

# 常用统计软件介绍

《概率论与数理统计》是一门实践性很强的课程。但是,目前在 国内,大多侧重基本方法的介绍,而忽视了统计实验的教学。这样既不 利于提高学生创新精神和实践能力,也使得这门课程的教学显得枯燥 无味。为此,我们介绍一些常用的统计软件,以使学生对统计软件有 初步的认识,为以后应用统计方法解决实际问题奠定初步的基础。

## 一、统计软件的种类

### 1. SAS

是目前国际上最为流行的一种大型统计分析系统,被誉为统计分 析的标准软件。尽管价格不菲, SAS 已被广泛应用于政府行政管理, 科研,教育,生产和金融等不同领域,并且发挥着愈来愈重要的作用。 目前 SAS 已在全球 100 多个国家和地区拥有 29000 多个客户群,直接 用户超过 300 万人。在我国,国家信息中心,国家统计局,卫生部, 中国科学院等都是 SAS 系统的大用户。尽管现在已经尽量“傻瓜化”, 但是仍然需要一定的训练才可以使用。因此,该统计软件主要适合于 统计工作者和科研工作者使用。

### 2. SPSS

SPSS 作为仅次于 SAS 的统计软件工具包,在社会科学领域有着 广泛的应用。SPSS 是世界上最早的统计分析软件,由美国斯坦福大 学的三位研究生于 20 世纪 60 年代末研制。由于 SPSS 容易操作,输 出漂亮,功能齐全,价格合理,所以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刊杂志纷纷就 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 30 余年的成长历史。全球

约有 25 万家产品用户，它们分布于通讯、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成文的规定，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，由此可见其影响之大和信誉之高。因此，对于非统计工作者是很好的选择。

### 3. Excel

它严格说来并不是统计软件，但作为数据表格软件，必然有一定统计计算功能。而且凡是有 Microsoft Office 的计算机，基本上都装有 Excel。但要注意，有时在装 Office 时没有装数据分析的功能，那就必须装了才行。当然，画图功能是都具备的。对于简单分析，Excel 还算方便，但随着问题的深入，Excel 就不那么“傻瓜”，需要使用函数，甚至根本没有相应的方法了。多数专门一些统计推断问题还需要其他专门的统计软件来处理。

### 4. S-plus

这是统计学家喜爱的软件。不仅由于其功能齐全，而且由于其强大的编程功能，使得研究人员可以编制自己的程序来实现自己的理论和方法。它也在进行“傻瓜化”，以争取顾客。但仍然以编程方便为顾客所青睐。

### 5. Minitab

这个软件是很方便的功能强大而又齐全的软件，也已经“傻瓜化”，在我国用的不如 SPSS 与 SAS 那么普遍。

## 6. Statistica

也是功能强大而齐全的“傻瓜化”的软件，在我国用的也不如 SAS 与 SPSS 那么普遍。

## 7. Eviews

这是一个主要处理回归和时间序列的软件。

### 二、SAS 简介

#### 1. 关于 SAS

SAS 是美国使用最为广泛的三大著名统计分析软件（SAS，SPSS 和 SYSTAT）之一，是目前国际上最为流行的一种大型统计分析系统，被誉为统计分析的标准软件。

SAS 为“Statistical Analysis System”的缩写，意为统计分析系统。它于 1966 年开始研制，1976 年由美国 SAS 软件研究所实现商品化。1985 年推出 SAS PC 微机版本，1987 年推出 DOS 下的 SAS6.03 版，之后又推出 6.04 版。以后的版本均可在 WINDOWS 下运行，目前最高版本为 SAS6.12 版。SAS 集数据存取，管理，分析和展现于一体，为不同的应用领域提供了卓越的数据处理功能。它独特的“多硬件厂商结构”（MVA）支持多种硬件平台，在大，中，小与微型计算机和多种操作系统（如 UNIX，MVS WINDOWS 和 DOS 等）下都可运行。SAS 采用模块式设计，用户可根据需要选择不同的模块组合。它适用于具有不同水平于经验的用户，初学者可以较快掌握其基本操作，熟练者可用于完成各种复杂的数据处理。

