

高等学校财经类核心课程教材

统计学实验

——SPSS和R软件应用与实例

费 宇 主编

高等学校财经类核心课程教材

统计学实验

——SPSS和R软件应用与实例

费 宇 主编
雷健敏 马云玲 鲁 筠 副主编

TONGJIXUE SHIYAN SPSS HE R RUANJIAN YINGYONG YU SHILI

内容简介

本书针对经济学类和工商管理类本科生统计学实验课程编写,结合两个国际通用的统计软件 SPSS 和 R 来介绍统计学的基本理论与方法,使读者能在最短的时间里掌握统计学基本理论和统计软件的基本操作。

本书的内容包括 SPSS 软件简介、R 软件简介、数据的描述、参数估计、假设检验、方差分析、回归分析、时间序列分析和抽样调查理论与方法。全书分两部分,第一部分“基于 SPSS 的统计学实验”(第 1 章至第 8 章)介绍如何使用 SPSS 软件实现以上统计分析,第二部分“基于 R 的统计学实验”(第 9 章至第 16 章)介绍如何使用 R 软件完成有关的统计分析。

本书可以作为经济学类和工商管理类本科生的统计学实验教材,也可以作为其他专业本科生和硕士研究生以及有关统计工作者学习 SPSS 和 R 统计软件的参考书。

为方便教学,本书配有教学辅助资源,内容包括教学课件 PPT、教材中的例题和习题的数据文件(下载网址:<http://www.hep.com.cn/sem>)。

图书在版编目(CIP)数据

统计学实验:SPSS 和 R 软件应用与实例/费宇主编.

—北京:高等教育出版社,2012.7

ISBN 978-7-04-035629-8

I. ①统… II. ①费… III. ①统计学—实验—软件包
—高等学校—教材 IV. ①G819—33

中国版本图书馆 CIP 数据核字(2012)第 132078 号

策划编辑 施春花

责任编辑 施春花

封面设计 王 晔

版式设计 余 杨

插图绘制 尹 莉

责任校对 杨雪莲

责任印制 田 甜

出版发行 高等教育出版社

社 址 北京市西城区德外大街 4 号

邮政编码 100120

印 刷 北京鑫海金澳胶印有限公司

开 本 787mm×1092mm 1/16

印 张 13.75

字 数 310 千字

购书热线 010-58581118

咨询电话 400-810-0598

网 址 <http://www.hep.edu.cn>

<http://www.hep.com.cn>

网上订购 <http://www.landaco.com>

<http://www.landaco.com.cn>

版 次 2012 年 7 月第 1 版

印 次 2012 年 7 月第 1 次印刷

定 价 23.70 元

本书如有缺页、倒页、脱页等质量问题,请到所购图书销售部门联系调换

版权所有 侵权必究

物 料 号 35629-00

前 言

统计学是一门关于数据的方法论科学,现代统计分析的一个显著特点就是运用计算机实现有关的统计计算与分析,因此统计软件成为学习统计学的必备工具。本书先介绍统计学基本理论与方法,然后结合实例通过 SPSS 和 R 软件完成统计分析,使读者能在最短的时间里掌握统计学基本理论和统计软件的基本操作,提高读者学习统计学的兴趣,帮助读者体会统计分析的乐趣;只要轻松点击鼠标或写几行命令就可以完成复杂的计算分析,得到美观漂亮的统计图表。

本书的主要特色是统计方法与应用结合,通过实例讲解统计软件的应用,具体特点可以概括为以下两点:

1. 内容丰富:采用目前最流行的两个统计软件 SPSS 和 R 作为计算软件来介绍统计学基本理论与方法,内容覆盖了本科统计学教学的主要内容,方便教师选择使用。

2. 简洁易学:每章首先给出本章的实验目的和主要内容,然后由实例引入本章的主要内容,介绍与例子相关的统计理论,以菜单方式和程序方式给出 SPSS 软件的具体操作,或给出 R 程序,最后对软件输出结果进行分析,方便教师讲授和学生学习。

为什么选用 SPSS 软件作为计算软件? 选用 SPSS 软件的原因主要是该软件是“傻瓜”式的软件,使用界面很友好,操作非常简单,初学者很容易掌握。所以,与很多学校一样,我校多年来一直选用 SPSS 软件作为统计教学软件。

