

Windows SAS 6.12 & 8.0



实用 统计

分析教程

胡良平 编著

军事医学科学出版社

Windows SAS 6.12 & 8.0

实用统计分析教程

胡良平 编著



北方工业大学图书馆

军事医学科学出版社

2001·北京

内容提要

本书较详细地介绍了 6.12 和 8.0 版 WINDOWS SAS 系统中常用窗口、各菜单条和图标等的含义和功能；介绍了 3 个非编程模块（即 SAS/ASSIST、SAS/INSIGHT 和 SAS/ANALYST APPLICATION）的功能和用法；用较多的篇幅介绍了各种实验设计类型、设计类型的识别技巧及常用实验设计（如多因素析因设计、分式析因设计、正交设计、反应曲面设计、混料设计、样本大小的估计等）的具体实现；用编程法和 3 个非编程模块实现各种统计计算（包括定量、定性资料的统计分析；直线回归、曲线拟合、多元线性回归和多元 Logistic 回归分析；生存分析和其他多元统计分析），并对输出结果给出了详细的解释；介绍了各种统计图形的绘制方法和各种报表的生成方法；还特别介绍了一些 DOS 环境下的 SAS 系统尚未涉及的计算方法、应用技巧和 SAS 8.0 版中新增添的内容。本书不仅有利于方便、快捷地学习 WINDOWS 版 SAS 软件的使用方法，更有利于灵活、正确地运用统计学知识解决各种复杂的统计学问题。

本书可做研究生、本科生的统计学教材，高等院校和科研机构的教师、科研人员、生物医学工作者、管理者等学习和应用统计方法的参考书；还可做为用 SAS 软件解决实验设计和统计分析方面问题的实用手册。

* * *

图书在版编目 (CIP) 数据

Windows SAS 6.12 & 8.0 实用统计分析教程/胡良平编著.

-北京: 军事医学科学出版社, 2001.1

ISBN 7-80121-275-4

I. W... II. 胡... III. 窗口软件, Windows SAS IV. TP316.7

中国版本图书馆 CIP 数据核字 (2000) 第 76990 号

* *

军事医学科学出版社出版

(北京市太平路 27 号 邮政编码: 100850)

新华书店总店北京发行所发行

潮河印刷厂印刷

*

开本: 787 mm×1092 mm 1/16 印张: 39 字数: 969 千字

2001 年 3 月第 1 版 2001 年 3 月第 1 次印刷

印数: 1-5000 册 定价: 52.00 元

(购买本社图书, 凡有缺、损、倒、脱页者, 本社发行部负责调换)

前言

无论是在企业还是在科研机构，如果没有变革，就不需要高水平的管理人才；在任何研究领域内，如果没有变异，也就不需要统计学。在生物医学研究领域内，统计学的作用和地位显得尤为突出。因为它的研究对象是具有极大变异性的生物，其研究目的是探索生物医学领域内未知的事物，而反映未知事物特征的指标常带有变异性。由于变异性的普遍存在，调查、实验或观测的结果就必然带有随机性。为了获得带有规律性的结果，人们就不得不进行大量的重复实验或观测。然而总体的容量很大，甚至是无限的，研究者的时间、精力、人力和物力却是十分有限的。客观实际迫使研究者不得不从总体中进行抽样研究，以期通过样本所提供的信息去推论总体的规律性。我们试想：一个在设计上漏洞百出、在施工上偷工减料的高层建筑，其后果将不堪设想；同理，一项科研课题，从统计研究设计、统计资料的收集和描述、统计分析方法的合理选择与实现，到计算结果的解释与应用，每一个环节，都涉及到正确与错误运用统计学的问题。

简而言之，在科学研究中必须正确运用统计学，这是科研工作的科学性所决定的。搞科研，首先必须尊重科学。借助统计学这个有用的工具，可以去探索未知事物，揭示和阐明客观事物变化的规律性。

