

# 第八章 列联分析

## 第一节 列联表

## 第二节 列联表中的 $\chi^2$ 检验

## 第三节 列联表中的相关系数



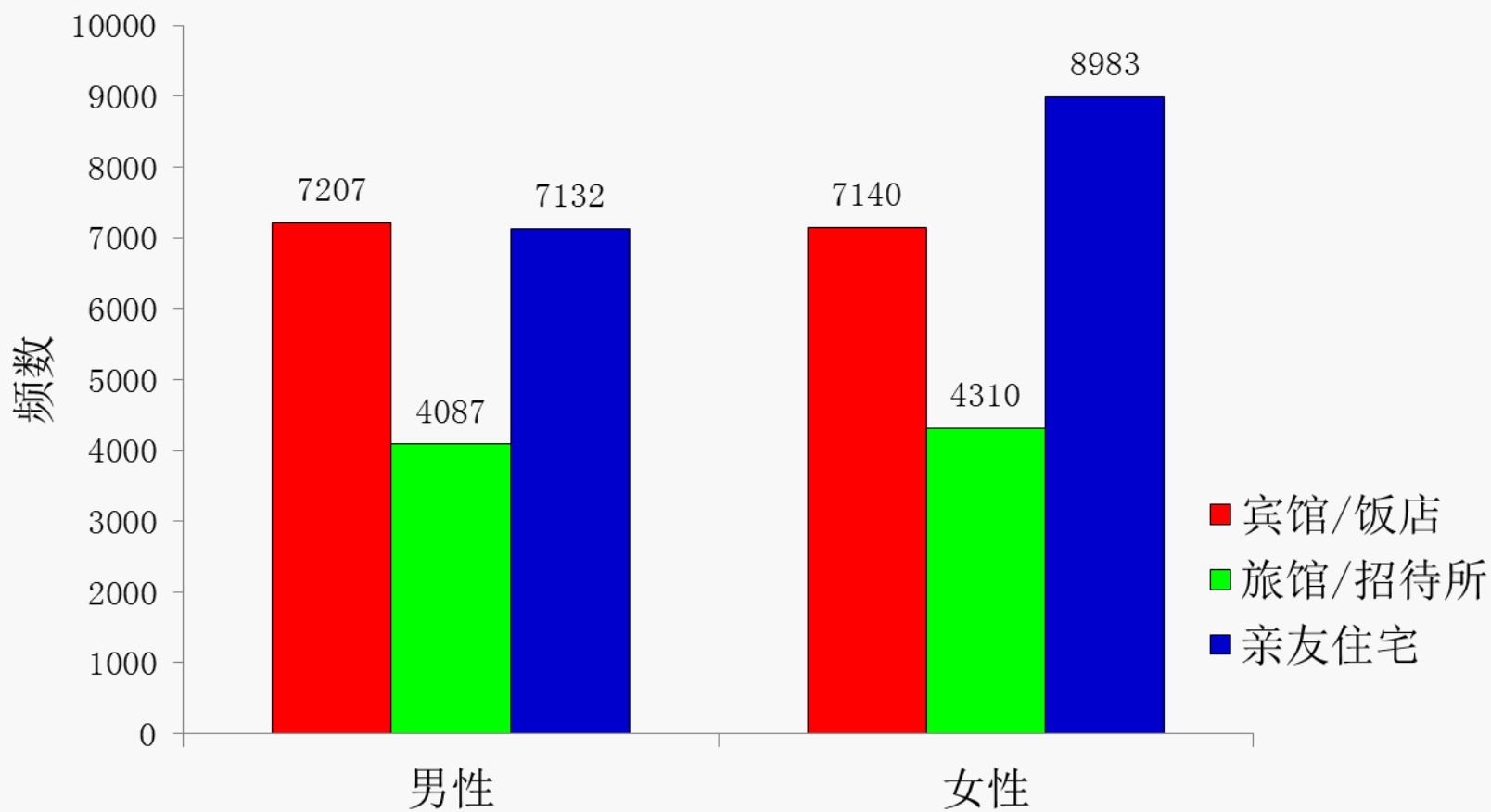
# 引例

旅游者对旅游住宿的选择倾向，直接影响不同种类旅游住宿设施的投资建设和营销策略。在一次针对我国居民选择旅游住宿设施是否存在性别差异的调查中，国家旅游局调查了38859位来自全国的旅游者，整理汇总后的不同性别旅游者对旅游住宿设施选择倾向的样本频数如下：

| 按性别分组 | 宾馆/饭店 | 旅馆/招待所 | 亲友住宅  | 合计    |
|-------|-------|--------|-------|-------|
| 男性    | 7207  | 4087   | 7132  | 18426 |
| 女性    | 7140  | 4310   | 8983  | 20433 |
| 合计    | 14347 | 8397   | 16115 | 38859 |



# 引例



# 引例

- 从样本数据的分布情况上看，不同性别的旅游者对于旅游住宿设施的选择倾向似乎是有差异的，那么是否可以根据此样本数据推断出全国旅游者总体上的选择偏好存在性别差异呢？
- **列联分析方法**，就是利用降维思想，通过分析原始数据结构，揭示分类变量之间以及分类变量各种状态之间相关关系的一种定量研究方法。



