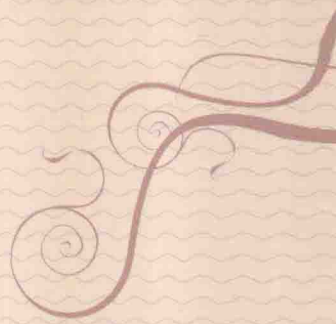




心理与教育 研究中的数据分析： SPSS应用教程

主 编 张姝玥
副主编 唐文清



科学出版社

心理与教育研究中的数据分析： SPSS 应用教程

主 编 张姝玥

副主编 唐文清

科 学 出 版 社

北 京

内 容 简 介

本书是在编者多年教学实践基础上编写而成。全书共 10 章内容, 分别是: 概述, 数据文件的建立、编辑与整理, 描述统计, 假设检验原理与单样本 t 检验, 相关分析, 双样本 t 检验, 单因素和多因素方差分析, 重复测量方差分析与混合设计, 线性回归, 非参数检验。全书以统计方法理论为主线索, 以 SPSS16.0 为软件操作版本, 以案例为示范, 阐述了每种统计分析方法的原理和实际应用过程, 配合大量的截图及注解, 直观呈现 SPSS 软件的操作, 使读者轻松掌握运用 SPSS 软件来解决现实的统计问题。

本书主要适用于心理学和教育学专业的本科生、研究生以及教师, 也适用于社会学、社会工作等专业的学生和教师。

图书在版编目 (CIP) 数据

心理与教育研究中的数据分析: SPSS 应用教程 / 张姝玥主编. —北京: 科学出版社, 2016

ISBN 978-7-03-048643-1

I. ①心… II. ①张… III. ①心理统计-统计分析-软件包-高等学校-教材②教育统计-统计分析-软件包-高等学校-教材
IV. ①B841.2②G40-051

中国版本图书馆 CIP 数据核字 (2016) 第 127587 号

责任编辑: 滕亚帆 王晓丽 / 责任校对: 郭瑞芝

责任印制: 张 伟 / 封面设计: 华路天然工作室

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>

北京中石油彩色印刷有限责任公司印刷

科学出版社发行 各地新华书店经销

2016 年 6 月第 一 版 开本: 787×1092 1/16

2016 年 6 月第一次印刷 印张: 13

字数: 310 000

定价: 39.00 元

(如有印装质量问题, 我社负责调换)

前 言

数据和信息是人类社会发展中极其宝贵的资源,数据本身多是杂乱无章、难以理解的,需要通过科学的方法对数据进行采集、存储、加工、转换和传输来获得有价值、有意义的信息,可以说,数据处理方法和技术的发展极大地影响着人类社会的发展过程。统计学是一门通过搜集、整理和分析数据资料获得数据蕴含的信息,进而对研究对象的本质进行解释和推论的学科,在社会科学和自然科学的各领域中都有广泛应用。心理与教育统计学是一门将统计学的理论方法应用于心理与教育科学的研究和实践的学科,是心理与教育科学体系中的重要分支。无论是心理学与教育学专业的本科生、研究生,还是从事心理学和教育学相关工作的教师、科研工作者,都需要掌握必要的统计分析方法和数据处理技术,以独立地处理数据资料。

然而,由于心理与教育统计学及相关软件的学习都要求学生具备一定的数理基础,对于数理基础水平不一的心理学与教育学专业的学生而言,心理与教育统计学也是一门让许多学生感到头疼的课程。学生们对心理与教育统计学课程的焦虑和恐惧往往会迁移到对 SPSS 统计软件的学习中,从而影响其对数据统计分析方法和技术的有效掌握。在长期的教学和实践过程中,编者发现学生在对数据进行统计分析时经常出现四大问题:一是面对复杂的数据无从下手,不知道该选择什么统计方法来分析数据;二是在实际应用过程中不能根据各类统计方法的适用条件选择合适的方法,导致误用和滥用统计方法,比如自变量为三个水平时用 t 检验进行分析;三是虽然会操作 SPSS 软件,但不能看懂或不能理解软件输出的结果;四是不知道如何报告统计分析结果,或者报告数据分析结果时不符合学术规范。针对这些问题,为了让读者有效地掌握心理和教育研究数据的分析方法和技术,编者精心编写了本书。

