

第3章

SPSS商业信息分析实验

本章由数据录入与基本统计分析、参数分析与方差分析、相关与线性回归分析 3 个实验构成。通过本章的学习,学生能够对商业数据进行基本处理和分析。本章的主要内容是把数据输入到统计分析软件 SPSS 中,并熟练掌握 SPSS 统计软件的参数分析、方差分析、相关分析和线性回归分析。

3.1 数据录入与基本统计分析实验

1. 实验目的和要求

1) 实验目的

本实验的教学目的是熟悉 SPSS 软件,能够把数据导入到 SPSS 中,并对这些数据进行基本描述统计与频率分析,提高学生信息分析的能力。

2) 实验要求

- ① 熟悉 SPSS 软件及界面。
- ② 掌握将数据输入到 SPSS 中。
- ③ 掌握 SPSS 数据描述统计与频率分析。
- ④ 本实验共 4 学时,为验证性实验。

2. 实验准备

1) 实验涉及的相关理论知识学习

SPSS 是软件英文名称 Statistical Product and Service Solutions 的首字母缩写,意为“统计产品与服务解决方案”。SPSS 是世界上最早的统计分析软件,由美国斯坦福大学的 3 位研究生于 20 世纪 60 年代末研制,同时成立了 SPSS 公司,并于 1975 年在芝加哥组建了 SPSS 总部。迄今 SPSS 软件已有 30 余年的成长历史。全球约有 25 万家产品用户,他们分布于通信、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。SPSS 是世界上最早采用图形菜单驱动界面的统计软件,它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,以对话框方式展示出各种功能选择项。用户只要掌握一定的 Windows 操作技能,略通统计分析原理,就可以使用该软件为特定的科研工作服务。SPSS 是非专业统计人员的首选统计软件。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、相关分析、回归分析、对数线性模型、聚类分析、数据简化、生存分析、时间序列分析、多重响应等几大类,每类中又分为多个统计过程。SPSS 也有专门的绘图系统,可以根据数据绘制各种图形。

(1) SPSS 界面

SPSS 主界面主要有两个,一个是 SPSS 数据编辑窗口,另一个是 SPSS 输出窗口。数据编辑窗口由标题栏、菜单栏、工具栏、编辑编辑区、窗口切换标签页和状态栏组成,数据编辑窗口由变量视图和数据视图组成,变量视图如图 3-1 所示,数据视图如图 3-2 所示。进行数据分析后,就会有 SPSS 输出窗口,输出窗口如图 3-3 所示。



图 3-1 变量视图窗口

(2) SPSS 数据结构

在如图 3-1 所示的变量视图中定义数据结构,每个数据变量包括如下栏目。

名称: 定义变量名称,是变量访问和分析的唯一标志。变量的命名有一定规则,如首字符应以英文字母开头,不能有“!、?、*”字符,下划线、圆点不能作为变量名的最后一个字母。变量名也不能同 SPSS 内部具有特定含义的保留字同名,如 ALL、BY、AND 等。默认的变量名为 VAR00001、VAR00002 等。

