

<<SPSS统计分析与数据挖掘>>

图书基本信息

书名：<<SPSS统计分析与数据挖掘>>

13位ISBN编号：9787121152191

10位ISBN编号：7121152193

出版时间：2012-1

出版时间：电子工业

作者：谢龙汉//尚涛

页数：505

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>

## <<SPSS统计分析与数据挖掘>>

### 内容概要

本书基于spss

19.0编写，结合大量的实例对spss各模块的统计分析功能及图形功能等进行了详细讲解。

每章均给出大量分析案例，具体内容为spss简介、spss数据挖掘系统介绍、spss数据文件管理、spss数据预处理、spss基本统计分析、多重反应分析、均值比较与检验、统计图制作、参数检验、回归分析、方差分析、相关分析、聚数分析、判别分析、因子分析、对应分析与结合分析、信度分析、生存分析、对数线性模型、时间序列分析、缺失值分析，以及spss在财务智能、数据预测、股市分析、社会经济分析、金融数据分析等方面的数据挖掘应用。

本书最大特点是抛弃了其他同类书籍中只介绍理论用法、缺乏案例分析的弊病，全书给出许多数据挖掘分析案例，为读者展示spss在数据分析、信用风险管理、直销分析、社会经济分析等实际项目中的应用技术。

本书适合自然科学和社会科学各领域、各专业的数据分析人员，也可供相关专业本科生、研究生、科技人员和事业单位工作人员，以及从事数据挖掘、金融分析、商业咨询、财务分析的人员使用。

## &lt;&lt;SPSS统计分析与数据挖掘&gt;&gt;

## 书籍目录

## 第1章 spss软件概述

## 1.1 spss简介

## 1.2 spss操作入门

## 1.2.1 软件安装、启动及退出

## 1.2.2 操作环境

## 1.2.3 系统参数的设置

## 1.3 spss的帮助系统

## 第2章 spss数据挖掘系统

## 2.1 数据挖掘概述

## 2.1.1 数据挖掘的含义

## 2.1.2 数据挖掘与olap

## 2.1.3 数据挖掘和统计学

## 2.1.4 数据挖掘的目的

## 2.1.5 数据挖掘应用

## 2.1.6 数据挖掘流程

## 2.2 成功的数据挖掘

## 2.2.1 crisp-dm方法论

## 2.2.2 选择数据挖掘工具

## 2.2.3 spss数据挖掘

## 2.3 spss数据挖掘的过程

## 2.3.1 商业理解

## 2.3.2 数据理解

## 2.3.3 数据准备

## 2.3.4 数据模型

## 2.3.5 评估

## 2.3.6 部署

## 第3章 数据文件、变量与函数

## 3.1 spss的变量类型

## 3.1.1 数据的输入

## 3.1.2 变量的编辑

## 3.2 数据文件的打开和保存

## 3.2.1 打开spss数据文件

## 3.2.2 打开其他格式的数据文件

## 3.2.3 数据文件保存

## 3.3 spss函数

## 3.3.1 算术函数

## 3.3.2 统计函数

## 3.3.3 逻辑函数

## 3.3.4 日期和时间函数

## 3.3.5 随机变量函数

## 3.3.6 反分布函数

## 3.3.7 累计分布函数

## 3.3.8 缺失值函数

## 3.3.9 字符串函数

## 第4章 数据预处理

## &lt;&lt;SPSS统计分析与数据挖掘&gt;&gt;

## 4.1 数据文件的整理

- 4.1.1 排序个案 (sort case) 过程
- 4.1.2 转置 (transpose) 过程
- 4.1.3 合并文件 (merge file) 过程
- 4.1.4 分类汇总 (aggregate) 过程
- 4.1.5 拆分文件 (split file) 过程
- 4.1.6 选择个案 (select cases) 过程
- 4.1.7 加权个案 (weight cases) 过程

## 4.2 数据变量的变换和计算

- 4.2.1 计算变量 (compute variables) 过程
- 4.2.2 计数 (count) 过程
- 4.2.3 重新编码 (recode) 过程
- 4.2.4 个案排秩 (rank cases) 过程
- 4.2.5 自动重新编码 (automatic recode) 过程