全书共 10 章,第 1 章概述,介绍了基本概念、术语和 SPSS 软件;第 2 章数据文件的建立、编辑与整理,介绍了如何在 SPSS 中生成数据文件和对数据进行管理;第 3 章描述统计,介绍了描述统计的主要集中趋势变量和离中趋势变量的基本原理及其在 SPSS 中的实现过程;第 4 章假设检验原理与单样本 t 检验,阐述了假设检验和单样本 t 检验的统计原理及其在 SPSS 中的实现过程;第 5 章相关分析,主要阐述线性相关分析、偏相关分析的统计原理及其在 SPSS 中的操作实现过程;第 6 章双样本 t 检验,阐述了独立样本 t 检验和配对样本 t 检验的统计原理、适用条件及在 SPSS 中的操作实现过程;第 7 章单因素和多因素方差分析,阐述单因素方差分析的统计原理及其在 SPSS 中的实现过程,多因素的主效应、交互效应、简单效应的统计原理和 SPSS 的统计分析过程;第 8 章重复测量方差分析与混合设计,阐述了被试内设计和混合设计方差分析的基本原理,通过单因素重复测量设计、两因素被试内设计、两因素混合设计的实际案例来阐述重复测量方差分析和混合设计方差分析的 SPSS 实现过程;第 9 章线性回归,阐述了一元线性回归和多元线性回归



的统计原理及其在 SPSS 中的实现过程；第 10 章非参数检验，主要介绍了卡方检验的基本原理及在 SPSS 中的统计分析实现，两独立样本、两相关样本的非参数分析在 SPSS 中的实现过程。

本书具有以下三个特点：

1. 以实践应用为导向。本书的重点在于介绍心理与教育研究数据分析的基本步骤及常用的统计方法攻略，强调各统计方法在实践中的应用，对于涉及到的统计学基础理论只做必要的介绍。通过详细的操作步骤，配合截图和注释，让学生直观地看到统计分析的实现过程。即使没有教师指导，学生也可通过自学有效掌握这些数据分析的方法和技术。

2. 以问题为中心，由案例引入。本书在讲解各类统计分析方法时，都以心理和教育研究中的实际问题为中心，以一个案例引入进行讲解。结合实际案例介绍统计方法的基本原理、SPSS 分析操作过程、结果分析和报告呈现。

3. 渐进式地全面讲解。全书内容从基本统计概念和 SPSS 常用命令简介，到数据文件建立、编辑与整理，再到描述统计和推理统计常用方法介绍，循序渐进地阐述 SPSS 软件应用于心理与教育研究数据分析的具体操作。章节内容从统计原理、统计方法适用条件，到数据格式、软件操作，再到结果解释和规范的结果呈现，配合大量的截图及注解，直观呈现 SPSS 软件的操作过程，使读者可轻松掌握。

为了让读者更好掌握各类统计方法，并学以致用，本书在编写过程中，在介绍基本的统计理论和 SPSS 软件功能模块的同时，侧重介绍统计分析方法在解决实际问题中的应用，引导读者不但掌握统计方法的基本原理和 SPSS 软件的操作过程，而且学会运用统计方法解决现实问题。另外，从第 4 章开始，根据现代统计理论的发展及应用的需要，在阐述到各类统计方法时均介绍了效应量的计算方法和报告规范。

本书是广西大学教材立项项目、广西高等学校优秀中青年骨干教师培养工程、广西高等教育教学改革工程的资助成果。本书的出版特别要感谢几位研究生的辛勤付出。陈卓豪参与了第 1 章、第 4 章及第 7 章到第 10 章的编写，朱重璋参与了第 5 章、第 6 章、第 7 章和第 10 章的编写，石荃参与了第 1 章和第 3 章的编写，广西建设职业技术学院的林艳参与了第 2 章的编写。

由于编者的水平和精力有限，书中难免存在疏漏和错误，敬请专家和读者批评指正！

张姝玥 唐文清

2016 年 5 月

目 录

第 1 章 概述	1
1.1 基本概念和术语	1
1.2 SPSS 简介	4
第 2 章 数据文件的建立、编辑与整理	10
2.1 生成数据文件	10
2.2 数据整理	18
练习题	34
第 3 章 描述统计	35
3.1 描述统计的基本原理	35
3.2 Frequencies 的应用与操作	38
3.3 Descriptives 的应用与操作	43
3.4 Means 的应用和操作	45
练习题	47
第 4 章 假设检验原理与单样本 t 检验	49
4.1 假设检验的基本思想	49
4.2 单样本 t 检验	52
练习题	56
第 5 章 相关分析	57
5.1 简单线性相关	58
5.2 偏相关	65
练习题	69
第 6 章 双样本 t 检验	71
6.1 独立样本 t 检验	71
6.2 配对样本 t 检验	76
练习题	80
第 7 章 单因素和多因素方差分析	81
7.1 单因素方差分析	81
7.2 多因素方差分析	90
练习题	104
第 8 章 重复测量方差分析与混合设计	105
8.1 重复测量方差分析的原理	105
8.2 单因素重复测量方差分析	109
8.3 两因素被试内设计的方差分析	116
8.4 混合设计的方差分析	127
练习题	138
第 9 章 线性回归	140
9.1 一元线性回归	140



