

全国高等院校医学实验教学规划教材

SPSS 软件与应用

主编 王在翔 崔庆霞



科学出版社

全国高等院校医学实验教学规划教材

SPSS 软件与应用

主 编 王在翔 崔庆霞
副主编 吕军城 李望晨 王胜男
编 委 (以姓氏笔画为序)
任艳峰 刘成凤 刘晓丽
范俊杰 唐云锋

科 学 出 版 社

北 京

· 版权所有 侵权必究 ·

举报电话:010-64030229;010-64034315;13501151303(打假办)

内 容 简 介

本书是全国高等院校医学实验教学规划教材,是一本面向21世纪融统计学方法知识与SPSS操作实现为一体的SPSS的综合实验教材。本书内容包括SPSS的数据编码、录入与保存,SPSS的数据预处理,数据资料的描述性分析,总体均值的参数假设检验,方差分析,非参数假设检验,列联表资料的检验,资料的相关性分析,回归分析,Logistic回归分析,生存分析,因子分析,信度分析与效度分析,聚类分析与判别分析,对应分析共15章,涵盖了绝大部分统计学方法及其SPSS的操作实现。本书不以详尽的演示教程为目标,而是以必要、实用、够用为原则,通过实际案例,结合作者多年的教学及科研实践经验安排内容,力争做到让读者了解各种统计学方法和原理,掌握其适用对象和基本操作,提高利用统计学方法解决实际问题的能力。

本书可作为高等院校公共卫生与预防医学、统计学(医药院校)、药学、生物学、生物医学工程、心理学等学科相关专业本科生或研究生的教材,也可作为具有初步统计学基础的各类工作人员、高年级本科生及研究生的学习与科研指导用书。

图书在版编目(CIP)数据

SPSS 软件与应用 / 王在翔,崔庆霞主编. —北京:科学出版社,2015.1

全国高等院校医学实验教学规划教材

ISBN 978-7-03-043057-1

I. S… II. ①王… ②崔… III. 医学统计—统计分析—软件包—医学院校—教材 IV. ①R195.1-39

中国版本图书馆CIP数据核字(2015)第013793号

责任编辑:王超 胡治国 / 责任校对:张怡君

责任印制:李利 / 封面设计:范璧合

版权所有,违者必究。未经本社许可,数字图书馆不得使用

科学出版社 出版

北京东黄城根北街16号

邮政编码:100717

<http://www.sciencep.com>

三河市骏杰印刷有限公司 印刷

科学出版社发行 各地新华书店经销

*

2015年1月第 一 版 开本:787×1092 1/16

2015年1月第一次印刷 印张:11

字数:257 000

定价:32.00元

(如有印装质量问题,我社负责调换)

前 言

SPSS 统计软件目前已经成为非统计专业人士进行统计分析的首选工具,是国际公认的最优秀的统计分析软件包之一,在国际学术界有条不紊的规定,凡是用 SPSS 和 SAS 等著名软件包分析的结果,在国际学术交流中可以不必说明算法,可见其权威性和信誉度之高。

统计学方法和 SPSS 软件在实际工作中是分工协作、优势互补的两种工具,统计学方法的作用主要体现在培养统计的思维能力,掌握统计学方法的适用条件和应用范围,学会正确解释与表达数据结果的统计学意义,并结合实际的专业问题得出恰当的结论;而 SPSS 软件的作用则是针对实际问题选择适当的统计学方法,进行简单操作设置,将统计学方法中复杂的统计公式及计算过程作为暗箱,直接实现数据结果的呈现。所以要处理好两者的关系,没有一定的统计学基础,SPSS 软件的学习就成了无源之水,如果不能熟练掌握一种统计软件,统计学方法的应用将会很困难。

理解统计学方法的原理、应用条件及适用范围,掌握其统计软件的操作实现方法及结果解读与呈现,是利用统计学方法解决实际问题必须具备的基本技能。为此,本书对每个统计学方法的内容都是按照统计学知识、实例分析、思考题及实验案例的模式设计的:统计学知识部分简要介绍统计学方法的基本原理、应用条件及适用范围等必备的统计学知识;实例分析部分是通过实验目的、操作步骤及主要结果与分析三个环节,对解决实际案例的目的、SPSS 的具体操作实现以及分析结果的解读等进行有针对性的介绍;思考题部分旨在为读者提出几个相应方法的思考点,强化对该方法的理解;实验案例部分是供读者操作实验用的实际案例,本书所选用的实例和实验案例主要来源于公共卫生、生物医学及心理学等领域,具有较强的示范性。