为什么选用 R 软件作为计算软件? 选用 R 软件作为计算软件的初衷,要从 R 软件的使用说起。2002 年我到英国曼彻斯特大学访问,在我的导师潘建新教授指导下撰写博士论文,我使用的统计软件是 S-PLUS,它是用 S 语言编写的商业统计软件,我利用 S-PLUS 软件顺利完成了博士论文的有关计算,即将回国之际,潘建新教授送给我一张光盘,告诉我这是他下载的 R1.0.0 软件,R 软件与 S-PLUS 类似,也是 S 语言编写的统计软件,但它是一个免费的软件。回国后我第一次尝试使用了 R 软件,发现 R 软件与 S-PLUS 和 SPSS 软件一样,都是非常优秀的统计软件,当时就产生了将 R 软件作为教学软件介绍给学生的想法,但是由于种种原因,一直没有落实。2008 年我们邀请中国人民大学统计学院吴喜之教授来我院,给我院 2007 级研究生开设了 R 软件课程,给学生系统介绍 R 软件的使用,取得了非常好的效果,研究生的硕士论文水平有了很大提高。这使我产生了编写一本使用 SPSS 和 R 统计软件介绍统计学基本理论和方法的想法。2009 年编写《统计学》(高等教育出版社,2010 年)教材时,就产生了编写这样一本统计学实验教材的想法,并开始编写这本教材。2011 年我再次访问曼彻斯特大学,详细考察了曼彻斯特大学数学学院应用 R 软件进行统计教学的情况,吸取了一些有用的经验,对书稿进行了改进,终于完成了本书的编写,希望能对我国统计实验课程的教学贡献一份力量。

本书由下列人员完成:费宇编写第 1、3、4、6、9、11、12、14 章,雷健敏编写第 2 章,马云玲编写

II 前言

第8章,鲁筠编写第5、7、10、13、15、16章。最后由费宇对全书进行了审阅、修改和定稿。

在本书出版之际,首先要感谢国家自然科学基金(项目编号:11061036)资助我2011年访问曼彻斯特大学,这对本书的写作有很大的帮助。感谢英国曼彻斯特大学数学学院李秋菊博士帮助编写了本书第11、12、14章的部分R程序。还要感谢高等教育出版社的编辑为本书出版付出的劳动。本书是云南省高等学校人才培养模式创新实验区项目——“统计学人才培养模式创新实验区”建设成果之一,感谢该项目的大力支持。

本书的编写参阅了许多国内外的文献资料,并引用了部分例题和习题,在此向这些文献的作者表示感谢。

本书的教学课件PPT、教材中的例题和习题的数据文件可以在高等教育出版社的网站(<http://www.hep.com.cn/sem>)下载,供读者使用。

由于我们水平有限,书中难免有错漏之处,诚恳地欢迎同行专家和读者批评指正,并提出宝贵意见(feiyukm@yahoo.com.cn)。

费 宇

2012年4月

于昆明云南财经大学统计与数学学院

目 录

第 1 章 SPSS 软件简介	1	5.2 双因素方差分析	48
1.1 SPSS 软件概述	1	习题 5	52
1.2 SPSS 软件的基本窗口	2	第 6 章 回归分析	55
1.3 SPSS 数据文件的建立	6	6.1 相关分析	55
习题 1	7	6.2 一元线性回归	59
第 2 章 数据的描述	8	6.3 多元线性回归	64
2.1 常见统计图的绘制	8	6.4 虚拟变量回归	71
2.2 描述性统计分析	18	6.5 logistic 回归	72
习题 2	19	6.6 回归曲线的估计	74
第 3 章 参数估计	22	习题 6	81
3.1 一个正态总体均值的估计	22	第 7 章 时间序列分析	85
3.2 两个正态总体均值差的估计	24	7.1 时间序列的图形化观察	85
3.3 总体比例的估计	25	7.2 指数平滑	89
3.4 总体比例之差的估计	27	7.3 时间序列的 Box-Jenkin	
习题 3	28	模型	91
第 4 章 假设检验	31	7.4 时间序列的季节分解	101
4.1 一个正态总体均值的检验——		习题 7	103
单样本 t 检验	31	第 8 章 抽样调查理论与方法	106
4.2 两个正态总体均值差的检验——		8.1 简单随机抽样调查方法	106
双样本 t 检验	33	8.2 分层抽样方法	110
4.3 两个正态总体方差比的检验——		8.3 系统抽样方法	119
F 检验	36	8.4 整群抽样方法	123
4.4 非正态总体参数的检验	37	习题 8	127
习题 4	40	第 9 章 R 软件简介	128
第 5 章 方差分析	43	9.1 R 软件的特点	128
5.1 单因素方差分析	43	9.2 R 软件的安装与运行	129

