第 1 章　SPSS Statistics 概述

1.1　SPSS Statistics 简介

SPSS(Statistical Product and Service Solutions，统计产品和服务解决方案，原名 Statistical Package for the Social Science，社会科学统计软件包)是由美国 SPSS 公司自 20 世纪 80 年代初开发的大型统计学系列软件。2009 年 4 月，SPSS 公司宣布重新包装旗下的 SPSS 产品线，定位为预测分析软件(PASW，Predictive Analytics Software)，包括统计分析(PASW Statistics 17.0)、数据挖掘(PASW Modeler，原名 Clementine)、数据收集(Data Collection 系列软件，原名 Dimensions)、结果发布(PASW Collaboration and Deployment Services，原名 Predictive Enterprise Services)四大部分，并支持多国语言。2009 年 8 月 IBM 宣布收购 SPSS 公司，自 2010 年 8 月发行 19.0 版本开始，SPSS 正式更名为 IBM SPSS Statistics(本书均简称“SPSS”)，目前最新版本是 2015 年 3 月发行的 SPSS 23.0 多国语言版，其用户界面和结果输出均可使用简体中文、繁体中文版和英文版等 12 种语言。多国语言版的推出消除了非英语国家用户在语言方面的很多困扰，人机界面特别友好，是广大用户的一大福音。

1.2　数 据 管 理

SPSS 具有强大的数据管理功能，共有 28 种，包括定义变量属性(Define Variable Properties)、设置未知测量级别(Set Measurement Level for Unknown)、复制数据属性(Copy Data Properties)、新建定制属性(New Custom Attribute)、定义日期(Define Dates)、定义多响应集(Define Multiple Response Sets)、验证(Validation)[包括加载预定义规则(Load Predefined Rules)、定义规则(Define Rules)和验证数据(Validate Data)]、标识重复个案(Identify Duplicate Cases)、标识异常个案(Identify Unusual Cases)、比较数据集(Compare Datasets)、排序个案(Sort Cases)、排序变量(Sort Variables)、变换(Transpose)、合并文件(Merge Files)[包括添加个案(Add Cases)、添加变量(Add Variables)]、重组(Restructure)、搜索权重(Rake Weights)、倾向得分匹配(Propensity Score Matching)、个案控制匹配(Case Control Matching)、汇总数据(Aggregate Data)、拆分为文件(Split Into Files)、正交设计(Orthogonal Design)、复制数据集(Copy Dataset)、拆分文件(Split File)、选择个案(Select Cases)及加权个案(Weight Cases)等。

1.3　数 据 变 换

SPSS 共提供 19 种数据变换功能，包括计算变量(Compute Variable)、可编程性变换(Programmability Transformation)、对个案内的值计数(Count Occurrences of Values within Cases)、转换值(Shift Values)、重新编码为相同变量(Recode into Same Variables)、重新编码为不同变量(Recode into Different Variables)、自动重新编码(Automatic Recode)、创建虚拟变量(Create

Dummy Variables)、可视分箱化(Visual Binning)、最优分箱化(Optimal Binning)、准备建模数据(Prepare Data for Modeling)[包括交互式数据准备(Interactive Data Preparation)、自动数据准备(Automatic Data Preparation)及逆转换得分(Backtransform Scores)]、个案等级排序(Rank Cases)、日期和时间向导(Date and Time Wizard)、创建时间序列(Create Time Series)、替换缺失值(Replace Missing Values)、随机数字生成器(Random Number Generators)及运行挂起的转换(Run Pending Transforms)。

1.4 统计分析

SPSS 统计分析(Analyze)方法共有 24 大类、120 个小类。

(1)报告(Reports):代码本(Codebook)、OLAP 多维数据集(OLAP Cubes)、个案汇总(Case Summaries)、按行汇总(Report Summaries In Rows)和按列汇总(Report Summaries In Columns)。

