

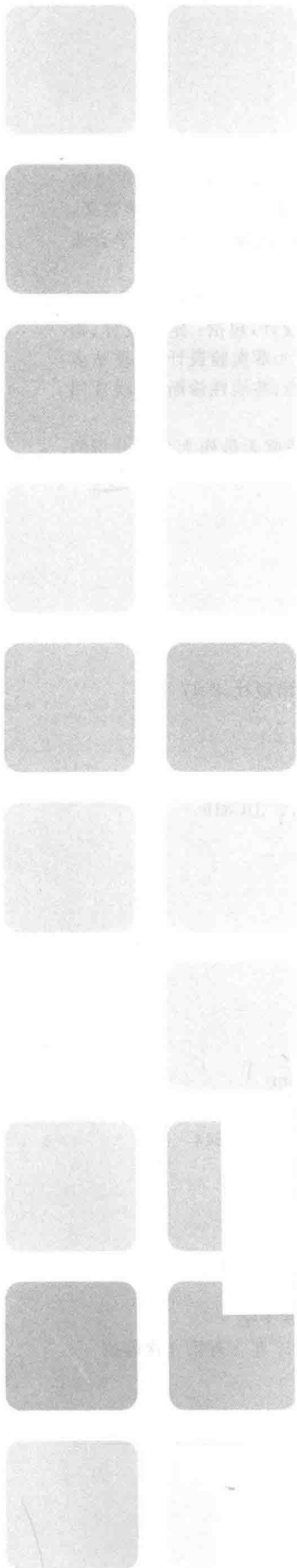
全国普通高等院校
信息管理与信息系统专业规划教材

SAS统计分析及应用 与JMP实验设计 (第2版)

黄 燕 编著

清华大学出版社





全国普通高等院校
信息管理与信息系统专业规划教材

SAS统计分析及应用 与JMP实验设计

(第2版)

黄 燕 编著

清华大学出版社
北京

内 容 简 介

本书主要介绍 JMP 的 DOE 试验设计和 SAS 与 JMP 两个软件包的多元统计分析方法。全书重点在于用实例讲解运用 SAS 与 JMP 两个软件包中的各种统计分析方法,并详细解释输出结果的统计学意义,比较两个软件相同统计方法的优劣与输出内容的异同,方便学习者从多角度应用,能够举一反三,学会选用科学合理的技能判断并获取科学的、有价值的统计分析结果,提高效率。

全书共 9 章,第 1 章介绍 SAS 与 JMP 软件基本操作;第 2 章和第 3 章介绍常用描述统计量、数据分布、正态性检验、t 检验与非参数检验等概念;第 4 章重点讲解 JMP 的 DOE 实验设计,包括:定制设计、筛选设计、完全析因设计、响应面设计、田口设计,补充介绍正交设计;第 5 章结合第 4 章实验设计,注重从多角度讲解方差分析应用;第 6 章重点介绍回归分析的多变量筛选、曲线直线化回归、共线性诊断、非线性回归、Logistic 回归及相关分析;第 7~9 章介绍聚类分析、主成分分析和因子分析。

本书可作为高校研究生、本科生的统计学教材,尤其适合非数理统计专业的研究人员作为数据分析的参考资料。

本书封面贴有清华大学出版社防伪标签,无标签者不得销售。

版权所有,侵权必究。侵权举报电话:010-62782989 13701121933

图书在版编目(CIP)数据

SAS 统计分析及应用与 JMP 实验设计/黄燕编著. —2 版. —北京:清华大学出版社,2017

(全国普通高等院校信息管理与信息系统专业规划教材)