目前 SAS 已在全球 100 多个国家和地区拥有 29000 多个客户群，直接用户超过 300 万人。在我国，国家信息中心，国家统计局，卫生部，中国科学院等都是 SAS 系统的大用户。SAS 已被广泛应用于政府行政管理，科研，教育，生产和金融等不同领域，并且发挥着愈来愈重要的作用。

## 2. SAS 的设计思想

SAS 的设计思想是为统计学家和科学工作者提供这样的一个工具，利用它可以完成包括从简单的描述性系统到复杂的多变数分析的各种运算，从而使人们从繁重的计算任务中解脱出来，有更多的时间和精力用于分析和解释计算的结果，而不必为如何获得这些结果花费过多的时间和精力。

## 3. SAS 的功能

SAS 是数据管理和分析软件包，能够完成各种统计分析，矩阵运算和绘图等。

SAS 的各项功能由功能模块完成。其中 BASA 模块为必需模块，其它模块可任选。供选择的模块包括统计 (STAS)，矩阵运算 (IML)，绘图 (GRAPH) 和全屏幕操作 (FSP) 等 20 余个。

基础模块 (BASE)，具有以下功能：进行数据存储，调入，追加，拷贝和文件处理；编写报告，打印图表；进行数据排序，分类等操作；完成一些基本统计数计算（如平均数和相关系数）；与一些软件包

(DBASE, LOTUS 等) 与大型机进行数据交换和通讯。BASE 模块为 SAS 系统的核心模块。

统计模块（STAT）提供一些高度可靠，完整的统计分析过程。主要有方差分析（包括一元，多元的单因素与多因素实验设计的方差分析），线性相关和回归分析（包括聚类分析，主成份分析，因子分析，典范相关分析）以及非参数测验等，共计 26 个过程。每个过程还提供多种不同算法和选项，从而 SAS 系统成为一个全面，细致，科学的统计分析方法集。STAT 模块为 SAS 系统的核心和精华。

矩阵运算模块（IML）是一种交互式矩阵语言。可直接进行矩阵运算（加法，乘法，求逆，计算特征值和特征向量等），适用于高级统计，工程运算和数学分析。

绘图模块（GRAPH）能在微机的绘图设备上绘制图形。可制作三维图形，地图和幻灯等。

全屏幕操作模块（FSP）为一交互式全屏幕软件。利用他可以建立，修改和浏览 SAS 数据集中的观察值，定义用户屏幕等。

#### 4. SAS 的特点

SAS 把数据存取，管理，分析和展现有机地融为一体。主要特点如下：

##### 1) 功能强大，统计方法齐，全，新

SAS 提供了从基本统计数的计算到各种试验设计的方差分析，相关回归分析以及多变量分析的多种统计分析过程，几乎囊括了所有最新分析方法，其分析技术先进，可靠。分析方法的实现通过过程调用完成。许多过程同时提供了多种算法和选项。例如方差分析中的多重比较，提供了包括 LSD，DUNCAN，TUKEY 测验在内的 10 余种方法；回

归分析提供了 9 种自变量选择的方法（如 STEPWISE, BACKWARD, FORWARD, RSQUARE 等）。回归模型中可以选择是否包括截距，还可以事先指定一些包括在模型中的自变量字组（SUBSET）等。对于中间计算结果，可以全部输出，不输出或选择输出，也可存储到文件中供后续分析过程调用。