9.2 多元线性回归	147
练习题	156
第 10 章 非参数检验	157
10.1 卡方拟合度检验	157
10.2 卡方独立性检验	161
10.3 其他非参数检验	166
练习题	177
附录 练习参考答案	178



第 1 章

概 述



1.1 基本概念和术语

在行为科学研究中,通过科学的实验或调查可获得数量化的测量结果,对这些数据进行统计分析和处理,可更准确地了解行为的特征和变化规律,并据此得出结论。因此,在诸如心理学和教育学等行为科学的研究中,数据统计分析方法和技术的学习显得十分重要。由于心理学、教育学等行为科学中都有专门阐述统计学理论的著作或教材,所以本书的重点在于用软件来实现研究中的数据分析和处理,而非统计理论与方法的数理证明。在本章中,1.1 节将对与研究设计及数据分析有关的基本概念和术语进行介绍,以方便后续章节的学习;1.2 节主要对 SPSS 分析软件做简单介绍。

1.1.1 变量

变量是与常量相对应的一个概念。常量是指不会随着外在条件的改变而改变,总是保持一定值的量,如圆周率和光速。变量则是会随着其他因素不同而变化的量,它既包括容易数值化、能直接测量的量,如身高和心率,也包括无法直接测量的量,如魅力和情绪,甚至更抽象的个性和价值观等。因为统计分析是通过数字形式来呈现事物的,所以我们要把变量数值化,也就是量化。变量的可能取值称为水平,以身高为例,某人的身高可能刚好是 1.7 米,1.7 米就是身高的一个水平,身高也可能被取值为 1.8 米或 1.9 米,这就是身高的其他水平。实际测量观察到的水平称为观测值,它是某个个体的实际数据,例如,上述的 1.7 米就是一个观测值。

分类依据不同,变量可被分成不同的类别。根据变量值是否连续,可以把变量分成连续变量和离散变量,连续变量在任何两个相邻的数据点之间都可以细分为无限个数值,如身高,在 1 米和 2 米之间还有无限个数值(1.1, 1.25, 1.366 等);离散变量在两个相邻的数据之间不能再划分为细小的单位,一般为整数,如人数,只能是 1 个人、2 个人,但是没有 1.5 个人。

根据量化程度的不同,我们还可以把变量分成称名变量、顺序变量、等距变量和等比变量四种类型。

称名变量/数据只说明某一事物与其他事物在属性上有所不同或类型上有差异。经常用到的称名变量是性别,我们可以给不同性别的水平赋值,把男性赋值为 1,女性赋值为 0,这里的 1 和 0 只能代表不同性别,却没有数学意义,不能比较大小,也不能进行加减乘除



等数学运算，仅代表不同水平之间的区别。称名变量也常被称为分类变量，因为一个水平相当于一个分类类别。

当变量的不同水平可按某个顺序排序，这类变量属于**顺序变量/数据**。如某次测验的名次，可按第一名、第二名、第三名……的顺序依次排列。顺序变量只能提供大小、等级和程度的顺序，不能提供量的差异，如我们不知道第一名跟第二名之间的距离和第二名跟第三名之间的距离的差距。当无法测量某个变量的准确数值但可以对它进行排序时，我们就可以把它归进顺序变量。例如，当没有量尺测量人们的身高，就可以通过比较身高得到身高顺序，并按身高顺序赋值 1、2、3……到最后一个。不过这些数字仍然不能进行加减乘除的运算。

量化程度更高的是**等距变量/数据**。等距变量水平之间的距离是相等的，若获得等距变量的数据，可对变量进行加减运算。如温度， 24°C 到 25°C 与 25°C 到 26°C 的距离是一样的，都是 1°C 。但是等距变量不能做乘除的运算，正如可以说 10°C 比 5°C 多 5°C ，但是不能说 10°C 比 5°C 多一倍。