ISBN 978-7-302-45055-9

I. ①S… II. ①黄… III. ①统计分析—应用软件—高等学校—教材 IV. ①C819

中国版本图书馆 CIP 数据核字(2016)第 218547 号

责任编辑:白立军 薛 阳

封面设计:常雪影

责任校对:焦丽丽

责任印制:沈 露

出版发行:清华大学出版社

网 址: <http://www.tup.com.cn>, <http://www.wqbook.com>

地 址:北京清华大学学研大厦 A 座 邮 编:100084

社 总 机:010-62770175 邮 购:010-62786544

投稿与读者服务:010-62776969, c-service@tup.tsinghua.edu.cn

质 量 反 馈:010-62772015, zhiliang@tup.tsinghua.edu.cn

课 件 下 载: <http://www.tup.com.cn>, 010-62795954

印 装 者:北京国马印刷厂

经 销:全国新华书店

开 本:185mm×260mm 印 张:28.75 字 数:679 千字

版 次:2006 年 1 月第 1 版 2017 年 1 月第 2 版 印 次:2017 年 1 月第 1 次印刷

印 数:1~1500

定 价:59.50 元

产品编号:069635-01

前言

随着科学技术的进步及信息技术的发展,计算机已成为人们科研工作的重要工具。科研工作所获取的各种数据都会借助各种计算机软件帮助研究者获得最终结果,统计分析软件是重要的工具和手段。SAS 软件自开发为人们使用以来,在统计分析方面一直处于权威领先地位。近几年,SAS 公司又开发出了一款小巧的统计分析软件 JMP,它兼有 SAS 的一些统计功能,同时又增加了 DOE 实验设计模块,中英文切换方便,中文输出统计结果一目了然,给研究者带来许多便利,其图形功能也很强大。

本书针对非数理统计专业的研究者缺乏系统的多元统计知识,时间紧、科研任务重,且需要在短时间内掌握所需要的统计分析方法并能应用到自己的科研数据分析中,编写了此教材。目标是使读者能快速掌握常用的多元统计分析方法,读懂输出结果,实用性强。

本书主要介绍 SAS 软件包的多元统计分析方法与 JMP 的 DOE 实验设计及相关统计分析方法。全书重点在于用实例讲解 SAS 与 JMP 软件包中的各种常用统计分析方法,并详细解释输出结果的统计学意义,比较两个软件相同统计方法输出内容的异同及优劣,方便读者从多角度应用。从实际应用出发,增加了 JMP 软件包中的 DOE 实验设计内容。

全书共 9 章,第 1 章主要介绍 SAS 与 JMP 软件基本操作;第 2 章主要介绍常用描述统计量、常用数据分布及参数检验概念;第 3 章讲述两个软件在 t 检验与非参数检验上的应用,并比较两个软件的优劣;第 4 章重点讲解 JMP 的 DOE 实验设计,包括:定制设计、筛选设计、完全析因设计、响应面设计、田口设计,补充介绍正交设计;第 5 章结合第 4 章的实验设计方法,从多角度讲解方差分析应用,尤其详细介绍了 SAS 软件在多因素且存在交互作用的方差分析上的优势及结果讨论与判断,同时介绍协变量在协方差分析中的重要作用;第 6 章重点介绍曲线直线化回归、回归模型的多变量筛选、多变量的共线性诊断、非线性回归、Logistic 回归,多方面介绍回归分析方法及相关分析,比较说明两种软件中 SAS 软件在回归分析中的优势;第 7~9 章介绍聚类分析、主成分分析和因子分析,比较两软件各自的优劣。

书中每种统计分析方法都以真实的科研数据为实例,从实验设计方法到数据整理、数据分析,详细解释各种统计分析方法的选项应用、输出内容的统计学意义,统计结果的判断依据,最优回归模型的筛选准则等,每种统计分析方法都运用 SAS 和 JMP 两种软件进行分析,比较两者的异同及优势,使学习者能够举一反三,学会选用科学、合理、简单方便的工具和方法,以判断并获取科学的、有价值的统计分析结果,尤其学会运用统计方法和思路。