## 第5章 基本统计分析

## 5.1 基本概念

- 5.1.1 基本的统计概念
- 5.1.2 描述性统计分析

## 5.2 频数分析

- 5.2.1 frequencies过程的操作界面
- 5.2.2 实例分析

## 5.3 描述性统计分析过程

- 5.3.1 descriptive过程的参数设置
- 5.3.2 实例分析

## 5.4 数据探索性分析过程

- 5.4.1 explore过程的参数设置
- 5.4.2 实例分析

## 5.5 列联表分析过程

- 5.5.1 crosstabs过程的参数设置
- 5.5.2 实例分析

## 第6章 参数检验

## 6.1 参数估计和假设检验概述

- 6.1.1 参数估计
- 6.1.2 假设检验

## 6.2 均值 (means) 过程

- 6.2.1 spss的means过程参数设置
- 6.2.2 means过程实例

## 6.3 单样本t检验

- 6.3.1 one-sample t test过程的参数设置
- 6.3.2 实例分析

## 6.4 独立两样本t检验

- 6.4.1 independent-sample t test过程的参数设置
- 6.4.2 实例分析

## 6.5 配对两样本t检验

- 6.5.1 paired-sample t test过程的参数设置
- 6.5.2 实例分析

## 第7章 基本图形的绘制

## &lt;&lt;SPSS统计分析与数据挖掘&gt;&gt;

- 7.1 统计图概述
- 7.2 条形图
- 7.3 线图
- 7.4 面积图
- 7.5 饼图
  - 7.5.1 spss设置
  - 7.5.2 实例分析
- 7.6 高低图
- 7.7 质量控制图
- 7.8 箱图
  - 7.8.1 spss参数设置
  - 7.8.2 实例分析
- 7.9 散点图
  - 7.9.1 spss参数设置
  - 7.9.2 实例分析
- 7.10 直方图
  - 7.10.1 p-p图和q-q图
- 7.11 时间序列图
  - 7.11.1 spss参数设置
  - 7.11.2 实例分析
- 第8章 非参数检验
  - 8.1 非参数检验概述
  - 8.2 检验
    - 8.2.1 检验的参数设置
    - 8.2.2 检验实例分析
  - 8.3 二项分布检验
    - 8.3.1 二项分布检验的参数设置
    - 8.3.2 实例分析
  - 8.4 游程检验
    - 8.4.1 游程检验的参数设置
    - 8.4.2 实例分析
  - 8.5 单样本k-s检验
    - 8.5.1 单样本k-s检验的参数设置
    - 8.5.2 实例分析
  - 8.6 两独立样本分布位置检验
    - 8.6.1 两独立样本分布位置检验的参数设置
    - 8.6.2 实例分析
  - 8.7 多个独立样本分布位置检验
    - 8.7.1 spss参数设置
    - 8.7.2 实例分析
  - 8.8 两相关样本分布位置检验
    - 8.8.1 spss参数设置
    - 8.8.2 实例分析
  - 8.9 多个配对样本分布位置检验
    - 8.9.1 spss参数设置
    - 8.9.2 实例分析
- 第9章 方差分析

