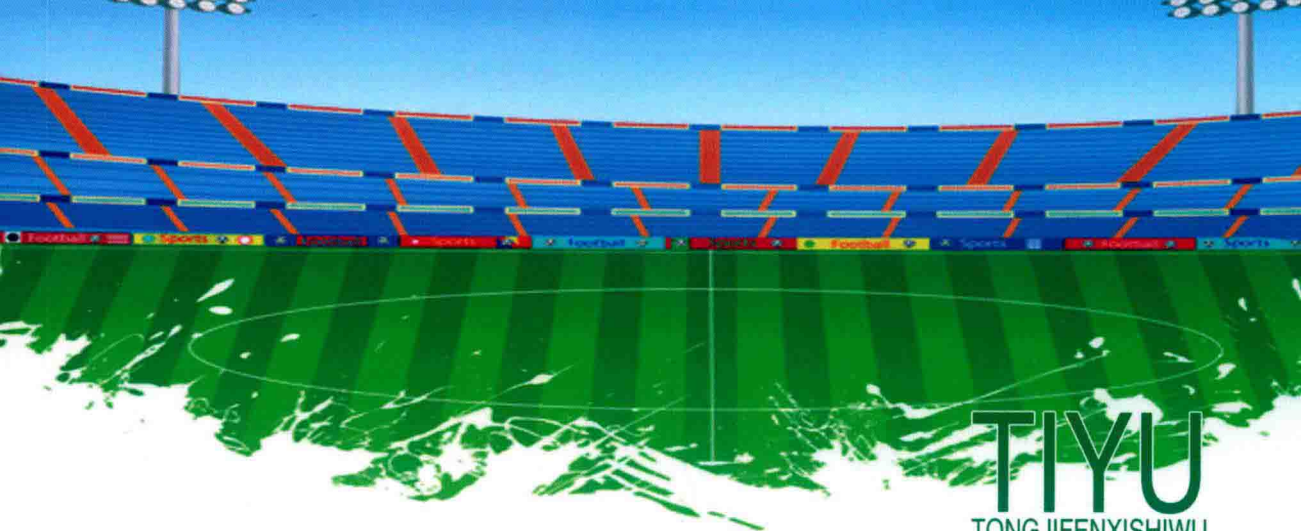




TIYU

TONGJIFENXISHIWU

SPSS CAOZUO JI YINGYONG

体育统计分析实务

SPSS操作及应用

苏 杨 · 著



东南大学出版社
SOUTHEAST UNIVERSITY PRESS

江苏高校品牌专业建设工程(TAPP)
南京体育学院运动人体科学专业资助项目

体育统计分析实务

——SPSS 操作及应用

苏 杨 著



东南大学出版社
SOUTHEAST UNIVERSITY PRESS

• 南京 •

内 容 提 要

本书采用 IBM SPSS Statistics 20 中文版,基于体育研究中的几类典型问题,在基本统计原理的解析基础上,从实务应用的角度出发,创造性地以体育领域真实的案例贯穿全书,介绍常见的数据处理及其 SPSS 统计分析流程。内容从体育实例到调查问卷,从参数检验到非参数检验,从描述性统计到推断性统计,主要包括:SPSS 的界面操作、数据文件建立、常见数据文件管理、连续变量和分类变量的统计描述、常用统计表制作、假设检验、方差分析、卡方检验、非参数检验、相关和回归分析、聚类分析、主成分和因子分析。

本书根据读者群数理基础薄弱的特点,完全从实际案例操作出发,讲解各类方法的综合运用,以协助读者提高实战能力,更加方便准确地分析体育相关数据。

本书可作为普通高等院校体育类各专业本科生与研究生的教材,也可作为体育教师、教练员、体育科研工作者、体育管理工作者的参考书。

图书在版编目(CIP)数据

体育统计分析实务:SPSS 操作及应用/苏杨著.