由于 JMP 软件的输出是表格形式,但又不同于一般正文中的表格,为了与软件的输出保持一致,这类表格保留原样,不做处理。

本书可作为高校研究生、本科生的统计学教材,也适合非数理统计专业的研究人员作为数据分析的参考资料。

作 者

2016 年 8 月



第 1 章 SAS 及 JMP 系统概述	1
1.1 SAS 及 JMP 系统简介	1
1.2 JMP 系统界面操作	2
1.2.1 JMP 安装、启动与退出	2
1.2.2 JMP 系统设置及中英文切换	2
1.2.3 JMP 与 SAS 数据文件管理	4
1.2.4 JMP 功能简介	7
1.3 SAS 系统概述及界面操作指南	9
1.3.1 SAS 的启动与退出	10
1.3.2 修改 SAS 系统的 SASV8.CFG	11
1.3.3 SAS 系统主要窗口简介	12
1.4 SAS 数据集及 SAS 的文件管理	15
1.4.1 SAS 数据集、逻辑库及程序文件	15
1.4.2 建立并调用 SAS 数据集	17
1.4.3 导出 SAS 数据集为 Excel 文件	22
1.4.4 数据集排序	24
1.5 SAS 的数据步与过程步简介	24
1.5.1 SAS 程序书写规范和运行方法	25
1.5.2 利用 DATA 步对已有数据集扩增新变量	26
1.5.3 DATA 步中常用函数	28
1.5.4 SAS 过程步及常用语句简介	31
第 2 章 常用统计学概念简介	36
2.1 统计学概述	36
2.1.1 统计学史概述	36
2.1.2 描述统计简介	36
2.2 统计量的抽样分布	38
2.2.1 概率和随机变量	39
2.2.2 二项分布	40
2.2.3 泊松分布	43
2.2.4 正态分布	44
2.2.5 t 分布	46
2.2.6 F 分布	47

2.2.7 χ^2 分布	48
2.3 统计假设检验	49
2.3.1 统计假设检验简介	49
2.3.2 t 检验	52
2.3.3 F 检验	52
2.3.4 χ^2 检验	52
习题	53
第 3 章 t 检验与非参数检验	54
3.1 t 检验及应用	54
3.1.1 t 检验适用条件及判断准则	54
3.1.2 UNIVARIATE 过程及正态性检验	56
3.1.3 配对样本均值比较及正态性检验应用实例	58
3.1.4 TTEST 过程及两组独立样本均值比较	68
3.1.5 SAS 与 JMP 两种软件 t 检验方法总结	79
3.2 非参数检验及应用	79
3.2.1 Kruskal-Wallis 检验	80
3.2.2 NPAR1WAY 过程(非参数检验过程)	81
3.2.3 NPAR1WAY 过程举例说明	82
3.2.4 利用 JMP 进行非参数检验	90
3.3 FREQ 过程及应用实例(频数分析)	93
3.3.1 卡方检验的基本思想	93
3.3.2 FREQ 过程	96
3.3.3 FREQ 过程及 JMP 频数分析应用实例	98
习题	106
第 4 章 JMP DOE 实验设计	109
4.1 完全析因设计	109
4.1.1 单因素多水平完全析因设计	110
4.1.2 多因素多水平完全析因设计	111
4.2 定制设计	114
4.3 筛选设计	118
4.4 响应面设计	122
4.5 混料实验设计	125
4.6 田口设计	129
4.6.1 田口设计简介	129
4.6.2 田口设计实例说明	129
4.7 正交设计	135
4.7.1 正交设计的概念和正交表的初步认识	135

4.7.2 含交互作用的正交设计	137
4.7.3 正交设计实例说明	139
习题	146
第 5 章 方差分析	148
5.1 方差分析简介	148
5.1.1 方差分析基本概念	148
5.1.2 方差的同质性检验	151
5.1.3 方差分析的基本假定和数据转换	153
5.2 ANOVA 过程和 GLM 过程简介	155
5.2.1 ANOVA 过程	155
5.2.2 GLM 过程	159
5.3 完全随机设计方差分析	168
5.3.1 完全随机设计——单因素 K 水平多重比较	168
5.3.2 完全随机设计——多因素 K 水平多重比较	179
5.4 析因设计及 CONTRAST 语句使用说明	207
5.4.1 CONTRAST 语句说明	207
5.4.2 实例说明	209
5.5 二次响应面设计的方差分析	219
5.6 正交设计实例说明	238
5.7 拉丁方设计及其统计分析	247
5.7.1 拉丁方设计及其线性模型	247
5.7.2 拉丁方设计应用实例	248
5.8 协方差分析	259
5.8.1 协方差分析简介	259
5.8.2 协方差分析的数学模型及基本假定	261
5.8.3 协方差分析应用举例	262
习题	277
第 6 章 相关与回归分析	282
6.1 相关与回归分析概述	282
6.1.1 相关与回归分析的概念与区别	282
6.1.2 直线回归与相关分析	284
6.2 相关、回归过程说明	286
6.2.1 CORR 相关过程	286
6.2.2 PLOT 过程	287
6.2.3 REG 回归过程	289
6.2.4 回归模型的共线性诊断及强影响点诊断	294
6.3 线性及曲线直线化回归模型	295