本书的撰写不以详尽的演示教程为目标,而是以必要、实用、够用为原则,结合作者多年的教学及科研实践经验安排内容。在实例部分的操作步骤模块中,安排了“变量基本要求”“菜单操作”及“界面参数设置”三个环节,将必要的、常用的内容放入其中,尽量做到简明扼要、易懂和实用。为便于读者熟悉统计学的英语词汇或直接采用英文界面操作,本书演示实例所用菜单及按钮等名称都附带英文。

SPSS 软件版本较多,考虑到其功能、占用空间、运行速度及汉化等因素,本

书选用各方面都较为出色的 SPSS 17.0 中文单机版为操作演示版本。

本书的撰写得到了潍坊医学院教务处和公共卫生学院领导和老师的大力支持和帮助,在此表示真诚的谢意。

本书的撰写参照或引用了本学科的有关文献资料,在此对这些参考文献的作者致以衷心感谢。同时,敬请读者对本书的不足之处给予指正。

王在翔

2014 年 6 月于潍坊

目 录

第 1 章 SPSS 数据编码、录入与保存	(1)
1.1 SPSS 简介	(1)
1.2 SPSS 的数据编码	(3)
1.3 SPSS 数据的录入和保存	(6)
第 2 章 SPSS 的数据预处理	(14)
2.1 数据菜单的预处理	(14)
2.2 转换菜单的预处理	(20)
第 3 章 数据资料的描述性分析	(28)
3.1 计量资料的描述性分析	(28)
3.2 计数资料的描述性分析	(35)
第 4 章 总体均值的参数假设检验	(39)
4.1 单样本 t 检验	(39)
4.2 两独立样本 t 检验	(41)
4.3 配对样本 t 检验	(43)
第 5 章 方差分析	(46)
5.1 单因素方差分析	(46)
5.2 单因变量双因素方差分析	(48)
5.3 单因变量协方差分析	(51)
5.4 重复测量方差分析	(56)
第 6 章 非参数假设检验	(62)
6.1 拟合优度检验	(62)
6.2 独立样本的非参数检验	(66)
6.3 相关样本的非参数检验	(70)
第 7 章 列联表资料的检验	(75)
7.1 双向无序列联表的检验	(75)
7.2 单向有序列联表的检验	(78)
7.3 双向有序列联表的检验	(80)
第 8 章 资料的相关性分析	(84)
8.1 计数资料的相关分析	(84)
8.2 计量资料的相关分析	(88)
第 9 章 回归分析	(91)
9.1 线性回归分析	(91)
9.2 曲线回归分析	(96)

9.3 非线性回归分析	(99)
第 10 章 Logistic 回归分析	(104)
10.1 二分类非条件 Logistic 回归分析	(104)
10.2 多分类非条件 Logistic 回归分析	(111)
第 11 章 生存分析	(120)
11.1 生命表分析	(120)
11.2 Kaplan-Meier 分析	(123)
11.3 Cox 回归分析	(126)
第 12 章 因子分析	(133)
第 13 章 信度分析与效度分析	(142)
13.1 信度分析	(142)
13.2 效度分析	(145)
第 14 章 聚类分析与判别分析	(151)
14.1 聚类分析	(151)
14.2 判别分析	(158)
第 15 章 对应分析	(165)
参考文献	(169)

第 1 章 SPSS 数据编码、录入与保存

1.1 SPSS 简介

一、SPSS 的发展

SPSS 是世界上最早的统计分析软件,开始是美国斯坦福大学的研究生 Norman H. Nie 为解决社会学研究中的统计分析问题,和其他两位合作者一起开发的一个用于大型机的统计软件包 Statistical Package for the Social Sciences,即“社会科学统计软件包”,简称 SPSS。其主要发展历史如下:

1968 年,于美国斯坦福大学诞生第一个用于大型机的统计软件包 SPSS。

1975 年,于芝加哥成立 SPSS 公司总部。

1984 年,推出用于个人计算机的 SPSS/PC+。

1992 年,推出 Windows 版本,并且自 SPSS 11.0 起,全称改为 Statistical Product and Service Solutions,即“统计产品和服务解决方案”,仍简称 SPSS。

