



语言研究应用SPSS

软件实例大全

杨端和 著

中国社会科学出版社

语言研究应用SPSS

云南师范大学

学术著作出版基金资助项目

软件实例大全

杨端和 著

中国社会科学出版社

M310/02

图书在版编目 (CIP) 数据

语言研究应用 SPSS 软件实例大全/杨端和著. —北京:
中国社会科学出版社, 2004.4

ISBN 7-5004-4300-5

I. 语… II. 杨… III. 语言统计-统计分析-软件包, SPSS IV. HO-05

中国版本图书馆 CIP 数据核字 (2003) 第 124685 号

责任编辑 程 东

责任校对 林 林

封面设计 新空气

版式设计 王炳图

出版发行 中国社会科学出版社

社 址 北京鼓楼西大街甲 158 号

邮 编 100720

电 话 010-84029453

传 真 010-84017153

网 址 <http://www.casspw.cn>

经 销 新华书店

印 刷 北京奥隆印刷厂

装 订 三河鑫鑫装订厂

版 次 2004 年 4 月第 1 版

印 次 2004 年 4 月第 1 次印刷

开 本 710×980 毫米 1/16

印 张 18.25

插 页 2

字 数 319 千字

定 价 32.00 元

凡购买中国社会科学出版社图书, 如有质量问题请与本社发行部联系调换
版权所有 侵权必究

前 言

政治经济学的一个基本观点就是生产工具的使用水平反映出社会生产力发展的水平。从这一观点出发,我们在语言研究中使用什么样的研究工具、研究手段,就体现了语言研究的发展水平。因此,我们新世纪的语言研究要走向现代化,首先就应该实现研究工具、研究手段的现代化,走应用计算机软件全方位处理语言数据之路。

SPSS (Statistical Package for Social Sciences) 这一社会科学统计软件就是一个功能强大,不断升级完善的数据处理软件,有“世界优秀统计工具”的盛名。不但在社会科学领域,而且在自然科学领域也得到越来越广泛的应用。但是这一优秀的统计软件在我们语言研究中的应用水平并不是很高。原因之一就是还没有一本专门为广大语言文字、中外文学、外语及民族语言研究者撰写,以语言研究为特定目的,全面论述这一软件各板块、各过程综合应用的参考手册。

本书用讲解众多实例的方式全面阐述这一软件最新 11.0 版本在语言研究各个领域、各个层面全方位的应用,并将统计学基本原理的描述与软件运行步骤的说明有机地结合起来,使广大语言工作者都能看懂系统所输出的统计图表的真实意义,使他们能够真正理解,弄通这一优秀的软件,并将其得心应手地应用于自己的研究实践中。

本书已为这一软件的英语统计学术语、各板块与分板块名称、说明与命令语句添加了汉语翻译,使不懂英语的研究人员也能够学习、应用并熟练掌握这一软件。应该说明,部分较新的术语及公式,还未见标准的汉语译名,作者的翻译仅供参考。

本书对于提高我们的语言研究水平、普及计算机在语言研究信息处理方面的应用具有积极意义。

本书作者谨向著名语言学家、外语教学专家、我国当代语言学的开拓者,广东外语外贸大学博士研究生导师桂诗春教授、何自然教授表示崇高的敬意。正是他们当年对我们早期硕士研究生的谆谆教诲,才使得作者从此与

现代语言学理论、统计学原理和计算机应用结下不解之缘。

同时，作者谨向中国社会科学出版社的领导和各位编辑表示诚挚的感谢，正是他们的亲切关怀、大力支持与辛勤劳动，才使得本书得以顺利出版。

本书作者系一普通外语教师，并非计算机专业人士，更非统计学专家。作者执著的追求，就是要帮助我们的语言工作者拿起 SPSS 这一强有力的研究工具，而绝非是要硬充内行，在统计学、计算机专家面前“班门弄斧”。因此，书中的错误、不妥、疏漏之处，敬请读者不吝赐教。最后，作者衷心希望读者能够与作者探讨书中提出的有关问题，大家共同努力，不负时代重托，把我们新世纪的语言研究不断推向前进。

作者的通讯地址：昆明云南师范大学外语学院（邮编：650092）。
E-mail: YDH53@Yahoo.com.cn.

