

普通高等教育“十二五”规划教材

· 医学教材系列 ·

医学统计学

YI XUE TONG JI XUE

何迎春◎主编

辽宁大学出版社

第2版 医学统计学 第2版 第2版 第2版

第2版 医学统计学 第2版 第2版 第2版

医学统计学

YIXUE TONGJIXUE

第2版 医学统计学 第2版 第2版 第2版

第2版 医学统计学 第2版 第2版 第2版

普通高等教育“十二五”规划教材
· 医学教材系列 ·

医学统计学

主 编 何迎春(湖南中医药大学)
主 审 刘明芝(湖南中医药大学)

辽宁大学出版社

图书在版编目(CIP)数据

医学统计学 / 何迎春主编. — 沈阳: 辽宁大学出版社, 2013. 10

普通高等教育“十二五”规划教材. 医学教材系列

ISBN 978-7-5610-7506-7

I. ①医… II. ①何… III. ①医学统计—医学院校—教材 IV. ①R195.1

中国版本图书馆 CIP 数据核字(2013)第 248731 号

出 版 者: 辽宁大学出版社有限责任公司

(地址: 沈阳市皇姑区崇山中路 66 号 邮政编码: 110036)

印 刷 者: 北京明兴印务有限公司

发 行 者: 辽宁大学出版社有限责任公司

幅面尺寸: 185mm×260mm

印 张: 13.5

字 数: 320 千字

出版时间: 2013 年 10 月第 1 版

印刷时间: 2013 年 11 月第 1 次印刷

责任编辑: 张琢石 黄 铮

封面设计: 可可工作室

责任校对: 齐 悦

书 号: ISBN 978-7-5610-7506-7

定 价: 28.00 元

联系电话: 86864613

邮购热线: 86830665

网 址: <http://WWW.lnupshop.com>

电子邮件: lnupress@vip.163.com

前言

《医学统计学》是应用概率论、数理统计的原理和方法,紧密结合医药的实际,设计、搜集、整理和分析资料,研究医药科学中不确定现象客观规律的新兴应用学科。

本书作为当代医药统计学教学内容和方法的知识载体,作为医药院校统计学教学的重要工具,共分 11 章,强调“三基”训练,重视实际应用。第一章绪论,介绍统计学中的基本概念、资料类型、相关统计软件,第二章介绍定量资料的统计描述(包含了正态分布),第三章介绍分类资料的统计描述(包括二项分布和 Poisson 分布),第四章介绍总体均数和总体率的估计,第五章介绍假设检验和 t 检验,第六章介绍方差分析,第七章介绍计数资料的假设检验,第八章介绍秩和检验与游程检验,第九章介绍简单相关与一元回归、协方差分析,第十章介绍统计图表和统计分析方法的选择,第十一章介绍医学研究设计的常用方法。

本书在坚持三基(基础理论、基本知识、基本技能)和五性(思想性、科学性、先进性、启发性、适用性)原则下,具有以下特色:(1)历来学生感觉统计难学,为改进教学,尽可能培养学生的统计思维,降低学习难度,面向医药统计学教学和科研所需,结合医药实际,内容新颖翔实。(2)将统计学理论和方法与多种软件操作优化衔接,兼具教材和工具功能。本书在介绍国际通用 SPSS19.0 统计软件的同时,介绍适用于中医药专业的优秀统计软件 DPS12.0,编排方法采用主要贯穿 SPSS19.0,辅以 DPS12.0 解难,直接在各例题后介绍操作。统计软件的应用侧重数据文件的格式,操作思路和结果的解释。授课时,教师只要讲授医药统计学的理论课,统计软件的操作只要稍作引导,学生很容易掌握,可以自行完成。(3)适应面广,本书内容具有良好的适应性,可作为高等院校各专业专科生、本科生统计课程的通用教材,亦可作为培训教育教材。

全书文字规范流畅,论例配合科学,重点突出、难点分解,可作为生物医药科研工作者自学和工具书使用。

本书吸收了众多专家学者的宝贵经验、智慧,凝聚了高等中医药院校统计教师的教学体会和潜心钻研中医药统计学的知识积累,为方便教学和应用,为用尽量少的篇幅容纳尽量多的实用知识,作者呕心沥血。限于我们的学识和能力,虽然我们力图提高质量,书中难免有错漏和不妥之处,希望同行专家和广大读者提出宝贵意见和建议。

