**独立性检验的基本思想及其初步应用易错点-高中数学选修1-2第一章**

**例1.**为了探究患慢性气管炎是否与吸烟有关，调查了339名50岁以上的人，调查结果如下表所示：

|  |  |  |  |
| --- | --- | --- | --- |
|  | 患病 | 不患病 | 合计 |
| 吸烟 | 43 | 162 | 205 |
| 不吸烟 | 13 | 121 | 134 |
| 合计 | 56 | 283 | 339 |

试问：50岁以上的人患慢性气管炎与吸烟习惯有关吗？

**分析：**最理想的解决办法是向所有50岁以上的人作调查，然后对所得到的数据进行统计处理，但这花费的代价太大，实际上是行不通的，339人相对于全体50岁以上的人，只是一个小部分，已学过总体和样本的关系，当用样本平均数，样本方差去估计总体相应的数字特征时，由于抽样的随机性，结果并不唯一。现在情况类似，我们用部分对全体作推断，推断可能正确，也可能错误。如果抽取的339个调查对象中很多人是吸烟但没患慢性气管炎，而虽不吸烟因身体体质差而患慢性气管炎，能够得出什么结论呢？我们有95%（或99%）的把握说事件image008 与事件image010 有关，是指推断犯错误的可能性为5%（或1%），这也常常说成是“以95%（或99%）的概率”是一样的。

**解：**根据列联表中的数据，得

image012 。

因为image014 ，所以我们有99%的把握说：50岁以上的人患慢性气管炎与吸烟习惯有关。

**评注：**对两个分类变量进行独立性检验，要对样本的选取背景、时间等因素进行分析。

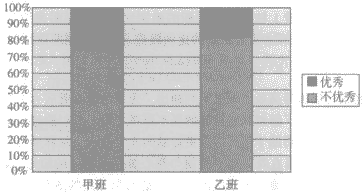
**例2．**甲乙两个班级进行一门考试，按照学生考试成绩优秀和不优秀统计成绩后，得到如下的列联表：

班级与成绩列联表

|  |  |  |  |
| --- | --- | --- | --- |
|  | 优秀 | 不优秀 | 总计 |
| 甲班 | 10 | 35 | 45 |
| 乙班 | 7 | 38 | 45 |
| 总计 | 17 | 73 | 90 |

画出列联表的条形图，并通过图形判断成绩与班级是否有关；利用列联表的独立性检验估计，认为“成绩与班级有关系”犯错误的概率是多少

**解：**列联表的条形图如图所示：



由图及表直观判断，好像“成绩优秀与班级有关系”；由表中数据计算得K2的观察值为k≈0.653>0.455。由下表中数据

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| P（K2≥k） | 0.50 | 0.40 | 0.25 | 0.15 | 0.10 | 0.05 | 0.025 | 0.010 | 0.005 | 0.001 |
| k | 0.455 | 0.708 | 1.323 | 2.072 | 2.706 | 3.841 | 5.024 | 6.635 | 7.879 | 10.828 |

得：P(K2≥0.455)≈0.50，

从而有50%的把握认为“成绩与班级有关系”，即断言“成绩优秀与班级有关系”犯错误的概率为0.5。

**评注：**

（1）画出条形图后，从图形上判断两个分类变量之间是否有关系。这里通过图形的直观感觉的结果可能会出错。

（2）计算得到K2的观测值比较小，所以没有理由说明“成绩优秀与班级有关系”。这与反证法也有类似的地方，在使用反证法证明结论时，假设结论不成立的条件下如果没有推出矛盾，并不能说明结论成立也不能说明结论不成立。在独立性检验中，在假设“成绩优秀与班级没有关系”的情况下，计算得到的K2的值比较小，且P(K2≥0.653)≈0.42，说明事件(K2≥0.653)不是一个小概率事件，这个事件的发生不足以说明“成绩优秀与班级没有关系”，即没有理由说明“成绩优秀与班级有关系”。这里没有推出小概率事件发生类似于反证法中没有推出矛盾。

**例3．**为考察某种药物预防疾病的效果，进行动物试验，得到如下的列联列表：

药物效果与动物试验列联表

|  |  |  |  |
| --- | --- | --- | --- |
|  | 患病 | 未患病 | 总计 |
| 服用药 | 10 | 45 | 55 |
| 没服用药 | 20 | 30 | 50 |
| 总计 | 30 | 75 | 105 |

请问能有多大把握认为药物有效？

**解：** 假设“服药情况与是否患病之间没有关系”，则K2的值应比较小；如果K2的值很大，则说明很可能“服药情况与是否患病之间有关系”。由题目中所给数据计算，得K2的观测值为k≈6.110，而P(K2≥5.024)≈0.025，所以有97.5%的把握认为“服药情况与是否患病之间有关系”，即大约有97.5%的把握认为药物有效。