2008 年,推出“统计产品和服务解决方案”的最终版本 SPSS 17.0(多国语言版)。

2009 年,SPSS 公司宣布重新包装 SPSS 产品线,产品定位为预测统计分析软件(Predictive Analytics Software, PASW),包括四部分: PASW Statistics, 统计分析(即 18.0); PASW Modeler, 数据挖掘; Data Collection Family, 数据收集; PASW Collaboration and Deployment Services, 企业应用服务。

2010 年,SPSS 被 IBM 公司以 12 亿美元收购,各子产品名称前面不再是 PASW,而是统一加上 IBM SPSS 字样,统计分析部分推出 IBM SPSS Statistics 19.0,目前最新版本是 22.0。

本教材以 SPSS 的单机标准版 SPSS 17.0 中文版为蓝本,介绍 SPSS 的应用与操作。

二、SPSS 的特点与功能

SPSS 软件包是国际公认的最优秀的统计分析软件包之一,在国际学术界有条不成文的规定,凡是用 SPSS 和 SAS 等著名软件包分析的结果,在国际学术交流中,可以不必说明算法,可见其权威性和信誉度之高。

SPSS 操作界面是世界上最早采用图形菜单驱动图形界面的统计软件,最突出特点就是操作界面友好,输出结果美观漂亮。它将几乎所有的功能都以统一、规范的图形界面展现出来,使用 Windows 的窗口方式展示各种管理和分析数据方法的功能,对话框展示出各种功能选择项。

SPSS 采用类似 Excel 表格的方式输入与管理数据,数据接口较为通用,能方便地从其他格式数据(如 Excel、txt、dbf)中读入数据;统计过程包括绝大部分常用的较为成熟的统计分析方法;输出结果清晰直观、易学易用,以专用的 SPO 格式存储,可以转存为 HTML 格式和文本格式。

用户只要掌握一定的 Windows 操作技能,精通统计分析原理,就可以使用该软件为特定的科研工作服务,所以 SPSS 是非统计专业人士的首选统计分析工具。当然,对于熟悉编程运行方式的用户,SPSS 还特别设计了语法生成窗口,用户只需在菜单中选择各个选项,然后单击“粘贴”按钮就可以自动生成标准的 SPSS 程序,极大地方便了中高级用户。

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。SPSS 统计分析过程包括描述性统计、均值比较、一般线性模型、广义线性模型、相关分析、回归分析、对数线性模型、分类分析、降维分析、生存分析、时间序列分析、多重响应等多个大类,每类中又分为多个统计过程,每个统计过程中允许用户选择不同的方法及参数。

SPSS 也有一些缺陷,如输出结果虽然简洁,但是很难与一般办公软件如 Office、WPS 等直接兼容,不能用 Word 等常用文字处理软件直接打开,只能采用复制、粘贴的方式加以交互;中文版的某些词汇翻译不恰当甚至有误;统计分析功能与 SAS 等统计软件相比也有一定欠缺。但 SPSS 软件包因其操作简单,便于掌握,已经在医药卫生领域、调查统计行业、市场研究行业、政府和企业的数据分析等领域占有重要地位,是一款经典的统计分析工具。

三、SPSS 17.0 界面与菜单

SPSS 17.0 中文版主界面如图 1-1 所示。



图 1-1 SPSS 17.0 中文版主界面

SPSS 17.0 主菜单共有 11 个,菜单名称及主要功能如下:

文件 (File): 文件新建、打开、保存、打印等。

编辑 (Edit): 文本内容的选择、复制、剪贴、寻找和替换等。

视图 (View): 主界面中工具视图的开关。

数据(Data):变量定义,个案的选择、排序、标识;变量的加权、汇总,文件的合并、拆分、转置,正交设计等。

转换(Transform):变量的数值计算、赋值处理、缺失值替代、随机数生成、时间序列的创建等。

分析(Analyze):这是软件的核心菜单,包括各种常用统计学分析方法,如描述性分析、各种假设检验方法、各种方差分析、相关性分析、各种回归分析、聚类与判别分析、因子分析与对应分析、生存分析、时间序列预测分析等。

图形(Graphs):各种统计图的制作,如散点图、饼图、条形图、直方图、箱图等。

实用程序(Utilities):提供用户的各种选项,如变量定义时的默认选项、变量注释、定义输出标题、窗口设计等。

附加内容(Adds-ons):一般为空。

窗口(Windows):窗口的排列、选择、显示等。

帮助(Help):帮助文件的调用、查寻、显示,软件版本信息等。

四、SPSS 工作步骤

SPSS 是一款专业性较强的统计软件,学习和应用时必须了解和掌握必要的统计学基础知识,只有这样才能避免滥用和误用,避免因引用偏差甚至错误的分析结果而作出错误决策。