何迎春

2013 年 5 月

《医学统计学》编委会

主 编 何迎春(湖南中医药大学)

主 审 刘明芝(湖南中医药大学)

副主编 马 莉(大连医科大学)

编 委 (按姓氏笔画排序)

孔丽娅(浙江中医药大学)

何 兰(湖南中医药大学)

李国春(南京中医药大学)

李 新(辽宁中医药大学)

汪旭升(广西中医学院)

张 波(湖南中医药大学)

姚 政(云南中医学院)

曹治清(成都中医药大学)

高德富(河南中医学院)

解 颖(黑龙江中医药大学)



目 录

第一章 绪 论	(1)
第一节 医学统计学的意义	(1)
第二节 统计学中的基本概念	(1)
第三节 医学统计中的资料类型	(4)
第四节 统计工作的基本步骤	(6)
第五节 SPSS 和 DPS 统计软件介绍	(8)
第二章 定量资料的统计描述	(14)
第一节 频数分布表和频数分布图	(14)
第二节 集中趋势的描述	(17)
第三节 离散趋势的描述	(19)
第四节 正态分布	(23)
第五节 医学参考值范围	(26)
第三章 分类资料的统计描述	(33)
第一节 常用相对数	(33)
第二节 率的标准化法	(37)
第三节 动态数列	(39)
第四节 二项分布和 Poisson 分布	(41)
第四章 总体均数和总体率的估计	(45)
第一节 常用的抽样分布	(45)
第二节 总体均数的估计	(48)
第三节 总体率的估计	(51)
第五章 假设检验和 t 检验	(56)
第一节 假设检验	(56)
第二节 正态性检验	(60)
第三节 方差齐性检验	(63)
第四节 单样本 t 检验	(65)
第五节 配对 t 检验	(66)
第六节 两独立样本 t 检验	(69)
第七节 假设检验应注意的问题	(73)
第八节 假设检验与可信区间的区别和联系	(74)
第六章 方差分析	(78)
第一节 方差分析的基本思想	(78)



第二节	完全随机设计资料的方差分析和多重比较	(80)
第三节	配伍组设计资料的方差分析	(85)
第四节	拉丁方设计资料的方差分析	(89)
第五节	交叉设计资料的方差分析	(91)
第六节	析因设计资料的方差分析	(92)
第七节	重复测量资料的方差分析	(93)
第七章	计数资料的假设检验	(101)
第一节	样本率与总体率比较	(101)
第二节	两样本比较的 Z 检验	(102)
第三节	$R \times C$ 表资料常用统计方法选择的思路	(104)
第四节	χ^2 检验的基本思想、四格表 χ^2 检验	(104)
第五节	四格表确切概率法	(109)
第六节	双向无序 $R \times C$ 表资料的独立性检验	(110)
第七节	方表资料的分析	(112)
第八节	率的多重比较	(118)
第八章	秩和检验与游程检验	(122)
第一节	秩和检验概述	(122)
第二节	配对设计和单样本比较的符号秩和检验	(123)
第三节	完全随机设计两组比较的秩和检验	(126)
第四节	完全随机设计两组或多组比较的秩和检验	(129)
第五节	配伍设计资料的秩和检验	(133)
第六节	游程检验	(136)
第九章	简单相关与一元线性回归	(139)
第一节	Pearson 相关	(139)
第二节	等级相关	(144)
第三节	一元线性回归	(145)
第四节	相关与回归的注意事项	(153)
第十章	统计图表和统计分析方法的选择	(156)
第一节	统计表	(156)
第二节	统计图	(159)
第三节	统计分析方法的选择	(169)
第十一章	医学研究设计	(175)
第一节	实验研究设计	(175)
第二节	调查研究设计	(185)
附录:统计用表		(195)
主要参考文献		(210)



