

# 《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

出版社：清华大学出版社

作者：张文彤,钟云飞

页数：501

版权说明：本站所提供下载的PDF图书仅提供预览和简介以及在线试读，请支持正版图书。

更多资源请访问：[www.tushu111.com](http://www.tushu111.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 and 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

##### 1.2 crisp-dm方法论介绍

#### 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

##### 2.2 数据的统计描述与参数估计

##### 2.3 常用假设检验方法

##### 2.4 多变量模型

##### 2.5 多元统计分析模型

##### 2.6 智能统计分析 / 数据挖掘方法

#### 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

##### 3.3 问卷数据分析

##### 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

##### 4.2 问卷录入

##### 4.3 问卷质量校验

##### 4.4 问卷数据分析

##### 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

##### 5.2 ibm spss modeler相关操作与技巧

##### 5.3 ibm spss modeler功能介绍

##### 5.4 案例分析：药物选择决策支持

##### 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

##### 6.2 数据理解

##### 6.3 不同品牌的评分差异分析

##### 6.4 两因素方差分析模型分析

##### 6.5 分析结论与讨论

#### 第7章 偏态分布的激素水平影响因素分析

##### 7.1 案例背景

##### 7.2 数据理解

##### 7.3 对因变量变换后的建模分析

##### 7.4 秩变换分析

##### 7.5 利用cox模型进行分析

##### 7.6 项目总结与讨论

#### 第8章 某车企汽车年销量预测案例

##### 8.1 案例背景

##### 8.2 数据理解

##### 8.3 变量变换后的线性回归

- 8.4 曲线拟合
- 8.5 利用非线性回归进行拟合
- 8.6 项目总结与讨论
- 第9章 脑外伤急救后迟发性颅脑损伤影响因素分析案例
  - 9.1 案例背景
  - 9.2 数据理解
  - 9.3 构建二分类logistic回归模型
  - 9.4 利用树模型发现交互项
  - 9.5 使用广义线性过程进行分析
  - 9.6 项目总结与讨论
- 第10章 中国消费者信心指数影响因素分析
  - 10.1 案例背景
  - 10.2 数据理解
  - 10.3 标准glm框架下的建模分析
  - 10.4 多元方差分析模型的结果
  - 10.5 最优尺度回归
  - 10.6 多水平模型框架下的建模分析
  - 10.7 项目总结与讨论
- 第三部分 信息浓缩、分类与感知图呈现
- 第11章 探讨消费者购买保健品的动机
  - 11.1 案例背景
  - 11.2 数据理解
  - 11.3 利用因子分析进行信息浓缩
  - 11.4 基于因子分析结果进行市场细分
  - 11.5 项目总结与讨论
- 第12章 1988年汉城奥运会男子十项全能成绩分析
  - 12.1 案例背景
  - 12.2 数据理解
  - 12.3 利用因子分析进行信息浓缩
  - 12.4 主成分回归
  - 12.5 将主成分回归方程还原回原始变量的形式
  - 12.6 项目总结与讨论
- 第13章 打败sars
  - 13.1 案例背景
  - 13.2 数据理解与数据准备
  - 13.3 “非典”信息关注倾向的多维偏好分析
  - 13.4 突发事件险种购买倾向的多重对应分析
  - 13.5 “非典”对未来生活方式的影响
  - 13.6 项目总结与讨论
- 第14章 住院费用影响因素挖掘
  - 14.1 案例背景
  - 14.2 数据理解与数据准备
  - 14.3 采用聚类分析寻找费用类型
  - 14.4 住院费用影响因素的神经网络分析
  - 14.5 不同疗法疗效与费用比较的神经网络分析
  - 14.6 项目总结与讨论
- 第四部分 数据挖掘案例精选
- 第15章 淘宝大卖家之营销数据分析
  - 15.1 案例背景

- 15.2 利用rfm模型定位促销名单
- 15.3 寻找有重购行为买家的特征
- 15.4 总结与讨论
- 第16章 超市商品购买关联分析
  - 16.1 案例背景
  - 16.2 数据准备
  - 16.3 商品购买关联分析
  - 16.4 结果应用
- 第17章 电信业客户流失分析
  - 17.1 案例背景
  - 17.2 商业理解
  - 17.3 数据理解与数据准备
  - 17.4 建立模型与模型评估
  - 17.5 模型的应用及营销预演
  - 17.6 总结与讨论
- 第18章 信用风险评分方法
  - 18.1 案例背景
  - 18.2 商业理解
  - 18.3 数据理解与数据准备
  - 18.4 建立模型与模型评估
  - 18.5 对若干问题的说明
- 第19章 医疗保险业的欺诈发现
  - 19.1 案例背景
  - 19.2 商业理解
  - 19.3 数据理解与数据准备
  - 19.4 建立模型
  - 19.5 结果发布
  - 19.6 进一步阅读
- 第20章 电子商务中的数据挖掘应用
  - 20.1 案例背景
  - 20.2 数据理解
  - 20.3 数据准备
  - 20.4 建立模型与模型发布
  - 20.5 进一步阅读
- 附录
  - 附录a 本书光盘内容介绍
  - 附录b spss软件的安装与激活
  - 附录c 书中统计方法、模型与知识点索引
  - 附录d ibm spss statistics函数一览表
  - 附录e ibm spss modeler节点功能简介
- 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  $\sigma$  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，理论上有  $x\%$  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。

