型变量,按照血型特征分了四个组,清点对象数可以得到 A 型血有 3 例,B 型血 4 例,O 型血 2 例和 AB 型血 1 例。其他如男性和女性、生存和死亡、患病和健康等都属于定性或分类变量资料。



### 3. 等级变量资料

有些分类变量资料兼有部分定量特征,呈等级依次排列,其分类顺序不能颠倒,因此又称为有序分类变量资料。其他如尿蛋白浓度(一、+、++、+++、++++),体型(消瘦、正常、超重、肥胖)等。

#### 1.2.3 统计资料分类的应用意义

统计资料种类不同,对应的统计分析方法或统计公式也不同。例如可使用平均数等描述数值变量资料,假设检验可采用  $t$  检验;使用“率”、“比”等指标来描述分类变量资料,假设检验可采用  $\chi^2$  检验。

统计资料种类不同,统计效率也不同。数值变量资料作统计分析相对而言最精细,统计效率最理想。在分类变量资料中,等级变量资料的统计效率优于非等级分类变量资料。

#### 1.2.4 变量类型的转换

根据实际需要,资料的变量类型是可以相互转换的。

第一种变量转换的目的是基于临床或统计应用的需要。常见的有以下两种。

##### 1. 数值变量的分类变量化

例如将一批调查对象分别按其血压测量值(定量),按健康人舒张血压的医学参考值范围(60~90 mmHg),可分别诊断为低血压患者、健康者和高血压患者(定性)。

##### 2. 等级资料的数值变量化

例如根据体质指数的医学参考值范围,将定性为消瘦、正常、超重和肥胖四种体型的对象,分别以“-1,0,1,2”替换。

##### 知识链接

第二种变量转换可以参见本书 2.4 节,其目的是为了更方便建立数据库和检索。

例如以“021”代替“上海市”,数据输入“021”相当于输入了汉字“上海市”。当给出指令为“021”时,SPSS 软件可以检索出所有居住在“上海市”的对象数据进行统计分析。

当然,将定性为消瘦、正常、超重和肥胖四种体型的对象,分别以“-1,0,1,2”替换。既兼顾了数据库建立和数据检索归类,同时也可满足特定的统计分析需要。



### 课堂活动 1

1. 某社区调查 1 000 名居民,得到 1 000 个血压测定值组成资料,形成了什么资料?
2. 如果将对象分为高血压患者 200 名,健康者 800 名,又形成了什么资料?
3. 如果再将 200 名患者分为轻度、中度和重度患者,此时形成了什么资料?

### 课堂活动 2

根据以下调查表建立所需的变量? 并思考相应的变量类型是什么?

1. 编号: \_\_\_\_\_ (1~100);
2. 性别: \_\_\_\_\_ (男、女);
3. 身高: \_\_\_\_\_ cm;
4. 体重: \_\_\_\_\_ kg;
5. 期末考试总分: \_\_\_\_\_;
6. 考试等级: \_\_\_\_\_ (优、良、中、差)。

### 课堂活动 3

打开数据库“学生情况.sav”,观察其中有多少个变量? 哪些是数值变量,哪些是分类变量,哪些是等级变量? 变量值各有怎样的特点?

## 1.3 统计学的基本概念

### 问题导向

1. 有位女青年,18 岁,身高 1.53 m,如何从身高数值评价她健康与否?
2. 某医疗保险公司调查 200 名成年人,他们的医疗开支平均为 860 元/人·年,在不考虑公司盈利和经营成本的前提下,按每人每年收费 860 元,能保证公司开展医疗保险业务并且收支平衡吗?
3. 某制药公司生产了一种新型的感冒药,服药一周感冒的治愈率达到 98%,是否可以说该药治疗效果非常好?
4. (继问题 3)某种已经在临床常规使用感冒药的治愈率为 72%,可以认为新感冒药的治疗效果更优吗?



统计学是以数学手段观察和比较事物的方法学。其研究对象是存在变异的事物,如果不存在变异事物,统计学就没有存在的必要。

由此,因变异可以引申出一系列的统计学重要概念。

### 1.3.1 变异和同质

变异是指同质研究对象或现象之间存在的差异。同质即性质相同。

事物性质相同但彼此间存在差异,似乎存在着矛盾,但实则不然。例如,同为 一群健康人(同质)却彼此在身高、血压、肺活量等方面均存在差异;又如,来源于同 一个人(同质)的血液样品,多次测定其中某一种成分如血细胞、蛋白质、血清胆固 醇等,每一次测定的结果也可能不同。这种同质事物彼此之间可能相同,更可能不 同,少数相同,大多数不同的现象,就是变异。无论是自然现象或者社会现象,变异 到处存在。

