

SPSS统计分析应用 案例教程

* 配教学资料

王周伟 龚秀芳 郭照蕊 傅毅 朱敏 编著



北京大学出版社
PEKING UNIVERSITY PRESS

大数据分析系列教材

SPSS统计分析应用 案例教程

王周伟 龚秀芳 郭照蕊 傅毅 朱敏 编著



北京大学出版社
PEKING UNIVERSITY PRESS

图书在版编目 (CIP) 数据

SPSS 统计分析应用案例教程 / 王周伟等编著. —北京: 北京大学出版社, 2018. 1
ISBN 978-7-301-29003-3

I. ① S… II. ①王… III. ①统计分析—软件包—教材 IV. ① C819

中国版本图书馆 CIP 数据核字 (2017) 第 302305 号

- | | |
|-------|--|
| 书 名 | SPSS 统计分析应用案例教程
SPSS Tongji Fenxi Yingyong Anli Jiaocheng |
| 著作责任者 | 王周伟 龚秀芳 郭照蕊 傅 毅 朱 敏 编著 |
| 责任编辑 | 旷书文 |
| 标准书号 | ISBN 978-7-301-29003-3 |
| 出版发行 | 北京大学出版社 |
| 地 址 | 北京市海淀区成府路 205 号 100871 |
| 网 址 | http://www.pup.cn 新浪微博: @北京大学出版社 |
| 电子信箱 | zpup@pup.cn |
| 电 话 | 邮购部 62752015 发行部 62750672 编辑部 62757857 |
| 印 刷 者 | 北京大学印刷厂 |
| 经 销 者 | 新华书店 |
| | 787 毫米 × 1092 毫米 16 开本 28 印张 590 千字
2018 年 1 月第 1 版 2018 年 1 月第 1 次印刷 |
| 定 价 | 69.00 元 |

未经许可, 不得以任何方式复制或抄袭本书之部分或全部内容。

版权所有, 侵权必究

举报电话: 010-62752024 电子信箱: fd@pup.pku.edu.cn

图书如有印装质量问题, 请与出版部联系, 电话: 010-62756370

前 言

大数据时代,为我们量化管理提供了便利,也需要我们全面掌握数据统计分析技术与方法。为适应这一趋势要求,培养能够系统掌握统计分析基本知识与软件实现的应用技术高级人才,我们整理了多年来的课程讲授内容,编写了本书。

本书特点可以概括为“2×2”。第一个“2”是指对每个统计分析技术或方法,都是“理论+应用”,先简要介绍相关知识原理,再以数据分析问题案例的方式,介绍用SPSS软件实现的步骤与结果说明及其解读;第二个“2”是指内容上进阶式地安排了简单数理统计与多元统计分析两门课程的技术方法和SPSS的应用。书中第1至第10章,主要介绍简单数理统计;第11章至第23章讲授了多元统计分析方法的原理与软件操作。相关数据、PPT可以链接到北京大学出版社网站(www.pup.cn)“下载专区”下载。

该书是编者集体合作完成的。全书体系由王周伟拟定,龚秀芳老师编辑整理了第1章至第8章、第10章、第16章;郭照蕊老师编写了第11章至第13章;傅毅、朱敏老师编写整理了第9章、第14章、第15章、第17章;王周伟编辑整理了第5章、第18章至第23章。在老师们的指导下,我们的研究生参与了数据收集、软件操作实现部分的编写与测试工作,具体分工如下:

- 第1章 SPSS 统计概述(王许利)
- 第2章 数据整理(金芦)
- 第3章 SPSS 数据的预处理(李伟利、向际森)
- 第4章 描述性统计分析(刘嘉、蒯欣欣)
- 第6章 参数检验(赵帅)
- 第7章 非参数检验(齐慧、高旭)
- 第8章 列联表分析(姜巧)
- 第9章 方差分析(周爱君)
- 第10章 相关分析(母亚萍)
- 第11章 回归分析(陈昌领、宋韞赟)
- 第12章 Logistic 回归分析(方璐)
- 第13章 因子分析(钱太红、李晶晶)
- 第14章 聚类分析(王余林)
- 第15章 对应分析(鲁文霞)
- 第16章 距离分析(胡少廷、李莉苹)