第一章 绪论

第一节 医学统计学的意义

医学统计学(Statistics for Traditional Medicine)是将数理统计的原理和方法应用于生物医药科研,收集、整理和分析资料,推断不确定现象客观数量规律的一门应用学科。主要内容包括统计学基本理论和统计分析方法。统计分析包括统计描述(statistical description)和统计推断(statistical inference)。

我国《药品临床试验的管理规范》(GCP)明确规定:“在临床试验的统计结果的表达及分析过程中都必须采用规范的统计学分析方法,并应贯彻于临床试验始终,各阶段均需有熟悉生物统计学的人员参与”;2002年颁布的《药品注册管理办法》也明确提出新药研究从方案设计开始就要有统计学家的加入,试验例数必须符合统计学要求,且满足法规规定的最低例数。

医药科研的基本步骤包括立项,设计,实践,统计处理和总结概括,均需用到医学统计学的思维和方法。医学统计学是现代医药科研定量分析的重要工具,是认识医药数量规律的必备知识。

第二节 统计学中的基本概念

一、总体与样本

总体是根据研究目的确定的,性质相同的研究对象的全体(population)。总体具有唯一性。个体(individual)是总体的基本单位(case)。个体可以是—个人、一个家庭、一个地区、一只动物、一个检测样品等。没有个体,总体就不存在;没有总体也就无法确定个体。总体又可分为有限总体和无限总体。有明确范围(如空间、时间)限制,个体数有限的总体称有限总体;不易划分确切范围,个体数无限的总体称为无限总体。

同一总体中个体之间的差异,称为个体差异(individual variation),简称变异(variation)。例如,同为正常发育的10岁男孩,各人身高有不同;又如,病情相同的患者服用相同的药物,其疗效也不尽相同。总体所要求的性质相同是指同一总体中的个体大同小异。变异是客观存在的,来源于一些未加控制或无法控制的甚至不明原因的因素。变异是统计学存在的基础,统计



学本质上是研究变异的科学。

总体的性质特征由其各个个体的性质而定,要研究总体的性质或特征,须对它的个体进行观测,由于有个体差异,最好是对每个个体都加以观察。但是,这样做要花费很大的人力和财力,而且有时是不允许的,例如带有破坏性的药品检验不可能逐一进行。

样本(sample)是从总体中抽取的部分个体。样本中所包含的个体数目称为样本含量(sample size)或样本大小、样本量,常用 n 表示。

从总体中抽取样本的过程称为抽样(sampling)。抽样的目的是用样本推断总体,所以样本要有足够的样本含量及可靠性、代表性。样本的可靠性是指样本中每一个观察单位是否确属于既定的同质的总体。同质应考虑时间、空间、非实验因素的干扰、诊断标准等条件尽量相同。样本的代表性是指样本是否能充分反映总体的真实情况。为得到有代表性的样本,要求抽样和分配必须遵循随机化(randomization)的原则,使总体中的每个个体都有同等的机会被抽到和分配(随机性),不受研究者的主观意愿或客观无意识偏性的影响。

二、参数与统计量、抽样误差

刻画总体特征的指标称为参数(parameter)。参数一般用小写的希腊字母表示,如用 μ 表示总体均数、 σ 表示总体标准差、 ρ 表示总体相关系数等。

样本的统计指标称为统计量(statistical variable)。统计量是不包含任何未知参数的样本观测值的函数。一般用英文字母表示, $\bar{x}$ 表示样本均数、 s 表示样本标准差、 r 表示样本相关系数等。

误差:统计上所说的误差泛指测量值与真值之差,样本指标与总体指标之差。主要有系统误差和随机误差两种:

系统误差指数据搜集和测量过程中由于仪器不准确、标准不规范等原因,造成观察结果呈倾向性的偏大或偏小。特点:具有累加性,必须消除。

随机误差指由于难以控制、无法避免的偶然因素使得结果或大或小,是不确定、不可预知的。特点:服从正态分布。

随机误差包括随机测量误差和抽样误差。

随机测量误差指由于各种偶然因素的影响造成同一对象多次测量的结果不完全一致。可以采取控制措施控制在一定的容许范围内。

由随机抽样造成样本指标(统计量)值与被推断的总体指标(参数)值的差异称为抽样误差(sampling error)。例如,从某一总体中,随机抽取样本含量为 n 的样本,算得样本均数 $\bar{x}$,用来估计总体均数 μ 。但 μ 与 $\bar{x}$ 有差异,这种差异是由于抽样误差造成的。由于个体差异客观存在,抽样误差是不可避免的。抽样误差的大小可以控制,并可用统计分析方法估计。