6.3.1	简单线性回归方程	295
6.3.2	线性回归方程应用举例	297
6.3.3	二次项直线化表示的回归方程	300
6.3.4	用 JMP 进行模型拟合	307
6.4	可直线化的曲线拟合	314
6.4.1	曲线拟合说明	314
6.4.2	曲线直线化应用举例	315
6.4.3	确定最佳模型的标准	320
6.5	RSREG 过程	334
6.5.1	RSREG 过程说明	334
6.5.2	二次响应曲面 RSREG 过程举例说明	337
6.6	非线性回归过程	359
6.6.1	NLIN 过程简介	359
6.6.2	NLIN 应用举例	361
6.7	Logistic 回归	371
6.7.1	Logistic 回归概述	371
6.7.2	Logistic 回归过程	372
6.7.3	用 SAS 进行 Logistic 回归应用实例	374
6.7.4	用 JMP 进行 Logistic 回归应用实例	379
	习题	381
第 7 章	聚类分析	386
7.1	聚类分析概述	386
7.2	TREE 过程	387
7.3	VARCLUS 过程及应用举例(对变量聚类)	388
7.3.1	用 VARCLUS 过程实现变量聚类分析	388
7.3.2	VARCLUS 过程应用实例	389
7.4	CLUSTER 过程及实例分析(对样本聚类)	397
7.4.1	CLUSTER 过程	397
7.4.2	CLUSTER 过程应用实例	399
7.5	JMP 的聚类分析	403
	习题	404
第 8 章	主成分分析	408
8.1	主成分的几何意义及数学模型	408
8.2	PRINCOMP 过程	412
8.3	主成分分析应用举例	414
	习题	422

第 9 章 因子分析.....	425
9.1 因子分析基础	425
9.2 PROC FACTOR 过程.....	427
9.3 因子分析应用举例	431
9.4 主成分与因子分析综合实例说明	434
9.5 用 JMP 进行主成分和因子分析.....	440
习题.....	444
参考文献.....	447

第1章 SAS及JMP系统概述

本章学习目标

- 了解 SAS 软件与 JMP 软件的特点及功能。
- 学习并掌握 JMP 软件的系统设置、基本功能及与 SAS 软件间的数据格式转换。
- 学习并掌握 SAS 软件的基本功能、SAS 数据集的建立、导入、导出。
- 介绍 SAS 软件的 Data 步与 Proc 过程步基本语法规则。

通过本章学习,初步了解 JMP 软件与 SAS 软件的不同特点与相互间的数据转换方式,学会利用 Excel 及 JMP 建立 JMP 和 SAS 数据文件,为以后的数据分析打好基础。

1.1 SAS 及 JMP 系统简介

SAS(Statistical Analysis System)是国际著名的、权威的统计分析软件之一。该系统软件于 1966 年开始研制,1976 年由美国 SAS 公司实现商业化,1985 年推出 SAS/PC 版本,1987 年推出 DOS 系统下的 6.03 版,目前广泛应用的是 Windows 系统支持的 8.2A 或 10.0 版,其核心的统计分析算法模块于 20 世纪 90 年代已成型,目前流行的 Windows 版在操作界面、操作窗口、菜单上做了一些改进,同时增加了与网络页面交互的相关功能以及数据挖掘功能,但 SAS 软件系统中的多元统计分析还是各功能的核心算法。本书仍采用 Windows/SAS 8.2 版,它可在现有 Windows 操作系统上正常运行。

