



SPSS基础培训

新浪微博：@数据挖掘与数据分析



目录

➤ 数据分析的流程

➤ 数据编辑

➤ 基本统计分析

➤ 交叉表



■ 易用性强

- 操作界面极为友好，操作简单
- 良好的帮助系统和自学功能
- 为高级用户提高编程功能

■ 功能强大

- 成熟的统计过程
- 完美的图形处理功能
- 提供多种数据准备技术

■ 兼容性好

- 数据输入: Excel, Lotus, Oracle, SQL Server, Access, dBASE, 文本
- 数据输出: Word, HTML, XML, Excel, Powerpoint , PDF.



数据的编辑

- 常量

- 数值型常量：除了普通写法外还可以用科学计数法，如：1.3E18；
- 字符型常量：用单引号或双引号括起来如果字符中包含单引号，则必须使用双引号；
- 日期常量：日期个数的数据，一般需要使用日期函数进行转换；

- 变量

- 变量名长度不能超过8；
- 三种基本的类型：数值、字符和日期；
- 可以在variable view界面设定变量的长度及小数位、变量的描述、变量值的描述、missing值、显示宽度、对齐方式和变量的测度方式；



数据的编辑

- 变量的测试方式
 - Scale：定距变量，如：身高、体重等；
 - Ordinal：定序变量，如：教育程度、级别等；
 - Nominal：定类变量，如：性别、民族等；
- 操作符与表达式
 - 三种基本的运算：数学、关系和逻辑
 - 数学运算符： $+$ $-$ $*$ $/$ $**$ $()$
 - 关系运算符： $>$ $>=$ $<$ $<=$ $=$ $\sim=$
 - 逻辑运算符： $\&$ (AND) $|$ (OR) \sim (NOT)
 - 三种运算对应三种表达式



- 常用的数据操作命令
 - Data->Sort Cases
 - Transform->Rank Cases
 - Transform->Count
 - Transform->Recode
 - Transform->Automatic Recode
 - Transform->Compute
 - Data->Transpose
 - Data->Split Files
 - Data->Merge Files



数据的编辑

- Compute
 - 数值型 : `compute num1=value.`
 - 字符型 : `String A(a11).`
`compute a='hello world'.`
 - 日期型 : `compute data1=date.mdy(month,day,year).`
- Rocode
 - `recode variable name(old value=new value).`
 - `recode variable name(old value=new value) into new variable name.`
 - 字符型变量使用 `auto recode`



数据的编辑

- Split file
 - 有的时候需要对变量做些分组分析，但一些分析方法并不提供分组变量的设置选项这就需要用到Split file命令；
 - 例如使用 Descriptives 做描述性分析，如果想分年龄做分析，这样就可以用年龄变量作为分组变量；
 - 可以看到这里的Split其实是分组，而不是拆分文件；
 - analyze all case分析所有的样本，不产生分组；
 - compare groups产生对比分析组；
 - output by groups分组输入分析结果；



数据的编辑

- Merge File
 - add cases 合并变量相同，但是case不同的文件；
 - add variables 合并变量不同，case相同的文件这里的变量不同可以是部分的变量不同，case相同也可以是一个文件的case是另外一个文件的子集；



数据的编辑

- 数据的分类汇总
 - 使用Aggregate命令
 - 指定分类变量对观测量进行分组，对每组观测量的各变量求描述统计量；
- 检查重复的数据
 - 使用identify duplicate cases
- 数据的加权
 - 使用weight case
- 选取一定的case进行分析
 - 使用select cases：在对数据的子集进行分析的时候需要用到这个命令；



- 常用的数学函数
 - 取绝对值：abs（数字型表达式）
 - 求余数函数：mod（数字型表达式，模数），模数不能为0该函数在需要对某一变量求模数的余数时使用，如果对一个顺序编号或自然数序列求模数的余数，可将该序列按模数等距分类，从而实行等距抽样；
 - 四舍五入函数：rnd（数字型表达式）
 - 开方函数：sqrt（数字型表达式）

基本的统计分析



基本统计分析

- SPSS统计分析概述：
 - 针对不同类型的数据选取不同的分析方法，正确的分析方法是得到正确结果的关键；
 - spss提供数字分析和图形分析两种分析形式；
 - 高级分析之前一般都需要做描述性统计分析，把握数据的规律对分析解释数据有很好的引导和帮助作用；



基本统计分析

- Descriptive Statistics
 - Frequencies : 频数分析
 - Descriptives : 描述统计
 - Explore : 探索分析
 - Crosstabs : 列联表分析
 - Ratio:比率分析



基本统计分析

- Descriptives
 - 可以对变量进行标准化；
- Explore
 - Explore是对连续性变量进行探索性分析最有效的工具；
 - 考察数据的奇异性和分布特征；
 - 箱盒图、茎叶图、正态检验图及方差齐次性检验；
- Crosstabs
 - 数据类型要求为分类变量；
 - 二维或多维交叉频数表（列联表），分析事物（变量）之间的相互影响和关系；
 - 可以做卡方检验，来分析行列变量之间是否存在相关性；



基本统计分析

- 分类变量统计描述常用指标
 - 统计量：
 - 频数、频率、累计频数、累计频率、众数
 - 比：任意两个变量之比
 - 性别比，货物/销售人员比
 - 构成比：部分占总体的比例
 - 率：事件的发生强度
 - 图形：
 - 条图、饼图



基本统计分析

- Spss操作
 - 单个变量的分析
 - Analyze...Descriptive Statistics...Frequencies
 - 多个变量的分析
 - Analyze...Descriptive Statistics...Crosstabs
 - 条图
 - Graph...(interactive...)bar
 - 饼图
 - Graph...(interactive...)pie



基本统计分析

- 连续变量的描述指标
 - 频数表Frequency
 - 直观的方法：分布类型分布特征
 - 集中趋势Central tendency
 - 均数mean 中位数median 众数mode
 - 离散趋势Dispersion tendency
 - 全距Range
 - 方差Variance 标准差std.deviation



基本统计分析

- 如何计算各个描述统计量
 - Analyze->Descriptive Statistics->Frequencies...
 - Analyze->Descriptive Statistics->Descriptives...
 - Analyze->compare means->means...
- 如何用图形描述连续变量
 - Graph...Interactive...Histogram
- 如何应用Explore对连续变量进行探索性分析
 - Analyze->Descriptive Statistics->Explore...



表

- Basic Tables过程：对分类/定量资料进行各种复杂格式的描述；
- General Tables过程：在同一张表格内同时对分类资料、连续资料和多选题数据进行汇总功能非常强大，但使用上相对复杂；
- Custom Tables过程：含有表格预览窗口，并可在制表过程中控制结果；
- Multiple Response Sets/Tables过程：专门为多选题数据设计的制表过程；
- Tables of Frequencies过程：在同一张表格中对多个分类变量同时输出频数表；

感谢大家

请大家多多关注：@数据挖掘与数据分析