1. 统计学的意义与作用

早在 16、17 世纪，英国著名哲学家培根说得好，“跛足而不迷路，能赶过虽健步如飞但误入歧途的人”。可见，研究方法对于科学研究是何等的重要。20 世纪 30~40 年代，医学界曾用金的化合物作为结核病的特效药，直到 40~50 年代，有统计学头脑的人运用随机对照的临床试验方法，才迅速作出了基本上是否定的结论。与此同时，运用科学的实验设计方法，只用了几个月的时间，就肯定了链霉素、对氨柳酸和异烟肼对结核病的疗效，并证实联合使用其中两种比单用一种更好；在第二次世界大战期间，为了检验军火的质量，人类创造出经济、高效的序贯试验设计方法，此法自 20 世纪 50 年代以后被广泛应用于医学领域，成为比较新、旧药疗效好坏的有效试验方法；自 20 世纪初开始，全球就有人陆续不断地研究吸烟与健康之关系。运用正确的统计分析方法，对大量的调查和实验数据进行科学分析，得出“吸烟有害于健康”的结论；美国于 1958 年开始在外科手术中采用麻醉剂氟烷，到 1962 年突然掀起一场风波——麻醉剂氟烷有严重的副作用，导致部分患者病情恶化、发热、肝脏大片坏死而死亡。后来利用 85 万例手术病例，经统计学方法校正了各种非

处理因素的干扰和影响之后，得出与氟烷、喷妥散、环丙烷、乙迷和其他麻醉剂相应的死亡率分别为 2.1%、2.0%、2.6%、2.0%和 2.5%。显然，完全是靠统计学平息了关于麻醉剂氟烷的这场风波；起始于 1942 年英国的一个关于马铃薯的肥料试验的正交试验设计，到 70 年代末由我国数学家王元和方开泰教授首创的均匀设计，为解决生产和科研中的多因素多水平的试验研究问题，作出了重大的贡献；各种多元统计分析方法的问世，为人类认识世界、改造世界奠定了良好的基础。限于篇幅，例子不胜枚举。

2. 统计学所遭受的偏见

大量事实表明：统计学在工农业生产、武器装备的研制和质量检验、医药卫生和流行病学等研究领域内，都发挥了很好的作用。然而，人们在运用统计学过程中，有意和无意中犯了很多错误。有些人始终把统计学当作自己所从事的科研工作的“陪衬物”，当作申报成果和发表论文的“敲门砖”；一些业务主管部门尚未认识到统计学对于科研工作的真正价值，总以为统计学不能产生巨大的直接经济效益，是可有可无的一门“美容”学科。因此，在科研课题的开题论证、中期检查考核、结束时的验收评审中，忽视统计学运用得是否正确；在名目繁多、数量与日俱增的杂志（尤其是医学期刊）中，许多有严重统计学错误的论文也可顺利地通过审稿而予以发表。

3. 统计学所面临的现状

在科学研究中对统计学知识运用得好坏，最突出的标志就是要看有无统计研究设计以及研究设计的质量如何。事实上，很多科研课题缺乏统计研究设计方案或设计方案漏洞百出，经不起推敲；有些试验研究缺乏必要的预试验，一旦试验中出现了异常现象，将措手不及；有些因收集数据的方法不当，而不可避免地导致误用统计分析方法；还有些不严格按实验设计的基本原则实施，用单因素设计取代多因素设计，无法研究因素之间的交互作用；误用 t 检验分析一切定量资料；误用 χ^2 检验分析一切定性资料；误用直线回归方程描述有明显曲线变化趋势的资料；用一元统计分析方法取代多元统计分析方法等等。

上述误用统计学的现象，在我国的医学期刊和科研成果的档案材料中出现的频率相当高，有些结论是丝毫站不住脚的。我们应当清醒地意识到：统计学本身不能创造规律，但当它被正确使用时，可以帮助人们揭示数据中已经存在的规律性，从而有助于医学科研的健康发展。

4. 统计学所面临的机遇和挑战

随着科学技术的飞速发展，无论是生产实践还是科学研究，现代化的仪器设备，使人类有能力探索 and 解决更加复杂的问题。同时，也给统计学提出了更加复杂的研究课题，使它面临着极大的发展机遇和严峻的挑战。多个边缘学科的合作研究，多中心临床试验，大

规模多层次的调查资料的出现等等，促进了 META 分析和多水平统计模型的诞生和发展；多因素多水平的试验研究问题的提出，促进了重复测量设计、分式析因设计、反应曲面设计、混料设计、均匀设计等多因素试验设计方法的发展；包含固定效应因素和随机效应因素的复杂课题的研究，促进了一般与广义、线性与非线性混合模型的研究和不断完善。许多大型通用统计软件包的出现，又使得复杂的试验设计和统计分析成为可能。