# 第一节 列联表

一、列联表的一般结构

二、列联表的分布



# 一、列联表的一般结构

列联表 是由两个分类变量进行交叉分类的频数分布表，用于定量描述两个分类变量及其各种状态的相互关系

- 在列联表中，行变量和列变量均为分类变量
- 通常用  $R_i$  表示行变量的第  $i$  个类别，用  $C_j$  表示列变量的第  $j$  个类别
- 行和列的每种组合为一个单元，用  $f_{ij}$  来表示其观察频数
- 一个  $i$  行  $j$  列的列联表称为  $i \times j$  列联表



# 一、列联表的一般结构

$i \times j$  列联表

观察频数、  
各行合计数、各列合计数、  
样本容量

| 行 \ 列     | $C_1$     | $C_2$     | $\Lambda$ | $C_j$     | 合计        |
|-----------|-----------|-----------|-----------|-----------|-----------|
| $R_1$     | $f_{11}$  | $f_{12}$  | $\Lambda$ | $f_{1j}$  | $R_{T1}$  |
| $R_2$     | $f_{21}$  | $f_{22}$  | $\Lambda$ | $f_{2j}$  | $R_{T2}$  |
| $\Lambda$ | $\Lambda$ | $\Lambda$ | $\Lambda$ | $\Lambda$ | $\Lambda$ |
| $R_i$     | $f_{i1}$  | $f_{i2}$  | $\Lambda$ | $f_{ij}$  | $R_{Ti}$  |
| 合计        | $C_{T1}$  | $C_{T2}$  | $\Lambda$ | $C_{Tj}$  | $n$       |



## 二、列联表的分布

### 观察值的频数分布

- 各单元的观察值频数称为条件频数，如  $f_{11}$ 、 $f_{21}$ ，反映了在变量X条件下变量Y的分布，或在变量Y的条件下变量X的分布；
- 行的合计数称为行边缘频数，如  $R_{T1}$ 、 $R_{T2}$
- 列的合计数称为列边缘频数，如  $C_{T1}$ 、 $C_{T2}$
- $n$ 表示总频数，即样本总数，它等于行边缘频数之和，也等于列边缘频数之和。



## 二、列联表的分布

不同性别旅游者对旅游住宿设施选择倾向的样本频数

| 条件频数  | 按性别<br>分组 | 宾馆/饭店 | 旅馆/招待<br>所 | 亲友住宅  | 合计    | 行边缘频数 |
|-------|-----------|-------|------------|-------|-------|-------|
|       | 男性        | 7207  | 4087       | 7132  | 18426 |       |
|       | 女性        | 7140  | 4310       | 8983  | 20433 |       |
|       | 合计        | 14347 | 8397       | 16115 | 38859 |       |
| 列边缘频数 |           |       |            |       | 总频数   |       |



## 二、列联表的分布

### 观察值的百分比分布

#### 行百分比

行的每一个观察值频数除以相应的行合计数 ( $f_{ij} / R_{Ti}$ )

#### 列百分比

列的每一个观察值频数除以相应的列合计数 ( $f_{ij} / C_{Tj}$ )

#### 总百分数

每一个观察值频数除以观测值的总个数 ( $f_{ij} / n$ )



## 二、列联表的分布

不同性别旅游者对旅游住宿设施选择倾向的样本频数

| 按性别分组 | 维度      | 宾馆/饭店 | 旅馆/招待所 | 亲友住宅  | 合计    |
|-------|---------|-------|--------|-------|-------|
| 男性    | 频数（人）   | 7207  | 4087   | 7132  | 18426 |
|       | 行百分比（%） | 39.1  | 22.2   | 38.7  | 100.0 |
|       | 列百分比（%） | 50.2  | 48.7   | 44.3  | 47.4  |
|       | 总百分比（%） | 18.5  | 10.5   | 18.4  | 47.4  |
| 女性    | 频数（人）   | 7140  | 4310   | 8983  | 20433 |
|       | 行百分比（%） | 34.9  | 21.1   | 44.0  | 100.0 |
|       | 列百分比（%） | 49.8  | 51.3   | 55.7  | 52.6  |
|       | 总百分比（%） | 18.4  | 11.1   | 23.1  | 52.6  |
| 合计    | 频数（人）   | 14347 | 8397   | 16115 | 38859 |
|       | 总百分比（%） | 36.9  | 21.6   | 41.5  | 100.0 |



## 二、列联表的分布

### 期望值的分布

- 列联表中的期望值，指在假定行变量和列变量相互独立情况下，根据既定的比例所计算出的理论数值。
- 观察值频数  $f_{ij}$  的期望频数  $e_{ij}$ ，是总频数的个数  $n$  乘以该观察值频数  $f_{ij}$  落入第  $i$  行和第  $j$  列的概率，即：

$$e_{ij} = n \times \frac{R_{T_i}}{n} \times \frac{C_{T_j}}{n} = \frac{R_{T_i} \times C_{T_j}}{n}$$



## 二、列联表的分布

不同性别旅游者对旅游住宿设施选择倾向的观察值和期望值频数对比表

| 按性别分组 | 维度  | 宾馆/饭店 | 旅馆/招待所 | 亲友住宅 |
|-------|-----|-------|--------|------|
| 男性    | 观察值 | 7207  | 4087   | 7132 |
|       | 期望值 | 6803  | 3982   | 7641 |
| 女性    | 观察值 | 7140  | 4310   | 8983 |
|       | 期望值 | 7544  | 4415   | 8474 |