类型: 定义变量数据类型。SPSS 具有的数据类型如图 3-4 所示,可分为 3 大类:字符串型、日期型、数值型。

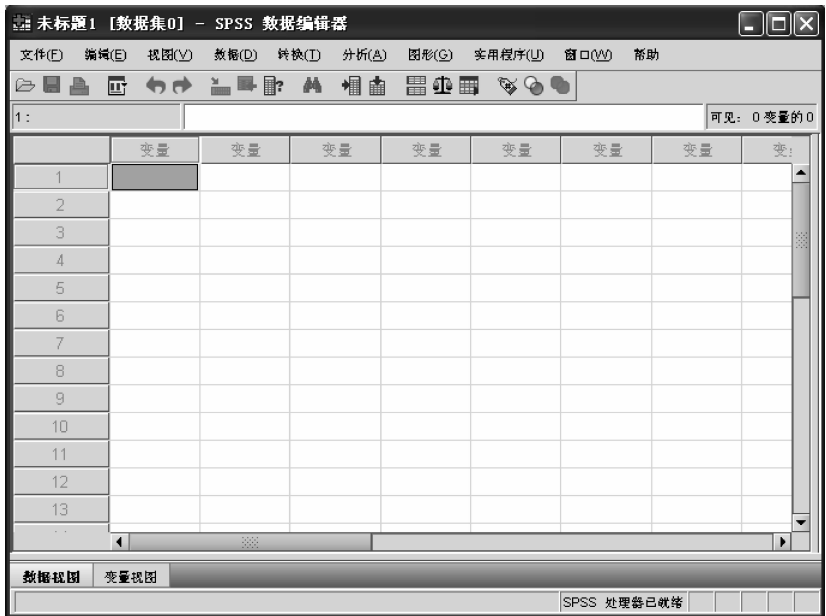


图 3-2 数据视图窗口

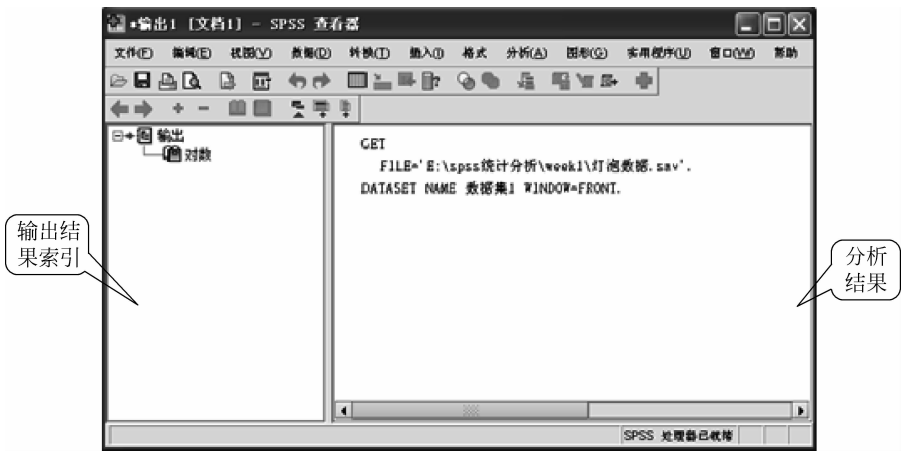


图 3-3 分析结果输出窗口

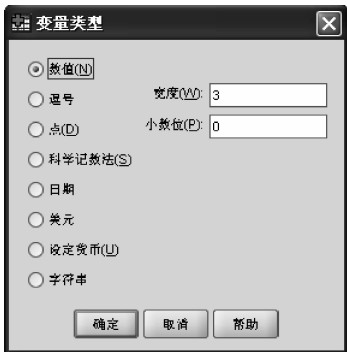


图 3-4 数据类型选择窗口



字符串型是 SPSS 较常用的数据类型,由一个字符串组成。如职工号码、姓名、地点等变量都可定义为字符型数据。

日期型用来显示日期或者时间,如生日、成立日期等可定义为日期型数据。SPSS 有很多日期类型,当选择了日期型数据时,系统会自动让用户选择日期类型。

数值型包括:

- ① 标准数值型,可以设置总宽度,小数位数等。
- ② 带逗号的数值型,在整数部分从个位开始每 3 位以一个逗号隔开。
- ③ 带圆点的数值型,在整数部分从个位开始每 3 位以一个圆点隔开,以逗号作为整数与小数的分隔符。
- ④ 科学记数型,如 $1.2E+02$ 。
- ⑤ 美元符号型,主要表示货币数据,它在数据前加美元符号 \$。

变量名称标签:对变量含义的进一步解释,它可增强变量名的可视性和统计分析结果的可读性。

变量值标签:是对变量取值含义的解释说明信息。如设数值 1 表示男,2 表示女,通过此设定不但明确了数据的含义,也增强了最后统计分析结果的可读性。定义变量值标签如图 3-5 所示。

缺失数据:输入数据时可能击错键,或者写错数字出现异常值,当用户注明这些值之后,在运算时,计算机会自动把这些值消去,作为缺失数据。“缺失值”对话框如图 3-6 所示。



图 3-5 变量值标签定义窗口

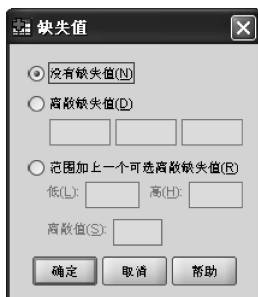


图 3-6 “缺失值”对话框

宽度:存储变量值的最大字符位数。

列:某些特殊数据类型的最大字符位数,如日期型。

对齐方式:有左对齐、右对齐、居中 3 种。

计量尺度:有定距型、定序型、定类型。定距型指身高、体重、血压、收入等连续型数据;定序型指具有内在大小或高低顺序的数据,如职称可分为低级、中级、高级 3 种,分别用数字 1、2、3 表示;定类型是指没有内在固有大小或高低数据,如性别可以分为男、女,用 1、2 表示。