然而，由于生物医学研究的复杂性，很多资料偏离统计模型的假设条件甚远，通常很难找到合适的变量变换方法。这就要求统计学工作者要进一步加强理论研究，开发出适用于分析多因素多指标的非参数统计方法或半参数统计方法，编出相应的统计分析软件，以提高统计学自身解决和驾驭复杂问题的能力。不仅如此，随着生物工程技术的发展，随着人类基因组计划的完成和后基因组计划的实施，将提出一些更深层次的统计学问题，如：不同人种各种基因的分布规律、各种基因的功能及其与疾病之间的联系、基因突变的原因和规律等的研究，这些问题仅靠传统的数学和统计学方法是难以解决的，统计学必须在理论和方法上有较大的创新和发展，这是生物信息时代发出的呼唤。

5. 提高医学科研质量的对策与建议

由于变异性在医学科研中表现得尤为突出，因此在这一领域内，运用统计学的场合比起其他研究领域可能要多得多。有远见卓识的学术和科研管理者，不应把人们对统计学的偏见代入 21 世纪。我们应该清醒地认识到：统计学是任何科研工作密不可分的重要组成部分，是保证科研工作的科学性、严谨性、经济、高效的基石和指南针；应该采取行之有效的措施，尽可能减少误用统计学的机会，不应将低档的统计学应用水平，带入 21 世纪的医学研究领域中去。为了提高医学科研的质量，笔者提出如下的对策和建议：

第一，政府和科研机构中分管业务的领导，应当努力增强自身的统计学意识。制定相应的政策和法规，重视医学科研工作的科学性。第二，有幸参加评审“科研成果”和“学术论文”等的专家们，应当努力提高自身的统计学素质，以便有能力识别常见的统计学错误。第三，要努力提高医学院校和科研院所统计学的教学质量，加强对有关人员统计培训的力度。第四，学术期刊的管理部门，应该清醒地意识到：现在的“稿约”已经过时了，因为它们对稿件的统计学质量几乎没提任何实质性的要求。认真修改“稿约”，是提高稿件的统计学质量的最关键的举措。第五，应大力提倡科研协作。医学统计学虽是一门应用性学科，但它以概率论和数理统计为理论基础，要想全面掌握它，尤其是正确、合理地运用它并非易事。一般性的问题，医学科研工作者可以自行解决，至于复杂的统计学理论和技术，专业性较强，不必人人掌握，需要时最好主动与统计学工作者开展科研协作，这样，不仅可以获得事半功倍的效果，而且可以大大地提高医学科研工作的科学性和严谨性。

6. 尾声

笔者虽感慨万千，但由于水平和“能量”有限，无力解决上述提及的诸方面的问题，真心希望所提出的问题能够引起广大读者的共鸣，从而为之“呐喊”和“效力”，使统计学在 21 世纪中为促进人类的文明和进步发挥更大的作用。笔者从事研究生医学统计学教学、科研统计咨询和统计研究已 16 年，现将自己所积累的有关学习和使用统计学的经验和体会汇集在一起写成本书，以报答党和人民对笔者多年的教育和培养之恩。本书共分 10 篇 40 章，较详细地介绍了 6.12 和 8.0 版 WINDOWS SAS 系统中常用窗口、各菜单条和图标等的含义和功能；介绍了 3 个非编程模块（即 SAS/ASSIST、SAS/INSIGHT 和 SAS/ANALYST APPLICATION）的功能和用法；用较多的篇幅介绍了各种实验设计类型、设计类型的识别技巧以及常用实验设计（如多因素析因设计、正交设计、分式析因设计、反应曲面设计、混料设计、样本大小的估计等）的具体实现；用编程法和 3 个非编程模块实现各种统计计算（包括定量、定性资料的统计分析；直线回归、曲线拟合、多元线性回归和多元 Logistic 回归分析；生存分析和其他多元统计分析），并对输出结果给出了详细的解释；介绍了各种统计图形的绘制方法和各种报表的生成方法；还特别介绍了一些 DOS 环境下的 SAS 系统尚未涉及的计算方法、应用技巧和 SAS 8.0 版中新增添的内容。本书的突出特点是：用编程法和非编程法实现各种统计分析，彻底淘汰了手工计算方法；注重从问题的原形入手去考虑统计分析方法的合理选择问题，使讲授方法更加贴近读者学习和运用统计知识的需要，易于化解其疑点和难点；通过“摆事实、讲道理”的方法展示人们在运用统计学的过程中可能遇到的各种误用和滥用统计学的实例，从“正反”两个方面引导人们努力学习和掌握统计学中的关键性问题，提高学习和运用统计学的有效性和准确性。本书不仅有利于方便、快捷地学习 WINDOWS 版 SAS 软件的使用方法，更有利于灵活、正确地运用统计学知识解决各种复杂的统计学问题。

