

行乘列表和卡方检验

行乘列表（列联表）用于分析分类数据，是观测数据按两个分类变量取值的交叉分布，如下表列出不同性别的不同文化程度的人数：

| | 文化程度低 | 文化程度中 | 文化程度高 | 合计 |
|----|-------|-------|-------|----|
| 男性 | N11 | N12 | N13 | R1 |
| 女性 | N21 | N22 | N23 | R2 |
| 合计 | C1 | C2 | C3 | N |

卡方检验

卡方检验是用于检验两个分类变量是相互关联还是彼此独立。无效假设是两变量相互独立（不相关）。在这种假设下，计算每个格子的期望频数，上表中 N11 的期望频数为： $E_{11} = R_1 \times C_1 / N$ ，然后比较实际观察值和期望值，计算卡方值。

X^2 值的计算公式为： $X^2 = \sum (O_i - E_i)^2 / E_i$

其中： O_i = 实际观察频数； E_i = 理论期望频数； n = 表中格子数

$X^2 = \text{Pearson's 统计值}$ ，近似服从 X^2 分布

如果 X^2 值的概率小于 5%，则拒绝无效假设，说明两个变量显著相关的。

Fisher 精确检验

Fisher 精确检验计算精确 P 值。卡方检验仅仅是一个近似法，因为样本分布只是近似等于理论的卡方分布。当样本量很小时或格子数据分布非常不均衡时，近似法就不适用了，此时应采用确切概率法计算精确 P 值。当样本量很大或数据平衡性好时，精确 P 值计算很困难。

配对四格表的卡方检验和 Kappa 一致性检验

比较两二分类变量（如两个诊断试验诊断结果）结果对称性，用配对四格表卡方检验。如根据两次测量血压值分别判断的是否患高血压结果，四格表结果为：

| | HBP2=0 | HBP2=1 | 合计 |
|--------|--------|--------|-----|
| HBP1=0 | a | b | a+b |
| HBP1=1 | c | d | c+d |
| 合计 | a+c | b+d | N |

两指标结果一致的格子是 a、d，不一致的是 b、c。检验 b 与 c 的差异，即检验两个结果哪个阳性率高。卡方计算公式：

$X^2 = (b-c)^2 / (b+c)$ ，当 b+c 较小时 (< 40)，宜用线性校正，公式改为：

$X^2 = (|b-c|-1)^2 / (b+c)$

配对 $\times 2$ 检验给出两种方法阳性（或阴性）检出率的差异是否具有统计学意义

Kappa 一致性检验： $Kappa = (p_0 - p_e) / (1 - p_e)$ ， $p_0 = (a+d) / n$ ，表示观察到的一致率， $p_e = ((a+b)(a+c) + (b+d)(b+c)) / n^2$ ，表示机遇一致率。

本模块可给出多个行变量与多个列变量，自动对行变量与列变量进行一一组合进行分析。如果给出的行变量数与列变量数相同，如勾选按顺序配对分析，则对每个行变量与相应顺序的列变量组合进行分析，不做交叉组合。

例 1，打开练习项目 DEMO 比较不同性别的文化程度与职业分布，输入界面：

(或下载练习数据: www.empowerstats.com/empowerStats/exdata/demol.xls)

行x列表和卡方检验 ?

标题:

选择分析对象:

列变量 行变量 列-行变量按序配对分析

变量

Education

Occupation

变量

SEX

选择分层变量:

检验方法:

输出结果:

Summary (Pearson Chi-squared test)

| # | Row var. | Column var. | Chi-square | df | P. value | Method |
|-------------------|----------|-------------|------------|----|----------|--|
| 1 | SEX | Education | 187.5805 | 2 | <0.0001 | Pearson's Chi-squared test |
| 2 | SEX | Occupation | 0.3059 | 1 | 0.5802 | Pearson's Chi-squared test |
| | | | 0.2339 | 1 | 0.6286 | Pearson's Chi-squared test with Yates' continuity correction |

1 : SEX * Education

N (行百分数)

| SEX | Education: elementary of lower | Education: middle | Education: high or above | 合计 |
|--------|--------------------------------|-------------------|--------------------------|------------|
| Male | 80 (19.28%) | 154 (37.11%) | 181 (43.61%) | 415 (100%) |
| Female | 261 (63.66%) | 104 (25.37%) | 45 (10.98%) | 410 (100%) |

卡方检验

| Chi-square | df | P. value | Method |
|------------|----|----------|----------------------------|
| 187.5805 | 2 | <0.0001 | Pearson's Chi-squared test |

Fisher's 精确检验

| P-value | Method |
|---------|-----------|
| <0.0001 | two.sided |

2 : SEX * Occupation

N (行百分数)

| SEX | Occupation: farmer | Occupation: others | 合计 |
|--------|--------------------|--------------------|------------|
| Male | 220 (52.88%) | 196 (47.12%) | 416 (100%) |
| Female | 211 (50.97%) | 203 (49.03%) | 414 (100%) |

卡方检验

| Chi-square | df | P. value | Method |
|------------|----|----------|--|
| 0.3059 | 1 | 0.5802 | Pearson's Chi-squared test |
| 0.2339 | 1 | 0.6286 | Pearson's Chi-squared test with Yates' continuity correction |

Fisher's 精确检验

| Odds ratio | 95%CI low | 95%CI high | P-value | Method |
|------------|-----------|------------|---------|-----------|
| 1.0798 | 0.8148 | 1.4313 | 0.6268 | two.sided |
| 1.0798 | 0.0000 | 1.3700 | 0.7332 | less |
| 1.0798 | 0.8512 | Inf | 0.3143 | greater |