三、概率与频率

对研究对象某项特征进行观察或测量得到的指标称为变量(variable)。变量的观察结果称为变量值(value of variable)或观察值(observed value)。由观察值构成资料(data)。统计学中研究的变量是随机变量(random variable),其特性是:①在一次观察中,取值具有不确定性。②在大量重复试验中,变量的各可能取值的发生具有一定的统计规律。

概率(probability)和频率(frequency)都是反映某一随机变量发生的可能性大小的数值。

若事件 A 在 n 次独立重复试验中发生了 f 次,则称 f 为频数(counts),称比值 f/n 为事件 A 在 n 次试验中出现的频率或相对频数(relative frequency)。医药工作中通常所说的有效率、治愈率等都是频率。显然, $0 \leq f/n \leq 1$ 。频率是统计量,反映是样本的性质,刻画样本中事件出现的可能性大小,样本不同,其值也不同,具有偶然性。

随着试验次数增多,频率呈现出稳定在一个确定常数的特征,这就是频率的稳定性。频率的稳定性说明随机事件出现的可能性是事物本身固有的一种客观属性。

概率的统计定义:频率所稳定的常数称为概率,常用 P 表示。概率 P 值介于 0 与 1 之间。必然事件的概率为 1,不可能事件的概率为 0。概率越接近 1,表明其发生的可能性越大,概率越接近 0,其发生的可能性越小。

概率反映总体的性质,刻画的是总体中随机事件出现的可能性大小,一个随机变量 A 发生的概率是一个参数,具有必然性,是一种客观存在。当样本量 n 够大时,可用频率作为概率的估计值。由频率认识概率,体现了统计学由样本认识总体的基本思想。

四、概率分布与抽样分布

任何一个随机变量各种可能取值的概率之和为 1。一个随机变量各种可能取值发生的概率所呈现的统计规律,称为该随机变量的概率分布(probability distribution),简称分布(distribution)。概率分布由随机变量各可能的取值与其对应的概率共同构造,完整地刻画随机变量的统计规律性。

常用的分布:连续型变量常用正态分布,离散型变量常用二项分布、Poisson 分布。

抽样研究中,总体中有很多个体,究竟哪些个体在抽样中被抽到,要依机会而定,因而样本具有随机性。因统计量是依赖于样本的,故统计量是随机变量,所以就有一定的概率分布,统计量的概率分布又称为抽样分布。常用的抽样分布有样本均数的分布、样本率的分布、 χ^2 分布、 t 分布、 F 分布。

统计学研究的基本方法是由样本构造检验统计量,按抽样分布的规律,对总体进行推断。

五、统计描述与统计推断

统计描述(statistical description)是对原始资料的一种概括,作用是描述数据的集中趋势、



离散趋势、分布规律,利用统计表和统计图(见第十章)给予正确的表达。

统计推断(statistical inference)包括参数估计(parameter estimation)与假设检验(hypothesis testing),都是对整理出的数据作进一步的分析,在考察和表明抽样误差的前提下,用样本信息推断总体特征,为研究者下专业结论时提供科学依据。

六、参数统计与非参数统计

总体分布类型已知,对未知参数进行估计或检验,如第四、五、六章的方法,这类统计推断方法称为参数统计(parametric statistics)法。参数统计的优点:对资料信息的利用较充分,统计分析的效率较高。缺点:对资料的要求高,适用范围有限。

非参数统计(nonparametric statistics)是不依赖总体分布类型的统计推断方法,它不推断总体参数,而是总体分布的位置,所以非参数检验又称不均匀分布或任意分布检验(distribution-free test)。如第八章的方法秩和检验,属于非参数统计方法。优点:对资料的条件要求相对宽松,适用于样本所代表的总体分布不易确定的资料、分布呈非正态而又无适当的数据转换方法的资料、一端或两端无确定数值的资料、各组离散程度相差悬殊的资料、等级资料等。但一般来说,非参数统计法的检验效能要低于相应的参数统计法。