- 第 17 章 多维尺度分析(胡少廷)
- 第 18 章 判别分析(习文杰)
- 第 19 章 信度分析(宫晨怡)
- 第 20 章 效度分析(黄盼)
- 第 21 章 生存分析(侯晗)
- 第 22 章 路径分析(杜丽)
- 第 23 章 时间序列分析(魏笑、刘静)

SPSS 统计分析,国内同行已经做出了很好的编著工作。在编写整理期间,我们参考了许多成果。我们在此表示最诚挚的感谢。

编 者
2017 年 4 月

目 录

第 1 章 SPSS 统计概述	1
1.1 SPSS 简介	1
1.2 SPSS for windows 的基本窗口	3
1.3 SPSS 数据分析的一般步骤	7
第 2 章 数据整理	10
2.1 SPSS 数据文件	10
2.2 SPSS 数据的结构和定义方法	11
2.3 SPSS 数据的录入与编辑	14
2.4 SPSS 数据文件的保存	16
2.5 SPSS 数据文件的合并	17
第 3 章 SPSS 数据的预处理	25
3.1 数据选取	25
3.2 数据的排序	27
3.3 计数	29
3.4 变量计算	31
3.5 分类汇总	39
3.6 数据分组	42
3.7 数据预处理的其它功能	48
第 4 章 描述性统计分析	52
4.1 频数分析	52
4.2 描述性分析	59
4.3 探索性分析	64
4.4 比率分析	73
4.5 多选项分析	78
第 5 章 参数的区间估计	89
5.1 区间估计的概念	89

5.2	正态总体均值的置信区间	89
5.3	正态总体方差的置信区间	93
5.4	两个正态总体均值差的区间估计	95
5.5	其他总体参数及参数的区间估计	98
第 6 章	参数检验	100
6.1	假设检验概述	100
6.2	单样本的总体参数假设检验	102
6.3	两个独立样本的总体参数假设检验	105
6.4	两个配对样本的总体参数假设检验	109
第 7 章	非参数检验	113
7.1	非参数检验概述	113
7.2	卡方检验	114
7.3	二项分布检验	119
7.4	游程检验	121
7.5	单样本 K-S 检验	124
7.6	两独立样本检验	127
7.7	多独立样本检验	132
7.8	两配对样本检验	137
7.9	多配对样本检验	141
第 8 章	列联表分析	146
8.1	列联表概述	146
8.2	独立性检验	148
8.3	变量的相关性测度与检验	150
8.4	列联表分析方法的实现	152
第 9 章	方差分析	159
9.1	方差分析概述	159
9.2	单因素方差分析	160
9.3	多因素方差分析	170
9.4	协方差分析	179
第 10 章	相关分析	185
10.1	相关分析概述	185
10.2	相关分析及 SPSS 的实现	186

10.3	偏相关分析	195
第 11 章	回归分析	200
11.1	回归分析概述	200
11.2	线性回归模型及 SPSS 的实现	201
11.3	曲线回归	222
第 12 章	Logistic 回归分析	226
12.1	Logistic 回归分析概述	226
12.2	二项 Logistic 回归分析	227
12.3	多项 Logistic 回归分析	239
12.4	多项有序回归分析	245
第 13 章	因子分析	250
13.1	因子分析概述	250
13.2	因子分析的基本原理	251
13.3	因子分析在 SPSS 中的操作过程及实例	254
第 14 章	聚类分析	263
14.1	聚类分析概述	263
14.2	聚类分析的基本原理	263
14.3	聚类分析在 SPSS 中的操作过程及实例	267
第 15 章	对应分析	278
15.1	对应分析概述	278
15.2	对应分析的基本步骤	280
15.3	对应分析方法的 SPSS 实现	281
15.4	对应分析方法实例	285
第 16 章	距离分析	294
16.1	距离分析概述	294
16.2	距离分析中的距离测度	294
16.3	变量距离分析的 SPSS 软件实现	298
16.4	样品距离分析的 SPSS 软件实现	301
第 17 章	多维尺度分析	304
17.1	多维尺度分析概述	304