(3) 数据的一般描述性统计

一般描述性统计包括计算均值、标准差、偏度、峰度等数据,通过这些描述把握数据的集中趋势、离散程度和分布形状等。



刻画集中趋势的描述性统计量有：

① 均值
$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

② 中位数，一组数据按升序排列后，处于中间位置的数据值。

③ 众数，一组数据中出现次数最多的数据值。

刻画离散程度的描述性统计量有：

④ 样本标准差
$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$$

⑤ 全距，也称极差，是数据的最大值与最小值之间的绝对值。

刻画分布形态的描述统计量有：

⑥ 偏度，描述变量取值分布形态对称性的统计量，其数学公式为：

$$\text{skewness} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^3 / s^3$$

⑦ 峰度，描述变量取值分布形态陡缓程度的统计量，其数学公式为：

$$\text{kurtosis} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^4 / s^4 - 3$$

(4) 频数统计

通过频数分析能够了解变量取值的状况，对把握数据的分布特征是非常有用的。通过频数分析，通常可以得到一个频数分布表，它包括某一个区间的频数、百分比、累计百分比及直方图。

对于定距型数据的频数分析，必须先分组；对于定序型和定类型数据可以直接做频数分析。

2) 实验相关软硬件准备

硬件要求：每人一台 PC。

软件：Windows 操作系统、SPSS 软件。

3. 实验内容

① 现有如表 3-1 所示数据，请输入 SPSS 中。

表 3-1 实验数据

编 号	性 别	年 龄	身高/厘米	体重/公斤
1	男	1985/03/24	156	47.5
2	男	1986/11/08	155	37.8
3	女	1987/12/15	157.9	49.2
4	男	1986/07/06	166	57

② 现有 Book1.xls 数据，请将它读入 SPSS 中。

③ 现有价格.sav 数据，请对价格做描述性统计，统计它的均值、标准差、众数、中位数、偏度、峰度。

④ 对价格.sav 数据中的某商品价格做频数统计。因该价格属于定距型数据，所以先要

对它进行分组,分为6组:0~9.93,9.93~9.96,9.96~9.99,9.99~10.02,10.02~10.05,10.05以上,然后做频数统计。

4. 实验步骤

1) 把数据输入 SPSS 中

① 数据结构设置如表 3-2 所示。

表 3-2 实验数据

变量名称	数据类型	宽 度	小数位数	标签	值标签	列宽度	计量尺度
编号	数值型	2	0			8	定距
性别	数值型	1	0		1 男 2 女	8	定类
年龄	日期型	Yyyy/mm/dd				20	定距
身高	数值型	5	1			8	定距
体重	数值型	4	1			8	定距