统计学方法和 SPSS 软件在实际工作中是分工协作、优势互补的两种工具,学习统计学方法旨在培养统计的思维能力,重点掌握基本统计学方法的适用条件和应用范围,学会正确解释与表达数据结果,并结合实际专业问题得出恰当的结论;而统计软件的作用是选择适当的统计学方法,通过简单操作设置,将统计学中复杂的数学公式及计算作为暗箱,直接实现数据结果的呈现,所以要处理好两者的关系。没有一定的统计学基础,统计软件的学习就成了无源之水,如果不能熟练掌握一种统计软件,统计学方法的应用将会很困难。

SPSS 处理实际问题的一般步骤如下:

(1) SPSS 数据的准备:对原始数据资料进行分析,制定数据的编码方案,录入数据,合并数据文件,保存数据文件等。

(2) 数据的加工整理:数据的筛选、缺失值处理、纠错及新数据变量的生成等。

(3) 数据的统计分析:根据加工整理后的数据特点及实际分析的需要,选择适当的统计学方法进行分析。

(4) 数据结果的解读和呈现:对统计分析的数据结果,解释其统计学意义,再结合问题的实际意义呈现出通俗表达的实际结论。

1.2 SPSS 的数据编码

一、数据编码

数据编码是指将原始数据资料(如问卷、试验结果等)的问题转化为 SPSS 能够识别的变量,并明确变量个数、变量名称、变量类型及变量赋值方法等的规划过程。

原始数据资料的问题有多种类型,主要由其答案类型决定,同时也决定了 SPSS 中变量

的个数及赋值方法。常见问题的类型如下:

- (1) 一般数值型问题:指每个个案中的问题答案是唯一的数值型数据,如身高、体重等。
- (2) 唯一选项数值型问题:指每个个案的问题答案是多选一的数值型数据,如家庭人口数是从 1、2、3、4 中选一个答案。值得注意的是,对于等级资料的多选一问题,一般也要归为这一类,如答案是优、良、中、差中的一个,这时在变量赋值时,通常根据实际需要确定一个赋值标准,对每个个案的答案赋予相应数值。
- (3) 一般字符型问题:指每个个案的问题答案是唯一的字符型数据,如姓名、籍贯等。
- (4) 唯一选项字符型问题:指每个个案的问题答案是多选一的字符型数据,如性别、学历等。
- (5) 限定多选项型问题:指每个个案的问题答案是由多个可选项中限定选择多于一个答案的数据。
- (6) 任意多选项型问题:指每个个案的问题答案是由所有选项中任意选定任何几个答案的数据。

(7) 日期型问题:指每个个案中的问题答案是唯一的日期型数据,如出生日期等。

变量个数是指在不丢失原始数据的情况下,每个问题转变为 SPSS 数据时,需要变量的最少个数,通常由问题的类型所决定。

变量名称是指根据问题实际意义及变量个数,给变量起的名字,变量名称可以是汉字、英文和字母等。变量名不宜太长,但最好便于识别。

变量类型是根据变量取值的特点和统计分析的需要选定的,通常也是由问题的类型所决定的。最常见的三种变量类型如下:

数值型(Numeric):可进行算术和逻辑比较运算的变量。

字符型(String):不能进行算术和逻辑比较运算的变量。

日期型(Date):变量取值是日期的变量,可进行加减运算。

变量赋值方法是指将原始数据资料的问题答案录入 SPSS 中作为变量取值的方法。

二、常见问题类型的数据编码

1. 一般数值型问题的编码 变量个数:1 个;变量类型:数值型;变量赋值方法:将每个个案的数值直接录入。

2. 唯一选项数值型问题的编码 变量个数:1 个;变量类型:数值型;变量赋值方法:按赋值标准将每个个案的选项赋值作为变量取值。

3. 一般字符型问题的编码 变量个数:1 个;变量类型:字符型;变量赋值方法:将每个个案的字符直接录入。

4. 唯一选项字符型问题的编码 变量个数:1 个;变量类型:数值型或字符型;变量赋值方法:按编码规则(一般以答案序号作为各答案的赋值),将每个个案的选项赋值作为变量取值。