最后，在本书即将问世之际，笔者非常真诚地感谢军事医学科学院研究生处和情报研究所多年来对笔者所从事工作的大力支持和帮助！感谢 SAS 研究所北京办事处曾给予笔者很多的技术支持和帮助！感谢家人对我的理解和支持！感谢我的研究生李子建对本书的校对、排版与编辑所付出的辛勤劳动！

由于笔者水平有限，缺点和错误在所难免，恳请广大读者批评指正！

胡良平

2001 年 1 月 15 日于北京

目 录

第 1 篇 预备知识及基本概念

第 1 章 绪论	(1)
1.1 引言	(1)
1.2 SAS 软件简介	(2)
第 2 章 SAS AWS 界面简介	(5)
2.1 引言	(5)
2.2 SAS AWS 窗口的组成	(5)
2.3 SAS AWS 窗口各部分的名称及作用	(6)
第 3 章 Globals 菜单中需要详细介绍的几个子菜单	(13)
3.1 引言	(13)
3.2 Globals 菜单中 Access 子菜单的内容及用法	(13)
3.3 Globals 菜单中 Manage 子菜单的内容及用法	(18)
3.4 Globals 菜单中 Present 子菜单的内容及用法	(21)
3.5 Globals 菜单中 Options 子菜单的内容及用法	(24)
3.6 Globals 菜单中 Accessories 子菜单的内容	(25)
第 4 章 用 SAS 需了解的几个基本概念	(26)
4.1 SAS 过程与 SAS 程序	(26)
4.2 SAS 数据文件与 SAS 数据集	(26)
4.3 SAS 程序的结构	(27)
4.4 SAS 程序的产生办法	(28)
4.5 SAS 程序的运行	(29)
4.6 SAS 数据库与库标记	(29)
第 5 章 SAS 数据集的种类	(31)
5.1 按 SAS 数据集在机器中保留的时间长短分	(31)
5.2 按 SAS 数据集名所包括的字数与特性分	(31)
第 6 章 创建 SAS 数据集的具体方法	(33)
6.1 通过程序编辑窗口 (PGM) 创建 SAS 数据集	(33)
6.2 通过引用某些 SAS 函数创建新 SAS 数据集	(33)

6.3	通过运行某些 SAS 过程步创建新 SAS 数据集	(33)
6.4	通过 VIEWTABLE 窗口创建 SAS 数据集	(34)
6.5	通过 SAS/ASSIST 提供的界面创建 SAS 数据集	(34)
6.6	通过 SAS/INSIGHT 提供的数据窗创建 SAS 数据集	(34)
6.7	通过“分析员应用”模块提供的数据窗创建 SAS 数据集	(34)
6.8	通过转换方式创建 SAS 数据集	(35)
第 7 章	帮助信息的查询与功能键的运用	(36)
7.1	引言	(36)
7.2	实例	(36)
7.3	功能键窗口及其用法	(36)

第 2 篇 三个常用非编程模块简介

第 8 章	SAS/ASSIST 模块的基本结构及功能	(38)
8.1	引言	(38)
8.2	SAS/ASSIST 模块的基本结构	(38)
8.3	SAS/ASSIST 模块的基本功能	(39)
8.4	SAS/ASSIST 模块中“数据管理”子窗口的概况	(40)
8.5	SAS/ASSIST 模块中“报表生成”子窗口的概况	(41)
8.6	SAS/ASSIST 模块中“绘图”子窗口的概况	(42)
8.7	SAS/ASSIST 模块中“统计分析”子窗口的概况	(43)
8.8	SAS/ASSIST 模块中“计划工具”子窗口的概况	(44)
8.9	SAS/ASSIST 模块中“结果管理”子窗口的概况	(45)
8.10	SAS/ASSIST 模块中“环境设置”子窗口的概况	(47)
8.11	SAS/ASSIST 模块中“任务索引”子窗口的概况	(47)
第 9 章	SAS/INSIGHT 模块的基本结构及功能	(49)
9.1	引言	(49)
9.2	进入 SAS/INSIGHT 模块的方法	(49)
9.3	进入 SAS/INSIGHT 模块后菜单条中的主要内容	(50)
第 10 章	SAS/AA 模块的基本结构与功能	(54)
10.1	引言	(54)
10.2	进入 SAS/AA 模块的方法	(54)
10.3	进入 SAS/AA 模块后菜单条中的主要内容	(55)