## 精彩短评

- 1、案例写得非常详实，而且都很有针对性，特别适合应用到工作中举一反三
- 2、此书其实很一般，书拿到手感觉有点厚，其实内容并不是很多，不过这好像也是市场上为数不多的有具体案例的SPSS书籍了，还是推荐一下
- 3、此书对学习帮助不大。很多问题，没有解释很清楚，原理不清，读者只能一知半解，或者死记硬背。看案例可以增长一下见识，但此书有点难学以致用。所以，还是不推荐买。
- 4、已经读完第一遍，在读第二遍。张文彤的书和课都是一级的精品，令我感恩世界一百遍，讲得实在是深入浅出，贴近实际，老师水平高又有丰富的实战经验，是一个很好的教师。
- 5、大量实战案例让本书精彩
- 6、此书只是囫囵吞枣地看过因为手上没IBM SPSS，也没有看明白里面的统计学术语，虽然糊涂带过但知道作者可是有真材实料的人，他对书中研究个案的数据理解（商业理解）非常有启发性（也许是我从不自己深思）。撇去统计分析部分，光看研究思路也很值了。以后会从新操起自己的统计学，向作者的专业水平进发！那才叫数据分析啊摔！
- 7、概述结合案例介绍数据分析与挖掘方法，同时介绍ibm spss modeler的使用，是本不错的书！值得推荐！
- 8、业界良心
- 9、经介绍购买了初级，高级和这个专家级系列，初略翻翻，内容还不错。不过纸张质量比较一般，有一本封面还有点小破损。
- 10、不错的书
- 11、从业务出发，其中模型建立描述的比较详细，入门推荐。
- 12、还不错，内容很详细。
- 13、老公已经挑好的专业书籍，网上买就是便宜。
- 14、超好的数据分析书，以业务为纲，每一章就是一个项目，陆陆续续读了近四个月，数据分析与业务分不开的，对业务好的数据理解，再选择相应的模型和算法，受益匪浅，值得重读~~
- 15、SPSS毕竟是商业软件，商业功能还是可以的。
- 16、很难得的经验分享，沙龙或专业会议都不可能分享这么全这么细
- 17、这本书是作者的SPSS基础篇，高级篇，专家篇系列三部曲中的专家篇，最好先看过SPSS统计分析基础教程再来看实战案例精粹，否则很多基础知识欠缺，难以迅速领会为什么在实战中会这样做或者那样做，学起来就会比较吃力了。
- 18、这本书是想学SPSS的同学的最佳选择，书中不仅有大量的案例参考，更有对SPSS的结果的讲解。我认为唯一的缺点是没有将各种方法的优劣分别说明
- 19、SPSS的高级应用版，有点难度的。
- 20、文彤老师说过，这本书不适合初学者。所以看这本书之前应该先看懂《SPSS统计分析基础教程（第2版）》。文彤老师的SPSS丛书目前共有3本，仔细看一遍就知道这3本书是一个完整的系统，初级=》高级=》实战。这种完整的体系是其他教材所不具备的。另外，作者还暗示了一个重要信息，目前SPSS的功能比较有限，3本书就足以概括SPSS的功能
- 21、个人感觉逻辑很清晰，阅读需要有基础的统计学知识，计量经济学知识，不一定要知道数学怎么算，但要清楚什么时候用什么的原理。现在想，当时看张文彤的基础教程简直就不是人看的，说的话都看不懂...
- 22、具体内容还没仔细看，大概翻了一下，感觉还不错，确实比较难，不适合初学者觉得比较遗憾的是，光盘里的stream的文件打不开，能不能告诉我一下这个的打开方式是什么？
- 23、一周看完，瞎了
- 24、案例挺好，附带的程序和数据也比较清楚！
- 25、锻炼分析方法，逻辑
- 26、逻辑很明晰，SPSS每一功能该采用怎样的思路应用到实战，很务实。值得一读的数据分析入门案例书，感觉挺有趣。数据处理、准备，理清思路是最重要的，使用只是操作手段。
- 27、很不错的一本参考书，可以用到的时候参考
- 28、一章一个案例思路清晰，从商业理解、数据准备、模型选择、建模调优再到项目总结。趋势、聚

类、因子分析几章受用了。

29、好书!!!!!!!!!!!!!!!!!!!!