作者

2003 年 6 月于昆明

目 录

第一章 用 SPSS 绘制语言研究常用统计图	(1)
1.1 条形图 (Bar)	(1)
1.2 线图 (Line)	(7)
1.3 面积图 (Area)	(8)
1.4 饼图 (Pie)	(9)
1.5 高一低图 (High - Low)	(11)
1.6 帕累托图 (Pareto)	(13)
1.7 控制图 (Control)	(15)
1.8 箱图 (Boxplot)	(17)
1.9 误差条图 (Error Bar)	(20)
1.10 散点图 (Scatter)	(22)
1.11 直方图 (Histogram)	(26)
1.12 P - P 概率图 (P - P Probability plots)	(28)
1.13 Q - Q 概率图 (Q - Q Probability plots)	(31)
1.14 序列图 (Sequence)	(31)
1.15 洛克曲线图 (Roc Curve)	(34)
1.16 时间系列图 (Time Series)	(38)
第二章 语言数据常用统计报表的编制	(42)
2.1 在线分析处理 (OLAP Cubes) 报表的应用	(42)
2.2 个案汇总报表 (Case Summaries) 的应用	(45)
2.3 行汇总报表 (Report Summaries in Rows) 的应用	(47)
2.4 列汇总报表 (Report Summaries in Columns) 的应用	(51)
第三章 描述统计学 (Descriptive Statistics) 板块的应用	(56)
3.1 频数分布 (Frequencies) 分板块的应用	(56)
3.2 描述统计 (Descriptives) 分板块的应用	(61)
3.3 探索性分析 (Explore) 分板块的应用	(62)

3.4	列联表 (Crosstabs) 分板块与卡方检验	(74)
3.5	比率 (Ratio) 分板块的应用	(92)
第四章	平均数间的比较 (Compare Means) 板块应用	(99)
4.1	平均数 (Means) 分板块的应用	(99)
4.2	单样本 T 测试 (One - Sample T Test) 分板块的应用	(105)
4.3	独立样本 T 测试 (Independent - Samples T Test) 分板块的应用	(108)
4.4	配对样本 T 测试 (Paired - Samples T Test) 分板块 的应用	(111)
4.5	单因素方差分析 (One - Way ANOVA) 分板块及其应用	(113)
第五章	一般线性模型 (General Linear Model) 板块的应用	(125)
5.1	单应变量 (Univariate) 分板块的应用	(125)
5.2	多应变量 (Multivariate) 分板块的应用	(159)
5.3	重复测量 (Repeated Measures) 分板块的应用	(173)
5.4	方差成分分析 (Variance Components) 分板块的应用	(182)
第六章	相关分析 (Correlate) 板块的应用	(188)
6.1	相关分析的基本概念	(188)
6.2	二变量相关分析 (Bivariate correlations) 的应用	(188)
6.3	偏相关分析 (Partial Correlations) 的应用	(193)
6.4	距离分析 (Distances) 分板块的应用	(196)
第七章	回归分析 (Regression) 板块的应用	(204)
7.1	线性回归 (Linear) 分板块的应用	(204)
7.1.1	一元线性回归模型	(204)
7.1.2	多元线性回归模型	(217)
7.2	曲线回归 (Curve Estimation) 分板块的应用	(227)
第八章	聚类分析与判别分析 (Classify) 板块及其应用	(234)
8.1	K - 均值聚类法 (K - means Cluster) 分板块的应用	(234)
8.2	分层聚类法 (Hierarchical Cluster) 分板块的应用	(242)
8.3	判别分析 (Discriminant) 分板块的应用	(249)
第九章	因子分析 (Factor Analysis) 板块的应用	(269)
第十章	信度分析 (Reliability Analysis)	(278)
	主要参考书目	(286)

第一章

用 SPSS 绘制语言研究常用统计图

统计图简单、直观,通过点的位置,线段的长短与条块的升降及圆面积的大小等手段来表达数据与变量间的关系,再辅以简洁的文字说明,就能更直观地反映统计结果所蕴涵的内在意义,诸如变化趋势,分布状态等,从而大大提高统计分析结论报告的明晰性与可读性,达到意想不到的效果,因而一直是描述统计的重要方法之一。而且,近年来,图形化分析方法在统计推断中也正在发挥更为重要的作用,有时候统计图的作用甚至超过某些统计量的作用。