第三节 医学统计中的资料类型

医学统计资料的类型及特点见表 1-1。

表 1-1 医药统计资料的类型及其特点

资料类型	特点	变量类型	实例
定量资料 (数值变量资料)	每个个体都能观察到一个观察指标的数值,有度量衡单位	连续型或离散型,数值变量	身高(cm),体重(kg),细胞数(个),人口数(人)
分类资料	每个个体的观察结果是属性或分类,不是数值,清点各属性或分类中个体的个数,才得出分类资料的数据	离散型	
无序分类资料 (计数资料)	每个个体观察结果的属性或分类间无大小顺序之分。二分类时为对立的两类属性,多分类时为不相容的多类属性,类与类之间界限清楚,不会错判。	离散型 名义变量	性别(男、女)为二分类,血型(A、B、O、AB)为多分类



资料类型	特点	变量类型	实例
有序分类资料 (等级资料)	多分类,按程度、档次等级顺序分别计数,各等级之间有程度的差别,按大小顺序排列,每个观察单位无确切定量。	离散型 顺序变量	疗效(痊愈、好转、有效、无效)

一、定量资料

定量资料(measurement data)又称数值资料(numerical data)、计量资料或尺度资料(scale data),是对观察对象测量指标的数值大小所得的资料,观察指标是定量的,表现为数值大小。

二、分类资料

将研究因素按属性类别或档次分组,清点各组个体的个数(频数),将分类变量的分组标志及各组频数同时给出,便得到了分类资料(categorical data)。

分类资料有无序分类和有序分类两种:

(一)无序分类资料(unordered categorical data),又称计数资料(enumeration data)。无序分类资料的属性或类型互斥,无次序之分。根据属性或类型分组的多少又可分为两种:①两分类资料:它是将观察单位按两种对立的属性分类,如死亡和生存、治愈和未愈、有效和无效等两种分类计数。②多分类资料:它是将观察单位按多种互斥的属性分类,清点各类观察单位的个数,如血型按A、B、AB、O四种分类计数。

(二)有序分类资料(ordered categorical data):将观察单位按某种属性的不同程度、档次或等级顺序分类或分组后,清点各顺序分组的观察单位数所得的资料为有序分类资料,亦称等级资料(ranked data)。例如,化验结果按-,+,++,++++分组计数等,虽有档次或等级的差别,可以用 $X=0,1,2,3$ 等区别表示,但等级之间的差别并非确切定量差别,可以是量的差别,也可以是质的差别,给人以“半定量”的概念,故又称半计量资料。

根据分析研究的需要,定量资料可以转化为分类资料,定量资料和有序分类资料可以互相转化,例如:白细胞计数属于定量资料;如果按白细胞数正常与不正常分为两组,然后清点各组人数,便成为两分类资料了;如果按白细胞总数过高、正常、低下分为三组,然后清点各组人数,就成为有序分类资料了;将有序资料的各个类别分别赋予一个合适的数量,可转化为定量资料,例如,将不咳嗽、轻度咳嗽、中度咳嗽、重度咳嗽的4个等级依次赋予0、1、2、3得分,就量化了。须指出的是:有序分类资料转化为定量资料,存在如何变换才合理的问题,例如,上面说的咳嗽4个等级,可以赋予0、1、2、3得分,也可以赋予0、1、3、6得分。研究者可依据自身的经验判断进行不同变换。



医学统计中的资料除定量资料和有序分类、无序分类资料之外,对呈周期性变化规律或表示在圆周上位置特点的资料,称为圆分布资料;随访慢性病例或保健对象远期效果的资料,常有未达到预定观察终点的不完全数据特点,称为随访资料;等等。

不同资料,统计分析方法有不同。所以统计分析之前,先要分辨其资料类型。

第四节 统计工作的基本步骤

统计工作包括统计设计,搜集资料,整理资料和分析资料四个基本步骤:

一、统计设计

设计就是制定完成科研目标的技术路线和实施方案。统计设计是整个统计研究实施的依据,是最关键的一环,在设计时应应对后三个步骤进行周密的设想和安排。统计设计要点见表 1-2。

表 1-2 统计学设计要点

依据	根据研究的目的,运用数理统计的原理和方法做出的周密的计划和安排
类别	主要分为:实验设计和调查设计(设计方法见第十一章)
目的	主要解决抽样误差的问题,保证样本的代表性和组间的可比性,实现调查或实验观察结果的可靠性、科学性、可重复性和经济性
意义	有效地控制误差和提高检验效率,是保证专业设计布局合理性和科研结果可靠性的关键。
内容	确定设计类型,确定研究总体及样本,拟定观察指标及测量方法,资料的可靠性及质量控制,数据的管理及统计分析计划等,包括:①对照形式;②选择调查或实验的设计方案;③随机抽样和均衡分配;④样本含量及抽样误差大小的估计;⑤盲法;等

二、收集资料

收集资料是根据研究目的和设计的要求,及时取得准确、完整、可靠的原始资料。收集资料是统计分析的前提与基础。收集方式有直接观察、采访、填表、通信、实验室等。

统计资料主要来自统计报表、经常性工作记录、专题调查或实验。

三、整理资料

整理资料是对原始资料进行审核、校正、整理并使之系统化、条理化,便于统计分析。包括:①对原始数据进行逻辑和计算审核。②设计分组,包括质量分组(如病人按性别、职业、病情轻重分组)和数量分组(如按年龄、体重、身高分组),往往结合使用,一般是在质量分组基础上进行



数量分组。③拟订整理表。整理表是统计表的前身,是过渡性表格,作归纳汇总用。④编制统计表。

整理资料存在的问题主要表现在:①编制整理表时,未将有联系的项目安排适当,不便说明事物之间相互的规律性。例如表 1-3,由于将性别与年龄这两个有联系的项目分割开来,计算不出不同性别、年龄的感染率,丧失了有价值的信息,为了克服上述缺点,应采用表 1-4 的整理表。②分组不合理。③有意将过高或过低数据筛选掉或无意舍弃掉,使原始资料所提供的信息变样。④统计表编制不规范。

表 1-3 ××病感染按性别、年龄整理表(问题表)

	性别		年龄(岁)				合计
	男	女	0~	5~	10~	15~	
调查人数							
感染人数							

表 1-4 ××病感染按性别、年龄整理表(正确表)

年龄组	男		女	
	调查人数	感染人数	调查人数	感染人数
0~				
5~				
10~				
15~				
合计				

四、分析资料

分析资料是根据研究目的、要求、资料的类型和分布特征选择正确的统计方法,进行统计描述和统计推断,反映数据的综合特征,阐明事物内在联系和规律,得出科学结论。常见的问题有:抽样研究未作统计检验,只凭表面数据大小即下结论;统计分析方法的选择不当;将偶然因素当作因果分析;忽视对比条件,未作标准化处理;将构成比当作率分析;发病率与患病率相混淆,死亡率与病死率相混淆等。

统计工作的四个基本步骤是紧密联系,不可分割的整体,缺一不可,忽视任何一步都将影响整个研究的结果。

强调指出:统计工作必须持严肃认真、实事求是的科学态度,反对伪造和篡改统计数据。



第五节 SPSS 和 DPS 统计软件介绍

一、关于统计软件的使用

统计软件能迅速完成统计的复杂计算。一般说来,使用统计软件,需要掌握“分析方法、数据格式、读取结论”三个方面的基本技能。只学会统计软件的操作方法,并不见得就会进行数据的统计处理,因为无论统计方法是否正确选择,无论数据格式是否满足运算要求,只要录入数据,统计软件都会输出一大堆计算结果。如果用统计软件乱套计算一通后,随意地从计算结果中挑出一部分感兴趣的東西,由此得出的统计结论缺乏科学性,统计就失去了意义。