17.2	多维尺度分析中的基本概念	305
17.3	古典多维尺度分析	305
17.4	非度量多维尺度分析	306
17.5	权重多维尺度分析	307
17.6	多维尺度分析 MDS 的 SPSS 软件实现	308
第 18 章	判别分析	315
18.1	判别分析概述	315
18.2	判别分析法的基本原理	316
18.3	判别分析的 SPSS 实现与结果	321
第 19 章	信度分析	332
19.1	信度分析的基本原理	332
19.2	重测信度与复本信度分析	335
19.3	内部一致性信度分析	342
19.4	评分者信度分析	348
第 20 章	效度分析	351
20.1	效度分析的基本原理	351
20.2	效度分析的 SPSS 实现	354
20.3	实例:效度分析的 SPSS 实现与结果分析	355
第 21 章	生存分析	359
21.1	生存分析概述	359
21.2	寿命表分析法	361
21.3	Kaplan-Meier 分析法	366
21.4	生存率比较分析	372
21.5	Cox 比例风险回归模型	374
第 22 章	路径分析	383
22.1	路径分析中的基本概念	383
22.2	路径分析的基本理论	384
22.3	路径分析的一般步骤	389
22.4	路径分析的 SPSS 实现	390
第 23 章	时间序列分析	405
23.1	时间序列分析概述	405

23.2	数据准备	407
23.3	时间序列的可视化描述及分析检验	408
23.4	时间序列的预处理	418
23.5	时间序列分析的简单回归分析法和趋势外推法	421
23.6	指数平滑法	423
23.7	谱分析	428
23.8	周期性分解	432
23.9	ARIMA 模型分析	434
参考文献		439

第 1 章 SPSS 统计概述

1.1 SPSS 简介

近几年,随着云计算、物联网等颠覆性革命的兴起,引发了社会各界对“数据”的重视。云计算等主要是为数据资产提供了一个保管、访问和便于提取与分享的场所,而其背后的“数据”才是真正有价值的资产。全球知名咨询公司麦肯锡最早提出“大数据”时代的到来:“数据,已经渗透到当今每一个行业和业务职能领域,成为重要的生产因素。人们对于海量数据的挖掘和运用,预示着新一波生产率增长和消费者盈余浪潮的到来。”

企业内部的财务信息、交易信息,互联网世界中的商品物流信息等组成了各种各样的数据,而这些数据对于当今处于竞争化异常激烈的环境中的企业来说有着至关重要的作用。但同时,企业不仅要及时地收集、存储和管理大量的市场和客户数据,建立企业级数据库管理信息系统,更要加强对这些数据的深层次挖掘工作。企业能获得大量的优质数据固然是好的,但是要想在竞争中立于不败之地,也必须加强自身数据处理方面软硬件设施的建设,而 SPSS 统计分析软件则是基于数据库基础上可以实现数据的深层次定性及定量分析并为企业决策提供积极帮助的有效工具。

SPSS 统计分析软件是 Statistical Product and Service Solutions 的英文缩写,意思是“统计产品与服务解决方案”,强调了其作为社会科学统计分析软件应用于实际需求的一面,而实际上 SPSS 统计分析软件以其强大的统计分析功能、简洁易用的用户操作界面、灵活的表格式分析报告和优良的图形展现形式,赢得了社会各界数据分析人员的喜爱。SPSS 凭借其自身独有的优势在当今众多的数据分析软件中开辟了一席之地,得到了社会各界的广泛应用。同时,SPSS 统计分析软件也越来越多地被各大院校作为统计类及财经类专业的必修课程之一,足见 SPSS 统计分析软件越来越高的社会地位。

1.1.1 SPSS 的发展历程及特点

SPSS 是世界上应用最早的统计分析软件,它最早于 20 世纪 60 年代末由美国斯坦福大学的三位研究生研发而成,并于 1975 年成立了专门用于研发和经营 SPSS 软件的公司。最初,由于计算机本身的不完善,SPSS 软件主要是在中小型计算机上运行,且面向企事业单位客户。随着计算机的不断完善和发展,SPSS 统计分析软件业抓准时机,紧跟信息时代的步伐,相继在 DOS 操作系统和 Windows 图形化操作系统上开发新版软件,以迎合 Internet 的广泛使用和用户在不同操作系统环境下工作的习惯。