SPSS 以豪华的操作界面为基础,可以制作出非常精美的统计图。作为专业统计软件,它提供了非常全面的统计图种类,其功能比其他软件都更强大而且产生的图形更容易修改。我们语言研究工作者首先应该学习掌握并充分利用 SPSS 的强大绘图功能,使我们的语言研究统计分析,实验报告,语言教学评估报告,文学文体分析结论等研究成果图文并茂,更具说服力。同时,我们从学习各类统计图入手,来熟悉 SPSS 的基本操作界面,掌握数据输入方法及主要命令语句,力求“理论与实践齐头并进”。这也许不失为一种帮助我们广大语言研究工作者迅速弄通,熟练应用这一优秀软件的简易方法。

SPSS11.0 版可以绘制以下 16 种统计图形:(1) Bar (条形图), (2) Line (线图), (3) Area (面积图), (4) Pie (饼图), (5) High-Low (高低图), (6) Pareto (帕累托图), (7) Control (控制图), (8) Boxplot (箱图), (9) Error Bar (误差条图), (10) Scatter (散点图), (11) Histogram (直方图), (12) P-P (P-P 概率图), (13) Q-Q (Q-Q 概率图), (14) Sequence (序列图), (15) 洛克曲线图 (Roc Curve), (16) Time Series (时间系列图)。

下面我们就结合语言研究实例——学习掌握这些图形的生成与制作。

1.1 条形图 (Bar)

在制作图形前,首先要输入数据,编制一个数据文件。现在准备将下面

英语专业某年级四个小班的听力测验成绩输入 SPSS。

例：四个小班英语听力测验成绩：

Class A:	39	42	47	31	33	40	44	37
	49	30	43	35	34	38	45	38
Class B:	45	38	32	48	39	32	36	45
	31	48	34	44	36	32	49	37
Class C:	36	28	46	30	31	42	43	36
	47	30	42	33	32	39	43	35
Class D:	43	36	27	47	36	32	34	43
	30	46	33	41	33	33	47	33

首先还应简单介绍一下 SPSS 的安装及数据录入方法。打开计算机，进入 Win98 或 Win2000，将光盘放入光驱后激活资源管理器，用鼠标双击光盘驱动器图标，然后在目录窗口中寻找 Setup（安装）应用程序，用鼠标双击它就启动了安装程序。或者在桌面上单击“我的电脑”，再双击光盘驱动器，也可进入安装界面。如同安装其他 Windows 软件一样，我们可以按照以下提示一步步完成：

(1) 首先阅读版权声明，看完后单击 Next(下一步)按钮，进入下一步。

(2) 选定安装位置。如不做改动，则安装程序会自动在 C 盘下的 Program files 下建立目录，然后将程序安装在此目录中。也可单击 Browse 按钮来选择安装位置。

(3) 填写用户名，单位名称，序列号。

(4) 选择安装的类型。共有三种安装类型：typical，compact 和 custom 三种。顾名思义，typical 是面向大多数用户的。Compact 则是内存需求最少的安装，如内存空间不够大时，应选用此类型。Custom 主要面向高级用户。我们进行语言研究统计，选择 typical 这一类型即可。现在一般计算机的内存都足够支持 SPSS11.0 版本。

(5) 选择所需的组件。可以根据需要选择自己所需的组件，还可以选择软件是安装在服务器上还是个人计算机上面。

(6) 用户要输入软件的正确序列号。如不能正确输入，安装过程就会终止。序列号一般都印在盘盒上，如果找不到，可以向 SPSS 中国公司 (WWW.SPSSBJ.COM.CN) 咨询。

(7) 单击 Next (下一步) 按钮就进入了安装过程。

(8) 安装完成。

软件安装好以后即可使用。打开该软件后，展示出的界面如下面的图 1.01 所示。

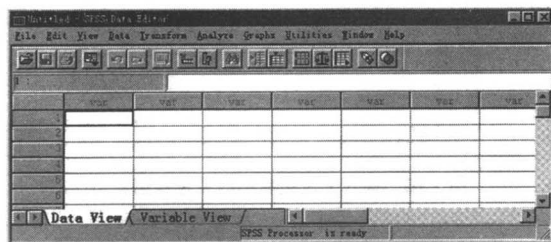


图 1.01 SPSS 软件的数据编辑窗口

窗口顶部显示出 SPSS Data Editor 字样，表示现在显示的是 SPSS 的数据编辑窗口。实际上这是典型的 Windows 的界面，有菜单栏与工具栏。工具栏下方是数据栏，数据栏下方则是数据编辑窗口主界面。由若干行和列组成。每行对应一条记录，每列则对应一个变量。现在没有输入任何数据，因而行、列都呈灰色。第一行第一列的单元格边框为深色，表明该数据单元格为当前单元格。数据录入方式应遵循下面三点：