5. 限定多选项型问题的编码 变量个数:限定选项的个数;变量类型:数值型或字符型;变量赋值方法:按编码规则(一般以答案序号作为各答案的赋值),依次将每个个案的各选项赋值分别作为各变量的取值。

6. 任意多选项型问题的编码 变量个数:所有选项的个数;变量类型:数值型或字符型;变量赋值方法:按编码规则(一般选中答案赋值为 1,不选中答案赋值为 0,若有多种选择需进行赋值编排),依次将每个个案的各选项赋值作为相应变量的取值。

7. 日期型问题的编码 变量个数:1个;变量类型:日期型;变量赋值方法:将每个个案的日期直接录入。

三、数据编码方案表的编制

对于一项调研指标较多、样本量较大的研究项目,在录入 SPSS 之前,先编制一张规范的数据编码方案表,将每个问题的变量个数、变量名称、变量类型及变量赋值方法等列在其中,是一项必需的工作程序,对保证数据录入的质量、统计分析的效率以及后期的保存与检索都有重要意义。

【实例分析】

[实例 1.1] 对下面的“新型农村合作医疗基本信息调查问卷”编制编码方案表。

新型农村合作医疗基本信息调查问卷(简化)

- 1. 您的姓名:王俊。
 - 2. 您的性别:√a. 男 b. 女
 - 3. 您的出生日期:1980.06.06。
 - 4. 您的学历:a. 初中及以下 b. 高中及中专 √c. 大专 d. 本科及以上
 - 5. 您参加新农合至今约 8 年。
 - 6. 您是通过何种途径了解新农合的(限选 3 项)?
√a. 电视 b. 广播 √c. 报纸 √d. 宣传 e. 其他
 - 7. 您平时是否有常发病?
√a. 是 b. 否
 - 8. 若有常发病,为哪些系统疾病?
√a. 呼吸系统 b. 消化系统 √c. 内分泌系统 d. 泌尿系统 e. 其他
 - 9. 您对新农合的总体满意度如何?
a. 很高 b. 高 √c. 一般 d. 差 e. 很差
 - 10. 您曾经对新农合抱怨过吗?
a. 经常抱怨 √b. 抱怨 c. 一般 d. 偶尔抱怨 e. 从没抱怨
- 新农合基本信息调查编码方案表如表 1-1 所示。

表 1-1 新农合基本信息调查编码方案表

题号	变量名称	数据编码	变量类型	个案选项	变量赋值
1	姓名	按实际姓名录入	字符型	王俊	王俊
2	性别	1-男,0-女	数值型	男	1
3	出生日期	按实际日期录入	日期型	1980.06.06	1980.06.06
4	学历	1-初中及以下,2-高中及中专,3-大专,4-本科及以上	数值型	大专	3
5	参加新农合年数	按实际年数录入	数值型	8	8
6	新农合途径 1	1-电视,2-广播,3-报纸,4-宣传,5-其他	数值型	电视	1
6	新农合途径 2	1-电视,2-广播,3-报纸,4-宣传,5-其他	数值型	报纸	3
6	新农合途径 3	1-电视,2-广播,3-报纸,4-宣传,5-其他	数值型	宣传	4
7	是否有常发病	1-是,0-否	数值型	是	1
8	呼吸系统疾病	1-有,0-无	数值型	呼吸系统	1
8	消化系统疾病	1-有,0-无	数值型		0

续表

题号	变量名称	数据编码	变量类型	个案选项	变量赋值
8	分泌系统疾病	1-有,0-无	数值型	内分泌系统	1
8	泌尿系统疾病	1-有,0-无	数值型		0
8	其他疾病	1-有,0-无	数值型		0
9	新农合满意度	10-很高,8-高,6-一般,4-差,0-很差	数值型	一般	6
10	新农合抱怨	0-经常抱怨,2-抱怨,4-一般,7-偶尔抱怨,10-没抱怨	数值型	抱怨	2

1.3 SPSS 数据的录入和保存

SPSS 数据的录入通常有两种方式,一是在 SPSS 数据管理窗口中新建数据文件,二是打开已存在的 SPSS 数据文件(.sav)或其他类型(如.xls、.txt、.dbf 等)的数据文件;同样,也可保存为 SPSS 的数据文件(.sav)和其他数据文件(如.xls、.txt、.dbf 等)。