30、关于数据挖掘的一些知识知识一笔带过，某些分析过程讲的不是很清楚，入门数据果然不应该选中国人写的书，总是喜欢把简单的事情复杂化，本来想给一星的，但是我确实从中学到了一些我之前所不知道的东西，所以多给了一星，但是真心觉得本书的作者从来没有站在读者和学习者的角度考虑如何写书。

31、社会学

32、本身不懂统计分析，但是看了书之后还是对自己启发很大。对数据分析的理解又进了一步，从案例当中学到了很多。建议做业务的可以看看。

33、这本书还没细看，不过翻了一下例子，很不错。看里一下光盘，里面的数据很全，操作性很强，正是我想要的。

34、与初级、高级一起买的。还不错。这个相对另外两本来说比较精要，能迅速上手做类似分析，了解步骤。当然讲解没有其他两本的全。

## 精彩书评

1、自13年5月7日购买本书开始，笔者有近四个月时间在试着读懂这本书，忘不了每一个上下班的路上，电脑包里放着这本书，让我感受到事业的重量，每一个成功都是心血的浇灌。闲话少说，回到数据分析上来，国内介绍spss的书已经有很多了，但大多是在讲解各种统计方法的时候，介绍spss软件是如何操作的。缺点是学完了之后，对数据的理解还不深入，以致于面对普通的数据，还是无从下手。其实这类书更适合当工具书来用。自从读了“小蚊子乐园”博主的《谁说菜鸟不会数据分析》，通俗的语言，生动的情景设置，有读小说的享受。但还是很不满足，因为那毕竟是入门级的读物。笔者继续在Amazon寻觅好的数据分析书，偶然之间看到了这本书，它让我眼前一亮，这本书的框架很完整，第一部分介绍了数据分析的方法论，相当于全书的纲要，总的来说，统计学方法需要严格的假设，更适合科学实验和学术研究；数据挖掘适用于商业。第二章的数据分析方法体系，分析了数据的一般特征，比如数据的离散和连续两种属性，平均数、方差、偏度，也是各种具体的分析方法的概要，方法体系很完备。统计分析方法较为详细，机器学习算法简略一些。有了上面两章“剑谱”的铺垫。接下来，两位数据分析的专家（分别是数据分析和DM）开始在一个个数据分析的项目中，施展他们的“剑术”了。有一些分析项目较难，比如Cox回归、广义线性过程，真的要随时准备一本《SPSS教程》在手头上，笔者一开始就想把它们当做项目，训练自己的能力，不求快，但求理解掌握。于是在看本书的同时，也看了两本SPSS和clementine的教材。过程蛮辛苦的，比如一些原案例中的clementine数据流相当复杂，要读懂模型图，没有别的好方法，只有对照clementine的说明书，按图索“理”，必要的时候还要自己动手在节点处增加一些“表”。最有收获的是第17章，电信客户流失的模型。衍生变量非常多，模型却不复杂，难怪一位前同事跟我说，有90%的精力花在数据处理和准备上了。用这种笨方法，我基本上读完了一遍这本书，收获了很多自信。值得一提的是，这本书的装帧设计也蛮大气的，尤其是IBM三个字母，赫然醒目，确实是表里如一的。要说读了这本书之后，数据分析的技能有突飞猛进的进步，也不是的，还要经过很多商业项目的锻炼，才能达到作者钱眼里“不滞于物，草木竹石皆可为剑”的境界。但这是一本出色的剑谱，以数据分析为志业的你们，不妨读一读它~~

2、我看书有个习惯，如果书籍附带作者靓照的话，我一般喜欢先看作者长得如何，籍此猜测作者是不是一个喜欢专研技术的人。但是很遗憾，看了张博士的靓照和简介，我心里就有了个大概的了解，接着翻阅本书三个多小时的时间里，我时常有这样的感觉，每当我耐着性子看完一节，过滤掉作者无用的话之后，开始有点懂这个软件一些操作的时候，就会看到这样的一段话：“因为篇幅的原因……”。唉，只是学习一个软件的操作而已，很容易的事情，软件背后的算法才是关键，作者总是喜欢两样都讲一点，然后两者都讲不清，最后这本书就变得这么厚了。考虑到张博士曾经在大学里任教，一直从事培训工作，可以理解啦，毕竟，大学的课堂，大家都懂得啦。

## 章节试读

### 1、《IBM SPSS数据分析与挖掘实战案例精粹》的笔记-第51页

为什么是 $s^2=28.5$ ,  $lf(age=2) S^2n=50$ ?赋值 $S^2n$ 的用途是什么? 为什么 $Q9-\&gt;Q9n$  是 $1=7,2=5.5,3=3.5$ ? 还是这些数本身就是随便捏造, 主要讲功能的? 刚开始啃, 已经开始抓狂了。

### 2、《IBM SPSS数据分析与挖掘实战案例精粹》的笔记-第58页

和英文翻译过来的比, 这本语句和逻辑上比较连贯, 但是还是有问题。。

## 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:[www.tushu111.com](http://www.tushu111.com)