(1) 不同观察对象的数据不能在同一条记录中出现，同一观察对象的数据应当单独占一行。

(2) 每一个测量指标/因素只能占据一列的位置，同一个指标的测量数值都应当录入到同一个变量中去。

(3) 最终的数据集应当能够包含原始数据的所有信息。

下面我们来定义变量。图 1.01 窗口左下角有两个标签，最左侧的 Data View 现在呈白色，表明我们当前的界面是数据视图窗口。单击紧挨在它右侧的 Variable View 标签，程序切换到变量视图窗口。这个视图的每一行代表了对一个变量的定义，每一列则代表定义该变量时用到的某种属性，如现在看到的名称 Name，变量类型 Type 等。现在就来输入例子中的数据。在 Name 这一列下分别输入 a、b、c、d 这四个字母，分别代表实例中的四个英语小班，请看下面的图 1.02。

现在我们切换回数据视窗，如图 1.03 所示，第一、二、三、四列的名称现在均由深色显示，就是刚才定义的班级名称 a、b、c、d，表明这四列已被定义为变量，其余各列为灰色的 Var（变量），表示这些是尚未使用的变量列。目前我们还未输入数据，这一数据集内没有记录。

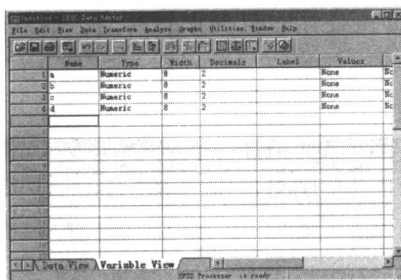


图 1.02 输入了班级名称的变量视图窗口

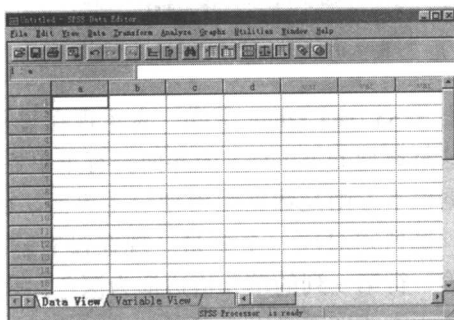


图 1.03 班级名称变量定义完毕的数据视图窗口

开始输入数据。用键盘键入 a 列的第一个数据 39，按→（回车）键，39 变为 39.00。然后输入 b 列第一个数据 45，输入 c 列第一个数据 36 及 D 列的 43。用同样方式输入各列各行的相应数据。最后完成的数据集为四列 16 行。数据集上半部分如图 1.04 所示。

	a	b	c	d
1	39.00	45.00	36.00	43.00
2	42.00	58.00	28.00	56.00
3	47.00	32.00	44.00	27.00
4	31.00	48.00	30.00	47.00
5	33.00	50.00	31.00	36.00
6	48.00	22.00	45.00	32.00
7	44.00	34.00	43.00	34.00
8	37.00	46.00	36.00	43.00
9	49.00	31.00	47.00	30.00
10	38.00	40.00	30.00	48.00
11	43.00	50.00	47.00	33.00
12	30.00	44.00	33.00	41.00
13	34.00	36.00	32.00	33.00
14	38.00	32.00	38.00	35.00
15	45.00	49.00	43.00	47.00

图 1.04 数据录入完毕的变量视图窗口上半部

数据核对无误后,应将其保存。选择菜单上的 File (文件) → Save (储存) 选项,弹出 Save Data As (储存数据) 对话框。该框下方有一个保存类型列表栏,单击后可看到在 SPSS 中数据可以保存的各种类型,有 DBF, FoxPro, Excel, Access 等。我们仍将这一文件存为 SPSS 的数据格式 (* .sav 文件)。将文件命名为 bar-1, 将其键入文件名框内并回车,这时看到系统回到数据编辑窗口,左上角原先的 Untitled (无标题) 字样已变为现在的 bar-1, 显示当前编辑文件的名称是 bar-1.sav。