(2)描述统计(Descriptive Statistics):频率分析(Frequencies)、描述性分析(Descriptives)、探索分析(Explore)、列联表(交叉表)分析(Crosstabs)、TURF 分析(Total Unduplicated Reach and Frequency,累积不重复到达率和频次分析)、比率统计(Ratio Statistics)、P-P 图(P-P Plots,proportion-proportion plot)、Q-Q 图(Q-Q Plots,Quantile-Quantile plot)。

(3)表格(Custom Tables):定制表(Custom Tables)和多响应集(Multiple Response Sets)。

(4)比较平均值(Compare Means):平均值(Means)分析、单样本 t 检验(One-Sample T Test)、独立样本 t 检验(Independent-Samples T Test)、配对样本 t 检验(Paired-Samples T Test)和单向方差分析(One-Way ANOVA)。

(5)一般线性模型(General Linear Model):单变量方差分析(Univariate Analysis of Variance)、多元方差分析(Multivariate Analysis of Variance)、重复测量方差分析(Repeated Measures Analysis of Variance)和方差分量分析(Variance Components Analysis)。

(6)广义线性模型(Generalized Linear Models):广义线性模型(Generalized Linear Models)和广义估计方程(Generalized Estimating Equations)。

(7)混合模型(Mixed Models):线性混合模型(Linear Mixed Models)和广义线性混合模型(Generalized linear mixed models)。

(8)相关(Correlate):双变量相关(Bivariate Correlation)、偏相关(Partial Correlation)和距离(Distances)相关。

(9)回归(Regression):自动线性建模(Automatic Linear modeling)、线性回归(Linear Regression)、曲线估计(Curve Estimation)、偏最小二乘回归(Partial Least Squares Regression)、二元 Logistic 回归(Binary Logistic Regression)、多元 Logistic 回归(Multinomial Logistic Regression)、有序回归(Ordinal Regression)、概率单位法(Probit,probability unit)、非线性回归(Nonlinear Regression)、权重估计法(Weight Estimation)、两步最小二乘回归(2-Stage Least Squares Regression)及分类回归(Categorical Regression)。

(10)对数线性模型(Loglinear):一般对数线性分析(General Loglinear Analysis),Logit 对数线性分析(Logit Loglinear Analysis)和模型选择对数线性分析(Model Selection Loglinear Analysis)。

(11)神经网络(Neural Networks):多层感知器(Multilayer Perceptron)和径向基函数(Radial Basis Function)。

(12)分类分析(Classify)：两步聚类分析(Two-Step Cluster Analysis)、逐步聚类分析(K-Means Cluster Analysis)、系统聚类分析(Hierarchical Cluster Analysis)、决策树(Decision Trees)、判别分析(Discriminant Analysis)及最近邻分析(Nearest Neighbor Analysis)。

(13)降维分析(Dimension Reduction)：因子分析(Factor analysis)，对应分析(Correspondence analysis)和最优尺度(Optimal Scaling)分析[多重对应分析(Multiple Correspondence Analysis，MCA)、分类主成分分析(Categorical Principal Components Analysis，CATPCA)、非线性典型相关分析(Nonlinear Canonical Correlation Analysis，OVERALS)]。

(14)尺度分析(Scale)：可靠性分析(Reliability Analysis)、多维尺度分析(Multidimensional Scaling Analysis，ALSCAL)和多维邻近尺度分析(Multidimensional Scaling Analysis，PROXSCAL)及多维展开分析(Multidimensional Unfolding Analysis，PREFSCAL)。