以第一行第一列单元格的期望频数的计算过程为例：

$$e_{11} = n \times \frac{R_{T1}}{n} \times \frac{C_{T1}}{n} = \frac{R_{T1} \times C_{T1}}{n} = \frac{18426 \times 14347}{38859} = 6803$$



## 第二节 列联表中的 $\chi^2$ 检验

一、列联表的自由度

二、一致性检验

三、独立性检验

四、使用SPSS软件实现列联表中的 $\chi^2$ 检验



# 一、列联表的自由度

## 列联表的自由度

- 列联表是关于两个定类变量的联合样本数据，其自由度指可以自由取值的观察频数的个数，也就是可以自由取值的单元格的个数；
- 一个  $i \times j$  列联表的自由度就是其行数减1再乘以其列数减1，即  $(i - 1) \times (j - 1)$  ；
- 列联表中的自由度与两个定类变量的取值个数及各个取值的交叉组合情况有直接联系。拥有同样单元格个数的列联表，其自由度也有可能是不同的。



# 一、列联表的自由度

3×4列联表（自由度为6）

|       | $C_1$    | $C_2$    | $C_3$    | $C_4$    | 合计       |
|-------|----------|----------|----------|----------|----------|
| $R_1$ | $f_{11}$ | $f_{12}$ | $f_{13}$ | $f_{14}$ | $R_{T1}$ |
| $R_2$ | $f_{21}$ | $f_{22}$ | $f_{23}$ | $f_{24}$ | $R_{T2}$ |
| $R_3$ | $f_{31}$ | $f_{32}$ | $f_{33}$ | $f_{34}$ | $R_{T3}$ |
| 合计    | $C_{T1}$ | $C_{T2}$ | $C_{T3}$ | $C_{T4}$ | $n$      |



# 一、列联表的自由度

2×6列联表（自由度为5）

|       | $C_1$    | $C_2$    | $C_3$    | $C_4$    | $C_5$    | $C_6$    | 合计       |
|-------|----------|----------|----------|----------|----------|----------|----------|
| $R_1$ | $f_{11}$ | $f_{12}$ | $f_{13}$ | $f_{14}$ | $f_{15}$ | $f_{16}$ | $R_{T1}$ |
| $R_2$ | $f_{21}$ | $f_{22}$ | $f_{23}$ | $f_{24}$ | $f_{25}$ | $f_{26}$ | $R_{T2}$ |
| 合计    | $C_{T1}$ | $C_{T2}$ | $C_{T3}$ | $C_{T4}$ | $C_{T5}$ | $C_{T6}$ | $n$      |