第3篇 数据管理与统计描述

第11章 数据管理	(61)
11.1 引言	(61)
11.2 录入数据与创建 SAS 数据集	(61)
11.3 SAS 数据集的编辑和浏览	(70)
11.4 不同类型数据文件与 SAS 数据集之间的相互转换	(75)
11.5 如何定义用户自己的输入与输出格式	(82)
11.6 如何查询一个指定的 SAS 数据集	(90)
11.7 如何将几个 SAS 数据集进行合并	(99)
11.8 如何将一个大 SAS 数据集拆分成几个小 SAS 数据集	(102)
11.9 如何比较两个 SAS 数据集的内容是否完全相同	(105)
第12章 统计描述	(108)
12.1 引言	(108)
12.2 计算常用统计量	(109)
12.3 编制常用统计表	(116)
12.4 绘制常用统计图	(123)

第4篇 定量资料的统计分析

第13章 定量资料假设检验方法合理选用的概念和基础	(152)
13.1 定量资料假设检验方法误用的原因和弊病	(152)
13.2 定量资料的结构与设计类型	(152)
13.3 定量资料假设检验方法的分类及其对资料的要求	(153)
13.4 正确分析定量资料的途径	(153)
13.5 参数检验法的种类及其理论依据	(154)
13.6 设计类型与定量资料假设检验方法的匹配	(154)
第14章 定量资料实验设计类型的辨析	(155)
14.1 引言	(155)
14.2 单因素 K 水平设计	(155)
14.3 配伍组设计	(159)
14.4 平衡不完全配伍组设计	(161)
14.5 拉丁方设计	(162)
14.6 交叉设计	(163)
14.7 析因设计	(165)

14.8	含区组因素的析因设计	(168)
14.9	正交设计	(169)
14.10	系统分组(或嵌套)设计	(170)
14.11	裂区(或分割)设计	(171)
14.12	重复测量设计	(173)
14.13	分式析因设计	(176)
14.14	通过对资料原形的转换去辨析其真正的设计类型	(178)
第15章	用编程法实现定量资料假设检验	(183)
15.1	引言	(183)
15.2	单因素K水平设计资料的统计分析	(183)
15.3	配伍组设计资料的统计分析	(197)
15.4	平衡不完全配伍组设计资料的统计分析	(201)
15.5	拉丁方设计资料的统计分析	(203)
15.6	交叉设计资料的统计分析	(206)
15.7	析因设计资料的统计分析	(207)
15.8	含区组因素的析因设计资料的统计分析	(210)
15.9	正交设计资料的统计分析	(213)
15.10	系统分组(或嵌套)设计资料的统计分析	(215)
15.11	裂区(或分割)设计资料的统计分析	(218)
15.12	重复测量设计资料的统计分析	(222)
第16章	用非编程法实现定量资料的假设检验	(242)
16.1	引言	(242)
16.2	单因素K水平设计资料的统计分析	(242)
16.3	配伍组设计资料的统计分析	(248)
16.4	平衡不完全配伍组设计资料的统计分析	(249)
16.5	析因设计资料的统计分析	(250)
16.6	系统分组(或嵌套)设计资料的统计分析	(253)
16.7	裂区(分割)设计资料的统计分析	(255)
16.8	重复测量设计资料的统计分析	(256)
第17章	用编程法与非编程法实现多元方差分析	(258)
17.1	引言	(258)
17.2	单组设计资料的多元方差分析	(258)
17.3	配对设计资料的多元方差分析	(259)
17.4	成组设计资料的多元方差分析	(261)
17.5	单因素K水平设计($K \geq 3$)资料的多元方差分析	(263)

17.6	配伍组设计资料的多元方差分析	(267)
17.7	析因设计资料的多元方差分析	(270)

