R×C列联表的统计分析方法选择与SPSS实现

作者：writer 来源：企鹅号 - 精鼎统计

本文原地址：https://www.iikx.com/news/statistics/7600.html

本文仅供学习交流之用，版权归原作者所有，请勿用于商业用途！

R×C列联表的统计分析方法选择与SPSS实现
。分类资料在医学统计中很常见，有些统计学书上称为计数资料，比如(有效、无效)，(发病、不发病)，(男、女)，血型(A、B、O、AB)等等。 分类资料一般根据频数整理成列联表的形式，一般的列联表多是二维的(也称行列表，或R×C列联表，高维列联表下次讨论)，列联表根据变量是否有序可以分为双向无序、单项有序、双向有序列联表，统计方法是不同的，分析如下：

一、双向无序列联表

(一)成组四格表

是指行、列变量均为无序的列联表，例如要研究吸烟和肺癌之间的关系，行变量为是否吸烟：吸烟、不吸烟，列变量为肺癌发病：发病，不发病，如下表：

对于这种数据，我们的统计目的是分析行列变量的独立性，即：肺癌发病是否与吸烟有关，可选用的方法有以下两种：

1、Pearson卡方检验：

基于卡方分布，H0为行、列变量相互独立，SPSS中“ 分析->描述性统计->交叉表 ”可实现。 四格表使用条件：专用公式①样本总数大于40;②各个单元格理论值均大于5。校正公式：①样本总数大于40;②理论值1

2、Fisher精确概率：

基于超几何分布，当数据不满足Pearson卡方检验时使用。SPSS中“ 分析->描述性统计->交叉表 ”可实现。注意SPSS仅提供了2×2表的精确概率，需要计算R×C列联表的精确概率，可以选择精确按钮中的蒙特卡罗近似法实现。

(二)成组R×C表(双向无序)

1.Pearson卡方检验

条件：不能有任何一个格子的理论频数T

如若不符合：可以增加研究样本量(通常少用);对理论频数较小的行或者列进行合并或者删除;采用R×C表的Fisher确切概率法(通常采用蒙特卡洛近似法)

2.R×C表Fisher确切概率法

操作：分析—描述—交叉表—设置好行列变量—点击精确—选择蒙特卡洛。

二、单项有序列联表

常见的情况是结果变量有序，而原因变量无序。比如要比较AB两种药物的治疗效果，药物分组(AB)是无序的，而结果变量是有序的(无效，显效，治愈)，可以整理成如下的表格：

可以选择的统计方法主要有：

1、Mann – Whitney U检验

基于卡方分布，H0为两组总体分布一致，SPSS中“ 分析->非参数检验->独立样本 ”中可实现。

注意：在SPSS中，如果是整理成了列联表资料，需要用频数进行数据加权。变量编码为：①药物(名义)：A=1，B=2，②疗效(度量)：无效=1，有效=2，治愈=3，③频数(度量)。

2、Kruskal-Wallis H检验：

用于分组数大于2的情况，比如要比较3种或以上药物的疗效。实现方法与Mann – WhitneyU类似。

注意：当行变量为有序时，通常当作无序处理。但若行变量为有序，列变量为二分类率时，根据研究目的，也可以选择趋势性卡方检验。

三、双向有序列联表

1、行、列变量有序但属性不同

比如这个例子：要比较某种药物对某种疾病的治疗效果，按年龄段的分组，要考察治疗效果是否与年龄段相关，整理成下表：

行、列都是有序的，这是我们主要关心的：行列变量之间是否有相关性，如果有相关性，是线性相关还是曲线关系。可以选择的方法如下：

(1)Spearman等级相关：

检验有无相关性，基于卡方检验，在SPSS中“ 分析->相关->双样本 ”中可实现。

(2)线性趋势卡方检验：

检验有无线性关系，基于卡方检验，SPSS“ 分析->交叉表 ”卡方结果表格中的“ 线性和线性组合

”就是。

2、行列有序且属性相同

(1)行列变量独立

通常是为了检验一致性。比如用两台仪器对同一样本进行检验，结果分为阴性、阳性，现在要比较两台仪器的结果是否据有一致性。整理成下表：

常用的方法为：

Kappa一致性检验：

H0为行列变量无一致性。在SPSS中“ 分析->描述性统计->交叉表 ”中可实现。

(2)配对行列表

行列变量为配对资料，比如有某种药物可以缓解某种疾病的某种症状，在同一个患者身上比较用药前后的症状，评价药物的资料效果，列成下表：

可以选用的统计方法：

(1)McNemar检验：

仅用于2×2列联表。基于卡方分布。在SPSS中“ 分析->描述性统计->交叉表 ”中可实现。本检验与Pearson卡方检验具有同一性，使用条件必须满足Pearson卡方检验的条件。如果条件不能满足，需要进行Yate校正。

(2)Bowker检验：

是McNemar检验的扩展，用于分类数目大于2的配对列联表分析。在SPSS中“ 分析->描述性统计->交叉表 ”中可实现。 SPSS中依然选择的是Mcnemar，结果输出的是Mcnemar-Bowker结果。

何为行变量、何为列变量？
设置原则一般为：**实验设计三要素中的干预因素为行变量，实验效应指标为列变量**。

更多 统计方法 请访问 https://www.iikx.com/news/statistics/

本文版权归原作者所有，请勿用于商业用途，爱科学iikx.com转发