量化程度最高的是**等比变量/数据**。像上述的身高，如果能借助量尺准确测量身高，身高就不是顺序变量，而是等比变量。等比变量与等距变量的区别在于是否存在绝对零点，等距变量没有数学上的绝对零点，所以无法进行乘除的运算。所谓的绝对零点是指数值“0”能代表实际意义上的“无”。如摄氏温度是等距变量没有绝对零点，因为它的 0 不是代表没有温度；身高可以成为等比变量，因为它的 0 确实代表没有高度（如果父亲身高两米，儿子长到一米高的时候，我们就可以说父亲的身高是儿子身高的两倍）。

心理学等行为科学中有一些常用的变量，这些变量常常无法直接去测量，如个体对某件事情的评价或态度。对于这类变量，我们可以通过一个李克特（Likert）五点等级量表来间接测量，给不同程度的态度赋上 1 到 5 的值，具体如下：

1	——	2	——	3	——	4	——	5
十分不同意		不同意		既不同意也不反对		同意		十分同意

我们假设等级之间的距离是一样的，也就是说从等级 1 到等级 2 的程度差异与从等级 2 到等级 3 的程度差异相同，这样可以最大限度地把一些抽象的变量近似地作为等距变量来使用。除了 5 点量表，还可以根据需要使用 4 点量表、7 点量表或者只有是和否的两点量表。

变量的类型很大程度上决定了数据分析过程中使用何种统计检验方法。一般的等距、等比数据可以用光滑的数学分布来表示（分布的概念将在后面的章节介绍），并通过几个参数来描述数据分布的特征，基于数学分布的统计检验方法叫做**参数统计**。对于称名变量、顺序变量或者满足特殊分布的等距、等比变量，则只能使用**非参数统计**的方法（第 10 章）。

1.1.2 自变量、因变量与第三变量

心理学和教育学的研究常常探讨几个变量之间的关系，例如，研究者想知道一个变量的变化是否会引起另外一个变量的变化。把引起其他变量变化的变量称为**自变量**

(independent variable), 自变量可以由研究者操纵; 另外一个称为因变量 (dependent variable), 是受到自变量影响的一方 (虽然在研究前并不知道这种影响是否真的存在)。自变量和因变量的说法一般出现在**实验研究**中, 因为不通过实验研究我们很难知道变量之间到底是谁在影响谁。如果只是研究两个变量之间是否有共变关系, 这种研究属于**相关研究** (在第5章会有更多的介绍), 不存在谁是自变量谁是因变量的问题。有时候即使不进行实验研究, 研究者也能知道哪个是自变量, 因为有些自变量非常稳定, 很少发生变化。像性别差异的研究, 我们几乎不会让自己的性别发生变化, 因此性别大多数时候只能是自变量。但是如果研究影响婴儿性别的因素, 这个时候性别就是因变量。

在研究中, 还有一类变量, 它不是研究的焦点, 但是又可能对研究过程和结果产生干扰。试想一下, 在进行一项室外宽敞环境是否比室内狭窄环境更容易产生愉悦情绪的实验, 结果发现室外活动时被试的情绪更高, 这似乎跟预期一致。但室外环境的温度比室内环境的温度要高, 也就是说被试的情绪变化可能是受到了温度而不是场所的影响, 此项研究的结果被温度这个变量混淆了。我们把研究中可能会混淆实验结果的变量称为**第三变量**、**额外变量**或者**混淆变量**。这些变量是需要实验中控制的, 所以也称**控制变量**。一个实验中可能会存在很多的第三变量需要控制, 有一些第三变量容易被发现, 有一些则很难被发现。在研究的过程中, 要尽可能让自变量以外的其他因素条件保持一致, 否则实验结果难免会受到质疑。

控制第三变量的方法通常是保持其恒定, 如上述例子中尽量让室内和室外温度保持一致。也有一些无法随意改变的第三变量, 如被试的性别、年龄、受教育程度等, 对此常用的方法是**随机化**, 即在被试分组时进行随机分配, 随机分配的结果是这些变量的不同水平也被随机分到不同的小组, 这样就能大大减少第三变量对研究结果的影响。

1.1.3 总体与样本

被研究的对象的集合称为**总体**, 组成总体的基本单位是**个体**。如果我们研究的是人的智力, 那全部人的智力就可以构成总体, 某个人的智力就是其中的一个个体。假如研究者真的想要研究人的智力, 由于无法对所有人的智力情况进行调查, 可把总体缩小到某个地区, 把某个地区的全部个体作为一个更小的总体。但是要收集某个地区全部人的数据还是十分困难, 这时, 只能从总体中抽取部分的观察对象, 形成一个**样本**, 用样本的信息进行统计推断, 分析总体的情况。