目前,数据分析对人类进步有着更重要的意义。一方面,随着科学技术的深入发展,科学研究所提取的大量实验数据需要进行科学、合理、高效的数据分析,从数据分析结果中获得重要的发展规律,进而推动科技进步,为人类服务,如人们在医学上研究出更好的新药以抵抗疾病,使人们获得健康;另一方面,计算机海量数据的出现,使人们越来越重视对前期历史数据的分析与挖掘,通过对原有数据资料的统计分析,进一步论证新方案的可行性。数据分析已成为各行各业工作中的重要手段之一。

当前流行的统计分析软件很多,各有各的优势,而 SAS 系统软件自商业化以来的三十多年,一直位于统计分析软件的权威地位,其优势在于它的统计算法的优秀性及权威性,且其软件功能的强大也一直处于领先地位。目前,国际上著名的学术期刊 *Nature*、*Science* 收录的学术论文中的数据分析结果要求必须使用 SAS 软件进行统计分析。这也是本书选择 SAS 软件的理由。

SAS 系统软件是模块式结构,最常用的三个模块分别是 SAS/BASE(基础)、SAS/STAT(统计)和 SAS/GRAPH(图形),另外还有 SAS/ETS(预测)、SAS/IML(矩阵运算)和 SAS/Enterprise Miner(数据挖掘)等二十多个模块。这些模块可单独购买并使用,也可互相配合起来使用。

本书仅就前两个模块的常用功能,从实例角度出发,讲述常用多元统计分析方法在 SAS 软件系统中的使用方法与技巧。通过科研实例贯穿讲述统计理论知识、统计软件功能、各种常用统计分析输出结果的统计学意义及选择判断方法和技巧,使大多数非数理统计专业的科研人员能够通过本书的学习,按书中实例做到举一反三,模仿应用到自身的研究分析中。

SAS/BASE 和 SAS/STAT 是 SAS 系统的核心平台和精华,是用来解决实际问题的主要模块。

SAS 于 1989 年开发了 JMP 软件,JMP 是 SAS 推出的一种交互式可视化统计分析软件系列之一,JMP 的算法源于 SAS,强调以统计方法的实际应用为导向,JMP 10.0 以上的版本的优势更突出了其 DOE 实验设计的功能,其可视性、互动性、综合性和可扩展性的特点,所提供的统计图形丰富多彩,使用更方便,尤其适合非统计专业背景的数据分析人员。

JMP 软件与 SAS 软件使用的数据集可以方便转换。目前的科研面临许多复杂对象,简单的实验设计方法很难解决,而复杂的实验设计对于许多非统计专业背景的科研人员来讲要想做好设计也很难。因此实验设计是许多研究者面临的重要难题之一。利用 JMP 的 DOE 功能会方便解决这一难题。如能将两款软件结合起来应用,各取所长,会给科研人员的研究带来很大的好处。

出于此目的,从科研工作的实际应用出发,本书将 JMP 与 SAS 软件结合起来介绍,各取所长,定会给实际科研应用者带来很大便利。

1.2 JMP 系统界面操作

1.2.1 JMP 安装、启动与退出

安装: JMP 软件的安装很简单,直接单击 JMP 系统软件中的  setup,按提示要求进行安装即可。

启动: 安装后会在桌面上有一个快捷方式图标 ,使用时直接双击即可启动 JMP 软件。打开时会有一些提示窗口,在右下角单击“关闭”按钮将其关闭。留下的主窗口如图 1-1 所示。

退出: 单击“文件”→“退出”命令;或直接单击窗口右上角的  按钮可退出或关闭软件。

1.2.2 JMP 系统设置及中英文切换

JMP 的功能很多,每次执行“分析”中的菜单命令后均能输出许多统计分析结果。用户可以根据自己的需求,通过“文件”→“首选项”命令(见图 1-2),打开如图 1-3 所示对话框,对常用输出统计信息项进行选择设置。

1. JMP 首选项设置

“首选项”对话框中提供了许多内容,如图 1-3 所示。在左侧“首选项组”列中单击“平台”,中间“平台”列中提供了各种分析方法名称,选择不同分析方法,右侧“选项”列中会显示



