

<<IBM SPSS数据分析与挖掘实战案>>

图书基本信息

书名：<<IBM SPSS数据分析与挖掘实战案例精粹>>

13位ISBN编号：9787302299547

10位ISBN编号：7302299544

出版时间：2013-2-22

出版时间：清华大学出版社

作者：张文彤,钟云飞

版权说明：本站所提供下载的PDF图书仅提供预览和简介，请支持正版图书。

更多资源请访问：<http://www.tushu007.com>

<<IBM SPSS数据分析与挖掘实战案>>

内容概要

《IBM SPSS数据分析与挖掘实战案例精粹》以IBM SPSS Statistics 20.0和IBM SPSS Modeler 14.1为工具，提供了医疗、金融、保险、汽车、快速消费品、市场研究、互联网等多个行业的数据分析/挖掘案例，基于实战需求，详细讲解整个案例的完整分析过程，并将模型和软件的介绍融于案例讲解之中，使读者在阅读时能突破方法和工具的局限，真正聚集于对数据分析精髓的领悟。

《IBM SPSS数据分析与挖掘实战案例精粹》所附光盘包括案例数据和分析程序/流文件，读者可完整重现全部的分析内容。

<<IBM SPSS数据分析与挖掘实战案>>

作者简介

张文彤，博士，数据挖掘、市场研究、统计软件教学与应用领域专家，现任全球第八大市场研究集团INTAGE中国公司全国技术总监。

曾在复旦大学任教数载，期间协助SPSS在中国建立并完善了其培训体系，是国内知名的SPSS培训教之一。

在数据挖掘、市场研究、医药数据分析等领域均经验丰富，曾负责为知名跨国公司完成了中国城市女性市场细分模型、销量预测模型、商圈选址模型等各类项目，并协助完成多项IT、电信、税务、银行等行业的数据挖掘项目。

钟云飞，资深数据分析专家，拥有超过10年的统计分析与数据挖掘在各行业的软件应用及咨询经验，历任SPSS、SAS软件公司首席咨询顾问，目前在国际商业机器（中国）有限公司软件部工作，从事SPSS软件企业应用的推广工作。

主要关注统计分析与数据挖掘在银行、电信、政府、保险、零售等行业的应用实践，致力于使用数据分析方法帮助企业和政府组织从数据中获取有价值的信息从而提高管理水平。

书籍目录

第一部分SPSS数据分析基础 第1章数据分析方法论简介 1.1三种数据分析方法论 1.1.1严格设计支持下的统计方法论 1.1.2半试验研究支持下的统计方法论 1.1.3偏智能化、自动化分析的数据挖掘应用方法论 1.2 CRISP—DM方法论介绍 1.2.1概述 1.2.2商业理解 1.2.3数据理解 1.2.4数据准备 1.2.5建立模型 1.2.6模型评价 1.2.7结果部署 第2章数据分析方法体系简介 2.1统计软件中的数据存储格式 2.1.1 二维数据表 2.1.2 变量的存储类型 2.1.3变量的测量尺度 2.2数据的统计描述与参数估计 2.2.1连续变量的统计描述 2.2.2连续变量的参数估计 2.2.3分类变量的统计描述和参数估计 2.2.4统计图形体系 2.3常用假设检验方法 2.3.1假设检验的基本原理 2.3.2单变量假设检验方法 2.3.3双变量假设检验方法 2.4多变量模型 2.4.1方差分析 / 一般线性模型 2.4.2广义线性模型和混合线性模型 2.4.3回归模型 2.4.4其他常见模型 2.5多元统计分析模型 2.5.1 信息浓缩 2.5.2变量间内在关联结构的探讨 2.5.3数据分类 2.5.4分析元素间的关联 2.6智能统计分析 / 数据挖掘方法 2.6.1树模型 2.6.2神经网络 2.6.3支持向量机 2.6.4贝叶斯网络 2.6.5最近邻元素分析 2.6.6关联规则与序列分析 第3章 IBM SPSS Statistics操作入门 3.1案例背景 3.2数据文件的读入与变量整理 3.2.1 SPSS的基本操作界面 3.2.2数据准备 3.3问卷数据分析 3.3.1生成频数表 3.3.2计算均值 3.3.3对多选题进行描述 3.4项目总结和讨论 第4章 IBM SPSS Statistics操作进阶 4.1案例背景 4.1.1项目背景 4.1.2分析思路 4.2问卷录入 4.2.1开放题的定义 4.2.2单选题的定义 4.2.3多选题的定义 4.3问卷质量校验 4.3.1去除重复记录 4.3.2发现异常值 4.3.3逻辑校验 4.4问卷数据分析 4.4.1 问卷加权 4.4.2业务分析 4.5项目总结和讨论 第5章 IBM SPSS Modeler操作入门 5.1 IBM SPSS Modeler概述 5.1.1 IBM SPSS Modeler的界面 5.1.2 IBM SPSS Modeler的架构与产品构成 5.2 IBM SPSS Modeler相关操作与技巧 5.2.1 IBM SPSS Modeler的基本操作 5.2.2 IBM SPSS Modeler中的表达式 5.2.3 IBM SPSS Modeler的若干使用技巧 5.3 IBM SPSS Modeler功能介绍 5.3.1 数据整理案例 5.3.2探索性数据分析案例 5.3.3建立模型、模型检验与模型应用案例 5.4案例分析：药物选择决策支持 5.4.1背景介绍 5.4.2数据说明 5.4.3商业理解 5.4.4数据理解 5.4.5数据准备 5.4.6模型建立和评估 5.4.7模型发布 5.5如何进一步学习IBM SPSS Modeler 第二部分影响因素发现与数值预测 第6章酸奶饮料新产品口味测试研究案例 6.1案例背景 6.1.1研究项目概况 6.1.2分析思路与商业理解 6.2数据理解 6.2.1研究设计框架复查 6.2.2均值的列表描述 6.2.3均值的图形描述 6.3不同品牌的评分差异分析 6.3.1单因素方差分析模型简介 6.3.2 品牌作用的总体检验 6.3.3组间两两比较 6.3.4方差齐性检验 6.4两因素方差分析模型分析 6.4.1两因素方差分析模型简介 第三部分信息浓缩、分类与感知图呈现 第四部分数据挖掘案例精选 附录 参考文献 后记