## 2) 使用简便，操作灵活

SAS 以一个通用的数据（DATA）步产生数据集，尔后以不同的过程调用完成各种数据分析。其编程语句简洁，短小，通常只需很小的几句语句即可完成一些复杂的运算，得到满意的结果。结果输出以简明的英文给出提示，统计术语规范易懂，具有初步英语和统计基础即可。使用者只要告诉 SAS “做什么”，而不必告诉其“怎么做”。同时 SAS 的设计，使得任何 SAS 能够“猜”出的东西用户都不必告诉它（即无需设定），并且能自动修正一些小的错误（例如将 DATA 语句的 DATA 拼写成 DATE, SAS 将假设为 DATA 继续运行，仅在 LOG 中给出注释说明）。对运行时的错误它尽可能地给出错误原因与改正方法。因而 SAS 将统计的科学、严谨、准确与方便使用有机地结合起来，极大地方便了使用者。

## 3) 提供联机帮助功能

使用过程中按下功能键 F1，可随时获得帮助信息，得到简明的操作指导

### 三、SPSS 简介

#### 1. 关于 SPSS 软件

SPSS 是“社会科学统计软件包” (Statistical Package for the Social Science) 的简称, 是一种集成化的计算机数据处理应用软件, 是世界上公认的三大数据分析软件之一 (SAS、SPSS 和 SYSTAT)。1968 年, 美国斯坦福大学 H. Nie 等三位大学生开发了最早的 SPSS 统计软件, 并于 1975 年在芝加哥成立了 SPSS 公司, 已有 30 余年的成长历史, 全球约有 25 万家产品用户, 广泛分布于通讯、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。伴随 SPSS 服务领域的扩大和深度的增加, SPSS 公司已决定将其全称更改为 Statistical Product and Service solutions (统计产品与服务解决方案)。目前, 世界上最著名的数据分析软件是 SAS 和 SPSS。SAS 由于是为专业统计分析人员设计的, 具有功能强大, 灵活多样的特点, 为专业人士所喜爱。而 SPSS 是为广大的非专业人士设计, 它操作简便, 好学易懂, 简单实用, 因而很受非专业人士的青睐。此外, 比起 SAS 软件来, SPSS 主要针对着社会科学研究领域开发, 因而更适合应用于教育科学研究, 是国外教育科研人员必备的科研工具。1988 年, 中国高教学会首次推广了这种软件, 从此成为国内教育科研人员最常用的工具。

## 2. SPSS 软件的特点

1) 集数据录入、资料编辑、数据管理、统计分析、报表制作、图形绘制为一体。从理论上说, 只要计算机硬盘和内存足够大, SPSS 可以处理任意大小的数据文件, 无论文件中包含多少个变量, 也不论数据中包含多少个案例。

2) 统计功能囊括了《统计学》中所有的项目, 包括常规的集中量数和差异量数、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，并能在屏幕(或打印机)上显示(打印)如正态分布图、直方图、散点图等各种统计图表。从某种意义上讲，SPSS 软件还可以帮助数学功底不够的使用者学习运用现代统计技术。使用者仅需要关心某个问题应该采用何种统计方法，并初步掌握对计算结果的解释，而不需要了解其具体运算过程，可能在使用手册的帮助下定量分析数据。

3) 自从 1995 年 SPSS 公司与微软公司合作开发 SPSS 界面后，SPSS 界面变得越来越友好，操作也越来越简单。熟悉微软公司产品的用户学起 SPSS 操作很容易上手。SPSS for Windows 界面完全是菜单式，一般稍有统计基础的人经过三天培训即可用 SPSS 做简单的数据分析，包括绘制图表、简单回归、相关分析等等，关键在于如何进行结果分析与解释，这一方面需要学习一些数理统计的基本知识，另一方面也要多进行实践，在实践中了解各种统计结果的实际意义。

### 3. SPSS 的版本与应用环境

从 SPSS/PC+ V2.0~V4.0 为 DOS 版；从 SPSS/PC V6.0~V14.0 为 Windows 版。运行 SPSS 的 Windows10.0 版（单机版），要求 P2 以上的 PC 机和 1G 以上硬盘（32M 内存、68M 硬盘储存空间和 32M 以上预留空间）；中文 Win95/98/2000 平台。国内 SPSS 正版（WINDOWS 英文单机版）价格约 12000 元（RMB/20XX）；网络版约为单机版的 10 倍以上。XX 有 SPSS 的中文版，中国大陆尚无。