## &lt;&lt;SPSS统计分析与数据挖掘&gt;&gt;

- 9.1 方差分析的基本原理
  - 9.1.1 自由度与平方和分解
  - 9.1.2 t检验
  - 9.1.3 多重比较
- 9.2 单因素方差分析
  - 9.2.1 单因素方差分析步骤
  - 9.2.2 判断与结论
  - 9.2.3 one-way anova过程的参数设置
  - 9.2.4 实例分析
- 9.3 多因素方差分析
  - 9.3.1 只考虑主效应的多因素方差分析
  - 9.3.2 存在交互效应的多因素方差分析
  - 9.3.3 univariate过程参数设置
  - 9.3.4 实例分析
- 9.4 协方差分析
  - 9.4.1 协方差分析概述
  - 9.4.2 实例分析
- 第10章 回归分析
  - 10.1 线性回归
    - 10.1.1 线性回归模型
    - 10.1.2 最小二乘估计
    - 10.1.3 回归方程的显著性检验
    - 10.1.4 预测问题
    - 10.1.5 spss线性回归分析设置
    - 10.1.6 回归分析模型的实例分析
  - 10.2 非线性回归
    - 10.2.1 非线性回归分析的基本原理
    - 10.2.2 非线性回归参数设置
    - 10.2.3 案例——销售数量和广告投入的非线性回归分析
  - 10.3 logistic回归
    - 10.3.1 logistic回归模型概述
    - 10.3.2 binary logistic回归模型参数设置
    - 10.3.3 案例——银行贷款的信用风险分析
- 第11章 相关分析
  - 11.1 相关分析概述
    - 11.1.1 相关关系
    - 11.1.2 相关图形和相关系数
    - 11.1.3 spss的相关分析功能简介
  - 11.2 bivariate过程
    - 11.2.1 两变量相关分析简介
    - 11.2.2 bivariate过程的参数设置
    - 11.2.3 案例——汽车销售中各变量之间的相关分析
  - 11.3 partial过程
    - 11.3.1 partial过程的参数设置
    - 11.3.2 案例——医疗门诊病人的流动情况分析
  - 11.4 distances (距离) 过程
    - 11.4.1 距离分析参数设置

