# 第 2 节 独立性检验（★★☆）

## 内容提要

本节归纳独立性检验问题. 假设两个分类变量 $X$ 和 $Y$ 的一组观测数据如下面的 $2\times2$ 列联表.

| 分类变量 $X$ | 分类变量 $Y$ | | 合计 |
|---|---|---|---|
| | $Y=0$ | $Y=1$ | |
| $X=0$ | $a$ | $b$ | $a+b$ |
| $X=1$ | $c$ | $d$ | $c+d$ |
| 合计 | $a+c$ | $b+d$ | $a+b+c+d$ |

根据小概率值 $\alpha$ 检验分类变量 $X$ 和 $Y$ 是否有关联的步骤：

第 1 步：写出零假设 $H_0$：分类变量 $X$ 与 $Y$ 没有关联；

第 2 步：根据列联表，由 $\chi^2=\dfrac{n(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)}$ 计算 $\chi^2$ 的值，并查表确定小概率值 $\alpha$ 的临界值 $x_\alpha$，与求得的 $\chi^2$ 比较；

| $\alpha$ | 0.1 | 0.05 | 0.01 | 0.005 | 0.001 |
|---|---|---|---|---|---|
| $x_\alpha$ | 2.706 | 3.841 | 6.635 | 7.879 | 10.828 |

第 3 步：作答.

①若 $\chi^2\ge x_\alpha$，则可以推断零假设 $H_0$ 不成立，即 $X$ 和 $Y$ 有关联，且该推断犯错误的概率不超过 $\alpha$；

②若 $\chi^2< x_\alpha$，则没有充分的证据推断 $H_0$ 不成立，可以认为 $X$ 和 $Y$ 没有关联.



## 典型例题

### 类型Ⅰ：独立性检验概念小题

【例 1】某中学为调查高一年级学生的选科倾向，随机抽取了 300 人，其中选考物理的有 220 人，选考历史的有 80 人，统计各选科人数如表所示，则下列说法中正确的是（　　）

| 选考类别 | 选择科目 | | | |
|---|---|---|---|---|
| | 思想政治 | 地理 | 化学 | 生物 |
| 物理类 | 80 | 100 | 145 | 115 |
| 历史类 | 50 | 45 | 30 | 35 |

参考公式：$\chi^2=\dfrac{n(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)}$，其中 $n=a+b+c+d$.

附表：

| $\alpha$ | 0.10 | 0.05 | 0.01 | 0.005 | 0.001 |
|---|---|---|---|---|---|
| $x_\alpha$ | 2.706 | 3.841 | 6.635 | 7.879 | 10.828 |

（A）选考物理类的学生中选择政治的比例比选考历史类的学生中选择政治的比例高

（B）选考物理类的学生中选择地理的比例比选考历史类的学生中选择地理的比例高

（C）参照附表，根据小概率值 $\alpha=0.1$ 的独立性检验，我们认为选择生物与选考类别无关

（D）参照附表，根据小概率值 $\alpha=0.1$ 的独立性检验，我们认为选择生物与选考类别有关

**解析：** A 项，选考物理类、历史类的学生中选择政治的比例分别为 $\dfrac{80}{220}=\dfrac{4}{11}$，$\dfrac{50}{80}=\dfrac{5}{8}$，因为 $\dfrac{4}{11}<\dfrac{5}{8}$，所以

A 项错误；

B 项，选考物理类、历史类的学生中选择地理的比例分别为$\frac{100}{220}=\frac{5}{11}$，$\frac{45}{80}=\frac{9}{16}$，因为$\frac{5}{11}<\frac{9}{16}$，所以 B 项错误；

C 项和 D 项，这两选项涉及独立性检验问题，先画列联表，两个分类变量分别为选考类别、是否选生物，

| | 选择生物 | 不选择生物 | 合计 |
|---|---|---|---|
| 物理类 | 115 | 105 | 220 |
| 历史类 | 35 | 45 | 80 |
| 合计 | 150 | 150 | 300 |