### 4. SPSS 软件的安装与汉化



SPSS 安装与其他 WIN 软件类似，在“安装向导”提示下完成。目前尚无汉化版。为了帮助学习，我们为大家提供了一个 14.0 版的“汉化补丁”（PSPSS10a.EXE），但仅能汉化菜单，尚不能汉化输出结果。

## 5. SPSS 的主要窗口与其功能

SPSS 系统提供两种操作运行方式，分别为窗口菜单方式和程序方式，我们主要介绍以窗口和对话框为主的窗口菜单运行方式。此种方式下用户在数据编辑窗口准备好数据后，利用窗口相应菜单项中的各功能进行管理和统计分析工作。此时用户通过点选某菜单项选择一种统计分析过程，系统用对话框接受用户设定的参数和选择项后即执行该过程，并将结果在输出窗口中显示出来。这种运行方式简便、直观，这也正是 SPSS 大受欢迎的原因。

既然是窗口操作方式，首先就要介绍 SPSS 系统的窗口。SPSS 的窗口都具有典型的 Windows 风格和功能，具备各种窗口控件，主要有三类：数据编辑窗口（Data Editor）、结果输出窗口（Viewer）和语法窗口（Syntax Editor），下面分别介绍各自的功能与特点。

### 5.1 数据编辑窗口——SPSS Data Editor

系统启动后自动打开的窗口就是数据编辑窗口，是 SPSS 系统的主窗口，主要用于准备、整理数据以及调用统计分析过程等。系统只能同时打开一个数据文件；当打开新的数据文件时，系统自动关闭前一个数据文件；一旦关闭数据窗口则 SPSS 系统也就相应退出。

#### 5.1.1 窗口结构



	手术次序	姓名	病号	gender	age	diameter	cs_invas	sell
1	1	王秀	E27124	0	55	45	1	
2	2	程秀	F17815	0	58	25	0	
3	2	余端	F27821	1	37	45	1	
4	1	冯广	E89391	1	46	30	1	
5	1	李召	D96224	1	65	20	0	
6	1	梁清	D87699	0	55	40	0	
7	1	叶榕	E05452	1	48	48	1	
8	1	罗睿	D90145	0	21	26	0	

图 1.1 SPSS 数据编辑窗的数据窗口

数据编辑窗口结构见图 1.1，在窗口标题栏中会显示当前打开的数据文件名，若是新建数据文件则系统默认文件名为 Untitled。

数据显示区是数据编辑窗口的主要区域，由两张工作表组成，即数据视窗 (Data View) 和变量视窗 (Variable View)，可以通过点选左下方的视窗标签进行视窗的切换，也可以通过菜单项 View→Data/Variables 进行切换。每张工作表都是一张可扩展的二维数据表格。其中数据视窗主要用于显示和编辑数据，所以专门提供了当前数据栏 (显示当前记录号和变量名) 和数据输入栏 (显示当前记录号和变量名对应的数据值)。在数据表中，系统自动取最左列显示记录号，最顶行显示变量名，要分析处理的数据存放在表格中。

变量视窗则用来定义编辑变量的有关属性。一个变量的属性占用一行，每种属性各占一列，包括变量名 (Name)、类型 (Type)、宽度 (Width)、精度 (Decimal)、标签 (Label)、值标签 (Values)、缺失值 (Missing)、列宽 (Columns)、对齐方式 (Align) 和测量类型 (Measure) 等。

### 5.1.2 窗口菜单和工具栏

窗口上方的菜单栏提供了 SPSS 系统全部可调用功能，共有 10 个选项，分别为：

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：<https://d.book118.com/388103110035007007>