对于非变异事物,只要观察一个个体,其结果就足以反映全部对象。但对于变 异事物的观察却要复杂许多,必须在群体观察的基础上科学地反映变异事物的 特征。

例如从一批对象(群体)的测量值中计算出平均数作为观察事物的代表;也可以 调查一群对象,经统计分析建立一个数值范围来反映事物的变异情况,临床诊断常 用的医学参考值范围就是实例。如舒张血压的医学参考值范围为  $8\sim 12\text{ kPa}$  ( $60\sim 90\text{ mmHg}$ ),这是正常人舒张血压的变异范围。凡血压测量值在这一范围内的个 人,虽然都可以被定性为血压正常,但因变异的存在,他们彼此的血压值高低不一。

### 1.3.2 总体与样本

能反映变异事物特征和变化规律的群体观察包括两个观察层面:总体和样本。

#### 1. 总体

指由研究目的确定的同质研究对象的全体。例如调查某地健康男童的出生体 重情况。研究对象或观察事物是体重,相应的同质总体需要 6 个定语,即所有的、 该地区的、健康的、男性的、儿童的、出生时的体重。

一般而言,经观察总体来了解变异事物没有必要或不可能。

#### 2. 样本

指对总体有代表性的部分研究对象。

要使样本对总体具有代表性,必须满足如下 3 个条件:①样本必须来源于欲反 映的总体;②抽样必须遵循“随机化”原则;③样本含量适宜。



观察样本只是手段,反映总体才是目的。通过样本观察结果反映总体情况时,通常存在着抽样误差。但只要是随机抽样形成的样本,抽样形成的误差就具有数学规律,可用统计学方法估计。

### 1.3.3 抽样误差

抽样误差指由于随机抽样而形成的误差。其表现形式有两种:一是反映在样本统计指标与所属总体统计指标可能存在的差异,例如某地若干名(样本)7岁男孩的身高平均数,一般不是恰好等于该地所有(总体)7岁男孩的身高平均数;二是表现在从同一个总体中抽取的样本与样本统计指标彼此之间可能存在的差异。例如来自同一个社区人群的两个调查样本,从中得到的糖尿病(样本之间)患病率一般不会恰好相等。

导致抽样误差的原因是事物存在变异,由于变异无法避免,因此抽样误差也是不可避免的。但随着样本含量的增加,抽样误差可以逐渐减少。

在样本层面看到的差异,在总体层面上就可能出现两种判定:①属于抽样误差;②属于本质区别。样本指标的数值差距越小,属于抽样误差的可能越大;反之数值差距越大,就越不属于抽样误差,而属于本质区别。例如一种新感冒药与传统的感冒药相比较,两药治愈率分别为98%和72%。运用统计推断才能更科学地判定,两样本相差26%,究竟属于抽样误差还是本质差别。

对于变异事物,在个体观察水平上,可以看到不同的个体之间存在着差异,称为变异;在样本观察水平上,可以看到的不同随机样本之间也存在着差异,称为抽样误差。只要所观察的事物来源于同一总体,无论变异或抽样误差都是表象,同质才是本质。

统计学就是一门利用变异或抽样误差的自然规律,使得样本结论能够推断或反映总体的应用数学。

### 1.3.4 概率

概率指事件发生的可能性,常用字符 $P$ 表示。

概率在数值上介于0与1之间,即 $0 \leq P \leq 1$ ,也可以用百分数表示,即 $0\% \leq P \leq 100\%$ 。 $P$ 越接近1,说明事件发生的可能性越大; $P$ 越接近0,说明某事件发生的可能性越小。

习惯上将发生概率 $\leq 0.05(5\%)$ 的事件称为小概率事件。由于出现的可能性极低,故在一次性观察中可以认为小概率事件不会发生,统计学将之称为小概率事



件原理。

小概率事件原理几乎应用于所有统计推断的结果判断。

例如,一位身高 1.53 m 的女青年,其身高健康吗? 科学判断应该是:经统计推断,如果判定为该女青年身高值位于健康女青年身高范围内的概率 $\geq 5\%$ ,则可以定性为身高健康;反之判定概率 $< 5\%$ ,则可以定性为不健康,即该女青年的身高属于健康的概率极低。

又如分别调查两地儿童的龋齿情况,农村和城市的患病率分别为达 56% 和 32%,两地患病率有区别吗? 科学的判断应该是:经统计推断,如果判定两样本患病率属于同一总体的概率 $\geq 5\%$ ,则可以认为样本的区别是由抽样误差造成的,从总体层面上看两地患病率相同;反之判定属于同一总体的概率 $< 5\%$ ,属于小概率事件,则可以认为从总体层面上看两地龋齿情况不同,农村患病率更高。