从总体抽取样本的过程称为**抽样**, 为了使样本能够代表总体, 科学的做法是使用**随机抽样**。随机抽样与随便抽样是不同的, 严格的随机抽样应该是保证总体里每个个体被抽到的概率都是一样的。随机抽样得到的样本是**随机样本**, 使用随机样本进行的统计检验可得出关于总体的有效结论。但是由于现实中的各种限制, 研究中常使用方便样本, 以致研究结果难以推广到更广泛的人群。

1.1.4 常用的参数和统计量及其字母表示

无论是总体还是样本, 都是数字的集合, 总体的数据特征值称为**参数** (parameter), 样本的数据特征值称为**统计量** (statistics) (研究中常通过样本统计量来推断总体参数)。数



据特征值可以有很多，有一些是比较常用的，表 1-1 列出了常用的参数和统计量及其字母表示。

参数和统计量之间最明显的区别是参数常用希腊字母表示，而样本统计量多用英文字母表示。如表示总体分散情况的统计指标标准差，用小写希腊字母 σ 表示，对应的样本标准差符号是 S 。在统计分析中，我们应当注意统计量是在描述样本还是描述总体，从而选择不同的符号。

表 1-1 常用参数和统计量及其字母表示

	总体参数表示	样本统计量表示
数据个数	N	n
均值	μ	$\bar{X}$
标准差	σ	S
方差	σ^2	S^2
相关系数	ρ	r

总体参数与样本统计量之间也存在一定的关系。从数值计算上看，当总体大小已知并与实验观察的总次数相同时，总体参数占样本统计量是同一统计指标。当总体无限时，统计量与总体参数不同，但统计量可在某种程度上作为总体参数的估计值。

1.2 SPSS 简介

1.2.1 SPSS 的启动与退出

在打开 SPSS 之前，必须先查看它是否安装在你的计算机里。从桌面上或者在程序菜单栏中找到相应 SPSS 图标，只需选取其中一种方法打开 SPSS 程序即可。

(1) 用屏幕上的图标打开 SPSS：找到桌面上的 SPSS 图标, 双击 SPSS 图标，打开 SPSS。

(2) 用程序菜单打开 SPSS：单击“开始”菜单，单击“所有程序”；选择“SPSS Inc”或者“SPSS for Windows”，单击电脑里安装的 SPSS。本书使用的是 SPSS16.0，详见图 1-1。

无论哪种方式打开 SPSS 程序，进入 SPSS 的初始界面后出现图 1-2 的提示框：“What would you like to do?” 该提示框有 6 个功能项和一个复选项，如果不需要用到提示框中的功能，可以单击 Cancel 按钮。

(1) Run the tutorial，是运行自学指导。单击此项并单击 OK 按钮，将打开 SPSS 自学指导，从中可以找到 SPSS 各类操作相关信息。

(2) Type in data，输入数据。单击此项并单击 OK 按钮，将出现数据编辑窗口，可以进行数据录入（数据编辑窗口和数据录入在第 2 章有详细讲解）。

(3) Run an existing query，打开一个已经存在的“.spq”文件。

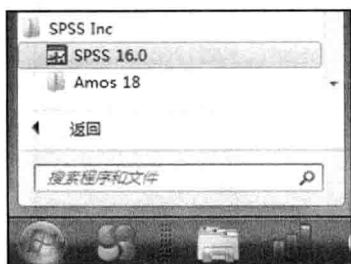


图 1-1 用程序菜单打开 SPSS16.0

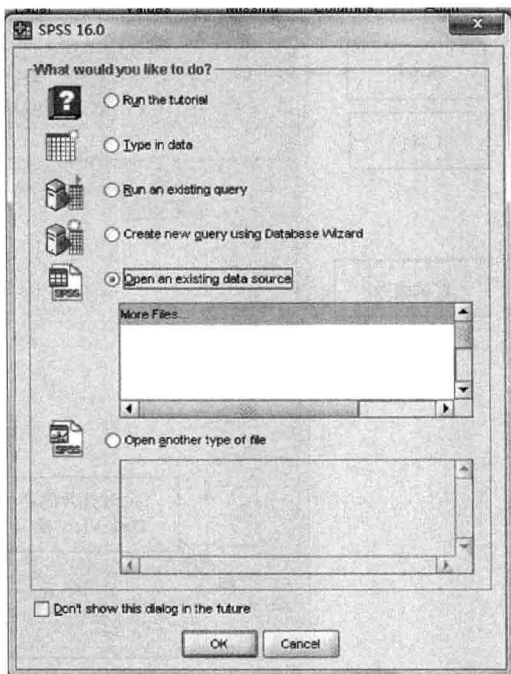


图 1-2 “What would you like to do?” 对话框

(4) Create new query using Database Wizard, 用数据转换建立新的文件。此项可以把 XLS 格式的 Excel 文件转换为 SPSS 数据文件。