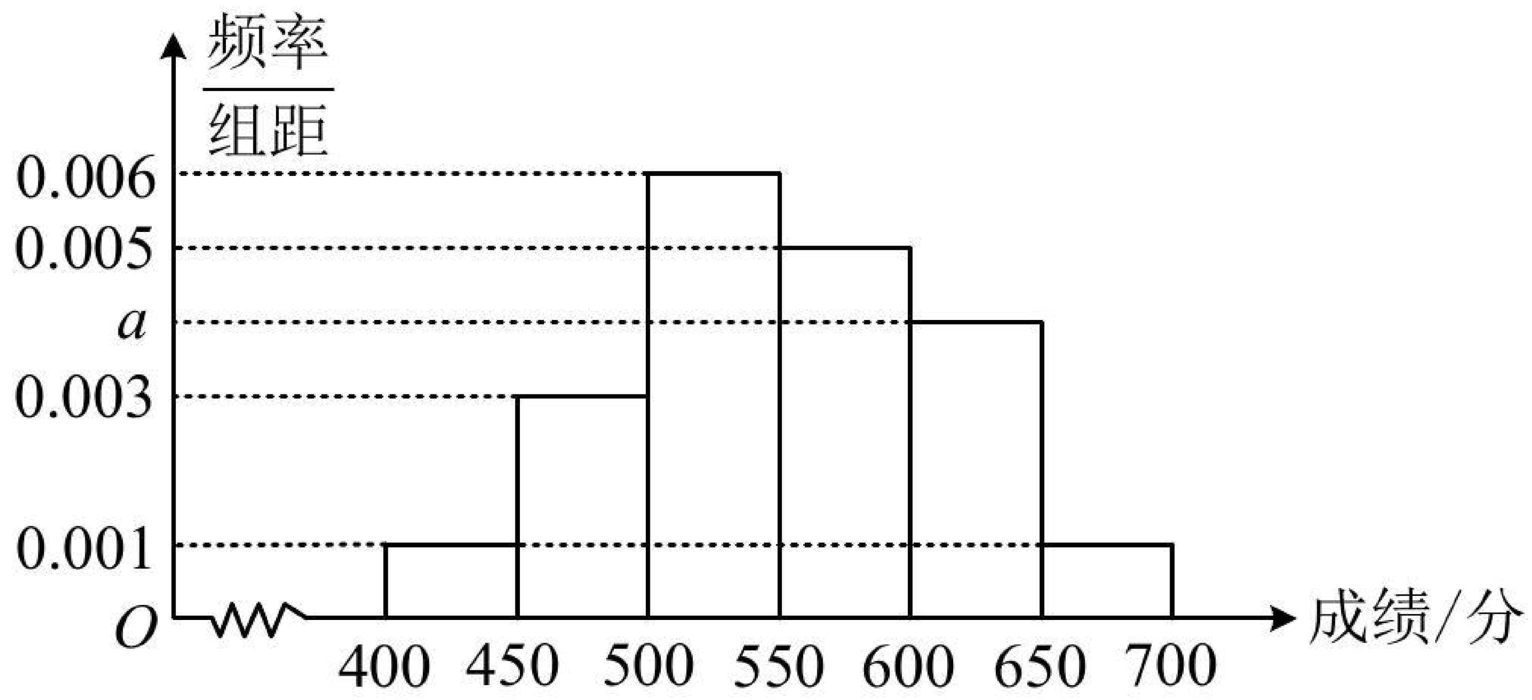
再计算$\chi^2$的值，$\chi^2=\frac{300\times(115\times45-105\times35)^2}{150\times150\times220\times80}\approx1.705<2.706$，所以我们认为选择生物与选考类别无关.

**答案**：C

【反思】在小题中，遇到独立性检验问题，根据列联表求出$\chi^2$的值后，直接与临界值比较大小，若大于等于临界值，则判断两个分类变量有关，否则判断为无关.

**类型Ⅱ：独立性检验综合大题**

【例 2】某省级综合医院共有 1000 名医护员工参加防疫知识技能竞赛，其中男性 450 人，为了解该医院医护员工在防疫知识技能竞赛中的情况，现按性别采用分层抽样（按比例分配）的方法从中抽取 100 名医护员工的成绩（单位：分）作为样本进行统计，成绩均分布在 400~700 分之间，根据统计结果绘制的医护员工成绩的频率分布直方图如图所示，将成绩不低于 600 分的医护员工称为优秀防疫员工.




（1）求 $a$ 的值，并估计该医院医护员工成绩的中位数 $x$；

（2）若样本中优秀防疫员工有女性 10 人，完成下面的$2\times2$列联表，并判断根据小概率值$\alpha=0.05$的独立性检验，能否认为该医院医护员工的性别与是否为优秀防疫员工有关联？

| | 优秀防疫员工 | 非优秀防疫员工 | 合计 |
|---|---|---|---|
| 男 | | | |
| 女 | | | |
| 合计 | | | |

参考公式：$\chi^2=\frac{n(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)}$，其中$n=a+b+c+d$.