南京:东南大学出版社,2019.3

ISBN 978-7-5641-8327-1

I. ①体… II. ①苏… III. ①体育统计-统计分
析-高等学校-教材 IV. ①G80-32

中国版本图书馆 CIP 数据核字(2019)第 046184 号

出版发行:东南大学出版社

社 址:南京市四牌楼 2 号 邮编:210096

出 版 人:江建中

网 址:<http://www.seupress.com>

电子邮箱:press@seupress.com

经 销:全国各地新华书店

印 刷:虎彩印艺股份有限公司

开 本:787 mm×1092 mm

印 张:11.5

字 数:288 千字

版 次:2019 年 3 月第 1 版

印 次:2019 年 3 月第 1 次印刷

书 号:ISBN 978-7-5641-8327-1

定 价:39.00 元

本社图书若有印装质量问题,请直接与营销部联系。电话(传真):025-83791830

前 言

体育统计是基础应用学科之一,它是数理统计方法在体育领域中的应用,为体育领域中的各类研究提供收集、整理与分析数据资料的统计方法。体育统计主要以体育领域中的随机现象为研究对象,研究其规律性。体育统计方法贯穿整个体育研究过程。

SPSS(Statistical Package for the Social Sciences)是世界上最早的统计分析软件,主要用于统计学分析运算、数据挖掘、预测分析和决策支持任务的完成。目前,国际常用统计软件有 SPSS、SAS、MATLAB 等。这些软件的功能和作用大同小异,但各有特色。其中,SPSS、SAS 是目前大多数企业、各类科研院所中较为流行的。SPSS 是为广大非专业人士设计的,是世界上最早采用图形菜单驱动界面的统计软件,具有操作简便、好学易懂的特点。

本书将体育领域中常见典型的数据统计问题进行分类,结合体育研究中的实际问题作为案例,说明如何用 SPSS 解决这些问题。内容主要介绍 SPSS 数据录入、数据管理、统计分析、报表制作等,包括常规的描述性统计、T 检验、方差分析、卡方检验、非参数检验、相关分析和回归分析;也包括近期发展的多元统计技术,如多元回归分析、聚类分析、主成分分析和因子分析等方法。

本书有以下特色:

(1) 全面 本书的内容覆盖了体育统计学科的各种基本统计分析方法及其 SPSS 所提供的基本统计分析功能。内容安排上以统计方法体系为主线,围绕描述性统计和推断性统计展开。

(2) 易懂 体育统计中多数统计模型都比较复杂,为降低读者学习的难度,本书中各章节大幅简化对体育统计理论的介绍和解释,重点选择较为简单易懂

的典型数据作为分析案例,突出各种体育统计模型的方法、SPSS 操作及结果解释。

(3) 注重理论联系实际 读者只需按照完整的操作步骤,即可完成相应的数据统计分析,强化理论的应用性,真正帮助读者学以致用。

(4) 内容丰富多元化 详细的结果解析与说明,让读者真正理解各种输出图表的意义。

本书希望通过典型体育案例的讲解和延伸讨论,帮助一部分专业技术强但数理基础薄弱的读者学习并运用合理的统计方法及技术分析数据,进一步对所得的结果进行准确的解释和应用,从而更好地满足读者的工作和学习需求。

运动健康科学系运动人体专业作为江苏省高校品牌专业建设工程资助项目,对本书的撰写、出版提供了支持。在此,非常感谢我的领导孙颀教授和宋雅伟教授给予支持;感谢我的同事叶强、吴文广、王阳阳老师的帮助,他们给本书提出了宝贵的意见。

编 者

2019 年 1 月

目 录

1	SPSS 与体育统计	1
1.1	SPSS 概述	1
1.1.1	启动 SPSS	2
1.1.2	SPSS 的数据编辑窗口	2
1.1.3	SPSS 结果输出查看器窗口	3
1.1.4	退出 SPSS	3
1.2	体育统计中的问题	4
1.2.1	体育统计中的基本概念	5
1.2.2	体育统计的常见问题类型	5
2	数据文件的建立	7
2.1	变量设置	7
2.1.1	定义变量名称	8
2.1.2	定义变量类型	9
2.1.3	变量宽度和小数位数	10
2.1.4	变量标签	11
2.1.5	变量值标签	11
2.1.6	缺失值定义方式	11
2.1.7	列宽和对齐方式	12
2.1.8	变量的度量标准	12
2.1.9	角色	12
2.2	调查问卷的直接录入	12
2.3	从其他文件导入数据	17
3	数据常用管理及整理	18
3.1	分类汇总	18
3.2	替换缺失值处理	19
3.2.1	锁定缺失值位置	20
3.2.2	填补缺失值	21
3.3	极端值的清理	22