9.3 R 软件的基本原理	129	13.2 双因素方差分析	166
9.4 R 软件的帮助	132	习题 13	168
习题 9	132		
第 10 章 数据的描述	134	第 14 章 回归分析	170
10.1 数据的整理	134	14.1 相关分析	170
10.2 描述统计分析	141	14.2 一元线性回归	173
习题 10	142	14.3 多元线性回归	176
		14.4 虚拟变量回归	180
第 11 章 参数估计	145	14.5 logistic 回归	181
11.1 一个正态总体均值的估计 ..	145	14.6 回归曲线的估计	183
11.2 两个正态总体均值差的估计	146	习题 14	190
11.3 总体比例的估计	147		
11.4 总体比例之差的估计	148	第 15 章 时间序列分析	194
习题 11	149	15.1 时间序列的图形化观察	194
		15.2 指数平滑	196
第 12 章 假设检验	152	15.3 Box-Jenkin 模型	197
12.1 一个正态总体均值的检验——单样本 t 检验	152	15.4 时间序列的季节分解	202
12.2 两个正态总体均值差的检验——双样本 t 检验	154	习题 15	203
12.3 两个正态总体方差比的检验— F 检验	158		
12.4 非正态总体参数的检验	158	第 16 章 抽样调查理论与方法	206
习题 12	161	16.1 简单随机抽样调查方法	206
		16.2 分层抽样方法	207
第 13 章 方差分析	164	16.3 系统抽样方法	209
13.1 单因素方差分析	164	16.4 整群抽样方法	210
		习题 16	212
		参考文献	213

第 1 章

SPSS 软件简介

1.1 SPSS 软件概述

SPSS 软件是美国斯坦福大学三位学生于 1968 年研制开发的统计软件,SPSS 是 Statistical Package for Social Science (社会科学统计软件包)的缩写,2000 年 SPSS 公司将英文全称更改为 Statistical Product and Service Solutions (统计产品与服务解决方案)。SPSS 软件现在已经成为一个统计功能强大的统计软件系统,广泛应用于经济学、管理学、生物学、医学、商业、金融业、农业、林业等领域。

SPSS 软件是普通统计用户的首选,具有如下特点:

1. 操作简单:可以采用菜单式操作,不需要编程序,只要简单点点鼠标,就可以完成统计分析,这种“傻瓜”式的操作方式,对于初学者来说非常方便,很容易入门。
2. 界面友好:SPSS 的标准界面类似于 Excel 的界面,方便初学者使用,SPSS16.0 及以下版本的界面是英文的,但从 SPSS17.0 版本开始推出了多国语言界面,包括简体中文界面,大大方便了中国普通用户的使用。
3. 数据编辑和传输方便:可以在数据编辑窗口中对打开的数据文件进行编辑,如增加、删除、复制、剪切和粘贴等常规操作;也可以与很多软件进行数据传输,如 DAT、SLK、DBF、Excel 和 Access 等多种文件都可以在 SPSS 中打开。
4. 输出的统计图表美观灵活:主要统计分析结果由表格和图形方式输出,表格和图形生成以后,可以进行编辑修改,可以复制到 Word 和 PowerPoint 文件里。
5. 统计功能强大:可以实现一般统计学教科书里的统计分析方法,如样本数据的描述分析和预处理,假设检验与参数估计,方差分析、回归分析、非参数检验、聚类分析、判别分析、因子分析、对应分析、时间序列分析、可靠性分析等。

基于上述特点,本书选用 SPSS 软件作为统计教学软件,在此选用了 SPSS16.0 版本而没有选择更高版本的原因有两个:一是 SPSS16.0 的标准界面是英文界面,这对于专业英语的学习是有帮助的,二是虽然界面是英文的,但输出语言可以设置为简体中文格式(具体见本章 1.2.1 节)。

1.2 SPSS 软件的基本窗口

SPSS16.0 的基本窗口有三种:数据编辑窗口、结果浏览窗口和程序编辑窗口。

1.2.1 数据编辑窗口(Data Editor)

为了说明方便,打开数据文件 li1.1.sav,即我国 2007 年 31 个地区城镇居民年人均可支配收入和年人均消费性支出数据,图 1.1 给出的是该数据文件的数据编辑窗口,它由标题栏、菜单栏、工具栏、数据编辑窗口和状态栏构成。