(5) Open an existing data source, 打开一个已经存在的数据文件(文件后缀名为“.sav”)。

(6) Open another type of file, 打开其他类型文件, 例如, 结果输出文件(后缀名为“.spo”)、语句文件(后缀名为“.sps”)等。

(7) Don't show this dialog in the future, 如果选择此复选框, 下次启动 SPSS 时, 将不再出现 “What would you like to do?” 的提示框。

退出 SPSS 非常简单, 最方便的方法是单击窗口右上角的 “X” 按钮。

1.2.2 Data View 窗口

打开 SPSS 软件, 将看到 Data Editor (数据编辑) 窗口, 这是 SPSS 中用于生成变量与录入数据的窗口如图 1-3 所示。界面左上角显示 SPSS 的文件名 (Untitled 表示这个文件尚未命名), 接下来是菜单栏、工具栏、数据单元格、被试编号等。数据编辑窗口中包含两个不同部分, 分别为 Data View (数据查看) 窗口和 Variable View (变量查看) 窗口, 通过单击界面左下角的 Data View 和 Variable View 按钮进行相互切换, 按钮显示为浅黄色背景时表示打开了此窗口 (图 1-3, 此时打开的是 Data View 窗口)。

菜单栏: 在菜单栏中有 SPSS 提供的可操作的大多数命令, 根据不同的操作需要单击相应按钮, 在跳出的系列选项中选择符合需要的命令即可。在这里介绍几个常用的命令。

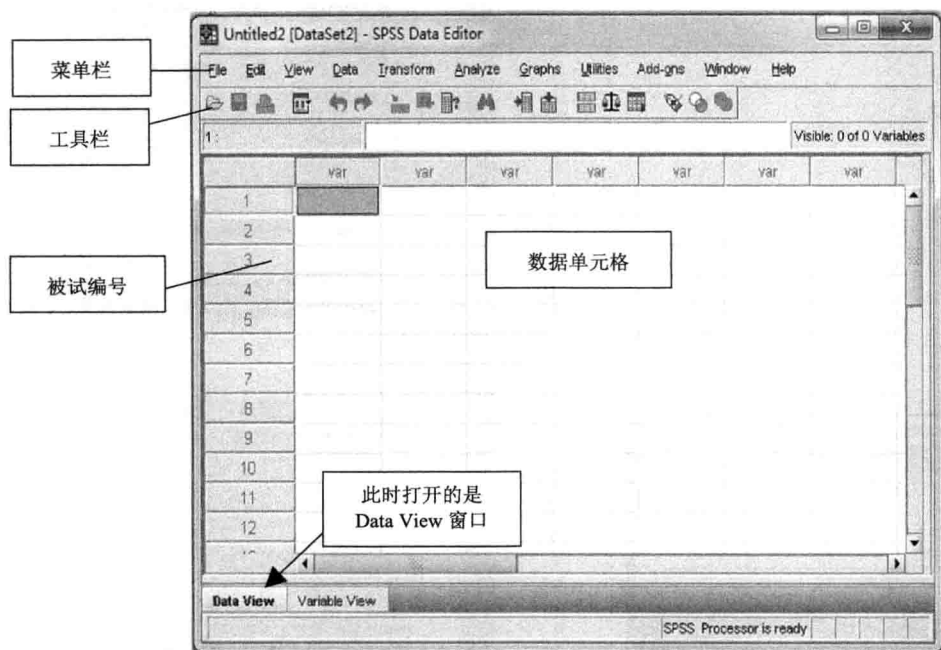


图 1-3 SPSS 的数据编辑窗口

◆ **File:**文件，处理和文件相关的各种功能，包括打开、读取、保存、退出 SPSS 等。

◆ **Edit:**若干编辑功能，包括撤销、复制、粘贴、查找、替换变量等。

◆ **View:**窗口外观控制，包括状态栏、工具栏的显示或隐藏，字体设置，表格线的显示或隐藏，变量值/变量标签的显示切换。

◆ **Data:**数据定义、设置、输入的相关操作，以及样本排列、文本合并、样本选择和加权。

◆ **Transform:**数据转换，包括重新编码、通过计算得出新变量、重新排序，以及缺失值处理等。

◆ **Analyze:**所有的数据分析都需要从单击 Analyze 开始。

◆ **Graphs:**创建图表可以先从单击 Graphs 命令开始，更经常的情况是在执行其他同级命令时单击图表选项。

工具栏：工具栏中有一些常用的工具图标按钮，如打开、保存、撤销、加权等，直接使用工具栏图标可简化某些鼠标操作，方便快捷。值得注意的是，工具栏中有些图标是亮的，有些是灰色的，灰色的图标表示目前不能使用，当录入数据或进行某些操作后，一些图标会变亮，这时就可以使用了。