## &lt;&lt;SPSS统计分析与数据挖掘&gt;&gt;

11.4.2 案例——全国各个省市自治区直辖市的农民家庭收支的分布研究

## 第12章 聚类分析

## 12.1 聚类分析的原理

12.1.1 一般原理

12.1.2 聚类分析步骤

12.1.3 系统聚类方法

## 12.2 快速样本聚类过程

12.2.1 快速聚类简介

12.2.2 spss快速聚类的设置

12.2.3 案例——2006年中国主要城市空气质量的比较分析

## 12.3 分层聚类过程

12.3.1 分层聚类简介

12.3.2 spss分层聚类设置

12.3.3 案例——中国经济地理区域的聚类分析

## 12.4 两阶段聚类分析

12.4.1 两阶段聚类简介

12.4.2 spss两阶段聚类的设置

12.4.3 案例——两阶段聚类分析应用于农村居民人均收入与生活消费支出研究

## 第13章 判别分析

## 13.1 判别分析的基本原理

13.1.1 判别分析简介

13.1.2 判别分析的数学模型与判别方法

## 13.2 一般判别分析

13.2.1 一般判别分析的参数设置

13.2.2 案例——上市公司类型的比较分析

## 13.3 逐步判别分析

13.3.1 逐步判别的参数设置

13.3.2 案例——全国各个省市地区的农民家庭收支的分布规律研究

## 第14章 因子分析

## 14.1 因子分析简介

14.1.1 因子分析的基本原理

14.1.2 因子分析的基本步骤和过程

## 14.2 spss软件进行因子分析

14.2.1 spss因子分析的参数设置

14.2.2 案例——汽车销售的数据中各变量的因子分析

## 第15章 对应分析

## 15.1 对应分析的基本原理

## 15.2 简单对应分析

15.2.1 correspondence analysis过程

15.2.2 案例——简单对应分析实例

## 15.3 optimal scaling过程

15.3.1 optimal scaling过程的spss参数设置

15.3.2 案例——最优尺度分析实例

## 第16章 可靠性和多维尺度分析

## 16.1 可靠性分析

16.1.1 可靠性分析的基本原理

16.1.2 可靠性分析的参数设置

## &lt;&lt;SPSS统计分析与数据挖掘&gt;&gt;

- 16.1.3 案例——电视节目调查数据可靠性分析
- 16.2 多维尺度分析
  - 16.2.1 多维尺度分析简介
  - 16.2.2 alsca过程的参数设置
  - 16.2.3 案例——多维尺度实例分析
- 第17章 生存分析
  - 17.1 生存分析简介
    - 17.1.1 生存分析的基本概念
    - 17.1.2 生存资料的特点
    - 17.1.3 生存分析方法
    - 17.1.4 spss中的生存分析过程
  - 17.2 life tables过程
    - 17.2.1 life tables过程的参数设置
    - 17.2.2 案例——电信公司客户流失分析
  - 17.3 kaplan-meier分析
    - 17.3.1 kaplan-meier分析的步骤
    - 17.3.2 案例——新药开发的数据分析
  - 17.4 cox模型回归分析
    - 17.4.1 cox回归模型
    - 17.4.2 cox模型参数设置
    - 17.4.3 案例——电信公司的客户流失的cox回归模型分析
- 第18章 对数线性模型
  - 18.1 对数线性模型概述
  - 18.2 general过程
    - 18.2.1 general过程的参数设置
    - 18.2.2 实例分析
  - 18.3 logit过程
    - 18.3.1 logit过程的参数设置
    - 18.3.2 实例分析
  - 18.4 模型 ( model ) selection过程
    - 18.4.1 模型 ( model ) selection的参数设置
    - 18.4.2 实例分析
- 第19章 时间序列分析
  - 19.1 时间序列概述
    - 19.1.1 时间序列的组成部分
    - 19.1.2 时间序列的数学模型
    - 19.1.3 时间序列的分析步骤
    - 19.1.4 spss时间序列分析功能
  - 19.2 时间序列数据的预处理
    - 19.2.1 缺失值替换
    - 19.2.2 定义时间变量
    - 19.2.3 时间序列的平稳化
  - 19.3 指数平滑模型过程
    - 19.3.1 指数平滑的基本原理
    - 19.3.2 指数平滑模型的参数设置
    - 19.3.3 案例——销售数据预测分析
  - 19.4 arima模型



## &lt;&lt;SPSS统计分析与数据挖掘&gt;&gt;

- 19.4.1 arima模型的基本原理
- 19.4.2 arima模型的参数设置
- 19.4.3 案例——上海证券交易所综合指数收益率模拟预测分析
- 19.5 季节分解模型过程
  - 19.5.1 季节分解模型的参数设置
  - 19.5.2 案例——具有季节因素的服装销售数据的预测分析
- 第20章 缺失值分析
  - 20.1 缺失值理论概述
    - 20.1.1 数据缺失方式
    - 20.1.2 缺失值处理方法
  - 20.2 spss缺失值分析操作
    - 20.2.1 缺失值分析的参数设置
    - 20.2.2 案例——数据集中缺失值的实例分析
- 第21章 决策树模型
  - 21.1 决策树模型概述
    - 21.1.1 chaid算法
    - 21.1.2 exhaustive chaid算法
    - 21.1.3 cart算法
    - 21.1.4 quest算法
  - 21.2 决策树的参数设置
    - 21.2.1 变量设置
    - 21.2.2 类别 ( categories ) 设置
    - 21.2.3 输出 ( output ) 设置
    - 21.2.4 验证 ( validation ) 设置
    - 21.2.5 保存 ( save ) 设置
    - 21.2.6 条件 ( criteria ) 设置
    - 21.2.7 chaid算法设置
    - 21.2.8 cart算法设置
    - 21.2.9 quest算法设置
    - 21.2.10 修剪 ( pruning ) 设置
    - 21.2.11 替代变量 ( surrogates ) 设置
    - 21.2.12 选项 ( options ) 设置
    - 21.2.13 误分类成本设置
    - 21.2.14 利润 ( profits ) 设置
    - 21.2.15 先验概率 ( prior probabilities ) 设置
    - 21.2.16 实例分析
    - 21.2.17 模型建立
    - 21.2.18 模型评估
- 第22章 神经网络
  - 22.1 神经网络概述
    - 22.1.1 历史及现状
    - 22.1.2 神经网络特点
    - 22.1.3 神经元模型
    - 22.1.4 神经网络模型
    - 22.1.5 神经网络的学习规则
    - 22.1.6 spss神经网络模型
  - 22.2 spss神经网络模型的设置