第5篇 定性资料的统计分析

第18章	定性资料假设检验方法合理选用的概念和基础	(275)
18.1	定性资料假设检验方法误用的原因和弊病	(275)
18.2	定性资料的分类与统计分析方法的合理选用	(276)
第19章	2×2表资料的假设检验	(279)
19.1	用编程法实现2×2表资料的假设检验	(279)
19.2	用非编程法实现非配对设计的2×2表资料的假设检验	(293)
19.3	用非编程法实现配对设计的2×2表资料的假设检验	(301)
第20章	用编程法与非编程法实现R×C表资料的假设检验	(303)
20.1	用编程法实现R×C表资料的假设检验	(303)
20.2	用非编程法实现R×C表资料的假设检验	(323)
第21章	用编程法实现高维列联表资料的统计分析	(325)
21.1	高维列联表资料的分类及统计分析方法的合理选用	(325)
21.2	对三维列联表资料进行压缩的可行性与潜在危险的分析	(325)
21.3	三维列联表资料的加权 χ^2 (或 $MH\chi^2$)检验	(330)
21.4	高维列联表资料的对数线性模型分析	(333)
21.5	应变量为二值变量的高维列联表资料的多元Logistic回归分析	(336)
21.6	自变量为有序变量的logistic回归模型	(344)
21.7	应变量为有序变量的logistic回归模型	(347)
21.8	用CATMOD过程实现具有重复测量的定性资料的分析	(352)
21.9	条件logistic回归模型	(355)
21.10	利用logistic回归方程正确解释比数比	(361)
第22章	用非编程法实现高维列联表资料的统计分析	(363)
22.1	概述	(363)
22.2	用非编程法实现 $MH\chi^2$ 检验	(363)
22.3	用非编程法拟合多元Logistic回归模型	(365)

第6篇 相关、回归和协方差分析

第23章	回归分析的种类与简单回归分析	(371)
23.1	回归分析的任务和种类	(371)

23.2	直线回归与相关分析的概念和要点	(372)
23.3	用编程法实现直线相关与回归分析	(377)
23.4	用非编程法实现直线相关与回归分析	(382)
23.5	可直线化资料的曲线拟合	(387)
第24章	多元线性回归分析和Logistic回归分析	(400)
24.1	多元线性回归分析的概述	(400)
24.2	多元线性回归分析的应用	(402)
24.3	多元Logistic回归分析	(407)
第25章	一元与多元协方差分析	(413)
25.1	协方差分析的概述	(413)
25.2	协方差分析的应用	(416)

第7篇 生存分析

第26章	生存分析的基本概念	(426)
26.1	生存资料的特点	(426)
26.2	生存时间函数	(426)
26.3	生存分析方法的分类	(427)
第27章	生存资料的非参数统计分析	(428)
27.1	统计描述与非参数统计分析概述	(428)
27.2	用编程法实现统计分析	(429)
27.3	用非编程法实现统计分析	(434)
第28章	C O X模型回归分析	(442)
28.1	C O X模型回归分析概述	(442)
28.2	用编程法实现C O X模型回归分析	(442)
28.3	用非编程法实现COX模型回归分析	(452)
第29章	参数模型回归分析	(465)
29.1	参数模型回归分析概述	(465)
29.2	参数模型回归分析应用举例	(465)

第8篇 多元统计分析

第30章	主成分分析	(470)
30.1	主成分分析的概念与数据结构	(470)
30.2	主成分的表达式与性质	(471)

30.3	用编程法实现主成分分析	(472)
30.4	用非编程法实现主成分分析	(476)
30.5	合成资料的主成分分析	(479)
第31章	因子分析	(483)
31.1	基本概念	(483)
31.2	因子模型	(483)
31.3	因子分析的基本定理与任务	(484)
31.4	用编程法实现因子分析	(486)
第32章	对应分析	(491)
32.1	对应分析方法的概述	(491)
32.2	对应分析中的变量变换方法	(491)
32.3	用编程法实现对应分析	(491)
第33章	聚类分析	(497)
33.1	聚类分析方法的概述	(497)
33.2	用编程法实现变量聚类分析	(497)
33.3	用编程法实现样品聚类分析	(504)
33.4	用编程法实现大样本样品聚类分析	(509)
33.5	用编程法对需作样品聚类分析的资料进行预处理	(509)
33.6	用SAS/GRAPH模块绘制样品聚类图的SAS程序	(511)
第34章	判别分析	(516)
34.1	判别分析方法的概述	(516)
34.2	定量资料的逐步判别分析(考虑变量筛选)	(516)
34.3	用编程法实现一般判别分析(不考虑变量筛选)	(520)
第35章	典型相关分析	(528)
35.1	典型相关分析方法的概述	(528)
35.2	用编程法实现典型相关分析	(528)
35.3	用非编程法实现典型相关分析	(532)