## 二、一致性检验

### 一致性检验

对多项总体（单个总体被划分为三个或更多类别），检验样本资料所显示的不同类别的比例差异在总体中是否也存在。

#### 步骤：

1. 建立原假设和备择假设。

$H_0$  : 总体各类别中的比例没有差异；

$H_1$  : 总体各类别中的比例存在差异。

2. 计算检验统计量，即  $\chi^2$  值。

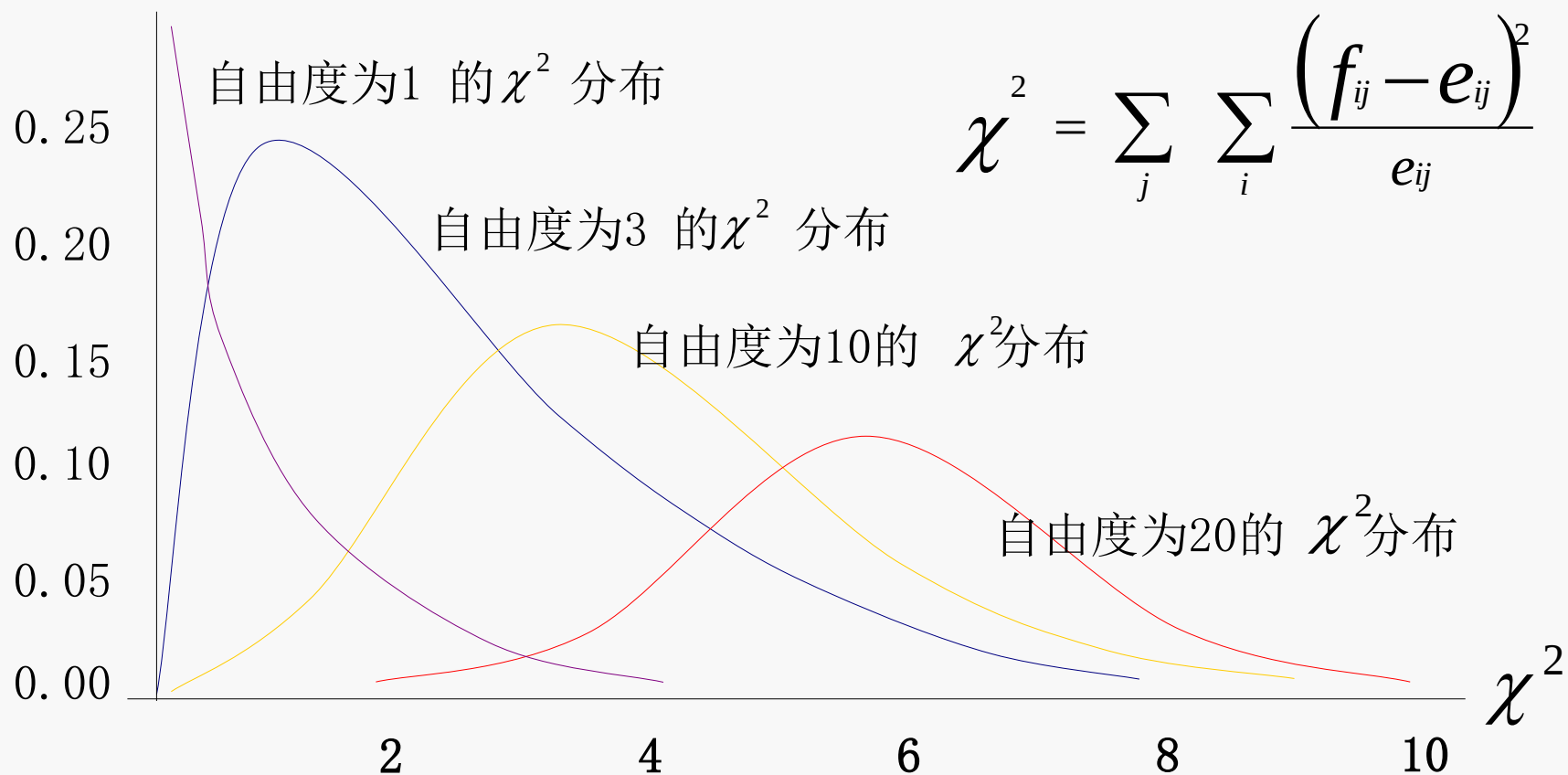
3. 确定自由度和显著性水平。

4. 依据检验规则拒绝或接受原假设。



## 二、一致性检验

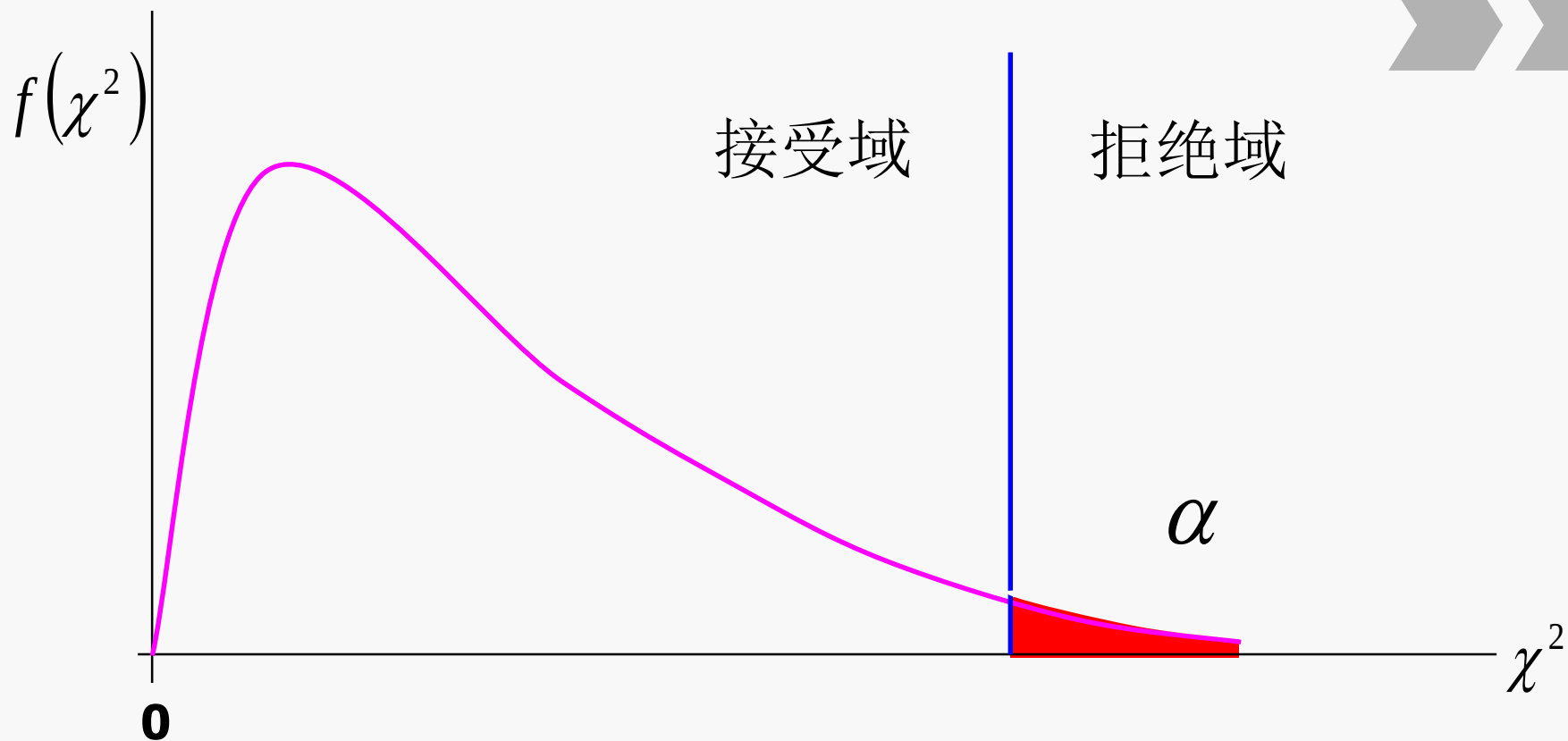
观察频数与期望频数差异的大小以  $\chi^2$  统计量衡量。该统计量服从自由度为  $(R-1) \times (C-1)$  的  $\chi^2$  分布。