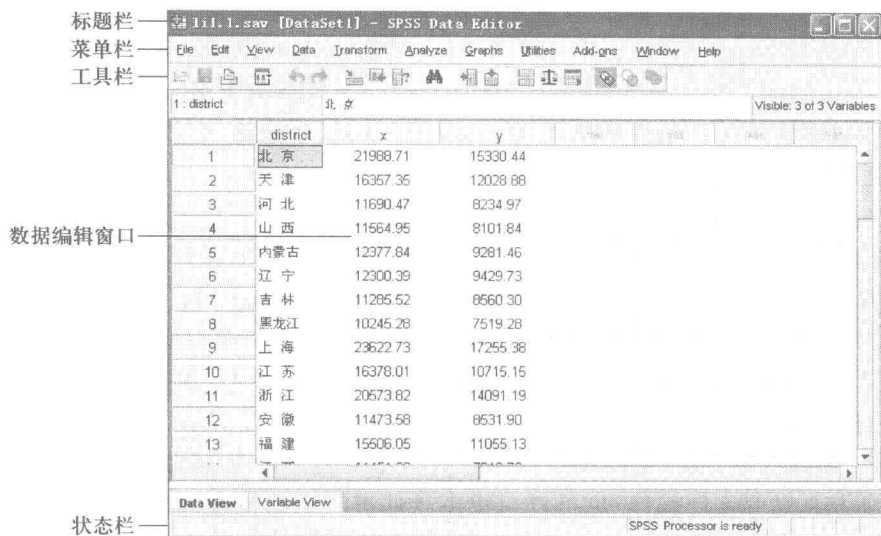


图 1.1 数据编辑窗口

1. 标题栏:显示当前打开的数据文件名,如文件 li1.1.sav。

2. 菜单栏:在标题栏下方,显示所有的一级菜单,点击这些一级菜单可以得到所包含的二级菜单(主要功能),下面简单介绍这些一级菜单的主要功能。

(1) File(文件):文件管理菜单,可以建立新文件,读入文件或数据库的内容,可以保存文件和将数据输出到数据库,也可以标记文件为只读文件,重命名数据文件和打印等。

(2) Edit(编辑):编辑菜单,可以对文件数据进行选择、复制、粘贴、删除、查找、插入变量、个案等操作,还可以利用 Options 项对 SPSS 系统参数进行设置和修改。比如,将输出结果由一般的英文方式修改为简体中文方式,可以在 Options 菜单的 General(通用参数设置)选项卡右边的 Output 栏中 Language(语言)选项框中选 Simplified Chinese(简体中文),如图 1.2 所示。

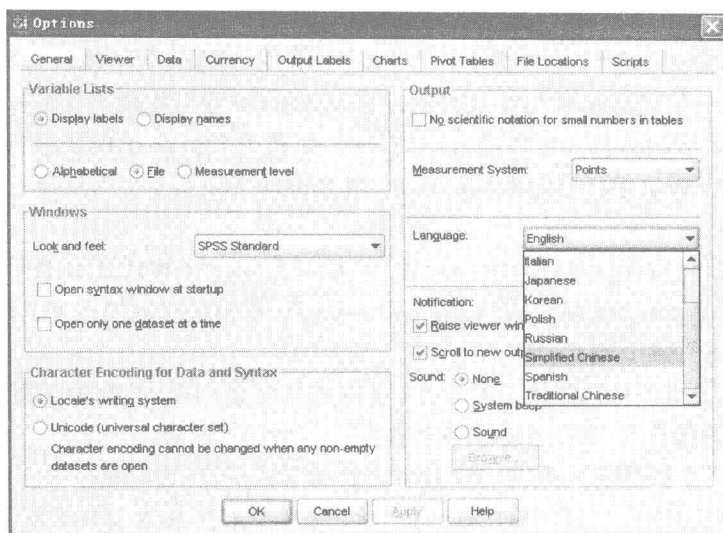


图 1.2 系统参数设置对话框

(3) View(视图):视图菜单,可以设置窗口的外观,自定义工具栏、字体设置、显示或隐藏格子、显示变量值标签以及在数据浏览和变量浏览之间切换等操作。

(4) Data(数据):数据管理菜单,可以进行数据变量的定义、拷贝数据或数据集、合并或拆分数据文件、对个案进行排序或加权以及个案选择等操作。

(5) Transform(转换):数据转换菜单,可以进行数值的计算、重新编码、缺失值替代、创建时间序列、产生随机数等操作。

(6) Analyze(分析):统计分析菜单,可以应用各种统计方法对数据进行统计分析。该菜单包括 Descriptive Statistics(描述统计)、Compare Means(比较均值)、General Linear Model(一般线性模型)、Correlate(相关分析)、Regression(回归分析)等 22 个分析模块,下面简要介绍后面几章将要用到的 5 个重要模块的主要功能。

