《2022年高考数学》

20.（12分）

一医疗团队为研究某地的一种地方性疾病与当地居民的卫生习惯（卫生习惯分为良好和不够良好两类）的关系，在已患该病的病例中随机调查了100例（成为病例组），同时在未患该疾病的人群中随机调查了100人（称为对照组），得到如下数据：

	$A=$不够良好	$\bar{A}=$良好
$B=$病例组	$a=40$	$b=60$
$\bar{B}=$对照组	$c=10$	$d=90$

$n=200$

附：

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

$\mathbb{P}\left(\chi_1^2\geq k\right)$	0.050	0.010	0.001
k	3.841	6.635	10.828

（1）能否有 99%的把握认为患该疾病群体与未患该疾病群体的卫生习惯有差异？

解：本题的统计学问题是：

随机变量

$$\text{患病} = \{\text{“病例组”; “对照组”}\}\text{(横行)}$$

是否与随机变量

$$\text{卫生习惯} = \{\text{“不够良好”; “良好”}\}\text{(纵列)}$$

相互独立。

χ^2 独立性假设检验可以回答这个问题。

第一步：声明原假设 H_0 与备择假设 H_A

$$\begin{cases} H_0 : \text{患病与卫生习惯互为独立。} \\ H_A : \text{患病与卫生习惯不独立。} \end{cases}$$

第二步：计算统计量 χ^2（套公式）

$$\begin{aligned} \chi^2 &= \frac{n(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)} \\ &= \frac{200 \times (40 \times 90 - 60 \times 10)^2}{(40+60) \times (10+90) \times (40+10) \times (60+90)} \\ &= 24\text{。} \end{aligned}$$

第三步：计算拒绝域

自由度 $\nu = (\text{行数} - 1) \times (\text{列数} - 1) = (2-1) \times (2-1) = 1$。

查表得 $\chi^2_{\text{临界}}(\nu = 1, \alpha = 0.010) = 6.635$，即拒绝域为 $(6.635, +\infty)$。

第四步：判断假设

由于统计量 $\chi^2 = 24$ 落入拒绝域 $(6.635, +\infty)$，故拒绝原假设，继而下结论：在 99%显著水平，有把握认定患该疾病群体与未患该疾病群体的卫生习惯有差异。

（2）从该地的人群中任选一人，A 表示事件“选到的人卫生习惯不够良好”，B 表示事件“选到的人患有该疾病”，

$$\frac{\mathbb{P}\left(B|A\right)}{\mathbb{P}\left(\bar{B}|A\right)} \text{ 与 } \frac{\mathbb{P}\left(B|\bar{A}\right)}{\mathbb{P}\left(\bar{B}|\bar{A}\right)}$$

的比值是卫生习惯不够良好对患该疾病风险程度的一项度量指标，记该指标为 R。

（i）证明：

$$R = \frac{\mathbb{P}\left(A|B\right)}{\mathbb{P}\left(\bar{A}|B\right)} \cdot \frac{\mathbb{P}\left(\bar{A}|\bar{B}\right)}{\mathbb{P}\left(A|\bar{B}\right)}。$$

证明：依题意：

$$R = \frac{\mathbb{P}\left(B|A\right)}{\mathbb{P}\left(\bar{B}|A\right)} \Bigg/ \frac{\mathbb{P}\left(B|\bar{A}\right)}{\mathbb{P}\left(\bar{B}|\bar{A}\right)}$$

R 中的每个因子都可以根据 Bayes 公式展开，例如：

$$\mathbb{P}\left(B|A\right) = \mathbb{P}\left(A|B\right) \cdot \frac{\mathbb{P}\left(B\right)}{\mathbb{P}\left(A\right)}$$

因此，R 的表达式可被展开为：

$$\begin{aligned}
R &= \frac{\mathbb{P}\left(B|A\right)}{\mathbb{P}\left(\bar{B}|A\right)} \Bigg/ \frac{\mathbb{P}\left(B|\bar{A}\right)}{\mathbb{P}\left(\bar{B}|\bar{A}\right)} \\
&= \frac{\mathbb{P}\left(B|A\right)}{\mathbb{P}\left(\bar{B}|A\right)} \cdot \frac{\mathbb{P}\left(\bar{B}|\bar{A}\right)}{\mathbb{P}\left(B|\bar{A}\right)} \\
&= \frac{\mathbb{P}\left(A|B\right) \cdot \frac{\mathbb{P}(B)}{\mathbb{P}(A)}}{\mathbb{P}\left(A|\bar{B}\right) \cdot \frac{\mathbb{P}(\bar{B})}{\mathbb{P}(A)}} \cdot \frac{\mathbb{P}\left(\bar{A}|\bar{B}\right) \cdot \frac{\mathbb{P}(\bar{B})}{\mathbb{P}(\bar{A})}}{\mathbb{P}\left(\bar{A}|B\right) \cdot \frac{\mathbb{P}(B)}{\mathbb{P}(\bar{A})}} \\
&= \frac{\mathbb{P}\left(A|B\right)}{\mathbb{P}\left(\bar{A}|B\right)} \cdot \frac{\mathbb{P}\left(\bar{A}|\bar{B}\right)}{\mathbb{P}\left(A|\bar{B}\right)} \text{（繁分式相互约分，然后对调分母位置）}
\end{aligned}$$

证毕。

(ii) 利用该调查数据，给出 $\mathbb{P}(A|B)$、$\mathbb{P}\left(A|\bar{B}\right)$ 的估计值，并利用 (i) 的结果给出 R 的估计值。

解：根据条件概率定义式：

$$\widehat{\mathbb{P}}(A|B)=\frac{\widehat{\mathbb{P}}(A\cap B)}{\widehat{\mathbb{P}}(B)}=\frac{40/200}{(40+60)/200}=0.4,$$

$$\widehat{\mathbb{P}}\left(A|\bar{B}\right)=\frac{\widehat{\mathbb{P}}\left(A\cap\bar{B}\right)}{\widehat{\mathbb{P}}\left(\bar{B}\right)}=\frac{10/200}{(10+90)/200}=0.1,$$

因而：

$$\widehat{R}=\frac{\widehat{\mathbb{P}}(A|B)}{\widehat{\mathbb{P}}\left(\bar{A}|B\right)}\cdot\frac{\widehat{\mathbb{P}}\left(\bar{A}|\bar{B}\right)}{\widehat{\mathbb{P}}\left(A|\bar{B}\right)}=\frac{0.4}{1-0.4}\cdot\frac{1-0.1}{0.1}=6。$$