一、在 SPSS 数据管理窗口新建文件

在 SPSS 数据管理窗口新建文件需要变量定义和数据输入两个环节:首先根据编码方案进行变量的定义,依次对变量的名称、类型、宽度、小数位数、值标签、缺失值等分别进行定义和设置;其次按照编码方案依次录入各个案对应变量的赋值。

(一) 变量定义

进入 SPSS 主界面,单击左下角“变量视图(Variable View)”项,进入变量定义界面,如图 1-2 所示。

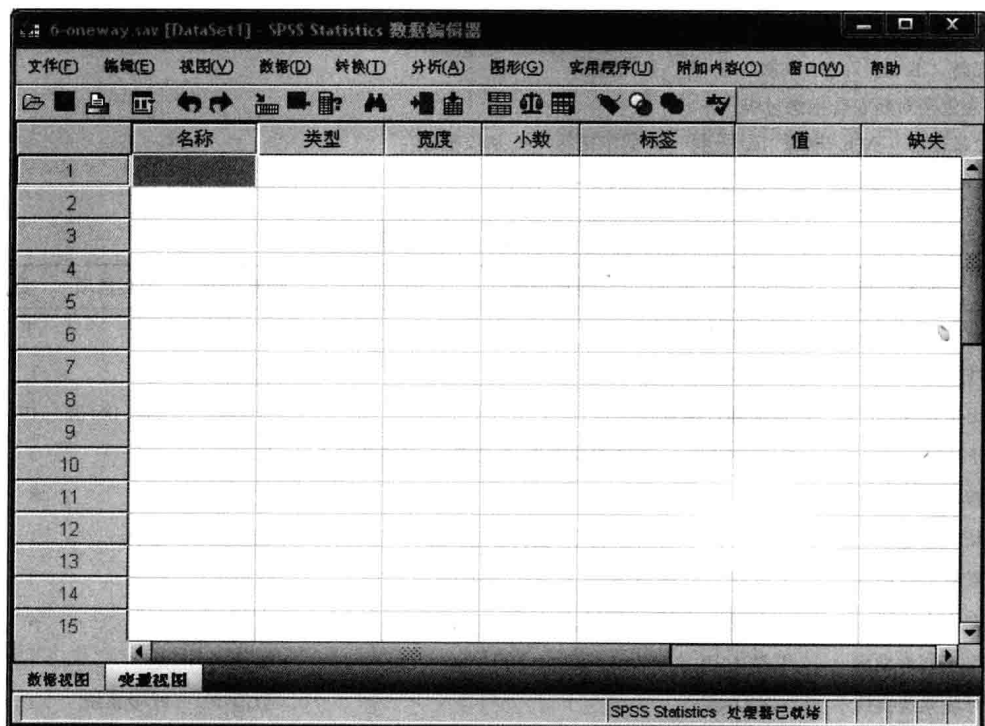


图 1-2 变量定义界面

1. 变量名称(Name) SPSS 17.0 对变量名称长度没有要求,但不宜太长,可以用汉字、英文、字母等作为文件名,但首位字符不能为数字或诸如“?”“、”等特殊字符。
2. 变量类型(Type) SPSS 提供了 8 种变量类型,如图 1-3 所示。

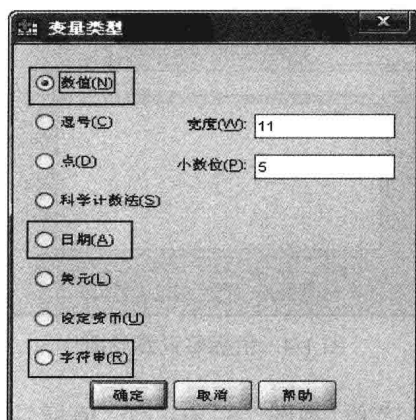


图 1-3 变量类型设置界面

最常用的变量类型有 3 种:

数值型 (Numeric): 可以同时定义变量取值的宽度 (Width) 和小数位数 (Decimal Places), 默认为 8 位宽度, 2 位小数。变量的宽度和小数位数都是根据变量取值的实际情况确定的。

字符型 (String): 可以定义字符长度以便输入字符。

日期型 (Date): 可从系统提供的日期格式中选择自己需要的。如选择 yy/mm/dd 类型, 则录入数据时必须按照该格式, 如录入 1990 年 6 月 12 日需要输入 90/06/12, 显示时也是这个格式。当然, 各种日期格式可以自动转换, 只需对日期型变量重新选择一个新格式, 则以前的日期格式可以自动转换成新格式, 新录入日期数据时应该按新的格式输入。