(15)非参数检验(Nonparametric Tests)：单样本非参数检验(One-Sample Nonparametric Tests)，两个或更多独立样本非参数检验(Two or More Independent Samples Nonparametric Tests)，两个或更多相关样本非参数检验(Two or More Related Samples Nonparametric Tests)，卡方检验(Chi-Square Test)，二项检验(Binomial Test)，游程检验(Runs Test)，单样本 Kolmogorov-Smirnov 检验(One-Sample Kolmogorov-Smirnov Test)，两独立样本非参数检验(Two-Independent-Samples Test)[Mann-Whitney U 检验(Mann-Whitney U test)、Moses 极端反应检验(Moses extreme reactions test)、Kolmogorov-Smirnov Z 检验(Kolmogorov-Smirnov Z test)、Wald-Wolfowitz 游程检验(Wald-Wolfowitz runs test)]，多个独立样本非参数检验(Tests for Several Independent Samples)[Kruskal-Wallis H 检验(Kruskal-Wallis H Test)、中位数检验(Median Test)和 Jonckheere-Terpstra 检验(Jonckheere-Terpstra Test)]，两相关样本非参数检验(Two-Related-Samples Tests)[Wilcoxon 符号秩检验(Wilcoxon Signed Ranks Test)、符号检验(Signed Test)、McNemar 检验(McNemar Test)和边际同质性检验(Marginal Homogeneity Test)]，多个相关样本非参数检验(Test for Several Related Samples)[Friedman 检验(Friedman Test)、Kendall W 检验(Kendall's W Test)和 Cochran Q 检验(Cochran's Q Test)]。

(16)预测(Forecasting)：时间序列建模器(Time Series Modeler)[专家建模器(Expert Modeler)、指数平滑法(Exponential Smoothing)]，综合自回归移动平均模型(ARIMA)，季节分解法(Seasonal Decomposition)，谱分析(Spectral Analysis)，序列图(Sequence Charts)，自相关(Autocorrelations)和互相关(Cross-Correlations)图。

(17)生存分析(Survival)：寿命表(Life Tables)、Kaplan-Meier 法(Kaplan-Meier)、Cox 回归(Cox Regression)和含时间依赖协变量的 Cox 回归(Time-Dependent Cox Regression)。

(18)多响应分析(Multiple Response)：定义多响应集(Define Sets)、多响应频率分析(Multiple Response Frequencies)和多响应交叉表(Multiple Response Crosstabs)分析。

(19)缺失值分析(Missing Value Analysis)。

(20)多重插补(Multiple Imputation)：分析模式(Analyze Patterns)和插补缺失数据值(Impute Missing Data Values)。

(21)复杂抽样(Complex Sample)：选择样本(Select a Sample)，准备分析(Prepare for Analysis)及统计分析的复杂抽样计划(Complex Sample Plan)[频率分析(Frequency)、描述性分析(Descriptive)、交叉表(Crosstabs)分析、比率分析(Ratios)、一般线性模型(General Linear Model)、Logistic 回归(Logistic Regression)、有序回归(Ordinal regression)和 Cox 回归(Cox Regression)]。

(22)模拟(Simulation)：用现有模型文件或手动输入方程创建应用于统计模型的数据。

(23)质量控制(Quality Control):控制图(Control Chart)[平均值、极差、标准差控制图(X-Bar, R, s Control Chart)],单值、移动极差控制图(Individuals, Moving Control Chart),不合格品率、不合格品数控制图(p, np Control Chart)和缺陷数、单位缺陷数控制图(c, u Control Chart),帕累托图(Pareto Chart)[简单帕累托图(Simple Pareto Chart)和堆积帕累托图(Stacked Pareto Chart)]。

(24)ROC 曲线(ROC Curve)。

1.5 直销分析

SPSS 提供 8 个直销分析(Direct Marketing)程序:交易数据 RFM 分析(RFM Analysis from Transaction Data)、客户数据 RFM 分析(RFM Scores from Customer Data)、聚类分析(Cluster Analysis)、潜在客户概要文件(Prospect Profiles)、邮政编码响应率(Postal Code Response Rate)、购买倾向分析(Propensity to Purchase)、控制包装检验(Control Package Test)及评分向导(Scoring Wizard)。

1.6 绘 图

绘图(Graphs)模块能简明生动、形象直观地表达统计资料。SPSS 的绘图功能非常强大,在统计分析过程中可选择多种图形,也可直接由绘图菜单产生,并加以修饰、编辑。

1)图表构建器(Chart Builder):条形图(Bar)、折线图(Line)、面积图(Area)、饼图/极坐标图(Pie/Polar)、散点图/点图(Scatter/Dot)、直方图(Histogram)、高低图(High-Low)、箱图(Boxplot)和双轴图(Dual Axes)等。