第9篇 单因素与多因素实验设计

第36章	实验设计的基本概念与方法	(541)
36.1	实验设计的意义、要素、原则和原理	(541)
36.2	误差及其分类	(543)
36.3	受试对象随机分组	(543)
36.4	对照组的几种常见形式和应用场合	(549)

36.5	组间均衡性的考虑	(550)
36.6	样本含量的估计	(550)
第 37 章 用分析员应用模块估计样本大小		(551)
37.1	分析员应用模块能估计哪些情形下的样本大小	(551)
37.2	单组或配对设计情形下实现三种统计目的时样本大小的估计	(552)
37.3	配对设计情形下实现三种统计目的时样本大小的估计	(557)
37.4	成组设计情形下定量资料实现三种统计目的时样本大小的估计	(557)
37.5	单因素 K 水平设计 ($K \geq 3$) 定量资料均值假设检验时样本大小的估计	(562)
第 38 章 多因素实验设计方案的制定		(563)
38.1	多因素实验设计方案的概念	(563)
38.2	用 ADX 产生多因素实验设计方案的入门	(563)
38.3	用 ADX 产生二水平分式析因设计	(566)
38.4	用 ADX 产生正交数组设计	(571)
38.5	用 ADX 产生反应曲面设计	(572)
38.6	用 ADX 产生混料设计	(583)

第 10 篇 WINDOWS SAS 8.0 版介绍

第 39 章 SAS 8.0 界面简介		(589)
39.1	六个常用窗口的界面	(589)
39.2	六个常用窗口功能简介	(591)
39.3	与 SAS 6.12 明显不同的菜单条中的内容简介	(591)
39.4	与 SAS 6.12 明显不同的工具栏中的内容简介	(594)
第 40 章 SAS 8.0 中有哪些新的内容		(595)
40.1	了解新内容的途径	(595)
40.2	SAS 8.0 中有哪些新的内容	(595)
40.3	SAS 8.0 中新内容的解释	(596)
40.4	获取有关 SAS 软件信息的重要网址	(596)
参考文献		(597)
附录 1 SAS 程序说明与统计咨询信息		(599)
附录 2 WINDOWS SAS 上机实习指导		(601)

第1篇 预备知识及基本概念

第1章 绪论

1.1 引言

1.1.1 科学研究与统计学之关系

科学研究的目的是探索各研究领域内未知的事物，而反映未知事物特征的指标常带有变异性。由于变异性的存在，实验或观测的结果就必然带有不确定性。为了获得带有规律性的结果，人们就不得不进行大量的实验或观测。然而总体的容量很大，有时甚至是无限的，而研究者的时间、精力、人力和物力却是十分有限的。客观实际迫使研究者不得不从总体中进行抽样研究，以期通过样本所提供的信息去推论总体的规律性。为了研究和解释自然界，乃至人类社会生产、生活等各个方面出现的随机现象的变化规律，产生了数学的一个重要分支——概率论和数理统计。其理论和方法在不同领域内灵活运用，便产生了工业统计学、农业统计学、环境统计学、生物统计学、医学统计学、卫生统计学、遗传统计学等。以专业知识为前提，正确合理地运用统计学知识，可以加速科学研究的进程，提高科学研究的质量。

如何定义总体、如何从特定总体中去抽样、如何确定样本大小、选择什么作为受试对象、观测哪些指标、如何合理安排和控制各种影响因素等一系列问题，就构成了在统计学中占有重要地位的核心内容——统计研究设计。在实验科学领域内，凡是忽视实验设计的研究，要么得出可笑、错误的结论；要么成百上千次地重复某项实验，方可得到预期的某种结果，以至“投入”与“产出”的比例严重失调。