随后,在 1994 年至 1998 年间,SPSS 公司对市面上出现的一些同行公司如 SYSTAT、ISL、BMDP 等进行大规模的收购,将这些公司的主打品牌收于麾下,扩大了 SPSS 公司的市场份额和客户群,使得 SPSS 统计分析软件由原来较为单一的产品定位转而面向为教育科研及政府机构等统计决策服务的综合性产品发展。2009 年 7 月 28 日,IBM 公司斥资 12 亿元美金收购了 SPSS 公司,因此,SPSS 公司也更名为 IBM SPSS Statistics,也称 PASW(Predictive Analysis Software)Statistics。SPSS 从原来的 SPSS/PC+ 版本已经升级到现在的 SPSS for Windows 22.0 版本,它最突出的特点是操作界面极其友好且输出结果美观漂亮。SPSS 采用类似 Excel 表格的方式输入和管理数据,且可以实现在 Excel 和 SPSS 数据窗口之间的方便复制,数据接口较为通用。

SPSS 的基本功能包括数据管理、统计分析、图表分析和输出管理等。它是一个组合式软件包,除了这些基本功能之外,用户可以根据自身需要选择模块,降低对系统硬盘的要求,有利于 SPSS 统计分析软件的推广。

SPSS 统计分析过程除了包括描述性统计、均值比较、相关分析、回归分析等基本的统计分析功能之外,还涉及到聚类分析、因子分析及信度分析等深层次的数据挖掘功能,并且 SPSS 的每个过程允许用户选择不同的方法和参数。SPSS 也有专门的绘图软件,图标功能强大且输出结果美观。

SPSS 凭借其方便易学的数据操作管理、成熟的统计分析方法、开放的数据接口及美观灵活的统计表格与图形,和 SAS、BMDP 并称为国际上最有影响力的三大统计软件。

1.1.2 SPSS Statistics 22.0 新特性

1. IBM developerWorks 可以释放 IBM SPSS Statistics 潜能,在发展技能、协作、下载扩展和定制过程以及保持最新技术趋势处于领先地位。

2. 对于 Cognos BI 用户而言,使用 SPSS Statistics 可以进一步加强统计分析能力,SPSS 22.0 可以将 Cognos BI 数据导入 SPSS Statistics,方便我们对数据进行深度挖掘。

3. 可以进行 Monte Carlo 模拟,通过 Monte Carlo 模拟方法,可以针对最重要的问题得到最可靠的解答。

4. 准备演示文稿部分可以直接在过程对话中标记表格中的重要值,并且具有可以自动完成输出文档中的常见编辑功能。

5. 使用 SPSS Statistics Server,可以更快捷地在数据库导向中制定新变量的计算并在数据库本身中执行此计算。

6. SPSS Statistics 提供了比较两个数据集或两个数据文件以确定二者间的差异的功能,可以用于对比数据文件,进行企业间简单的质量控制。

7. 最新的数据合并方法,可以合并多个数据集而不用预排序,使用字符串键值时更灵活。

8. 可以在 SPSS Statistics 输出,在浏览器或任何智能设备(智能手机、平板电脑

等)上查看此输出,方便 SPSS 图标和表格的随身查看和使用。

9. 透视表更加强大,使用诸如排序、按值范围搜索单元格、插入新的列和行等的功能来处理透视表。

1.2 SPSS for windows 的基本窗口

了解 SPSS 的基本窗口是进行 SPSS 操作的基础,也是学习和熟练运用 SPSS 的入门点。了解 SPSS 有哪些基本窗口,各窗口的特点和功能是什么,窗口之间有什么联系,用户便能够清楚地理清 SPSS 软件的基本框架。本节主要介绍 SPSS 22.0 的常用窗口,包括数据编辑窗口、结果输出窗口和命令语句窗口。

1.2.1 SPSS 数据编辑窗口

数据编辑窗口(SPSS Data Editor)包括数据视图(Data View)和变量视图(Variable View)两种视图。该窗口是 SPSS 的主程序窗口,主要功能是定义 SPSS 数据的结构特征、录入和编辑数据及管理和分析数据等。其中,数据视图用于录入、查看和修改数据,而变量视图用于输入和修改变量的定义。