$$\chi^2 = \sum_j \sum_i \frac{(f_{ij} - e_{ij})^2}{e_{ij}}$$



## 二、一致性检验



当  $\chi^2 > \chi_\alpha^2$  时，拒绝原假设，说明多项总体各类别中的比例存在差异；当  $\chi^2 < \chi_\alpha^2$  时，接受原假设，说明多项总体各类别中的比例没有差别，调查数据中存在的差异来自样本的随机性



## 二、一致性检验

【例8.1】在引例所述的调查项目中，对旅游者的旅游资源选择偏好及其文化程度也进行了调查。试以  $\alpha = 0.01$  的显著性水平，检验不同程度的旅游者对旅游资源的选择偏好是否存在差异。

| 按偏好分组 | 大专及以上 | 中专及高中 | 初中   | 小学及以下 | 合计    |
|-------|-------|-------|------|-------|-------|
| 自然景观  | 9453  | 7751  | 4726 | 2582  | 24512 |
| 人文景观  | 8094  | 3866  | 1514 | 873   | 14347 |
| 合计    | 17547 | 11617 | 6240 | 3455  | 38859 |



## 二、一致性检验

**解：** 如果不存在差异，对不同类型旅游资源的选择偏好，在不同文化程度旅游者中的比例应该是一致的，即倾向于选择游览自然景观的旅游者比例为 $24512/38859=0.631$ ，因此原假设和备择假设分别为：

$$H_0 : P_1 = P_2 = P_3 = P_4 = 0.631$$

（不同文化程度旅游者的旅游兴趣一致）

$$H_1 : P_1, P_2, P_3, P_4 \text{不全相等}$$

（不同文化程度旅游者的旅游兴趣不一致）

根据列联表中的观测值，计算相应的期望值，进而求出  $\chi^2$  统计量的值。



## 二、一致性检验

### $\chi^2$ 统计量的计算过程

| $f_{ij}$ | $e_{ij}$ | $f_{ij} - e_{ij}$ | $(f_{ij} - e_{ij})^2$ | $(f_{ij} - e_{ij})^2 / e_{ij}$ |
|----------|----------|-------------------|-----------------------|--------------------------------|
| 9453     | 11069    | -1616             | 2609942               | 235.80                         |
| 7751     | 7328     | 423               | 178991                | 24.43                          |
| 4726     | 3936     | 790               | 623862                | 158.50                         |
| 2582     | 2179     | 403               | 162094                | 74.38                          |
| 8094     | 6478     | 1616              | 2609942               | 402.86                         |
| 3866     | 4289     | -423              | 178991                | 41.73                          |
| 1514     | 2304     | -790              | 623862                | 270.79                         |
| 873      | 1276     | -403              | 162094                | 127.07                         |
| 合计       | —        | —                 | —                     | 1335.55                        |



## 二、一致性检验

显著性水平  $\alpha = 0.01$

自由度 =  $(2 - 1) \times (4 - 1) = 3$

查表得临界值：

$$\chi^2_{0.01}(3) = 11.34$$

$$\text{经计算：} \chi^2 = \sum_j \sum_i \frac{(f_{ij} - e_{ij})^2}{e_{ij}} = 1335.55$$

结论：由于  $\chi^2 > \chi^2_{\alpha}$ ，拒绝原假设，认为不同文化程度旅游者的旅游兴趣存在差异。



# 三、独立性检验

## 独立性检验

利用样本资料来检验两个变量的独立性，即样本资料所显示的两个分类变量的相关性在总体中是否存在，如果相关性不存在就称为独立。

## 检验步骤：

1. 建立原假设和备择假设。

$H_0$  : 行变量和列变量之间相互独立（不相关联）；

$H_1$  : 行变量和列变量之间不相互独立（相关联）。

2. 计算检验统计量，即：

$$\chi^2 = \sum_j \sum_i \frac{(f_{ij} - e_{ij})^2}{e_{ij}}$$



# 三、独立性检验

## 独立性检验

利用样本资料来检验两个变量的独立性，即样本资料所显示的两个分类变量的相关性在总体中是否存在，如果相关性不存在就称为独立。

### 检验步骤：

3. 确定自由度和显著性水平。
4. 依据检验规则拒绝或接受原假设：

当  $\chi^2 > \chi_\alpha^2$  时，拒绝原假设，说明多项总体各类别中的比例存在差异；当  $\chi^2 < \chi_\alpha^2$  时，接受原假设，说明多项总体各类别中的比例没有差别，调查数据中存在的差异来自样本的随机性。