附表：

| $\alpha$ | 0.10 | 0.05 | 0.01 | 0.005 | 0.001 |
|---|---|---|---|---|---|
| $x_\alpha$ | 2.706 | 3.841 | 6.635 | 7.879 | 10.828 |

**解**：（1）由图可知，$(0.001+0.003+0.006+0.005+a+0.001)\times 50=1$，解得：$a=0.004$；

注意到前三组的频率和为$(0.001+0.003+0.006)\times 50=0.5$，所以估计该医院医护员工成绩的中位数$x=550$.

（2）由题意，抽取的 100 人中，男性的人数为$\frac{450}{1000}\times 100=45$，女性的人数为$100-45=55$，

由频率分布直方图可知优秀防疫员工的人数为$100\times(0.004+0.001)\times 50=25$，其中女性 10 人，男性 15 人，

所以$2\times 2$列联表如下：

| | 优秀防疫员工 | 非优秀防疫员工 | 合计 |
|---|---|---|---|
| 男 | 15 | 30 | 45 |
| 女 | 10 | 45 | 55 |
| 合计 | 25 | 75 | 100 |

（此为独立性检验大题，接下来的作答分三步，第一步，写出零假设）

零假设为$H_0$：性别与是否为优秀防疫员工无关联，

（第二步，由列联表计算$\chi^2$的值，并与$\alpha$临界值比较）$\chi^2=\frac{100\times(15\times 45-30\times 10)^2}{25\times 75\times 45\times 55}\approx 3.030<3.841$，

（第三步，结合比较的结果作答）根据小概率值$\alpha=0.05$的独立性检验，没有充分的证据推断$H_0$不成立，

所以可以认为性别与是否为优秀防疫员工无关联.

【总结】独立性检验问题的解题流程基本是套路化的，记住上述作答的三个步骤即可.

## 强化训练

1.（2023·上海模拟·★★）某地政府调查育龄妇女生育意愿与家庭年收入高低的关系时，随机调查了当地 3000 名育龄妇女，用独立性检验的方法处理数据，并计算得 $\chi^2=7.326$，则根据这一数据以及临界值表，判断育龄妇女生育意愿与家庭年收入高低有关系的可信度（　　）

（A）低于 1%　　（B）低于 0.5%　　（C）高于 99%　　（D）高于 99.5%

参考数据：$P(\chi^2\ge 10.828)\approx 0.001$，$P(\chi^2\ge 7.879)\approx 0.005$，$P(\chi^2\ge 6.635)\approx 0.01$，$P(\chi^2\ge 3.841)\approx 0.05$，$P(\chi^2\ge 2.706)\approx 0.1$.

2.（2023·云南统考·★★）党的二十大胜利召开后，某校为调查性别因素对党史知识的了解情况是否有影响，随机抽查了男女教职工各 100 名，得到如下数据：

| | 不了解 | 了解 |
|---|---|---|
| 女职工 | 30 | 70 |
| 男职工 | 20 | 80 |

根据小概率值 的独立性检验，能否认为对党史知识的了解情况与性别有关？



参考公式：$\chi^2=\dfrac{n(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)}$，其中 $n=a+b+c+d$.

附表：

| $\alpha$ | 0.010 | 0.005 | 0.001 |
|---|---|---|---|
| $x_\alpha$ | 6.635 | 7.879 | 10.828 |

3．（2023 • 湖北模拟 • ★★★）某数学兴趣小组为研究本校学生数学成绩与语文成绩的关系，采取有放回的简单随机抽样，从学校抽取样本量为200的样本，将所得数学成绩与语文成绩的样本观测数据整理如下：

| | | 语文成绩 | | 合计 |
|---|---|---|---|---|
| | | 优秀 | 不优秀 | |
| 数学成绩 | 优秀 | 50 | 30 | 80 |
| | 不优秀 | 40 | 80 | 120 |
| 合计 | | 90 | 110 | 200 |

（1）根据小概率值 $\alpha=0.01$ 的独立性检验，能否认为数学成绩与语文成绩有关联？

（2）在人工智能中常用 $L(B\mid A)=\dfrac{P(B\mid A)}{P(\bar{B}\mid A)}$ 表示在事件 $A$ 发生的条件下事件 $B$ 发生的优势，在统计中称为似然比. 现从该校学生中任选一人，$A$ 表示“选到的学生语文成绩不优秀”，$B$ 表示“选到的学生数学成绩不优秀”，请利用样本数据，估计 $L(B\mid A)$ 的值.

参考公式：$\chi^2=\dfrac{n(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)}$，其中 $n=a+b+c+d$ .

附表：

| $\alpha$ | 0.01 | 0.005 | 0.001 |
|---|---|---|---|
| $x_\alpha$ | 6.635 | 7.879 | 10.828 |