好了。一切准备工作都已完成,我们可以绘制条形图了。点击窗口上部专门用于统计绘图的 Graph (统计图) 栏,即弹出 Graph 菜单,显示出我们前面列举的那 15 种图形名称。点击 Bar (条形图) 栏,系统立即弹出一个预定义对话框。我们可以先绘制一个简单 (Simple) 条形图。点击该框,框缘变黑,表示要制作简单条形图。然后点击该对话框下部的 Summaries of separate variables (独立变量之和) 前面的白色圆圈,该圈中心出现了黑点,如图 1.05 所示。

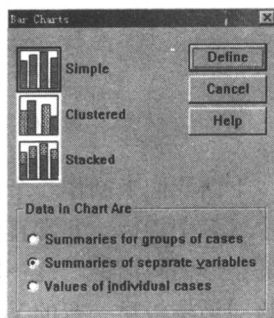


图 1.05 条形图的已定义对话框

现在点击该对话框的 Define (限定) 按钮,弹出一个新的对话框 Define Simple Bar: Summary of Separate Variables (限定简单条形图: 独立变量之和)。该对话框中有左右两个白栏。在左栏中可看到我们的四个班的名称 a、b、c、d 排成一列。分别点击它们再点击两白栏之间的 ◀ 按钮将它们依次移入 Bars Represent (用条形表示) 标题下的右栏内,如图 1.06 所示。栏内的 MEAN (a) 等字样表示系统要生成显示各班平均分的条形图。

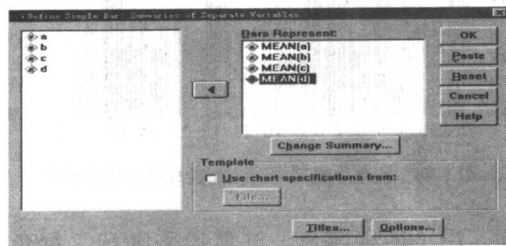


图 1.06 选项完毕的四个小班英语听力测验成绩简单条形图主对话框

现在可以制图了。点击该框右上角的 OK 按钮,一组鲜红的条形图出

现, 见图 1.07。从图中可看出, B 班平均分最高, 而 C 班平均分最低。AB 班与 CD 班的平均分差异明显。

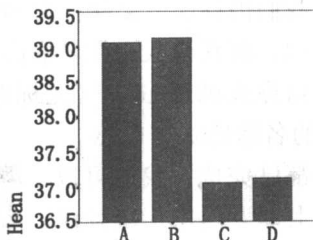


图 1.07 四个小班英语听力测验成绩平均分条形图

下面我们来制作一个 Clustered (分组) 条形图, 这种条形图可同时表现两个或多个指标, 非常直观。选择 Graphs→Bar 选项后, 又弹出如图 1.05 所示的条形图对话框。这一次点击 Clustered (分组) 框, 该框边缘变黑。再点击该对话框最后一行的 Values of individual cases (各个案值) 前的白圈, 白圈中间出现黑点, 表示系统要生成展示所有数据的分组条形图。点击右上角 Define (限定) 键, 弹出 Define Clustered Bar: Values of Individual cases (限定分组条形图: 各个案值) 对话框。同上面制作简单条形图的操作方法一样, 依次将左栏内的 a、b、c、d 选入右边标题下的白栏。点击右上角 OK 按钮后, 绘制出的分组条形图如下面的图 1.08 所示。

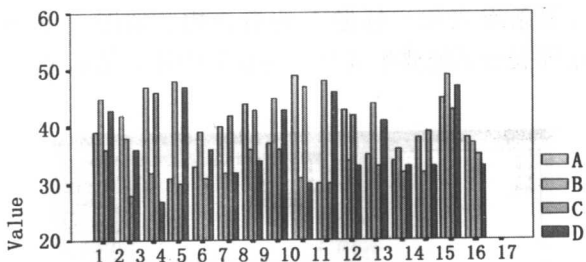


图 1.08 四个小班英语听力测验成绩分组条形图

这个图很漂亮。用了四种颜色。红色代表 A 班, 绿色代表 B 班, 蓝色是 C 班, 而紫色是 D 班。各个班第 1—16 个分数都自成一组。共分为 16 组。每组间还有白色间隙以便于观察。每一个数据都一目了然。