① Descriptive Statistics(描述统计):可以进行描述性统计分析,如利用 Frequencies 子模块进行频数分析(详细操作说明见第 2 章),用 Descriptives 子模块计算均值、标准差等统计量,用 Explore 子模块进行探索性分析,即绘制茎叶图、直方图、箱线图、构造置信区间等(详细操作说明见第 3 章)。

② Compare Means(比较均值):可以进行均值比较分析,如利用 One Sample T Test 子模块进行单样本 t 检验,用 Independent Samples T Test 子模块进行独立样本 t 检验、用 Paired - Samples T Test 子模块进行配对样本 t 检验(详细操作说明见第 4 章),用 One Way ANOVA 子模块进行单因素方差分析(详细操作说明见第 5 章)。

③ General Linear Model(一般线性模型):可以利用一般线性模型进行统计分析,比如利用

Univariate 子模块进行多因素方差分析(详细操作说明见第5章)和虚拟变量回归分析(详细操作说明见第6章)。

④ Correlate(相关分析):可以进行相关分析,如利用 Bivariate 子模块计算相关系数进行相关分析,用 Partial 子模块计算偏相关系数进行相关分析(详细操作说明见第6章)。

⑤ Regression(回归分析):可以进行回归分析,如利用 Linear 子模块进行线性回归分析,用 Curve Estimation 子模块进行回归曲线的估计,用 Binary Logistic 子模块进行 logistic 回归分析(详细操作说明见第6章)。

(7) Graphs(图表):可以绘制各种统计图,比如利用 Bar 子模块绘制条形图,用 Pie 子模块绘制饼图,用 Histogram 子模块绘制直方图,用 Boxplot 子模块绘制箱线图,用 Line 子模块绘制线图(详细操作说明见第2章)。

(8) Utilities(实用工具):实用工具菜单,可以进行变量列表、控制输出管理系统、输出文件信息、定义和使用变量集合、运行菜单编辑等操作。

(9) Add-ons(增加模块):该菜单中列出注册的其他模块,选择相应项会链接到相应的帮助模块,注册相应模块功能后,其选项将从 Add-ons 菜单中移至其他相应菜单中。

(10) Window(窗口):窗口管理菜单,可以进行窗口拆分、最小化、切换窗口等操作。

(11) Help(帮助):帮助菜单,提供 SPSS 帮助、在线指南、统计分析指导等。

3. 工具栏:在菜单栏的下方列出了常用任务的快捷按钮,可以利用这些按钮快速进入相应的任务对话框,如第一个按钮是打开一个数据文件。

4. 数据编辑窗口:在工具栏的下方,该窗口的下方有两个标签,当选择“Data View”标签时,当前窗口为数据编辑窗口,可进行变量值的录入和修改,如果选择“Variable View”标签,则当前窗口为变量编辑窗口,可以显示数据窗口中变量的基本属性,也可以创建和修改变量属性。图 1.1 显示的是数据文件 li1.1.sav 在数据编辑窗口中“Data View”下的视图,而图 1.3 则是该数据在“Variable View”下的视图,它给出了该数据中 3 个变量的基本属性,如第 1 个变量的 Name(名称)是 district,它的 Type(类型)是 String(字符型),Width(宽度)是 6 个字符,Decimals(小数点位数)为 0,Label(变量标签)为地区,Values(变量值标签)为 None(没有),Missing(缺失值)为 None(没有),Columns(列宽)为 6(变量在 Data View 中所占的列宽为 6 个字符),Align(对齐方式)为 Left(左对齐),Measure(测度方式)为 Nominal(定类型变量);第 2 个变量的 Name(名称)是 x,它的 Type(类型)是 Numeric(数值型),Width(宽度)是 11 个字符,Decimals(小数点位数)为 2,Label(变量标签)为可支配收入,Values(变量值标签)为 None(没有),Missing(缺失值)为 None(没有),Columns(列宽)为 11(变量在 Data View 中所占的列宽为 11 个字符),Align(对齐方式)为 Center(中间对齐),Measure(测度方式)为 Scale(定距型变量)。