本书介绍国际通用统计软件 SPSS 19 的同时,考虑其价格不菲很多人使用不起,特介绍国产优秀统计软件 DPS12,国产统计软件具有价格低廉、建立数据文件和操作方便的优点。

为简便计,本书在说明统计软件的操作方法和输出结果时,采用:

“→”表示用“用鼠标左键点击”的操作;采用系统默认的选项时,一般不予写出。

二、SPSS 19 简介

SPSS 的全称 Statistical Program for Social Sciences,即社会科学统计程序。IBM 在收购 SPSS 后发布了新版本 IBM SPSS 18(2009),最新版本为 IBM SPSS Statistics 19(2010)。本书介绍 IBM SPSS Statistics 19。

(一)数据编辑器

SPSS 启动后,关闭导航对话框,自动打开数据编辑器。

数据编辑器的数据视图与变量视图可以互相切换,状态栏显示“IBM SPSS Statistics Processor 就绪”,见图 1-1。



图 1-1 SPSS 19 的简体中文界面数据编辑器数据视图

1. 数据视图

数据视图用来输入和编辑数据,顶部是菜单栏和工具栏,中部是工作表。



菜单栏或工具栏,可以选择命令完成操作。菜单栏的命令后面标记圆点表示有对话框,标记箭头表示有下拉菜单。对话框中,主要功能按钮有:确定(执行操作),粘贴(将命令粘贴到语法编辑器),重置(重新设置),取消(取消操作),帮助(帮助信息)。→文件→新建或打开→数据,可以打开新的数据编辑器,新建或打开已有的数据文件。在多个数据编辑器打开时,可以通过任务栏的图标切换当前数据编辑器。

工作表,可以指定单元格直接输入数据,也可以从 Word、Excel 文档复制、粘贴数据。工作表每一列为一个变量,每一行为一个案(记录)。工作表上方的当前坐标栏以“个案编号:变量名”显示当前单元格的位置,当前数据栏显示当前单元格的值。左右及上下滚动条,可以移动工作表。

2. 变量视图

变量视图用来定义和修改变量的名称、类型、宽度、小数、标签、值、缺失、列、对齐、度量标准、角色(SPSS18 以前的各版本没有)等 11 项属性,见图 1-2。

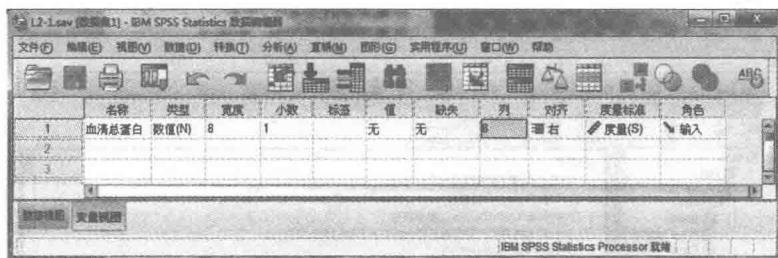


图 1-2 SPSS 19 的简体中文界面的数据编辑器的变量视图

名称、标签分别是变量名与别名,可以在该列的单元格直接键入或粘贴输入。名称的首字符必为字母或汉字(低版本不能超过 8 个字符)。若不定义名称,则系统依次默认为“VAR00001”、“VAR00002”……

类型、值、缺失分别是变量类型、取值说明(标签)、缺失值。点击该列的单元格时,其右侧会出现灰色按钮,点击它,弹出对话框,可以选择数值(系统默认)、字符串等 8 种类型,设置值标签,定义缺失值。

宽度、小数、列分别是变量的宽度、小数位、显示的列宽,系统默认数值型变量的宽度为 8 位、小数为 2 位,列为 8 位。点击该列的单元格时,其右侧会出现上下双向箭头,点击一次可以增加或减少 1 位。

对齐、度量标准分别是变量的对齐方式、度量标准。点击该列的单元格时,其右侧会出现下拉箭头,可以选择左、右、居中三种对齐方式,度量、有序、名义三种度量标准。

角色是数据的状态,有输入、目标、两者都、无、分区、拆分六种。

3. 数据文件

在数据视图输入数据,在变量视图定义属性,形成数据集。可以由编辑菜单、数据菜单或转