当成功启动 SPSS 后,屏幕会出现如图 1.1 所示的窗口,即为数据编辑窗口。

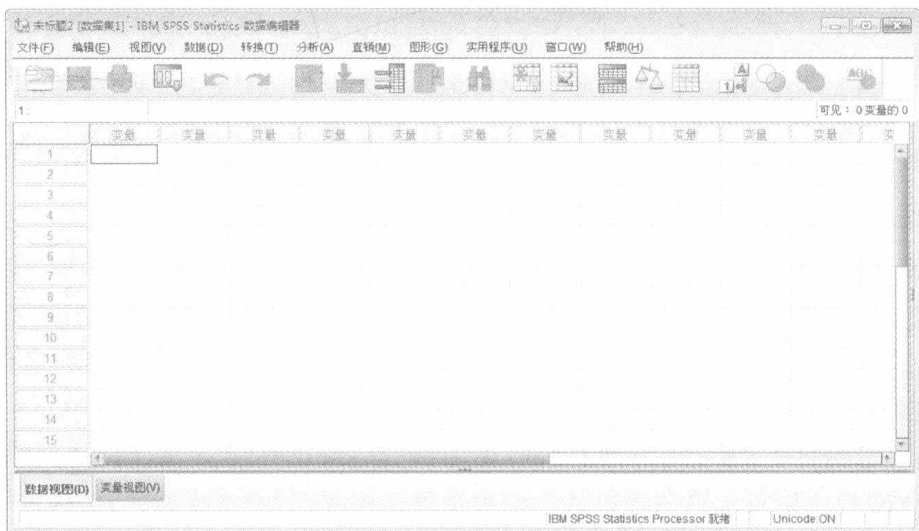


图 1.1 SPSS 数据编辑窗口

数据编辑窗口包括标题栏、主菜单栏、工具栏、数据编辑区和状态栏。

1. 标题栏:显示当前的文件名。
2. 主菜单栏:包括文件、编辑、视图、数据、转换、分析、直销、图形、实用程序、窗口和帮助,用户可以通过单击菜单完成相应的操作。菜单栏对应的功能见表 1.1。

表 1.1 主窗口菜单及功能

菜单名	功能	功能解释
文件	文件操作	新建、保存文件,打开数据库,读取数据,重新命名数据集等基本文件管理工作。
编辑	数据编辑	对数据进行复制、撤销等工作,插入变量、个案,设置 SPSS 的语言、图标外观等功能。
视图	窗口外观状态管理	对窗口外观如状态栏、工具栏、字体、网格线、值标签等进行设置。
数据	数据基本操作和管理	对原始数据的加工处理,如数据排序、转置、拆分、分类汇总及加权等。
转换	数据基本处理	对编辑窗口的数据进行基本处理,如计算变量、生成新变量、分组等。
分析	数据统计分析	对数据进行统计分析和建模,如基本统计分析、回归、相关分析及非参数检验等。
直销	客户关系管理	识别最佳客户(RFM 分析)、聚类分析、潜在客户概要模型、购买倾向等。
图形	统计图形制作	对数据编辑窗口选中的数据生成各种图形,如直方图、条形图、散点图等。
实用程序	实用程序	SPSS 其他辅助管理,如合并模型、定义变量集、菜单编辑器等。
窗口	窗口管理	对多窗口管理,如窗口拆分、窗口最小化等。
帮助	帮助	实现 SPSS 的联机帮助。

3. 工具栏:除了保留数据编辑窗口的某些功能按钮,还增加了一些特有的按钮,用户可以直接单击工具栏上的按钮完成相应功能,使操作更加方便和快捷。

4. 数据编辑区:用于显示和管理数据结构与内容的区域。在编辑区的最左边是每条个案的顺序编号。数据编辑区中的表格线可以通过【视图】菜单下的【网格线】选项设置成显示或不显示状态。【数据】视图和【变量】视图都位于数据编辑区。

5. 状态栏:查看系统的当前运行状态。

6. 另外,可以通过【编辑/选项】,实现中英文界面的切换。在图 1.2 中,在【语言】标签页上面,输出选择如图示的“英语(E)”,用户界面也选择“英语(E)”,再按【应用/确定】就切换到英文界面;反之,在英文主界面,可以通过【Edit/Options】选项,在【Language】标签页上面,output 选择“Chinese(simplified)”,User Interface 也选择“chinese(simplified)”,再按【Apply/Ok】就切换到简体中文界面。这种切换是经常性的,因为有时中文的统计专业名词翻译有点欠缺,看英文更加明确。



图 1.2 SPSS 中英文切换窗口

SPSS 允许同时打开多个数据编辑窗口,分别显示不同的数据文件内容,数据窗口中的数据也称为数据集,这些数据通常以 SPSS 数据文件的形式保存在计算机上,其文件扩展名为 .sav。