5. 状态栏:在数据编辑窗口的下方,主要显示当前数据文件中个案选择情况、加权状况、数据拆分情况和 SPSS 此时的工作状态,只有当 SPSS 程序的状态是“SPSS processor is ready”时,统计分析才能正常进行。

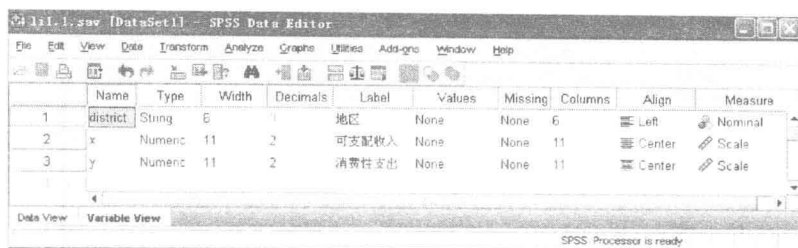


图 1.3 数据编辑窗口的“Variable View”下的视图

1.2.2 结果浏览窗口(Output Viewer)

在第一次分析完成后,系统会自动打开结果浏览窗口,图 1.4 是数据文件 li1.1.sav 做了描述统计分析后得到的结果。SPSS 所有的统计分析结果都显示在结果浏览窗口里,在结果浏览窗口里可以浏览、编辑输出结果,改变输出显示顺序等,输出结果可以保存为“*.SPV”格式的文件(SPSS16.0 以前的格式为“*.SPO”格式)。

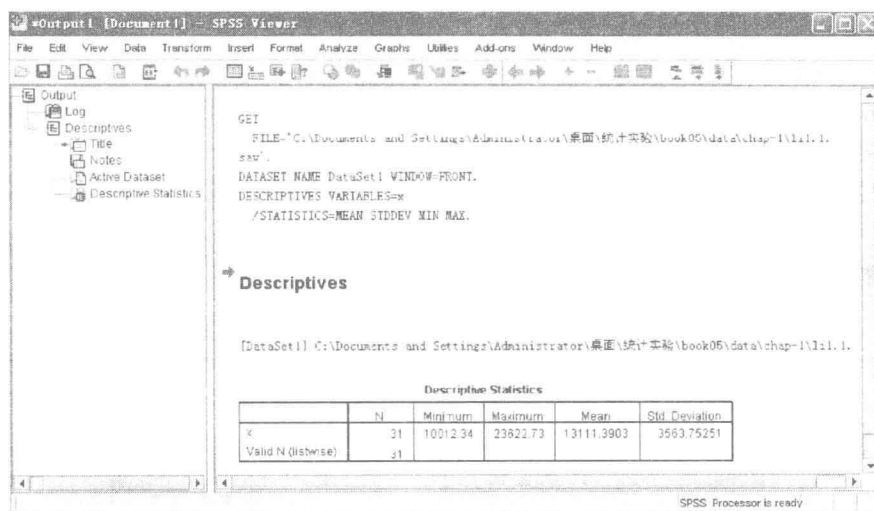


图 1.4 结果浏览窗口

1.2.3 程序编辑窗口(Syntax Editor)

程序编辑窗口是编写、调试和运行 SPSS 程序的窗口,如图 1.5 所示。SPSS 是菜单操作和程序控制并重的统计分析系统,大多数 SPSS 的功能可以利用菜单方式完成,但也有少数 SPSS 的功能只能通过编写程序实现,完成一个统计分析可以采用菜单方式或程序方式两种等价方法(详细操作说明见第 2 章至第 8 章)。