**例4．**在一次恶劣气候的飞行航程中调查男女乘客在机上晕机的情况如下表所示，根据此资料你是否认为在恶劣气候中男人比女人更容易晕机？

|  |  |  |  |
| --- | --- | --- | --- |
|  | 晕机 | 不晕机 | 合计 |
| 男人 | 24 | 31 | 55 |
| 女人 | 8 | 26 | 34 |
| 合计 | 32 | 57 | 89 |

**分析：**这是一个image050 列联表的独立性检验问题，根据列联表的数据求解。

**解：**由条件中数据，计算得：

image020 ，

因为image022 ，所以我们没有理由说晕机是否跟男女性别有关，尽管这次航班中男人晕机的比例image024 比女人晕机的比例image026 高，但我们不能认为在恶劣的气候飞行中男人比女人更容易晕机。

**评注：**在使用image042 统计量作image050 列联表的独立性检验时，要求表中的4个数据大于等于5，为此，在选取样本的容量时一定要注意这一点，本例中的4个数据都大于5，且满足这一要求的。　　**本周练习：**

1．在一次独立性检验中，其把握性超过了99%，则随机变量image042 的可能值为（　　　　 ）　　A．6.635　　　　 B．5.024　　 C．7.897　　 D．3.841

2．把两个分类变量的频数列出，称为（　　　　 ）

A．三维柱形图　　　　 B．二维条形图　　　　 C．列联表　　　　 D．独立性检验

3．由列联表

|  |  |  |  |
| --- | --- | --- | --- |
|  | image034 | image036 | 合计 |
| image038 | 43 | 162 | 205 |
| image040 | 13 | 121 | 134 |
| 合计 | 56 | 283 | 339 |

则随机变量image042 的值为　　　　　　　　　　　　　　　　　　　　 。

4．某大学希望研究性别与职称之间是否有关系，你认为应该收集哪些数据？　　答：　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　 。

5．某高校“统计初步”课程的教师随机调查了该选修课的一些学生情况，具体数据如下表：

|  |  |  |
| --- | --- | --- |
|  | 非统计专业 | 统计专业 |
| 男 | 13 | 10 |
| 女 | 7 | 20 |

为了检验主修专业是否与性别有关系，根据表中的数据，得到　　image044 。

因为image046 ，所以断定主修统计专业与性别有关系。这种判断出错的可能性为　　　　　　　　　 。

6．在对人们休闲的一次调查中，共调查了124人，其中女性70人，男性54人。女性中有43人主要的休闲方式是看电视，另外27人主要的休闲方式是运动；男性中有21人主要的休闲方式是看电视，另外33人主要的休闲方式是运动。

（1）根据以上数据建立一个image050 的列联表；

（2）检验性别与休闲方式是否有关系。

7． 调查某医院某段时间内婴儿出生的时间与性别的关系，得到下面的数据表。试问能以多大把握认为婴儿的性别与出生的时间有关系。

|  |  |  |  |
| --- | --- | --- | --- |
| 出生时间  性别 | 晚上 | 白天 | 合计 |
| 男婴 | 24 | 31 | 55 |
| 女婴 | 8 | 26 | 34 |
| 合计 | 32 | 57 | 89 |

**参考答案：**　　1．C 　　2．C　　3．7.469　　4．女教授人数，男教授人数，女副教授人数，男副教授人数（或高级职称中女性的人数，高级职称中男性的人数，中级职称中女性的人数，中级职称中男性的人数。）　　5．5%（或0.05）　　6．答案：　　（1）image050 的列联表：

|  |  |  |  |
| --- | --- | --- | --- |
|  | 看电视 | 运动 | 合计 |
| 女 | 43 | 27 | 70 |
| 男 | 21 | 33 | 54 |
| 合计 | 64 | 60 | 124 |

（2）假设休闲方式与性别无关，计算　　image052 ；

因为image054 image056，所以有理由认为假设休闲方式与性别无关是不合理的，即我们有97.5%的把握认为休闲方式与性别无关。

7．由所给数据计算得K2的观测值为k≈3.689，而由

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| P（K2≥k） | 0.50 | 0.40 | 0.25 | 0.15 | 0.10 | 0.05 | 0.025 | 0.010 | 0.005 | 0.001 |
| k | 0.455 | 0.708 | 1.323 | 2.072 | 2.706 | 3.841 | 5.024 | 6.635 | 7.879 | 10.828 |

知P(K2≥2.706)=0.10

所以有90%的把握认为“婴儿的性别与出生的时间有关系”。