请读者自己用同样的方法制作 Stacked (堆积) 条形图。这种条形图常

用来表达各组成部分在整体中所占的比例，每一段之间没有间隙，也用不同颜色显示。限于篇幅，对于交互式条形图与更加漂亮的三维立体条形图，本书不作介绍了。

1.2 线图 (Line)

线图和条形图真像两个兄弟，各个地方都相像，简单条图对应于单线图，分组条图则类似于多线图。惟一不同的就是前者用直条长短来显示其关系，而后者用的是线条的高低。

选择 Graph (统计图) 后点击 Line (线图) 选项，弹出如图 1.09 所示的线图预定义框。这次我们制作一个 Drop - Line (垂线图) 就可说明问题。点击 Drop - Line 框，框缘变黑，再选择最后一行 Values of individual cases (各个案值)，然后点击右上角的 Define (限定) 按钮，弹出 Define Drop - Line: Values of Individual Cases (限定垂线图：各数据个案值) 主对话框。如制作条形图时一样，将该框左栏内的 a、b、c、d 依次选入右边的 Points Represent (用点表示) 标题下的白栏。点击 OK 按钮，系统即刻绘制出一幅垂线图，如图 1.10 所示。

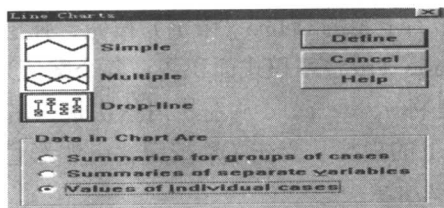


图 1.09 四个小班英语听力测验成绩线图预定义框

与分组条形图一样，红、绿、蓝、紫四种颜色的小方格分别代表 A、B、C、D 班的分数。在垂线图中，各班处于同一位置的分数由一条垂线串起。可以看出，第 3、4、9、10 组分数差异较大，而第 13、15 组差异较小，且各集中于高、低两端。这类垂线图使我们更容易看清各班处于相同位置的分数之间的差异。

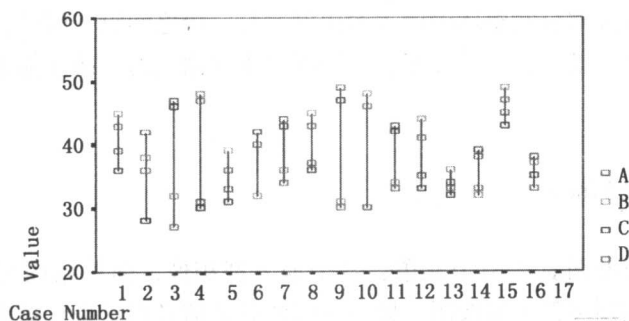


图 1.10 四个小班英语听力测验成绩垂线图

1.3 面积图 (Area)

面积图其实与上面的条形图和线图差别不大,只不过面积图是用面积板块的大小来进行比较罢了。面积图更为简单,只有 Simple (简单)与 Stacked (堆积)式两种。

当然还是先选择 Graphs→Area (面积图)选项,弹出如图 1.11 所示的面积图预定义框。这次我们来绘制一幅堆积式面积图。点击 Stacked 方框,此时框缘变黑。移下鼠标再点击最后一行 Values of individual cases (各个案值)前的白圈。最后点击右上角的 Define (限定)按钮,弹出 Define Stacked Area: Values of Individual Cases (限定堆积式面积图:各个案值)主对话框。其实大家都熟悉了,还是将 a、b、c、d 从左栏内依次选入右边的 Areas Represent (用面积表示)标题下的白栏。点击 OK 按钮,一幅彩色堆积式面积图出现在屏幕上,如图 1.12 所示。