### 三、独立性检验

【例8.2】就引例中的调查数据表8-1中的数据，试以 $\alpha = 0.05$ 的显著性水平检验旅游者的性别是否与对不同旅游住宿设施的选择偏好存在关联。

解：假设检验形式为：

$H_0$  : 性别与对旅游住宿设施的偏好相互独立(不相关联)

$H_1$  : 性别与对旅游住宿设施的偏好不相互独立(相关联)



### 三、独立性检验

根据公式  $e_{ij} = \frac{R_{Ti} \times C_{Tj}}{n}$  逐一计算各单元格的期望值，以第一行第一列单元格为例：

$$e_{11} = \frac{18426 \times 14347}{38859} = 6803$$

期望频数分布表

| 按性别分组 | 宾馆/饭店 | 旅馆/招待所 | 亲友住宅  | 合计    |
|-------|-------|--------|-------|-------|
| 男性    | 6803  | 3982   | 7641  | 18426 |
| 女性    | 7544  | 4415   | 8474  | 20433 |
| 合计    | 14347 | 8397   | 16115 | 38859 |



### 三、独立性检验

期望频数分布表中各单元值均大于5，可以计算统计量  $\chi^2$ ，  
计算过程如下：

| $f_{ij}$ | $e_{ij}$ | $f_{ij} - e_{ij}$ | $(f_{ij} - e_{ij})^2$ | $(f_{ij} - e_{ij})^2 / e_{ij}$ |
|----------|----------|-------------------|-----------------------|--------------------------------|
| 7207     | 6803     | 404               | 163215                | 23.99                          |
| 4087     | 3982     | 105               | 11098                 | 2.79                           |
| 7132     | 7641     | -509              | 259431                | 33.95                          |
| 7140     | 7544     | -404              | 163215                | 21.64                          |
| 4310     | 4415     | -105              | 11098                 | 2.51                           |
| 8983     | 8474     | 509               | 259431                | 30.62                          |
| 合计       | —        | —                 | —                     | 115.49                         |



### 三、独立性检验

显著性水平  $\alpha = 0.05$

$$\text{自由度} = (2 - 1) \times (3 - 1) = 2$$

查表得临界值：

$$\chi_{0.05}^2(2) = 5.99$$

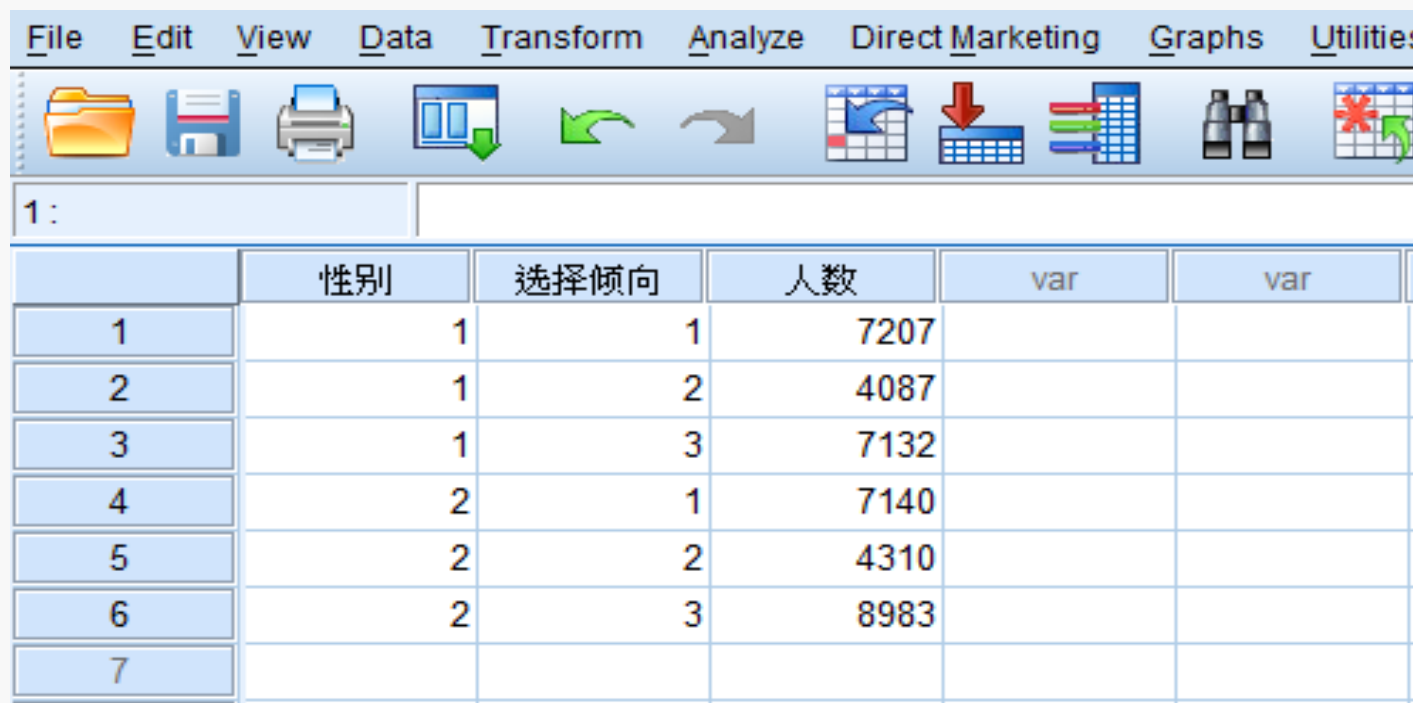
$$\text{于是有： } \chi^2 = \sum_j \sum_i \frac{(f_{ij} - e_{ij})^2}{e_{ij}} = 115.49 > 5.99$$