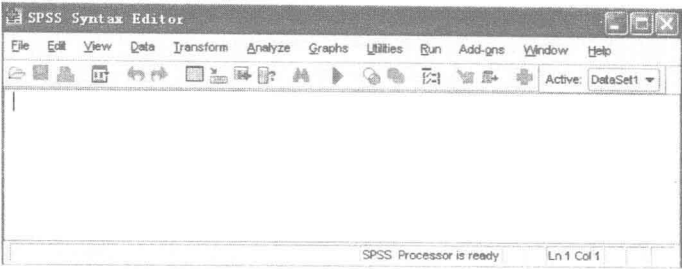


图 1.5 程序编辑窗口

1.3 SPSS 数据文件的建立

要进行统计分析,必须有数据文件,建立 SPSS 数据文件一般有以下三种方法。

1. 从键盘输入:即在 SPSS 数据窗口直接从键盘输入数据,SPSS 数据窗口的每一列表示一个变量,每一行代表一个个案。
2. 粘贴方式输入:即采用复制粘贴的方式将数据粘贴到数据窗口中,生成数据文件。
3. 读取方式输入:SPSS 支持多种格式的数据文件的读入,如果已经有其他格式的数据文件,可以采用读取的方式将数据读入生成 SPSS 数据文件。下面以读入 Excel 数据文件为例,说明如何读取数据。

【例 1.1】 (数据文件为 li1.1.xls) 表 1.1 给出了我国 2007 年 31 个地区城镇居民年人均可支配收入和年人均消费性支出数据(单位:元),该数据文件是 Excel 格式的文件,请采用读取的方式将数据读入 SPSS 生成相应的 SPSS 数据文件。

表 1.1 城镇居民年人均可支配收入和年人均消费性支出数据

地区	可支配收入(元)	消费性支出(元)	地区	可支配收入(元)	消费性支出(元)
北 京	21 988.71	15 330.44	湖 北	11 485.80	8 701.18
天 津	16 357.35	12 028.88	湖 南	12 293.54	8 990.72
河 北	11 690.47	8 234.97	广 东	17 699.30	14 336.87
山 西	11 564.95	8 101.84	广 西	12 200.44	8 151.26
内蒙古	12 377.84	9 281.46	海 南	10 996.87	8 292.89
辽 宁	12 300.39	9 429.73	重 庆	12 590.78	9 890.31
吉 林	11 285.52	8 560.30	四 川	11 098.28	8 691.99
黑龙江	10 245.28	7 519.28	贵 州	10 678.40	7 758.69

续表

地区	可支配收入(元)	消费性支出(元)	地区	可支配收入(元)	消费性支出(元)
上 海	23 622.73	17 255.38	云 南	11 496.11	7 921.83
江 苏	16 378.01	10 715.15	西 藏	11 130.93	7 532.07
浙 江	20 573.82	14 091.19	陕 西	10 763.34	8 427.06
安 徽	11 473.58	8 531.90	甘 肃	10 012.34	7 875.78
福 建	15 506.05	11 055.13	青 海	10 276.06	7 512.39
江 西	11 451.69	7 810.73	宁 夏	10 859.33	7 817.28
山 东	14 264.70	9 666.61	新 疆	10 313.44	7 874.27
河 南	11 477.05	7 826.72			

先运行 SPSS 软件进入数据编辑窗口,选择 File→Open→Data,将数据文件 lil.1.xls 选入 Open Data 对话框,点 Open 弹出对话框 Opening Excel Data Source,选择 Read variable names from the first row of data(从数据的第一行读取变量名),然后点 Continue 即可。

习 题 1

1.1 SPSS 软件的主要特点有哪些?

1.2 SPSS 软件的三种主要窗口是什么?每种窗口的主要功能是什么?

1.3 运行 SPSS 软件,选择 Help 菜单,打开 Tutorial(统计向导)获得入门向导帮助;打开 Case Studies(案例分析)查看各种统计分析过程的操作指导;打开 Statistics Coach(统计教练)获得基本统计方法的指导。

第 2 章

数据的描述

一、实验目的

通过运用 SPSS 进行频数分布表制作、图形绘制以及描述统计量计算,熟悉 SPSS 分析(Analyze)菜单里描述统计(Descriptive Statistics)中频数分布表子菜单(Frequencies)、探索性分析子菜单(Explore)、描述统计量子菜单(Descriptives);熟悉 SPSS 绘图(Graphs)菜单里条形图子菜单(Bar)、饼图子菜单(Pie)、直方图子菜单(Histogram)、箱线图子菜单(Boxplot)以及线图子菜单(Line)等模块的主要功能,掌握运用 SPSS 进行数据整理、图形绘制以及描述统计分析的基本操作过程,并能读懂 SPSS 输出的结果。

二、实验环境

1. 系统软件 Windows2000 或 WindowsXP 或 Windows7;
2. 统计软件 SPSS16.0 或更高版本。

三、实验内容

运用 SPSS 进行数据整理、图形绘制以及描述统计分析,包括对分类数据和顺序数据进行频数分布表展示及图形展示;对数值型数据进行频数分布表展示及图形展示;对数值型数据进行描述统计分析等内容。

