对2022年数学新高考Ⅰ卷第20题的思考
——一类与条件概率及全概率相关习题的解法探究

尹健萍(江苏省常州市第二中学 213000)

【摘 要】 概率和统计的综合问题是近几年的高考热点问题之一,其中2×2列联表经常出现在独立性检验的问题中.列联表除了能描述两个分类变量分布的频数表,还能描述两个分类变量分布的频率表,而利用表中的频率可以解决一些条件概率与全概率的问题.本文通过2022年新高考Ⅰ卷的第20题(2)的第(ⅰ)小问来研究列联表这一用途.

【关键词】 列联表;条件概率;全概率公式

1 高考题再现

例1 (2022年新高考Ⅰ卷第20题)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,在已患该疾病的病例中随机调查了100例(称为病例组),同时在未患该疾病的人群中随机调查了100人(称为对照组),得到如表1数据:

表1

	不够良好	良好
病例组	40	60
对照组	10	90

(1)略;

(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".$\frac{P(B\mid A)}{P(\bar{B}\mid A)}$与$\frac{P(B\mid \bar{A})}{P(\bar{B}\mid \bar{A})}$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:$R=\frac{P(A\mid B)}{P(\bar{A}\mid B)}\cdot\frac{P(\bar{A}\mid \bar{B})}{P(A\mid \bar{B})}$;

(ⅱ)略.

解析 这里探究第二问第一小问的解答.笔者尝试重新列列联表,让表中的数据一般化,将表1改写为表2,这样就能用字母直接表示出R值,从而进行了证明.具体过程如下:

表2

	不够良好A.$m\%$	良好$\bar{A}$.$(1-m)\%$
病例组B	$p\%$	$(1-p)\%$
对照组$\bar{B}$	$q\%$	$(1-q)\%$

(2)(ⅰ)**证明** 设该地卫生习惯不够良好的居民占$m\%$,在卫生习惯不够良好的居民中已患该种疾病的居民占$p\%$,在卫生习惯不够良好的居民中未患该种疾病的居民占$q\%$,则

$$R=\frac{P(B\mid A)}{P(\bar{B}\mid A)}\cdot\frac{P(\bar{B}\mid \bar{A})}{P(B\mid \bar{A})}$$

$$=\frac{P(AB)}{P(A)}\cdot\frac{P(A)}{P(A\bar{B})}\cdot\frac{P(\bar{A}\bar{B})}{P(\bar{A})}\cdot\frac{P(\bar{A})}{P(\bar{A}B)}$$

$$=\frac{mp}{mq}\cdot\frac{(1-m)(1-q)}{(1-m)(1-p)}=\frac{p(1-q)}{(1-p)q}.$$

因为$\frac{P(A\mid B)}{P(\bar{A}\mid B)}\cdot\frac{P(\bar{A}\mid \bar{B})}{P(A\mid \bar{B})}$

$$=\frac{P(AB)}{P(B)}\cdot\frac{P(B)}{P(\bar{A}B)}\cdot\frac{P(\bar{A}\bar{B})}{P(\bar{B})}\cdot\frac{P(\bar{B})}{P(A\bar{B})}$$

$$=\frac{mp}{(1-m)(1-p)}\cdot\frac{(1-m)(1-q)}{mq}$$

$$=\frac{p(1-q)}{(1-p)q},$$

所以$R=\frac{P(A\mid B)}{P(\bar{A}\mid B)}\cdot\frac{P(\bar{A}\mid \bar{B})}{P(A\mid \bar{B})}$.

2　解决问题

例2　设验血诊断某种疾病的误诊率为5%，用A表示验血为阳性，B表示受验者患病，则$P(\bar{A} \mid B)=P(A \mid \bar{B})=0.05$，若已知受检人群中有0.5%患此病，即$P(B)=0.005$，求一个验血为阳性的人的确患此病的概率.

最初让学生做这道题时，他们感觉思路不太清楚，各个数据不知如何处理. 笔者引导学生，梳理已知数据，利用列联表来描述题中变量分布的频率表，如表3：

表3

	阳性A	阴性$\bar{A}$
患病B　0.5%(I)	0.95(Ⅲ)	0.05
不患病$\bar{B}$　99.5%(Ⅱ)	0.05(Ⅳ)	0.95

接着可以让学生反思，表格中各个量的位置为什么这样放？先根据要求的问题，把“患病”当成“主元”列在(I)的位置，相对应的阳性概率列(III)的位置上，接着用贝叶斯公式写出计算公式
$$P(B|A)=\frac{P(B)\cdot P(A|B)}{P(B)\cdot P(A|B)+P(\bar{B})\cdot P(A|\bar{B})}$$
$=\frac{(\mathrm{I})\times(\mathrm{III})}{(\mathrm{I})\times(\mathrm{III})+(\mathrm{II})\times(\mathrm{IV})}$，最后把表格中的数据相应地代入公式即可.

解　由题意，可得：
$$P(B|A)=\frac{P(B)P(A|B)}{P(B)P(A|B)+P(\bar{B})P(A|\bar{B})}$$
$$=\frac{0.005\times0.95}{0.005\times0.95+0.995\times0.05}=\frac{19}{218}.$$

变式　两台车床加工同样的零件，第一台车床出现废品的概率是0.03，第二台车床出现废品的概率是0.02，加工出来的零件放在一起，并且已知第一台加工的零件比第二台加工的零件多一倍. 如果任意取出的零件是废品，求它是第二台车床加工的概率.

分析　由已知条件求出两台车床加工的零件的百分比，然后把第二台车床当成“主元”列在(I)的位置，相对应的废品的概率列(III)的位置上，得出如表4的列联表：

表4

	废品	正品
第二台$\frac{1}{3}$(I)	0.02(III)	0.98
第一台$\frac{2}{3}$(II)	0.03(IV)	0.97

解　设事件A_i=“第i台机床加工的零件”$(i=1,2)$；事件B=“出现废品”；事件C=“出现合格品”；

则$P(B|A_1)=0.03$，$P(B|A_2)=0.02$，

$P(A_1)=\frac{2}{3}$，$P(A_2)=\frac{1}{3}$.

则$P(A_2|B)$
$$=\frac{P(A_2)\cdot P(B|A_2)}{P(A_1)\cdot P(B|A_1)+P(A_2)\cdot P(B|A_2)}$$
$$=\frac{\frac{1}{3}\times0.02}{\frac{2}{3}\times0.03+\frac{1}{3}\times0.02}=0.25.$$

3　结语

通过以上两道例题的分析可以发现解决此类问题按照以下步骤进行：

一、合理画出列联表，根据题意及所求问题列出列联表的位置(I)和位置(III)的内容，然后确定位置(II)和位置(IV)的内容，最后完成表格的第二列，要注意的是第一行和第二行的概率和为1，表格如表5所示.

表5

	第一列	第二列
第一行$m\%$(I)	$p\%$(III)	$(1-p\%)$
第二行$(1-m\%)$(II)	$q\%$(IV)	$(1-q\%)$

二、根据公式$P=\frac{(\mathrm{I})\times(\mathrm{III})}{(\mathrm{I})\times(\mathrm{III})+(\mathrm{II})\times(\mathrm{IV})}=\frac{m\%\times p\%}{m\%\times p\%+(1-m\%)\times q\%}$即可算出概率.

参考文献：

[1][1]王文相.条件概率与贝叶斯公式在教学中的难度层次[J].考试周刊，2018(8).