图 1-1 JMP 主窗口



图 1-2 JMP 功能设置菜单——首选项



图 1-3 JMP 首选项——系统设置

对应的该分析方法输出的统计分析信息。将一些常用项勾选上,以后在执行对应分析时会直接输出这些内容,否则需要单击输出窗口左上角的红色倒三角 , 从下拉菜单中一项一项选择,比较麻烦。

“首选项”的设置,方便了用户对统计输出内容的显示方式。如图 1-3 所示是单因子方差分析中的输出选项。打勾项会在进行方差分析时直接显示在输出窗口,而没有打勾的选项,如果想显示输出,需要用户单击输出窗口左上方红色倒三角 , 在弹出的下拉菜单中一项一项选择,这种方式虽然较灵活,但相比较“首选项”一次设置完成要麻烦很多。但对于不常用的选项,在“首选项”中不必都选上,否则输出窗口中输出的统计信息太多,用户看着较烦杂。即“首选项”中的设置方便用户自己设置常用统计信息的输出。

2. JMP 系统中英文切换

JMP 软件对中文的兼容性很好,但对于许多统计专用词语,翻译成中文后不是很合适,不利于研究人员发表英文 SCI 论文。JMP 可通过菜单“文件”→“首选项”中“Windows 特定”(见图 1-4)页面上方“显示语言”中的下拉菜单进行方便切换。当切换为“English”,并单击下方的“确定”按钮后,软件显示的菜单、输出信息均转换为英文。

1.2.3 JMP 与 SAS 数据文件管理

进行数据分析的重要前提是将实验数据整理并建成统计分析软件认可的数据格式,也称为数据整理或数据初始化管理。计算机已是人们工作、学习、生活中不可缺少的工具,而人们最熟悉的 Office 软件,其 Excel 工作表是最简单方便的整理数据工具。

1. 用 Excel 工作表整理数据

(1) 将原始数据文件按要求录入 Excel 工作表,变量名放在 Excel 工作表的第 1 行,第

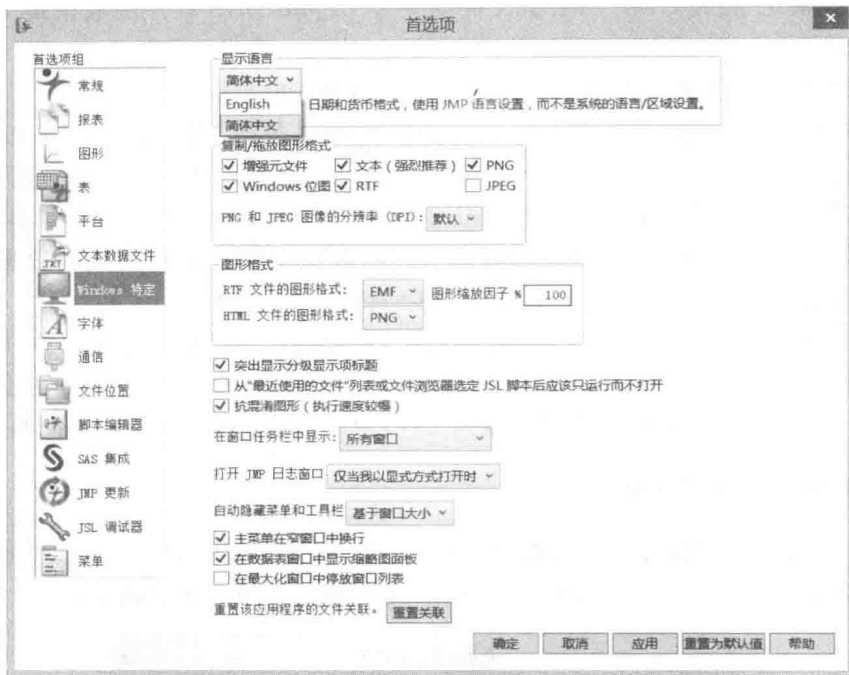


图 1-4 首选项-Windows 特定——JMP 中英文切换

2 行开始为数据。保存的 Excel 文件名最好是英文字符串或英文加数字的字符串。由于 SAS 软件对中文的兼容性较差,为了保证各软件间数据格式转换正确无误,建议变量名为小于 8 字符长的英文或英文加数字的字符串。