2.1 常见统计图的绘制

2.1.1 分类数据和顺序数据的频数分布表

【例 2.1】 (数据文件为 li2.1.sav)某高校经济系 30 名教师性别及职称登记结果如表 2.1 所示,试用 SPSS 分别编制教师性别及职称的频数分布表。

表 2.1 某高校 30 名教师性别及职称情况统计表

序号	性别	职称	序号	性别	职称	序号	性别	职称
1	男	讲师	11	男	教授	21	男	副教授
2	女	助教	12	女	副教授	22	女	副教授
3	女	副教授	13	女	副教授	23	男	讲师
4	女	副教授	14	男	讲师	24	女	助教
5	男	助教	15	男	讲师	25	男	副教授
6	男	教授	16	男	副教授	26	男	讲师
7	女	教授	17	女	讲师	27	女	教授
8	男	讲师	18	男	助教	28	男	讲师
9	女	副教授	19	女	副教授	29	男	副教授
10	男	教授	20	女	副教授	30	女	教授

数据来源：费宇，等. 统计学. 北京：高等教育出版社，2010：25.

【统计理论】 对分类数据和顺序数据，计算出每一类别出现的频数或频率后，可通过频数分布表展示。将分类数据或顺序数据各分组及其相应的频(次)数全部罗列出来，就是频数分布(Frequency Distribution)(或次数分布)；将频数分布用表格的形式表现出来，就是频数分布表(或次数分布表)；分布在各组的单位数称为频数(Frequency)，也称次数；各组次数与总次数之比称为频率，也称比重。一组资料中，各组频率之和等于 100%。

【软件操作】 首先将教师性别用代码 0、1 表示；将教师职称用代码 2、3、4、5 表示，然后在数据文件的 Variable View 窗口 Values 栏定义变量值标签：0 表示女性，1 表示男性；2 表示助教，3 表示讲师，4 表示副教授，5 表示教授。SPSS 软件描述统计(Descriptive Statistics)功能可以完成频数分布表的制作，下面给出菜单方式和程序方式的操作步骤。

【菜单方式】 打开数据文件 li2. 1. sav，选择 Analyze→Descriptive Statistics→Frequencies→将“性别”与“职称”选入 Variable 框，点击 OK，可得表 2.2 及表 2.3。

表 2.2 某高校 30 名教师性别分组频数分布表
性别

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	女	14	46.7	46.7	46.7
	男	16	53.3	53.3	100.0
	Total	30	100.0	100.0	

表 2.3 某高校 30 名教师职称分组频数分布表
职称

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	助教	4	13.3	13.3	13.3
	讲师	8	26.7	26.7	40.0
	副教授	12	40.0	40.0	80.0
	教授	6	20.0	20.0	100.0
	Total	30	100.0	100.0	

【结果分析】 由表 2.2 可看出,30 名教师中男教师占 53.3%,女教师占 46.7%,教师性别比例差异不大;由表 2.3 可看出,30 名教师职称分布为钟形分布,其中副教授比重最大,占 40%;副教授以上职称的教师占 60%。

【程序方式】 等价的子程序如下:
FREQUENCIES VARIABLES=XB ZHCH
/ORDER=ANALYSIS.

2.1.2 分类数据和顺序数据的条形图及饼图

【例 2.2】 (数据文件为 li2. 2. sav) 根据表 2.3 资料,用 SPSS 绘制条形图。

【统计理论】 对分类数据和顺序数据,计算出每一类别出现的频数或频率后,还可通过统计图展示,如条形图。条形图是用条形高度表示数值大小的图形。这些条形的宽度相同,因而它能直观反映不同类别数据的多少及其分布状况。

【软件操作】 SPSS 软件条形图(Bar)功能可以完成条形图的绘制,下面给出菜单方式和程序方式的操作步骤。

【菜单方式】 打开数据文件 li2. 2. sav,选择 Graphs→Legacy Dialogs →Bar→Simple→在 Data in chart Are 框中选 Summaries for groups of cases→Define→在 Bars Represent 框中选 Other statistic,然后将“人数”选入 Variable(纵轴),将“职称分类”选入 Category Axis(横轴),单击 OK,可得图 2.1。

【程序方式】 等价的子程序如下:
GRAPH
/BAR(SIMPLE)=MEAN(RSH) BY ZHCHFL.