**结论：**由于  $\chi^2 > \chi_{\alpha}^2$ ，拒绝原假设，即可以认为性别与对旅游住宿设施的偏好有关联，不同类型旅游住宿设施服务单位，应根据旅游者的性别采取有针对性的营销策略。



## 四、使用SPSS软件实现列联表中的 $\chi^2$ 检验

在SPSS中，可以利用Crosstabs过程来实现对列联表的 $\chi^2$  检验过程。以【例8.2】中对两个分类变量独立性检验的操作为例。



The screenshot shows the SPSS software interface with a data table. The table has columns for '性别' (Gender), '选择倾向' (Choice Tendency), and '人数' (Count). The data is as follows:

|   | 性别 | 选择倾向 | 人数   | var | var |
|---|----|------|------|-----|-----|
| 1 | 1  | 1    | 7207 |     |     |
| 2 | 1  | 2    | 4087 |     |     |
| 3 | 1  | 3    | 7132 |     |     |
| 4 | 2  | 1    | 7140 |     |     |
| 5 | 2  | 2    | 4310 |     |     |
| 6 | 2  | 3    | 8983 |     |     |
| 7 |    |      |      |     |     |



## 四、使用SPSS软件实现列联表中的 $\chi^2$ 检验

### 操作步骤:

- 打开数据文件，先用变量“number”进行加权：  
Data → Weight cases → Weight cases by → 将变量  
“number”置入Frequency Variable → OK;
- Output → Analyze → Descriptive statistics → crosstabs →  
将变量“gender”和“choice”分别选入Row(s) 和Column(s)  
→在Statistics选项中勾选Chi-square
- 返回crosstabs → OK。



## 四、使用SPSS软件实现列联表中的 $\chi^2$ 检验

Chi-Square Tests

|                                 | Value                | df | Asymp. Sig.<br>(2-sided) |
|---------------------------------|----------------------|----|--------------------------|
| Pearson Chi-Square              | 115.495 <sup>a</sup> | 2  | .000                     |
| Likelihood Ratio                | 115.611              | 2  | .000                     |
| Linear-by-Linear<br>Association | 110.120              | 1  | .000                     |
| N of Valid Cases                | 38859                |    |                          |

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 3981.65.

**结论：**由分析结果可知， $\chi^2$  统计量为115.495，对应的假设检验概率值为小于0.001，即小于0.05的显著性水平，故拒绝原假设，认为性别与对旅游住宿设施的偏好有关联，与前面的计算结果一致。



## 第三节 列联表中的相关系数

一、 $\phi$  相关系数

二、C相关系数

三、V相关系数

四、使用SPSS软件计算



### 第三节 列联表中的相关系数

- 列联表中的独立性检验，判断的是两个分类变量之间是否相互关联。如果变量相互不独立，说明它们之间存在着联系，不过  $\chi^2$  检验并不能给出相关程度的高低。
- 为进一步测度两个分类变量之间的相关程度，通常要计算相关系数，包括 $\phi$ 相关系数、C相关系数和V相关系数等。
- 列联表的类型不同，适合采用的相关系数也有所不同。
- 列联表中的变量反映研究对象的不同品质类别，因此分类变量之间的相关也称为品质相关。



# 一、 $\phi$ 相关系数

## $\phi$ 相关系数

- 是描述 $2 \times 2$ 列联表数据相关程度最常用的一种相关系数
- 计算公式：

$$\phi = \sqrt{\frac{\chi^2}{n}} = \frac{ad - bc}{\sqrt{(a + b)(c + d)(a + c)(b + d)}}$$

式中， $\chi^2$  值的计算方法与进行独立性检验时的相同； $n$  为列联表中的总频数； $a$ 、 $b$ 、 $c$ 、 $d$  均为条件频数。

- $\phi$  系数的取值范围在  $(0, 1)$  之间，且  $\phi$  的绝对值越大，说明变量  $X$  与  $Y$  的相关程度越高。



# 一、 $\phi$ 相关系数

2×2列联表

| 因素Y   | 因素X     |         | 合计      |
|-------|---------|---------|---------|
|       | $X_1$   | $X_2$   |         |
| $y_1$ | $a$     | $b$     | $a + b$ |
| $y_2$ | $c$     | $d$     | $c + d$ |
| 合计    | $a + c$ | $b + d$ |         |