4 连续变量的描述统计分析	25
4.1 描述性统计的指标体系	25
4.1.1 集中量数	25
4.1.2 离散量数	27
4.2 描述性统计分析的 SPSS 案例分析	29
4.2.1 频率统计	29
4.2.2 描述性统计	31
4.2.3 探索分析与参数估计	32
5 分类变量的描述统计	35
5.1 单个分类变量的描述统计	35
5.2 多个分类变量的联合描述统计	36
5.3 多选题的描述统计	38
5.3.1 多选题的频数列表	38
5.3.2 多选题的交叉列表分析	39
6 统计表	42
6.1 OLAP Cubes(在线分析处理)	42
6.2 个案汇总	45
6.3 分类变量制表	49
6.4 多选题的统计报表制作	53
7 T 检验	58
7.1 假设检验概述	58
7.1.1 统计假设	58
7.1.2 假设检验	58
7.1.3 假设检验的基本思想	58
7.1.4 假设检验的两类错误	60
7.2 单样本 T 检验	60
7.2.1 基本概念和方法	60
7.2.2 案例分析	60
7.3 配对样本 T 检验	62
7.3.1 基本概念和方法	62
7.3.2 案例分析	62
7.4 独立样本 T 检验	64
7.4.1 基本概念和方法	64
7.4.2 案例分析	64
8 方差分析	67
8.1 基本概念和方法	67
8.2 单因素方差分析	67

8.2.1	问题提出	67
8.2.2	单因素方差分析的方法步骤	68
8.2.3	案例分析	69
8.3	双因素方差分析	75
8.3.1	问题提出	75
8.3.2	双因素方差分析的方法步骤	75
8.3.3	无重复试验的双因素方差分析案例分析	77
8.3.4	析因双因素方差分析案例分析	84
9	非参数检验	89
9.1	二项分布检验	89
9.1.1	基本概念和方法	89
9.1.2	案例分析	89
9.2	单样本 K-S(Kolmogorov-Smirnov)检验	91
9.3	两独立样本非参数检验	93
9.3.1	基本概念和方法	93
9.3.2	案例分析	94
9.4	两配对样本非参数检验	96
9.4.1	基本概念和方法	96
9.4.2	案例分析	97
9.5	多独立样本非参数检验	99
9.5.1	基本概念和方法	99
9.5.2	案例分析	100
9.6	多配对样本非参数检验	102
9.6.1	基本概念和方法	102
9.6.2	Friedman 检验案例分析	104
9.6.3	Kendall's W 检验案例分析	105
9.6.4	Cochran's Q 检验案例分析	107
10	卡方检验	109
10.1	卡方检验解决的常见问题类别	109
10.2	卡方检验的基本概念和方法	109
10.3	单样本案例分析	110
10.4	一般卡方案例分析	111
10.5	配对卡方案例分析	119
10.6	分层卡方案例分析	122
11	相关分析	126
11.1	两种不同的变量关系	126
11.2	连续变量的相关分析	127

11.2.1	基本概念和方法	127
11.2.2	案例分析	127
11.3	等级变量相关分析	129
11.3.1	基本概念和方法	129
11.3.2	案例分析	130
11.4	偏相关分析	131
11.4.1	基本概念和方法	131
11.4.2	案例分析	132
11.5	距离相关分析	133
11.5.1	基本概念和方法	133
11.5.2	案例分析	134
12	回归分析	137
12.1	一元线性回归分析	137
12.1.1	基本概念和方法	137
12.1.2	案例分析	138
12.2	多元线性回归分析	141
12.2.1	基本概念和方法	141
12.2.2	案例分析	142
12.3	曲线拟合	146
12.3.1	基本概念和方法	146
12.3.2	案例分析	146
12.4	二分类 Logistic 回归分析	149
12.4.1	基本概念和方法	149
12.4.2	二分类自变量案例分析	150
12.4.3	多分类自变量案例分析	152
13	聚类分析	154
13.1	层次聚类分析	154
13.1.1	基本概念和方法	154
13.1.2	案例分析	155
13.2	快速聚类分析	158
13.2.1	基本概念和方法	158
13.2.2	案例分析	158
14	因子分析与主成分分析	163
14.1	基本概念和方法	163
14.2	案例分析	164
	参考文献	173