其余 5 种变量类型实质上都属于数值型, 只不过显示的格式不同。

货币型 (Dollar): 可从系统提供的格式中选择自己需要的, 并可以定义数值宽度和小数位数, 显示形式为数值前带有 \$, 如 \$669.38。

加显逗号型 (Comma): 整数部分每 3 位数加一个逗号, 其余同数值型, 如 13896.126 显示为 13,896.126。

三位加点型 (Dot): 以整数形式显示, 每 3 位加一点 (但不是小数点), 可定义小数位数, 但都显示 0, 且小数点用逗号表示。如 8.2645 显示为 8.2645,0000 (实际是 82645E-4)。

科学记数型 (Scientific Notation): 可以定义数值宽度和小数位数, 在数据管理窗口中以指数形式显示。例如, 定义数值宽度为 10, 小数位数为 3, 945.698 显示为 9.457E+002。

常用型 (Custom Currency): 显示为整数部分每 3 位加一个逗号, 用户可定义数值宽度和小数位数。例如, 72745.678 显示为 72,745.678, 同加显逗号型。

3. 变量标签 (Label) 用于定义变量名称的标签, 一般用于说明变量名称的实际含义, 在某些结果表格中变量名称将会以变量标签显示, 所以要慎用。

4. 值标签 (Values) 用于定义变量值及其代表的含义, 一般当变量取值是多项选择时, 常要设置值标签, 以减少录入数据时出错的概率及方便数据的检查和纠错, 同时在结果

表格中变量取值也将会显示值标签。这也是直接利用 SPSS 录入数据相较导入其他格式数据的优势之一。值标签设置界面如图 1-4 所示。

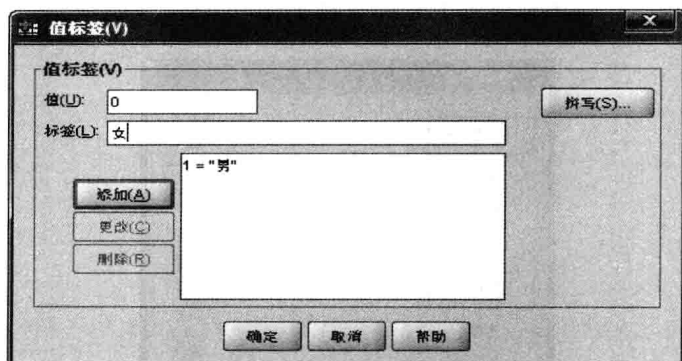


图 1-4 值标签设置界面

需要说明一点,在 SPSS 主菜单“视图”下,有一个“值标签”开关功能菜单,在数据编辑视图界面下有效,默认是关闭的,打开时输入的变量值显示设置的值标签。

5. 缺失值 在实际工作中,出于各种原因会出现部分个案数据缺失的现象,尤其对于数值型数据,处理不好可能会极大地影响最后的分析结果。在录入数值型数据时,如果数据框中不输入任何字符,则系统默认该值为缺失值,若输入任何数值,则系统将其作为正常值看待,除非进行缺失值设置。缺失值设置其实就是将缺失的数据用不同于变量正常值的数值标识出来,而且可以分类标识,系统分析时能够加以识别,如身高变量,若有缺失值,可用-1 标识。缺失值设置界面如图 1-5 所示。

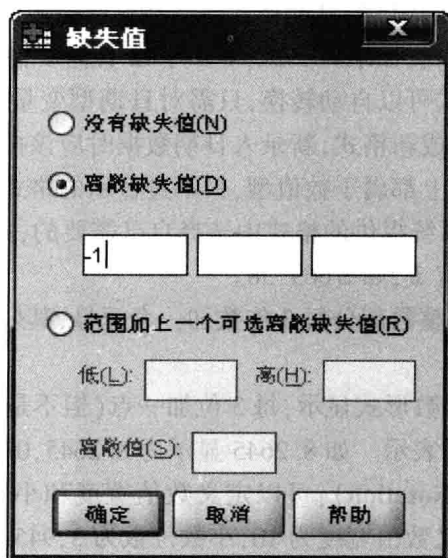


图 1-5 缺失值设置界面

【实例分析】

[实例 1.2] 按[实例 1.1]编制的“新型农村合作医疗基本信息调查问卷”编码方案进行变量定义。

[实验目的] 理解各种问题类型及其数据编码,掌握各种类型变量定义的 SPSS 操作。

[操作步骤]