如何使用统计学方法作科学推断,可详见后续章节。统计学的上述基本概念及其联系,是与统计学应用技能密切相关的重要理论基础,贯穿于统计方法学习的始终。初学者可能在开始阶段会感到有些抽象,但随着后续学习的循序渐进,结合具体实例反复加深理解,自然会水到渠成。

## 第 2 章 SPSS 应用基础

### 2.1 SPSS 简介

#### 2.1.1 SPSS 基本介绍

数据的统计分析处理是一项非常复杂的数据演算过程,往往需要大量的计算过程,在计算机高速发展的今天,我们完全可以利用计算机和统计分析软件来代替繁琐的手工计算,提高统计效率,降低统计分析的难度。

社会科学统计软件包(statistical package for the social science,SPSS)是世界著名的统计分析软件之一。该软件迄今为止已有 30 多年的发展历史,适用于通信、医疗、银行、证券、保险、商业、市场调研等多个领域的统计分析,是世界上应用最广泛的专业统计软件之一。

SPSS 使用 Windows 的窗口方式展示各种管理和分析数据的方法,使用对话框展示出各种功能选择项,只要掌握一定的 Windows 操作技能,并了解统计分析原理,就可以使用该软件为特定的科研工作服务。

SPSS 的基本功能包括数据管理、统计分析、图表分析、输出管理等。其过程包括描述性统计、均值比较、相关分析、回归分析等,使用该软件可以避免复杂的统计分析计算过程,快速准确地得到计算结果,使得统计分析简单化、人性化。此外,SPSS 中还有专门的绘图系统,可以根据数据绘制各种图形。

本书以 SPSS17.0 中文版为基础,介绍与医学相关的统计分析在 SPSS 中的实现过程。

#### 2.1.2 SPSS 的基本界面

SPSS 安装完毕后,系统会自动在 Windows 菜单中及桌面上创建快捷方式“SPSS Statistics”。双击该快捷方式即可启动 SPSS 并出现 SPSS 数据编辑器的主界面,如图 2-1 所示。

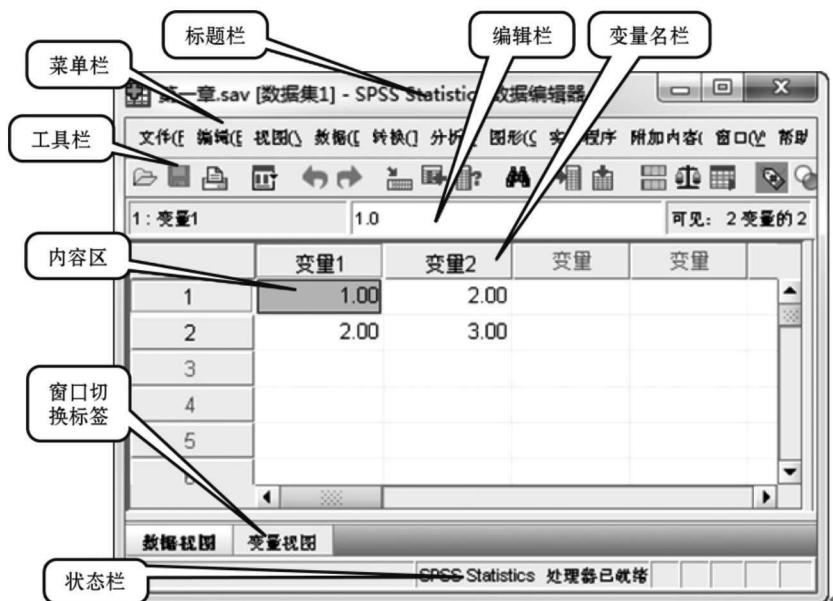


图 2-1 SPSS 数据编辑器的主界面

SPSS 的数据编辑窗口与微软公司的 Excel 窗口有些相似,并且有一些功能也相同。不过 SPSS 的数据统计功能要比 Excel 强得多。

数据编辑窗口由标题栏、菜单栏、工具栏、编辑栏、变量名栏、内容区、窗口切换标签页和状态栏组成。

标题栏中显示编辑的数据文件名,本图中所编辑的数据文件为“第一章.sav”。

菜单栏中列出了 SPSS 的命令菜单,每个菜单对应一组相应的功能。

“文件”菜单是对 SPSS 文件的操作;

“编辑”菜单可以对数据文件进行编辑操作;

“视图”菜单可以对用户界面进行设置;

“数据”菜单是数据文件的建立和编辑菜单;

“转换”菜单是数据的基本处理菜单;

“分析”菜单是统计分析菜单,主要的统计功能都集中在该菜单中;

“图形”菜单中可以根据数据绘制基本统计图形;

“实用程序”是相关应用和设置菜单;

“窗口”菜单可用于 SPSS 各窗口的切换;

“帮助”菜单可以为你的操作提供必要的帮助。

工具栏中列出了一些常用的快捷图标。