

## 交叉(卡方)算法

## 目录

Pearson 卡方 .....	1
------------------	---

交叉分析用于计算研究定类和定类数据之间的关系情况，在 SPSSAU 中支持：

- ✓ 批量对‘标题’进行分析；
- ✓ 提供交叉汇总数据，包括汇总数字、百分比等；
- ✓ 支持按行或者按列汇总百分比等；
- ✓ 提供 Pearson 卡方值。



## Pearson 卡方

在交叉分析时，SPSSAU 默认提供 Pearson 卡方值，如果需要其它指标比如 fisher 卡方，yates 校正卡方或者似然比卡方，效应量等，请使用‘医学研究’模块里面的卡方研究或者 fisher 卡方。

首先得到交叉列联表，类似如下：

	Column1	Column2	Column3	Total
Row1	10	20	30	60
Row2	15	25	40	80
Total	25	45	70	140

上表格为 2\*3 结构的列联表，共 6 个类别，接着计算 Pearson 卡方，如下：

$$\chi^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i}$$

其中：

$\chi^2$ : Pearson 卡方统计量

$O_i$ : 第  $i$  类别的观察频数 (observed frequency)

$E_i$ : 第  $i$  类别的期望频数 (expected frequency)

$k$ : 类别的总数

期望频数 $E_{ij}$ :

$$E_{ij} = \frac{R_i \times C_j}{n}$$

其中:

$E_{ij}$ : 第 $i$ 行第 $j$ 列的期望频数

$R_i$ : 第 $i$ 行的总频数 (行合计)

$C_j$ : 第 $j$ 列的总频数 (列合计)

$n$ : 总频数 (所有单元格的总和)

卡方检验的自由度 $df$ :  $df = (r - 1)(c - 1)$

其中:  $r$ 是行数,  $c$ 是列数