章节摘录

版权页：插图：截尾均数（Trimmed Mean）：由于均数较易受极端值的影响，因此可以考虑按照一定比例去掉两端的数据，然后再计算均数。

如果截尾均数和原均数相差不大，则说明数据不存在极端值，或者两侧极端值的影响正好抵消。

常用的截尾均数有5%截尾均数，即两端各去掉5%的数据。

其他集中趋势描述指标：除了上述最常用的几种指标外，还会遇到众数、调和均数等，前者是指样本数据中出现频次最大的那个数字，后者是指观察值 X 倒数之均数的倒数，这些指标的实际应用都比较少见。

2.离散趋势的描述指标 显然，仅仅反映数据的集中趋势是远远不够的，还要反映数据的离散趋势，即数据的波动范围，描述该趋势的统计量称为尺度统计量（Scale Statistic），常用的尺度统计量有标准差、方差、四分位间距等。

全距（Range）：全距又称为极差，即一组数据中最大值与最小值之差，它是最简单的变异指标，但因其过于简单，因此一般只用于预备性检查。

方差（Variance）和标准差（Standard Deviation）：总体和样本的标准差分别用 σ 和 s 来表示，方差即标准差的平方，这两个指标是应用最广泛的离散程度描述指标，由于标准差和方差的计算利用到每个原始变量值，所以它们反映的信息在离散指标中是最全的，因此也是最理想、最可靠的变异描述指标。

但也正是由于标准差和方差的计算用到每一个变量值，所以它们会受到极端值的影响，当数据中有较明显的极端值时不宜使用。

实际上，方差和标准差的适用范围应当是服从正态分布的数据。

百分位数、四分位数与四分位间距：百分位数（Percentile）是一种位置指标，用 P_x 表示。

一个百分位数 P_x 将一组观察值分为两部分，理论上 $x\%$ 的观察值比它小，有 $(100-x)\%$ 的观察值比它大。

前面学习过的中位数实际上就是一个特定的百分位数，即 P_{50} 。

除中位数外，常用的百分位数还有四分位数，即 P_{25} 、 P_{50} 和 P_{75} 分位数的总称。

这3个分位数正好能够将总体单位按标志值的大小等分为四部分，且 P_{25} 和 P_{75} 这两个分位数间包括中间50%的观察值，因此四分位间距既排除了两侧极端值的影响，又能够反映较多数据的离散程度，它是当方差、标准差不适用时较好的离散程度描述指标。

变异系数（Coefficient of Variation）：当需要比较两组数据离散程度大小的时候，如果两组数据的测量尺度相差太大，或者数据量纲不同，直接使用标准差来进行比较不合适，此时就应当消除测量尺度和量纲的影响，而变异系数可以做到这一点，它是标准差与其平均数的比。

CV显然没有量纲，同时又按照其均数大小进行了标准化，这样就可以进行客观比较了。

<<IBM SPSS数据分析与挖掘实战案>>

编辑推荐

《IBM SPSS数据分析与挖掘实战案例精粹》适合从初学者到专家各个级别的数据分析人员阅读，尤其适合于以下读者群：需要提升实战能力的数据分析专业人员；在市场营销、金融、财务、人力资源管理中需要应用数据分析的人士；从事咨询、科研等工作的专业人士；同时也适合于各专业的本科和研究生作为学习数据分析应用的参考书。

版权说明

本站所提供下载的PDF图书仅提供预览和简介，请支持正版图书。

更多资源请访问:<http://www.tushu007.com>