2)图形画板模板选择程序(Graphboard Template Chooser):表面(Surface)图、饼图(Pie)、参考地图上的箭头(Arrows on a Reference Map)、参考地图上的坐标(Coordinates on a Reference Map)、带有正态分布的直方图(Histogram with Normal Distribution)、带状图(Ribbon)、地图上的饼图(Pie on a Map)、地图上的计数饼图(Pie of Counts on a Map)、地图上的计数条形图(Bar of Counts on a Map)、地图上的条形图(Bar on a Map)、地图上的线图(Line Chart on a Map)、点图(Dot Plot)、点状重叠地图(Point Overlay Map)、多边形重叠地图(Polygon Overlay Map)、二维点图(2-D Dot Plot)、和分区图(Choropleth of Sums)、和分区图上的坐标(Coordinates on a Choropleth of Sums)、计数饼图(Pie of Counts)、计数的分区图(Choropleth of Counts)、计数分区图上的坐标(Coordinates on a Choropleth of Counts)、计数条形图(Bar of Counts)、聚类箱图(Clustered Boxplot)、平均值分区图(Choropleth of Mens)、平均值分区图上的坐标(Coordinates Choropleth of Mens)、分箱化散点图(Binned Scatterplot)、六边形分箱化散点图(Hex Binned Scatterplot)、路径(Path)图、面积图(Area)、平行(Parallel)图、气泡图(Bubble Plot)、热图(Heat Map)、三维饼图(3-D Pie)、三维密度(3-D Density)图、三维面积图(3-D Area Chart)、三维散点图(3-D Scatterplot)、三维条形图(3-D Bar)、三维直方图(3-D Histogram)、散点图(Scatterplot)、散点图矩阵(SPLOM)(Scatterplot Matrix)、条形图(Bar)、线图(Line)、线形重叠地图(Line Overlay Map)、箱图(Boxplot)、直方图(Histogram)、值分区图(Choropleth of Values)、值分区图上的坐标(Coordinates on a Choropleth of Values)、中位数分区图(Choropleth of Medians)及中位数分区图上的坐标(Coordinates on a Choropleth of Medians)。

3)比较子组(Compare Subgroups)。

4)回归变量图(Regression Variable Plots)。

5)旧对话框(Legacy Dialogs):此菜单栏兼容 SPSS 17.0 及以前版本的对话框。

(1)条形图(Bar Chart):简单条形图(Simple Bar Chart)、复式条形图(Clustered Bar Chart)和堆积条形图(Stacked Bar Chart)。

(2)三维条形图(3-D Bar Charts)。

(3)线图(Line Chart):简单线图(Simple Line Chart)、多线图(Multiple Line Chart)和下降线图(Drop-line Line Chart)。

(4)面积(区域)图(Area):简单面积图(Simple Area Chart)和堆积面积图(Stacked Area Chart)。

(5)饼图(Pie Chart)。

(6)高低图(High-Low Chart):简单高低收盘图(Simple high-low-close Chart)、复式高低收盘图(Clustered high-low-close Chart)、差别面积图(Difference Area Chart)、简单极差图(Simple range bar Chart)和复式极差图(Clustered range bar Chart)。

(7)箱图(Boxplot):简单箱图(Simple Boxplot)和复式箱图(Clustered Boxplot)。

(8)误差条图(Error Bar):简单误差条图(Simple Error Bar)和复式误差条图(Clustered Error Bar)。

(9)人口金字塔(Population Pyramids)图。

(10)散点图/点图(Scatterplot):简单散点图(Simple Scatterplot)、重叠散点图(Overlay Scatterplot)、散点图矩阵(Scatterplot Matrix)、简单点图(Simple Dot)和三维散点图(3-D Scatterplot)。

(11)直方图(Histogram)。

总之,SPSS 22.0 比以往版本具有更丰富的统计和绘图功能,可读性更强,易学易用。

练习题

(请访问 www.hxedu.com.cn 下载。)