界面选择:SPSS 主界面→变量视图。

界面操作:按上面介绍的方法依次在各行分别输入变量名称;然后设置变量类型及宽度,数值型变量还要设置小数位数;若变量取值是多项选择,一般要设置值标签;数值型变量若有缺失值,可以设置缺失值。

[主要结果与分析] 变量定义结果如图 1-6 所示。

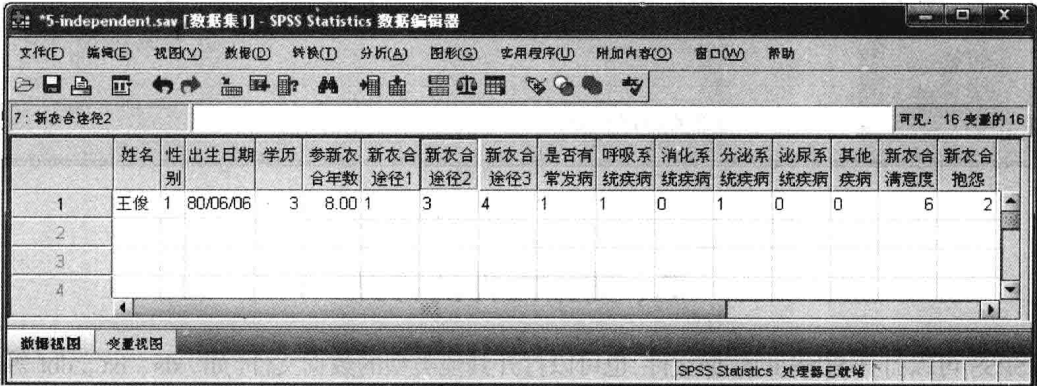


	名称	类型	宽度	小数	标签	值	缺失
1	姓名	字符串	8	0		无	无
2	性别	数值(N)	4	0		{0, 女}...	无
3	出生日期	日期	8	0		无	无
4	学历	数值(N)	4	0		{1, 初中及以...	无
5	参新农合年数	数值(N)	6	2		无	-1.00
6	新农合途径1	数值(N)	4	0		{1, 电视}...	无
7	新农合途径2	数值(N)	4	0		{1, 电视}...	无
8	新农合途径3	数值(N)	4	0		{1, 电视}...	无
9	是否有常发病	数值(N)	4	0		{0, 否}...	无
10	呼吸系统疾病	数值(N)	4	0		{0, 无}...	无
11	消化系统疾病	数值(N)	4	0		{0, 无}...	无
12	分泌系统疾病	数值(N)	4	0		{0, 无}...	无
13	泌尿系统疾病	数值(N)	4	0		{0, 无}...	无
14	其他疾病	数值(N)	4	0		{0, 无}...	无
15	新农合满意度	数值(N)	4	0		{0, 很差}...	-1

图 1-6 [实例 1.1]变量定义结果

(二) 数据的录入与保存

进入 SPSS 主界面,单击“数据视图”项进入数据编辑界面,将[实例 1.1]编码方案表中个案选项的变量赋值录入,结果如图 1-7 所示。



	姓名	性别	出生日期	学历	参新农合年数	新农合途径1	新农合途径2	新农合途径3	是否有常发病	呼吸系统疾病	消化系统疾病	分泌系统疾病	泌尿系统疾病	其他疾病	新农合满意度	新农合抱怨
1	王俊	1	80/06/06	3	8.00	1	3	4	1	1	0	1	0	0	6	2
2																
3																
4																

图 1-7 [实例 1.1]编码方案表中个案变量赋值录入结果

打开“视图”菜单,单击“值标签”开关,将显示值标签,如图 1-8 所示。



图 1-8 [实例 1.1] 编码方案表中举例变量赋值标签显示结果

数据录入完毕或录完一部分都可以随时进行保存,保存界面与 Word、Excel 等软件基本一致,默认数据文件保存类型为 SPSS Statistics(*.sav),其他数据文件类型需要先选择,本例保存数据文件名为实例 1-2. sav,如图 1-9 所示。



图 1-9 SPSS 数据保存界面

二、数据文件的打开

SPSS 可以打开 SPSS 的数据文件,也可以打开其他类型的数据文件(如 .xls、.txt、.dbf 等)。

1. SPSS 的数据文件(.sav)的打开 执行“文件(File)”→“打开(Open)”→“数据(Da-