1 SPSS 与体育统计

1.1 SPSS 概述

SPSS(Statistical Package for the Social Sciences)是世界上最早的统计分析软件,主要用于统计学分析运算、数据挖掘、预测分析和决策支持任务的完成。SPSS 由美国斯坦福大学的三位研究生于 1968 年研究开发成功,同时成立了 SPSS 公司,并于 1975 年成立法人组织,在芝加哥组建了 SPSS 总部。1984 年 SPSS 总部首先推出了世界上第一个统计分析软件微机版本 SPSS/PC+,开创了 SPSS 微机系列产品的开发方向,极大地扩充了它的应用范围,并使其很快地应用于自然科学、技术科学、社会科学的各个领域。世界上许多有影响的报纸杂志纷纷就 SPSS 的自动统计绘图、数据的深入分析、使用方便、功能齐全等方面给予了高度的评价与称赞。迄今 SPSS 软件已有 50 余年的成长历史。全球约有 25 万家产品用户,它们分布于通信、医疗、银行、证券、保险、制造、商业、市场研究、科研教育等多个领域和行业,是世界上应用最广泛的专业统计软件。在国际学术界有条不紊的规定,即在国际学术交流中,凡是用 SPSS 软件完成的计算和统计分析,可以不必说明算法,由此可见其影响力之大和信誉之高。2009 年 7 月 28 日,IBM 公司宣布将用 12 亿美元现金收购统计分析软件提供商 SPSS 公司。目前 SPSS 的最新版本为 25,而且更名为 IBM SPSS Statistics。SPSS 是软件英文名称的首字母缩写,原意为 Statistical Package for the Social Sciences,即“社会科学统计软件包”。但是随着 SPSS 产品服务领域的扩大和服务深度的增加,SPSS 公司已于 2000 年正式将英文全称更改为 Statistical Product and Service Solutions,意为“统计产品与服务解决方案”,标志着 SPSS 的战略方向正在做出重大调整。

目前,国际常用统计软件有 SPSS、SAS、BMDP、GLIM、GENSTAT、EPILOG、MATLAB 等。这些软件的功能和作用大同小异,但各有特色。其中,SPSS、SAS 是目前大多数企业、各类科研院所中较为流行的,尤其是 SPSS 的总体印象分最高。SPSS 是世界上最早采用图形菜单驱动界面的统计软件,最突出的特点就是操作界面极为友好,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。用户只要掌握一定的 Windows 操作技能,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。该软件易学易用、功能强大,是非专业统计人员的首选统计软件。

SPSS 集数据录入、资料编辑、数据管理、统计分析、报表制作、图形绘制于一体,包括常

规的集中量数和离散量数、相关分析、回归分析、方差分析、卡方检验、T 检验和非参数检验；也包括近期发展的多元统计技术，如多元回归分析、聚类分析、判别分析、主成分分析和因子分析等方法。

SPSS 自 SPSS 16.0 起支持 Windows 8/10、Mac OS X、Linux 及 UNIX。SPSS for Windows 是一个组合式软件包，它集数据录入、整理、分析功能于一身。本书以 SPSS 20 for Windows 标准版为案例讲解，并简称为 SPSS。

1.1.1 启动 SPSS

软件安装完毕，Windows 系统自动在“开始”菜单中创建了 SPSS 软件的快捷方式。打开“开始”菜单中的“所有程序”菜单项，单击“IBM SPSS Statistics 20.0”启动 SPSS，如图 1.1 所示。

1.1.2 SPSS 的数据编辑窗口

SPSS 主要的界面有两个：SPSS 数据编辑窗口和 SPSS 结果输出查看器窗口。首次启动 SPSS 直接进入数据编辑窗口，如图 1.2 所示。该窗口包含数据视图和变量视图两种视图方式。与电子表格类似，数据编辑窗口的左下方是视图切换标签，包括“数据视图”和“变量视图”。在新建数据文件时，首先进入变量视图进行变量类型的定义，其次进入数据视图录入并处理数据（见第 2 章）。