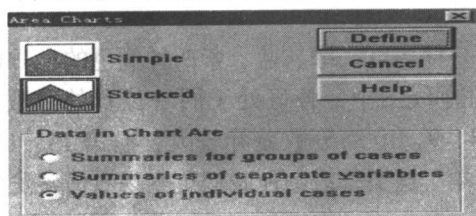


图 1.11 四个小班英语听力测验成绩面积图预定义框

A 班在最下面, B、C、D 班依次叠加在其上。纵轴分值以累加形式出现。可看出第 15 组的 4 个分数在面积表现形式上都呈尖峰状,说明这 4 个

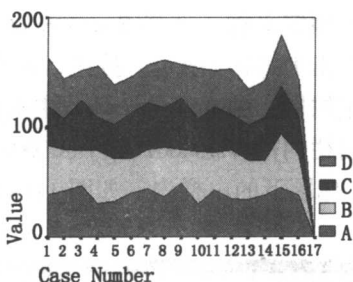


图 1.12 四个小班英语听力测验成绩堆积式面积图

分数都集中于各班分数分布曲线的高端。

应该说, 条形图、线图与面积图基本相同, 我们可根据数据特征与自己的爱好随意选择应用。

1.4 饼图 (Pie)

饼图用来表示数据的部分与整体之间的比例关系。整个圆面积代表整体 (100%), 以圆内各扇形面积代表整体中各部分所占的比例。以圆心角度大小表示所占构成比的大小。每 1% 相当于 3.6 度。饼图只有一种模式。从 Graphs 菜单选择 Pie (饼图) 选项, 进入其对话框就可看出这一点。我们引用一个诗歌文体分析的例子来证明饼图在文学研究中的价值。

例: 英国诗人华兹华斯诗歌所用修辞格百分比: (1) 明喻 29.1%, (2) 暗喻 24.3%, (3) 拟人法 16.5%, (4) 委婉法 8.4%, (5) 换喻 6.9%, (6) 提喻 5%, (7) 其他 9.8%。(该实例选自杨端和、李强《语言统计学》, 云南大学出版社 1998 年版, 第 19 页)

先输入数据。将百分比转换成整数 (29.1% → 0.291; 9.8% → 0.098 等)。为了简便起见, 可直接将数据输入在基本界面各 Var (变量) 列下第一行即可, 不用进行数据转换。输入完成后的形式如下:

Var 001	Var 002	Var 003	Var 004	Var 005	Var 006	Var 007
.29	.24	.17	.08	.070	.05	0.10

可以看出, SPSS 系统仅取两位小数并四舍五入。先将这组数据取名为 Pie 并按本章开头的描述方法存入 SPSS 数据库。现在就可以绘制饼图了。进

入 Pie charts (饼图绘制) 预定义框, 选择 Summary of separate variables (独立变量合计) 这一栏, 再点击 Define (限定) 按钮, 弹出 Define Pie: Summaries of Separate Variables (限定饼图: 独立变量合计) 主对话框。按照制作前面几类图形的方法, 将左边白栏内的 Var00001 至 Var00007 依次移入右边的 Slices Represent (用扇形表示) 标题下的白栏内, 如图 1.13 所示。

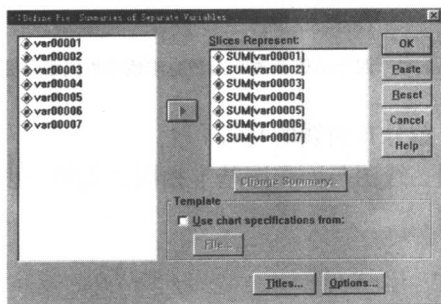


图 1.13 华兹华斯诗歌修辞格百分比饼图生成主对话框

为了将饼图制作得更完美, 可添加标题与说明词。点击该框下方 Titles (标题) 按钮, 弹出 Title (标题) 子对话框。将输入法转换为输入中文的任何一种模式, 在 Title 栏下键入相应的中文标题; 在 Footnote (脚注) 栏下输入文字说明, 如图 1.14 所示。

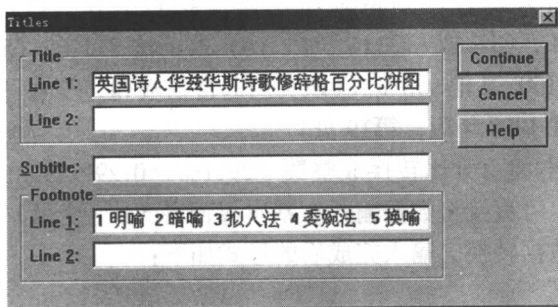


图 1.14 华兹华斯诗歌修辞格百分比饼图标题输入子对话框

输入完毕, 点击该框右上角的 Continue (继续) 按钮, 又回到由图 1.13 所示的主对话框。点击该主对话框右上角的 OK 按钮, 一幅漂亮的彩色饼图就大功告成。请看图 1.15。