	打开文件		查找任意变量中的特定数据
	保存文件		插入新被试
	打印文件		插入新变量
	提取最近使用过的命令		把文件分组
	撤销及恢复撤销		加权
	根据编号来查找样本		选择样本
	查找变量		在多水平变量的值和标签之间转换
	显示变量的相关信息		用户定义的变量集

Data View 窗口是 SPSS 中用于输入数据（数据通常指数字、字母、文字或符号）的窗口，如图表 1-4 所示即为完整的 Data View 窗口界面。其最左侧的数字符号表示被试编号，每一行都代表一个被试，而每一列代表一个变量，初始状态下，所有的列名称都显示为 var。此时第一个单元格显示为深色，意味着它是激活的或者准备接受输入。

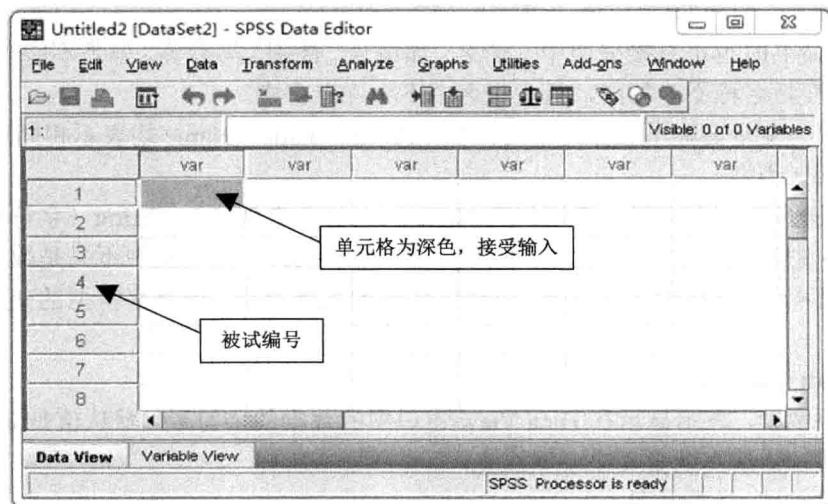


图 1-4 SPSS 的 Data View 窗口

1.2.3 Variable View 窗口

Variable View 窗口是 SPSS 中用于定义变量（生成变量和设置数据文件信息）的窗口，我们需要在这个窗口中为每个变量命名，设置变量便签或添加其他信息。单击界面左下角



的 Variable View 按钮即可切换到 Variable View 窗口。如图 1-5 所示，窗口左侧显示的数字为变量编号，每一行表示一个变量，而每一列有不同的名字和功能，用于为每个变量设置信息。