有了良好的统计研究设计，并不代表科研工作一定能取得成功。如果研究者对某设计的本质认识不清，在收集实验数据时就易犯错误，以至不能按原先实验设计所定义的数学模型去分析实验资料。例如：许多医学研究工作者常自发地运用重复测量设计，而收集资料时却是按独立重复实验设计方式进行的。把本来只有30只小鼠分别被重复观测5次而得到的150个实验数据，错误地视为取自150只小鼠的实验数据。本该用重复测量设计的方差分析，却被迫使用析因设计的方差分析，其结论的可靠性必然要受到影响。

统计研究设计和资料收集都正确，是否就万事大吉了呢？事情并非如此简单。数据的统计描述与表达方面的错误，往往会降低研究论文的质量；而不用统计分析方法或误用、滥用统计分析方法，往往直接导致研究者得出错误的结论，这方面的事例，在医学论文中却是常见的。

简而言之，在科学研究中必须正确运用统计学，这是科研工作的科学性所决定的。搞科研，首先必须尊重科学。借助统计学这个有用的工具，可以去探索未知事物，揭示和阐明客

观事物变化的规律性。

1.1.2 学习和运用统计学的难点与障碍

学习和运用统计学的难点在于将实际问题转化成统计问题,进而选择最合适的统计处理方法(包括统计研究设计和数据分析方法)予以解决。而学习和运用统计学的最大障碍是各种复杂的统计计算。

1.1.3 攻克“难点”和跨越“障碍”的办法

攻克“正确运用统计学的难点”的办法是:在学习统计学时,就应当树立正确的统计思想:合理分解统计学内容,逐一消化吸收;注意研究和推敲各种统计研究设计方法和统计分析方法之间的区别与联系,并将统计学知识与科研实际紧密联系在一起,从而提高运用统计学的准确率和正确率。要想做到这一点,首先,研究者必须具备扎实的专业理论知识和基础知识,以便在进行实验设计时不会因专业知识的缺乏而犯统计学上的III型错误(即研究结果未能正确回答拟解决的问题);其次,研究者还应当具备较为系统的统计学知识,以便在选择实验设计或统计分析方法时不会因统计知识的缺乏而误用或滥用统计学。只有同时具备这两方面知识的研究者或科研协作组,才能科学地、合理地运用好统计学,使之更好地为科研工作服务。

跨越“统计计算障碍”的办法是:运用现成的统计分析软件包,实现各种复杂的统计计算。要想学会运用一个大型统计分析软件包处理实际问题,若从读懂说明书,并利用软件包所特有的语言编制引导程序入手,则不是一件容易的事。其捷径是:利用软件开发编制好的引导程序集,在“调用和修改指南”的指导下,可以较为迅速地学会软件的使用。

多年来,笔者一直运用此方法,借助 SAS 软件对医学研究生进行统计教学和面向全国举办多期 SAS 软件与统计分析培训班。学员通过一个名叫 HLP 的菜单程序,只需输入两个“编号”,就可方便、快捷地调用 200 多个常用的 SAS 引导程序,更新例中的数据,并在程序修改指导下对现成的程序略加修改,就可获得正确的统计分析结果。使 SAS 软件的应用成了一件轻而易举之事。

1.1.4 大型统计软件包的种类与特点

国际上大型统计软件包的种类有:美国加利福尼亚州大学研制的生物医学计算机程序集——BMDP (Biomedical Computer Programs)、美国北卡罗来纳州 Raleigh 的 SAS 软件有限公司研制的统计分析系统——SAS (Statistical Analysis System)、美国斯坦福大学研制的社会科学统计程序包——SPSS (Statistical Package for the Social Science)。这些大型统计软件包具有如下共同特点:其一,自成体系,力求解决尽可能多的问题;其二,有各自的指令、函数、语言和完成特定任务的程序块;其三,选择项多,适用面宽,功能强大且齐全;其四,尤其是 SAS 软件,还配有开发工具。用户运用其丰富的语言和有效的开发工具,可以编制出界面十分友好的应用软件。

1.2 SAS 软件简介

1.2.1 SAS 软件简史

早在 60 年代,美国北卡罗来纳州 Raleigh 的 SAS 软件有限公司就开始研制统计分析系统——SAS 软件包。于 1976 年研制成功,并作为产品推向市场。目前已被 120 个国家和地区的两万九千多个机构所采用,直接用户超过三百万人。起初, SAS 是建立在 DOS 环境下