# 一、 $\phi$ 相关系数

完全相关的两种情况

|       | $C_1$ | $C_2$ | 合计  |
|-------|-------|-------|-----|
| $R_1$ | a     | 0     | a   |
| $R_2$ | 0     | d     | d   |
| 合计    | a     | d     | a+d |

|       | $C_1$ | $C_2$ | 合计  |
|-------|-------|-------|-----|
| $R_1$ | 0     | b     | b   |
| $R_2$ | c     | 0     | c   |
| 合计    | c     | b     | b+c |

$$|\phi| = \left| \frac{ad - bc}{\sqrt{(a+b)(c+d)(a+c)(b+d)}} \right| = 1$$



# 一、 $\phi$ 相关系数

## $\phi$ 取值的意义

$\phi = 0 \rightarrow X$ 与 $Y$ 相互独立

$|\phi| = 1 \rightarrow X$ 与 $Y$ 完全相关

由于在列联表中，变量的位置可以任意变换，因此  $\phi$  的正负符号没有实际意义

$|\phi|$ 越大  $\rightarrow X$ 与 $Y$ 的相关程度越高



## 二、C 相关系数

### C 相关系数

➤ 是列联相关系数的简称，主要用于大于 $2 \times 2$ 列联表的情况

➤ 计算公式：

$$C = \sqrt{\frac{\chi^2}{\chi^2 + n}}$$

- C的取值范围是  $0 \leq C < 1$
- $C=0$ ，表明列联表中的两个变量相互独立
- C的数值大小取决于列联表的行数和列数，并随行数和列数的增大而增大
- C相关系数适合于比较行数和列数一致的列联表



## 二、C 相关系数

- C 的最大值依赖于列联表的行列数，并随行数和列数的增大而增大；
- 根据不同的行和列计算的列联相关系数不能直接比较，除非两个列联表中行数和列数一致。

不同行和列的C 相关系数的最大值表

| 列联表   | $2 \times 2$ | $3 \times 3$ | $4 \times 4$ |
|-------|--------------|--------------|--------------|
| C的最大值 | 0.7071       | 0.8165       | 0.87         |



### 三、V 相关系数

#### V 相关系数

$$V = \sqrt{\frac{\chi^2}{n \times \min[(i-1), (j-1)]}}$$

式中： $\min[(i-1), (j-1)]$ 表示取 $(i-1)$ 和 $(j-1)$ 中较小的一个

- 当两个变量相互独立时， $V=0$ ；当两个变量完全相关时， $V=1$ ； $V$ 的取值在 $(0, 1)$ 之间。
- $V$ 相关系数避免了 $\varphi$ 相关系数和 $C$ 相关系数的缺陷。
- 当列联表中的行数或列数为2，即 $\min[(i-1), (j-1)]=1$ ， $V$ 值等于 $\varphi$ 值。



### 第三节 列联表中的相关系数

【例8.3】在本章第三节独立性检验中，已经针对引例的表8-1中男女旅游者选择旅游住宿设施偏好的样本数据，得出旅游者性别与选择旅游住宿设施偏好之间具有相关性的结论。要求：进一步计算  $\varphi$  相关系数、C相关系数和V相关系数，判断相关程度的大小。



### 第三节 列联表中的相关系数

解：已知 $n=38859$ ， $\chi^2 = 115.49$ ，为 $2 \times 3$ 列联表，

$\min[(i-1), (j-1)] = \min(3-1, 2-1) = \min(2, 1) = 1$ ，则

$$\varphi = \sqrt{\chi^2 / n} = \sqrt{\frac{115.49}{38859}} = 0.055$$

$$C = \sqrt{\frac{\chi^2}{\chi^2 + n}} = \sqrt{\frac{115.49}{115.49 + 38899}} = 0.054$$

$$V = \sqrt{\frac{\chi^2}{n \times \min[(i-1), (j-1)]}} = \sqrt{\frac{115.49}{38859 \times 1}} = 0.055$$



### 第三节 列联表中的相关系数

#### 判断及结论：

- 对于 $\phi$ 相关系数，当列联表规模大于 $2 \times 2$ 时，其值可能大于1，此处为0.055，不能认为数值大；对于C相关系数，C值总是小于1，此处为0.054，数值也并不大；V相关系数为0.055，数值也较小。
- 综合来看，虽然 $\chi^2$ 检验表明，旅游者性别与旅游住宿设施的选择倾向两变量之间存在相关关系，但这种关系的密切程度不高。
- 可以认为，除了旅游者的性别因素，还有其他因素对选择旅游住宿设施的倾向起着更重要的作用。



### 第三节 列联表中的相关系数

在SPSS中，计算列联表中的相关系数的过程是Crosstabs。

#### 操作步骤：

- 打开数据文件，先用变量“number”进行加权  
Data → Weight cases → Weight cases by → 将变量“number”置入Frequency Variable → OK;
- Output → Analyze → Descriptive statistics → crosstabs → 将变量“gender”和“choice”分别选入Row(s) 和Column(s) → 在Statistics选项中勾选Chi-square
- Statistics → 勾选 Contingency coefficient 和 Phi and Cramer's V
- 返回crosstabs → OK。



### 第三节 列联表中的相关系数

| Symmetric Measures |                         |       |              |
|--------------------|-------------------------|-------|--------------|
|                    |                         | Value | Approx. Sig. |
| Nominal by Nominal | Phi                     | .055  | .000         |
|                    | Cramer's V              | .055  | .000         |
|                    | Contingency Coefficient | .054  | .000         |
| N of Valid Cases   |                         | 38859 |              |

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

#### SPSS计算列联表相关系数的输出结果

在输出结果中，Phi是 $\varphi$ 相关系数，Cramer's V是V相关系数，Contingency Coefficient是C相关系数。由分析结果可知， $\varphi$ 相关系数、V相关系数和C相关系数分别等于0.055、0.055和0.054，和【例8.3】中的计算结果一致。