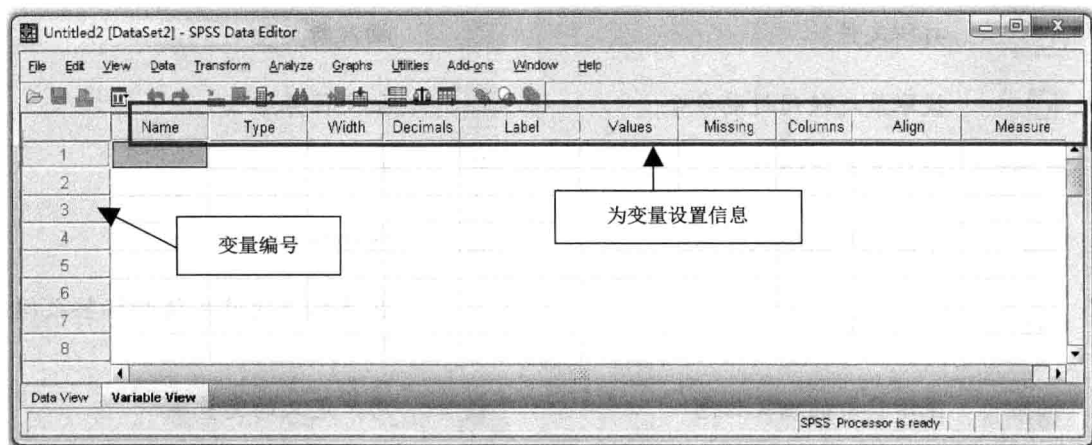


图 1-5 SPSS 的 Variable View 窗口

Variable View 窗口中用于定义变量的各个按钮的名称及功能简要介绍如下：

(1) Name (名称)。

① 用于给变量命名，一般而言，每个变量名需以字母开头，之后可以使用数字、字母、符号等，如 name、gender、age……或 a1、A2、a#……。

② 变量命名时也可直接采用中文命名，如编号、性别、年龄等，但是不能以数字开头。

③ 两个变量名称不能相同，且变量名中不允许有空格。

④ SPSS 中变量名称不区分大小写，如 NAME、name、Name 均表示相同意思。

(2) Type (类型)。

① 表示变量的数据类型，常用的有 Numeric (标准数值型)、String (字符型)。

② SPSS 中默认的数据类型为标准数值型 (Numeric)，当变量中不只是数字，而含有文字、字母或其他时，这时需在 Type 中设置相应的数据类型，但是字符型的变量不能参与统计分析。

(3) Width (宽度)。

变量的总宽度，表示显示在 Data View 窗口中的数据的字符数，默认值是 8，也就是可以显示 8 个阿拉伯数字或者 4 个中文汉字。

(4) Decimals (小数位数)。

表示显示在 Data View 窗口中的小数点位数，数值型变量默认为 2 位即小数点后显示两位数字，字符型变量默认为 0 位。

(5) Label (标签)。

① 用于给变量设置标签，当一个变量的含义不那么明了时，可通过设置标签帮助理解变量代表的意义、特征等。

② 标签的内容会显示在输出中，可帮助我们理解结果。

③ 标签的单元格为文本形式，最大字符长度为 256。

(6) Values (值)。

① 用于对分类变量进行编码 (如对于性别变量, 1=男, 2=女), 对变量的不同水平设置值标签 (Values Labels), 便于理解输出结果。

② 值标签 (Values Labels) 的最大字符长度为 60。

(7) Missing (缺失)。

① 用于定义缺失值, 表示所设的值是缺失的 (如可用 99、999 等表示缺失值)。

② 系统默认的缺失值为没有任何数据录入的单元格。

(8) Columns (列)。

用于设置 Data View 窗口的列宽, 默认值为 8 个字符。通常不需要自行设置。

(9) Align (对齐)。

用于设置 Data View 窗口中数据的对齐方式, 数值型变量默认为右对齐 (Right), 字符型变量默认为左对齐 (Left)。通常不需要自行设置。

(10) Measure (度量)。

用于描述变量的度量水平, 共有三个选项, 分别为 Nominal (称名变量, 性别等变量可以选择此项)、Ordinal (顺序型变量)、Scale (数量型变量), 系统默认类型为 Scale。通常不需要自行设置。

第 2 章我们将以具体数据为例, 讲解如何在 SPSS 中进行数据录入。

第 2 章

数据文件的建立、编辑与整理

2.1 生成数据文件

2.1.1 定义变量

生成 SPSS 数据文件前，需在 Variable View 窗口中对变量加以定义，包括定义名称、数据类型、宽度、小数点位数等。现有一项关于大学生生活满意度的调查，测量量表共包含 5 个项目，采用七级评分方式，从 1~7 表示“非常不符合~非常符合”（表 2-1），现获得 10 位大学生的测量数据（表 2-2）。

表 2-1 生活满意度调查问卷

非常不符合	比较不符合	有些不符合	不确定	有些符合	比较符合	非常符合				
1	2	3	4	5	6	7				
1. 在大多数方面，我的现实生活很接近于理想生活				1	2	3	4	5	6	7
2. 我生活得很好				1	2	3	4	5	6	7
3. 我对我的生活感到不满意				1	2	3	4	5	6	7
4. 迄今为止，我已经得到了生活中我想要的重要的东西				1	2	3	4	5	6	7
5. 如果可以再活一次，我也不想做什么改变				1	2	3	4	5	6	7

注：每一项目只能选择一个选项

表 2-2 样本数据集

编号	性别	年龄	生源地	第一题	第二题	第三题	第四题	第五题
1	男	19	长沙	3	3	5	2	1
2	女	18	郑州	4	4	5	2	3
3	男	18	福建	2	5	2	7	7
4	男	20	广西	4	5	2	2	4
5	女	19	常德	4	6	1	2	4
6	男	19	山东	5	5	5	5	1
7	女	17	重庆	6	7	1	5	5
8	女	19	山东	1	4	2	2	5
9	女	18	山东	3	5	5	2	1
10	男	20	武汉	4	5	2	3	3