## <<SPSS统计分析与数据挖掘>>

- 22.2.1 多层感知器 ( mlp ) 的设置
- 22.2.2 径向基函数 ( rbf ) 的设置
- 22.3 实例分析
  - 22.3.1 参数设置
  - 22.3.2 结果分析
- 第23章 信用风险分析
  - 23.1 信用风险概述
    - 23.1.1 信用风险基本概念
    - 23.1.2 信用风险度量方法
  - 23.2 案例分析
    - 23.1.3 spss中信用风险分析模块
    - 23.2.1 二元logistic过程
    - 23.2.2 决策树 ( tree ) 过程
    - 23.2.3 discriminant过程
- 第24章 spss在社会经济综合评价中的应用
  - 24.1 沿海省市经济综合指标的主成分分析
  - 24.2 中国城镇居民消费结构的聚类分析
  - 24.3 我国内地可支配收入和消费性支出之间的回归分析

## &lt;&lt;SPSS统计分析与数据挖掘&gt;&gt;

## 章节摘录

版权页：插图：数据挖掘过程中，数据挖掘的工具同样也是很重要的，下面举几个例子。

寻找一个已被证实的可用于解决项目所陈述商务问题的数据挖掘工具：即选择一种所知道的，可以用于解决公司问题，并在计划应用方面有成功记录的工具。

选择用于在商务理解和数据挖掘技术方面起到沟通作用的工具：确保工具所用的步骤与数据挖掘的商务需要相匹配。

· 工具是否可清晰地表达数据挖掘概念；· 工具是否与项目管理软件或其他可能用的工具相结合？如果不能，是否不得不新建应用软件以弥补此不足。

确保工具可对现有的数据资源和格式进行操作：如果能选择一种能提取和合并多来源、多格式数据，将会节省时间和金钱，并最大可能地得到可靠结果的工具。

这一点很重要，尤其是在数据挖掘过程的后期发现不得不从新的来源加入新的数据时。

寻找交互式开发和可视化能力：选择一种可提供交互式可视技术的工具会使开发和理解数据变得容易。

这些技术会使你通过在图内变化及根据不同的数据尺度产生新的图表更快地获得直觉。

选择一种可高效、易解的进行数据准备的工具：选择一种可高效进行数据准备（从初始步骤到模型建立），且以易于理解的方式表达数据准备步骤的工具会节省时间和资源。

这会使不同经验水平的项目成员获得有效的结果。

确保工具可自动地提取数据：选择一种可为不同数据步骤自动提取数据的工具，可以避免耗时的人工书写查询。

该工具是否可在合理的时间内建立有效的模型：寻找一种工具，它可使分析家快速找到最有效的模型。

这种工具应该支持有效的建立和检测多个模型。

选择一种含宽范围技术的工具：为了确保最好的结果，确保工具能为可视化、分类、聚类、相关和回归提供一个宽范围的技术或运算法则。

例如，可能发现，对某一数据而言，一种技术比另一种更好。

你需要能灵活地试用多种技术以获得准确、有效的结果。

这种工具应该能联合应用在不同情况下可获最佳结果的多种技术。

该工具是否可利用现有的数据和设备：选择一种数据挖掘工具，它能利用现存的数据——或数据库中或文件中，也能与现有的分析和可视化工具相兼容。

你不会愿意因为不能利用现有的数据库而浪费时间和资源再新建一个。

选择一种可发送一致的、高质的结果的工具：要得到准确的结果，需利用在各种情形下都能很好工作的、适应性强的数据挖掘环境和各种数据的工具，而不是单一地为某一类型的数据或环境而设计的工具。

工具应该能管理任何可能有助于阐释商务问题的数据。

## <<SPSS统计分析与数据挖掘>>

### 编辑推荐

《SPSS统计分析与数据挖掘》：SPSS-全球使用最为广泛的统计分析软件，SPSS-金融、科研、数据挖掘领域必备之软件，基础知识-实训实例-工程实例，实例操作视频教学，轻松学习。

<<SPSS统计分析与数据挖掘>>

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>