(2) 打开 JMP 软件,单击菜单“文件”→“打开”,打开磁盘中已经有的 Excel 文件,然后单击菜单“文件”→“另存为”,见图 1-5,从“保存类型”中选择要保存的文件类型,在文件名



图 1-5 另存为不同类型数据文件

中起名,单击“确定”按钮(JMP 的数据文件类型是 *.jmp,SAS 软件的数据集文件类型为 *.sas7bdat)。

保存好相应的数据文件类型后,以后在进行统计分析时可以直接调用。

JMP 软件的数据管理与 Excel 工作表能够很好地兼容,当用户将要分析的数据用 Excel 表整理好后,可通过 JMP 菜单“文件”→“打开”直接打开 Excel 数据表文件,一个 Excel 工作簿中若有多个工作表 Sheet,JMP 打开此文件时可一次性同时打开多个数据表,然后根据需要将每个工作表数据分别存为不同的数据文件或不同类型的数据文件。如图 1-5 所示,“保存类型”下拉列表显示可存为多种类型数据文件。

(3) 当原始数据需要大的修改和调整时,用户最好利用自己最熟悉的软件,如 Excel 工作表进行操作,调整修改后保存好 Excel 工作表,再利用 JMP 软件打开 Excel 工作表并另存为需要的 *.jmp 或 *.sas7bdat 类型的数据文件,为以后进行统计分析时方便调用。

2. 用 JMP 直接建立数据文件

(1) 单击“文件”→“新建”→“数据表”弹出一个空白表,如图 1-6 所示。可直接双击空白列头,列名变为第 2 列、第 3 列,再双击“第 2 列”可输入变量名定义新变量。



图 1-6 建立新列——新变量

(2) 在列名处单击右键,弹出图 1-6 中的列菜单,选择“列信息”命令弹出如图 1-7 所示窗口,在此可以定义新变量名、变量类型等。

(3) 定义完变量名后可直接输入数据。也可随时调出“列信息”窗口更改变量名和变量类型。这方面比 SAS 灵活方便,SAS 软件对已建好的数据集文件保护性较强,打开一个已存在的 SAS 数据集,它处于浏览模式,若想修改数据集中的数据,需选择 Edit 编辑模式后

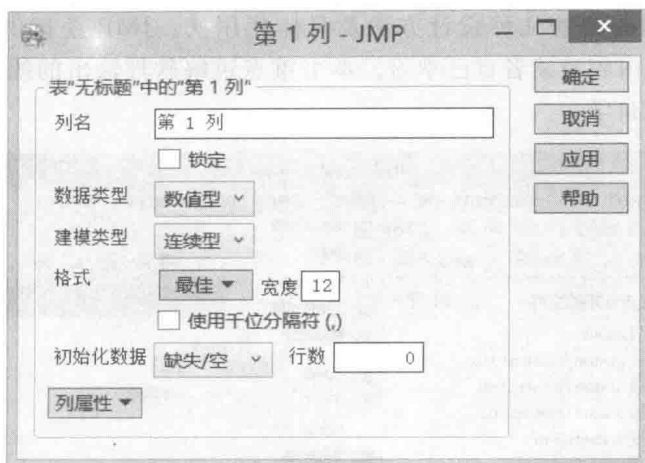


图 1-7 定义新变量名、变量类型

才能修改,变量名及变量类型不能修改。

现在计算机各种软件功能丰富,各有所长,利用一个软件完成全部或大部分工作不是个好方法。为了提高工作效率及质量,应发挥自己对各种软件熟悉的操作及各种软件功能特点及操作简单方便,多个软件同时使用,大大提高工作效率。

数据分析的第一步是对原始数据进行整理,并保存为所要使用的统计软件所需要的数据文件格式,然后调用统计分析软件相应的分析功能进行数据分析。

1.2.4 JMP 功能简介

JMP 是一款交互性很好的软件,而且自带了初级教程,用户完全可以通过自学方式学会软件的常规操作。其主要功能之一为 DOE 实验设计,见图 1-8。DOE 包含 11 种设计方



图 1-8 “实验设计”菜单