1.2.2 SPSS 结果输出窗口

SPSS 结果输出窗口是 SPSS 的另一个主要窗口,主要用于显示和管理 SPSS 统计分析结果(包括文本、表格和图形)的窗口,如图 1.3 所示。输出结果通常以 SPSS 输出文件的形式保存在计算机磁盘上,其文件扩展名为 .spv。

此窗口在 SPSS 启动时会自动创建并打开,若要在 SPSS 运行中打开新的结果输出窗口,可以通过【文件/新建 or 打开/输出】。同数据编辑窗口类似,SPSS 也可以同时创建或打开多个输出窗口,并且多个输出窗口中只有一个是主输出窗口或称当前输出窗口。输出窗口由菜单栏、工具栏及分析结果显示区等组成。

此处需要注意的是对于输出窗口中的分析结果显示区,是专门显示统计分析结果的区域,可以分为左右两个区域,左边是目录区,右边是内容区。目录区和内容区是独立的形式,两者区域大小也可自由调整。目录区和内容区各有一个红色箭头,它们所指示的内容是一一对应的,用户也可以对这两个区域的内容进行增、删、改等编辑管理工作。



图 1.3 结果输出窗口

1.2.3 SPSS 语句窗口

SPSS 语句窗口即语法编辑窗口,是编辑和运行命令文件的窗口,如图 1.4 所示。该窗口可能更适用于具有编程基础的人员使用,它不仅可以编辑不能实现的特殊过程的命令语句,还可以将大量的分析过程汇集在一个命令文件中。用户还可以根据自己的需求对语句窗口中的命令进行修改、编辑、复制及粘贴等,也可以编写针对当前数据文件的命令程序。该窗口的内容可存为 .sps 为扩展名的 SPSS 文件。



图 1.4 语法编辑窗口

若要打开新的命令语句窗口,可单击【文件/新建/语法】,编写好命令文件后,可以单击菜单栏上的【运行】按钮,显示输出窗口,得到分析结果。

有时,经常将菜单运行与语句命令运行结合起来使用。如对成绩数据中的“发展经济学”成绩进行描述统计,单击【分析/描述统计/描述】,出现如图 1.5 的界面,然后按【粘贴】按钮,就出现图 1.4 的语法界面。这样就可以在菜单操作中学习语法语句。



图 1.5 语法编辑窗口

1.3 SPSS 数据分析的一般步骤

SPSS 统计分析软件是一款专业性较强的统计软件,因此我们学习和应用它必须了解、掌握统计学基本专业知识与数据分析的一般步骤和原则,这样才可以避免在用 SPSS 进行数据分析时发生一些统计常识的错误,进而得出错误的统计分析结果,引发错误的决策。

1.3.1 数据分析的一般步骤

数据分析一般包括收集数据、整理和加工数据、分析数据、解释数据分析结果等四个方面。在统计学中对此有着严格的定义,而在应用 SPSS 软件时,用统计学的理论来指导数据分析工作是必不可少的。

1. 明确分析目标

进行数据分析的前提是先明确确定一个分析目标。明确分析目标就是要明确本次分析要研究的主要问题和预期目标等。例如,分析企业中不同类型客户的消费偏好特征;分析不同高校之间科研能力的差异是否显著;分析不同城市间儿童身高差异是否显著,等等。分析目标一旦明确,我们便可以根据目标来收集相关数据,以及考虑怎样搜集与目标相关的数据。

2. 正确收集数据

一般情况下,我们所收集到的原始数据中常常会包含多种干扰项,这种情况下数据虽然是合理数据但并非有效数据,若是使用这种数据进行统计分析,将会导致统计分析结果的偏差或错误,因此正确有效的数据对于实现数据分析的目的起到关键性的作用。

如何正确地收集数据,这在医药应用方面相对有着更广泛且更重要的作用,比如研究中西医结合治疗方案治疗 H7N9 的例子;研究某种降压药是否有显著的降压效果,等等,必须排除数据中那些与目标不相关的干扰因素,收集那些能迎合数据分析目标的数据,才能达到较好的分析结果。为了正确收集数据,可以采用统计学中更加专业的数据收集方法,比如采取随机双盲式实验,通过合理地随机分组来对影响分析结论的诸多因素加以控制等。