数据视图界面如图 1.2 所示。从上而下主要由标题栏、菜单栏、常用工具栏、编辑栏、变量名称栏、内容区域、窗口切换标签页及状态栏组成。

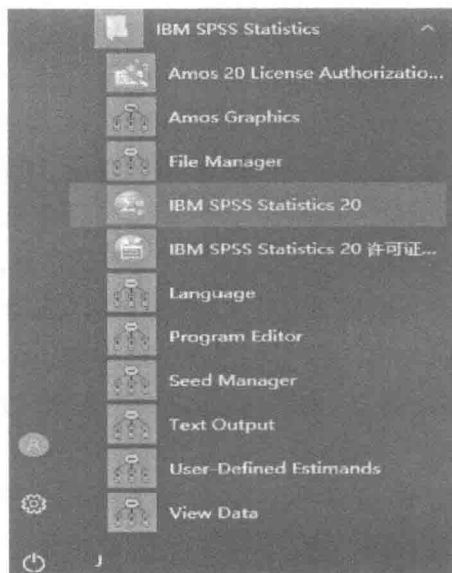


图 1.1 启动 SPSS



图 1.2 SPSS 数据编辑窗口

1) 标题栏

标题栏通常显示默认数据文件名为“未标题 1[数据集 0]”，数据文件的后缀名为 .sav。SPSS 的各类命令菜单在菜单栏中列出，如“文件”“编辑”“帮助”等常规命令菜单。“数据”菜单实现数据基本处理的相关功能，主要的统计功能命令集中在“分析”菜单中，“图形”是构建图表的菜单，“窗口”菜单实现各窗口的切换。常用工具栏中的快捷图标通常表示一些常用命令，如排序、选择个案等。

2) 编辑栏

编辑栏的功能与电子表格类似，显示当前数据单元格的内容并可同步修改。

3) 内容区域

内容区域的左侧是行号，用阿拉伯数字表示，每一行表示一个个案。列号默认为“变量”，通常在变量视图中根据需要进行自定义(见第 2 章)。

4) 状态栏

状态栏显示 SPSS 当前的运行状态。“IBM SPSS Statistics Processor 就绪”表示 SPSS 等待用户操作。

1.1.3 SPSS 结果输出查看器窗口

当用户在数据编辑窗口中完成变量定义、数据输入并进行统计分析后，SPSS 将自动弹出结果输出查看器窗口，如图 1.3 所示。该窗口主要显示数据统计分析的过程及其结果，其输出结果分左、右两个部分，左侧窗口为索引输出区，主要是分析结果的标题及其内容索引；右侧窗口为结果输出区，显示统计分析的所有结果，包括图表等。SPSS 结果输出文件默认名为“输出 1[文档 1]”，结果文件以 .spv 的类型保存。



图 1.3 SPSS 数据结果输出查看器窗口

1.1.4 退出 SPSS

选择数据编辑窗口的“文件”菜单，单击“另存为”命令即可。关闭 SPSS 文件时不仅要保存数据编辑窗口的 .sav 数据文件，如图 1.4 所示，而且要保存结果输出查看器窗口的 .spv 文件，默认文件名为“输出 1”，如图 1.5 所示。



图 1.4 保存数据编辑窗口的 .sav 数据文件



图 1.5 保存结果输出查看器窗口的 .spv 文件

1.2 体育统计中的问题

统计学作为一门应用科学,在收集、整理和分析数据资料的基础上,揭示其数据规律。作为决策科学的重要工具,统计学已渗透入多门学科,如医学、社会学等。体育统计学是统计理论及方法在体育学中的应用,基本划分为描述性统计和推断性统计两大类。体育统计基于概率统计中的大量定律及相关统计理论基础,旨在研究并揭示体育教学训练、科研及管理中的数量规律,从而为体育领域的各类行为做参考决策,促进体育科学化发展。

体育统计的研究过程通常如下:

- (1) 根据研究目的,进行体育统计设计,如设计相应的调查问卷。
- (2) 结合统计的抽样调查方法和特定的体育调查手段,采集客观数据并整理。

- (3) 利用描述性统计方法和推断性统计方法对获取的数据进行分析判断。
- (4) 得出统计结论并为体育决策提供依据。

1.2.1 体育统计中的基本概念

1) 总体和样本

总体是指所研究对象的全体。样本是指从总体中抽取的部分研究对象的集合。例如,研究某市 18 周岁男生的速度素质现状,选择 100 m 跑成绩为测试项目。总体是该市 18 周岁男生 100 m 跑步成绩,由于人数过多工作量太大,实际操作中通常抽取部分该年龄男生(如 800 名)进行测试,则被抽取的 800 名 18 周岁男生 100 m 跑步成绩构成了该研究的一个样本,其中样本含量为 800,通常记作 $n=800$ 。该样本中的每名男生 100 m 跑步成绩称为个体。

体育统计中推断性统计占很大比例,推断性统计即通过样本的信息推断总体信息。因此,首先要保证一定的样本含量,以便减小抽样误差;其次必须注意抽样方法的合理性,以保证样本有较好的代表性。

2) 指标或变量

指标是指反映研究对象特性的具体描述,即 SPSS 中的变量。变量的取值即为数据。例如,身高和体重是反映青少年形态发育状况的常用指标,因此,在 SPSS 中设置“身高”变量后,160 cm、170 cm 等是采集的该变量的数据。

根据不同的数据类型,将变量分为不同的类型,如离散型变量和连续型变量;如分类变量、序号变量、度量变量。这部分内容将在后续 SPSS 变量设置中详细阐述(见第 2 章)。

1.2.2 体育统计的常见问题类型

体育统计方法大致分为描述性统计和推断性统计两大类。本节主要将体育科学研究中的典型问题进行归类,给出解决此类问题的相应统计方法。具体方法如下:

(1) 对体育领域的数据进行描述性统计。

例如,抽样调查某地区学生的体质状况,分别从形态、力量、速度、耐力等项目进行测试,对上述指标进行描述性统计,如频数统计、频率、方差等。

(2) 两个未知总体之间特征数据的差异显著性比较。

例如,研究田径专业和足球专业学生的 100 m 跑成绩有无显著性差异。

(3) 多个未知总体之间特征数据的差异显著性比较。

例如,研究三种不同教学方法对跳远成绩有无影响。

(4) 两个变量数据的相关程度。

例如,肌肉拉力与肌肉收缩速度之间的相关关系。

(5) 对于有较强相关性的变量,探究回归函数关系。

例如,研究某校 20 岁女生的身高和体重的回归方程。

(6) 判断变量是否服从指定分布。

例如,某市 60 名 45 岁女性身高是否服从正态分布。

(7) 某变量是否服从特定分布。

例如,某省体质监测中性别比例是否服从正态分布。

(8) 将多个研究对象进行聚类。

例如,将体育爱好者的 9 项特征指标凝聚成三大类。

(9) 如何将不同量纲的多个指标归结为少数几个综合因素(因子)。

例如,根据优秀跳高运动员各项成绩的测试数据,如原地纵跳、助跑摸高成绩等,研究优秀跳高运动员应具备的综合能力。

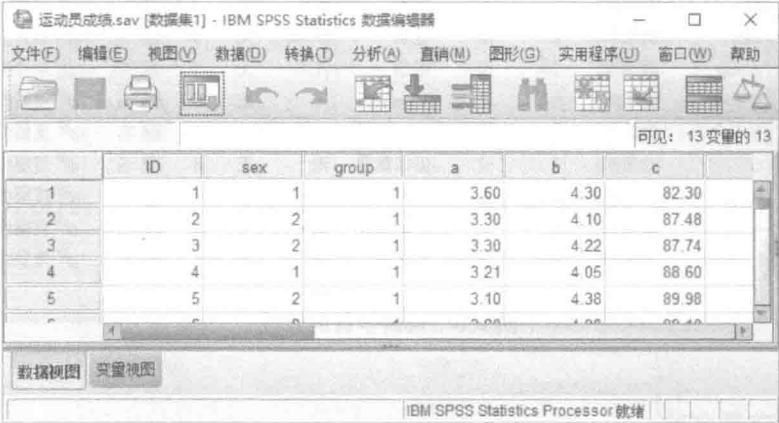
2 数据文件的建立

数据文件的建立是进行统计分析的基础,因此,本章主要围绕如何将科研、教学过程中采集的各种数据以文件方式建立并储存起来。在 SPSS 中建立数据文件的途径大致分为两种:一是根据原始数据资料(如数据图片、调查问卷等),录入数据建立文件;二是将已有其他格式(如电子表格)的数据文件导入 SPSS。

针对上述两类方式,本章将展开说明。一般来说,直接录入法包括变量设置、数据录入、保存数据文件等几个步骤。调查问卷中不同题型的不同录入方法值得重视。

2.1 变量设置

【案例 2-1】 对运动员多项素质练习成绩的记录,包括 30 m 跑、投掷小球、挺举重量、抛实心球、前抛铅球等,建立“运动员成绩.sav”数据文件。如图 2.1 所示。



	ID	sex	group	a	b	c						
1	1	1	1	3.60	4.30	82.30						
2	2	2	1	3.30	4.10	87.48						
3	3	2	1	3.30	4.22	87.74						
4	4	1	1	3.21	4.05	88.60						
5	5	2	1	3.10	4.38	89.98						

图 2.1 运动员成绩数据文件

案例 2-1 中 1 号选手的 30 m 跑成绩为常量 3.6 s,但 20 名选手 30 m 跑的成绩是变化的。变量是相对常量而言的,可以用变量 a 表示 30 m 跑成绩。

SPSS 建立数据文件的第一步,通常不是直接进入数据视图进行录入数据,而是进入变量视图进行变量设置。启动 SPSS 后,打开 SPSS 数据编辑窗口,点击左下角的“变量视图”即可,如图 2.2 所示。

在变量视图中,每一行对应一个变量。首先要定义每个变量的属性,包括名称、标签、类型及度量标准等,如图 2.3 所示。

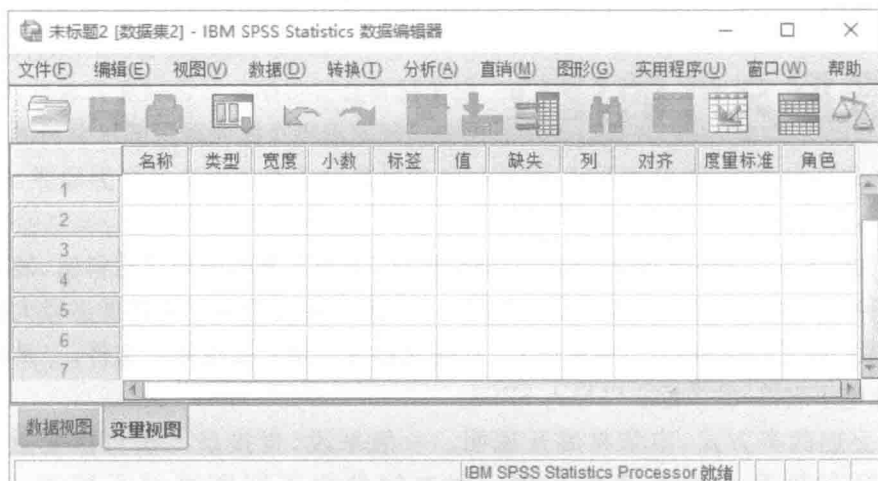


图 2.2 变量视图窗口



图 2.3 设置变量的属性

2.1.1 定义变量名称

1) 变量的名称

SPSS 默认的变量名称为“变量”，可根据自己的需要来命名变量，通常用英文单词或字母表示。案例 2-1 中主要变量名称如图 2.3 所示，分别为：

ID——编号 sex——性别 group——组别 a——30 m 跑 b——投掷小球
c——挺举重量 d——抛实心球 e——前抛铅球 f——五级跳

2) 变量的命名原则

在 SPSS 中，定义变量名称时应遵循以下原则：

(1) 变量名不能相同，即唯一。