② 然后把数据输入 SPSS 界面中,变量视图如图 3-7 所示,数据视图如图 3-8 所示。



图 3-7 变量的数据结构视图



图 3-8 变量的数据视图

2) 把 Book1.xls 读入 SPSS 中

① 通过菜单“文件”→“打开”→“数据”，出现如图 3-9 所示界面，选择正确的文件类型和文件名称，然后单击“打开”按钮，出现如图 3-10 所示界面。



图 3-9 选择文件类型与文件名称

② 在图 3-10 界面上，选中“从第一行读取变量名称”复选框，然后单击“确定”按钮，则数据就读到 SPSS 中。

3) 对价格.sav 中的数据进行分析

现有价格.sav 数据，请对价格做描述性统计，统计它的均值、标准差、众数、中位数、偏度、峰度。

① 通过菜单“分析”→“描述统计”→“描述”得到如图 3-11 所示的界面，把“某商品价格”选入“变量”框内。

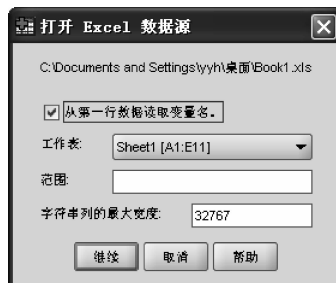


图 3-10 读取变量名与数据



图 3-11 描述性统计界面

② 单击“选项”按钮，得到如图 3-12 所示的界面，按图选中复选框，单击“继续”按钮回到如图 3-11 所示界面。

③ 在图 3-11 界面中单击“确定”按钮，得到如图 3-13 所示的描述性统计结果。



图 3-12 描述性统计选项

描述统计量

	N	极小值	极大值	均值	标准差	偏度		峰度	
	统计量	统计量	统计量	统计量	统计量	统计量	标准误	统计量	标准误
某商品价格(单位: 元)	30	9.93	10.06	10.0030	.03098	-.212	.427	-.165	.833
有效的 N(列表状态)	30								

图 3-13 描述性统计结果

4) 对价格.sav 数据中的某商品价格进行分组

分为 6 组: 0~9.93, 9.93~9.96, 9.96~9.99, 9.99~10.02, 10.02~10.05, 10.05 以上, 然后做频数统计。

① 通过菜单“转换”→“重新编码为其他变量”得到如图 3-14 所示界面, 把“price”选入“数字变量→输出变量(V):”框内, 在“输出变量”组合框中的“名称”框中输入新的变量名称后, 单击“更改(H)”按钮。



图 3-14 数据重新编码

② 单击图 3-14 中的“旧值和新值”按钮,出现如图 3-15 所示界面,在“旧值”框内输入范围,在“新值”框内输入新值,然后单击“更改”按钮。重复输入,最后如图 3-15 所示,单击“继续”按钮,离开此界面。



图 3-15 给变量分组

③ 单击如图 3-14 中的“确定”按钮,则会出现相应的新变量。

④ 通过菜单“分析”→“描述统计”→“频率”得到如图 3-16 所示的界面,选中“显示频率表格”,把已分组的新变量 price1 选入“变量”框。



图 3-16 对变量做频率分析

⑤ 单击图 3-16 中的“统计量”按钮,出现如图 3-17 所示界面,按图选中统计量后单击“继续”按钮离开。

⑥ 单击图 3-16 中的“图表”按钮,出现如图 3-18 所示界面,选中“带正态曲线”的直方图后单击“继续”按钮离开。

⑦ 单击图 3-16 中的“确定”按钮,得到如图 3-19 所示的结果。



图 3-17 频率分析的描述性统计

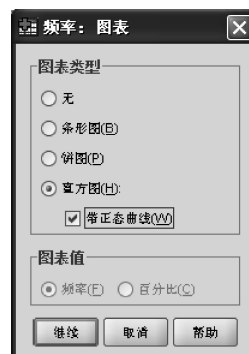


图 3-18 频率分析选中的图表

price1				
		频率	百分比	有效百分比
有效	1	1	3.3	3.3
	2	2	6.7	10.0
	3	9	30.0	40.0
	4	10	33.3	73.3
	5	7	23.3	96.7
	6	1	3.3	100.0
	合计	30	100.0	100.0

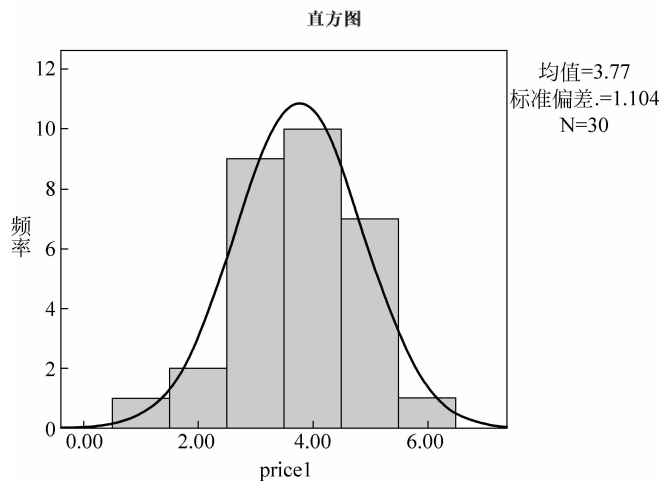


图 3-19 频率分析的结果

5. 实验思考

通过本次实验,思考如下问题。

- ① SPSS 基本统计与 Excel 操作和分析方面有哪些不同?
- ② 在 SPSS 中,如何对数据进行描述性统计分析和频数分析?