**独立性检验的基本思想及其初步应用题库及答案-高中数学选修1-2第一章**

**►达标训练**

**1．在研究两个分类变量之间是否有关时，可以粗略地判断两个分类变量是否有关的是(　　)**

**A．散点图　　　　　　　B．等高条形图**

**C．2×2列联表 D．以上均不对**

**答案：B**

**2．在等高条形图形图中，下列哪两个比值相差越大，要推断的论述成立的可能性就越大(　　)**

**A.与 B.与**

**C.与 D.与**

**答案：C**

**3．对分类变量*X*与*Y*的随机变量*K*2的观测值*k*，说法正确的是(　　)**

**A．*k*越大，“ *X*与*Y*有关系”可信程度越小**

**B．*k*越小，“ *X*与*Y*有关系”可信程度越小**

**C．*k*越接近于0，“*X*与*Y*无关”程度越小**

**D．*k*越大，“*X*与*Y*无关”程度越大**

**答案：B**

**4．下面是一个2×2列联表：**

|  |  |  |  |
| --- | --- | --- | --- |
|  | ***y*1** | ***y*2** | **总计** |
| ***x*1** | ***a*** | **21** | **73** |
| ***x*2** | **2** | **25** | **27** |
| **总计** | ***b*** | **46** | **100** |

**则表中*a*、*b*的值分别为(　　)**

**A．94、96 B．52、50**

**C．52、54 D．54、52**

**答案：C**

**5．性别与身高列联表如下：**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **高(165 cm以上)** | **矮(165 cm以下)** | **总计** |
| **男** | **37** | **4** | **41** |
| **女** | **6** | **13** | **19** |
| **总计** | **43** | **17** | **60** |

**那么，检验随机变量*K*2的值约等于 (　　)**

**A．0.043 B．0.367**

**C．22 D．26.87**

**答案：C**

**6．给出列联表如下：**

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

**根据表格提供的数据，估计“成绩与班级有关系”犯错误的概率约是(　　)**

**A．0.4　　　B．0.5　　　C．0.75　　　D．0.85**

**答案：B**

**►素能提高**

**1．在调查中发现480名男人中有38名患有色盲，520名女人中有6名患有色盲，下列说法中正确的是(　　)**

**A．男人、女人中患有色盲的频率分别为0.038、0.006**

**B．男人、女人患色盲的概率分别为、**

**C．男人中患色盲的比例比女人中患色盲的比例大，患色盲是与性别有关的**

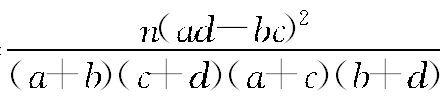
**D．调查人数太少，不能说明色盲与性别有关**

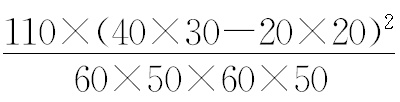
**解析：男人患色盲的比例为，比女人中患色盲的比例大，其差值为≈0.067 6，差值较大．**

**答案：C**

**2．通过随机询问110名性别不同的大学生是否爱好某项运动，得到如下的列联表：**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **男** | **女** | **总计** |
| **爱好** | **40** | **20** | **60** |
| **不爱好** | **20** | **30** | **50** |
| **总计** | **60** | **50** | **110** |

**由*K*2＝算得，**

***K*2＝≈7.8.**

**附表：**

|  |  |  |  |
| --- | --- | --- | --- |
| ***P*(*K*2≥*k*0)** | **0.050** | **0.010** | **0.001** |
| ***k*0** | **3.841** | **6.635** | **10.828** |

**参照附表，得到的正确结论是(　　)**

**A．有99%以上的把握认为“爱好该项运动与性别有关”**

**B．有99%以上的把握认为“爱好该项运动与性别无关”**

**C．在犯错误的概率不超过0.1%的前提下，认为“爱好该项运动与性别有关”**

**D．在犯错误的概率不超过0.1%的前提下，认为“爱好该项运动与性别无关”**

**答案：A**

**3．若由一个2×2列联表中的数据计算得*K*2＝4.013，那么在犯错误的概率不超过0.05的前提下认为两个变量\_\_\_\_\_\_(填“有”或“没有”)关系．**

**答案：有**

**4．(2013·韶关二模)以下四个命题：**

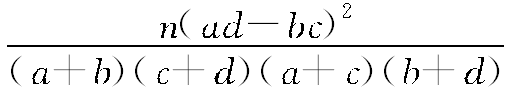
**①在一次试卷分析中，从每个试室中抽取第5号考生的成绩进行统计，是简单随机抽样；**

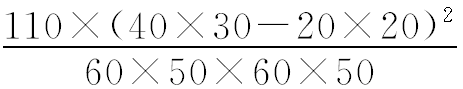
**②样本数据：3,4,5,6,7的方差为2；**

**③对于相关系数*r*，|*r*|越接近1，则线性相关程度越强；**

**④通过随机询问110名性别不同的行人，对过马路是愿意走斑马线还是愿意走人行天桥进行抽样调查，得到如下列联表：**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **男** | **女** | **总计** |
| **走天桥** | **40** | **20** | **60** |
| **走斑马线** | **20** | **30** | **50** |
| **总计** | **60** | **50** | **110** |

**由*K*2＝**

**可得，*K*2＝＝7.8，则有99%以上的把握认为“选择过马路方式与性别有关”，其中正确的命题序号是\_\_\_\_\_\_\_\_．**

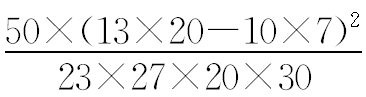
**答案：②③④**

**附表**

|  |  |  |  |
| --- | --- | --- | --- |
| ***P*(*K*2≥*k*0)** | **0.05** | **0.010** | **0.001** |
| ***k*0** | **3.841** | **6.635** | **10.828** |

**5.某学校为了调查喜欢语文学科与性别的关系，随机调查了一些学生情况，具体数据如下表：**

|  |  |  |
| --- | --- | --- |
| **类别**  **性别** | **不喜欢语文** | **喜欢语文** |
| **男** | **13** | **10** |
| **女** | **7** | **20** |

**为了判断喜欢语文学科是否与性别有关系，根据表中的数据，得到*K*2的观测值*k*＝≈4.844，因为*k*≥3.841，根据下表中的参考数据：**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ***P*(*K*2≥*k*0)** | **0.50** | **0.40** | **0.25** | **0.15** | **0.10** | **0.05** | **0.025** | **0.010** | **0.005** | **0.001** |
| ***k*0** | **0.455** | **0.708** | **1.323** | **2.072** | **2.706** | **3.841** | **5.024** | **6.635** | **7.879** | **10.828** |

**判定喜欢语文学科与性别有关系，那么这种判断出错的可能性为\_\_\_\_\_\_\_\_．**

**答案：5%**

**6．某学校课题组为了研究学生的数学成绩与物理成绩之间的关系，随机抽取高二年级20名学生某次考试成绩(满分100分)如下表所示：**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **序号** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** |
| **数学成绩** | **95** | **75** | **80** | **94** | **92** | **65** | **67** | **84** | **98** | **71** |
| **物理成绩** | **90** | **63** | **72** | **87** | **91** | **71** | **58** | **82** | **93** | **81** |

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **序号** | **11** | **12** | **13** | **14** | **15** | **16** | **17** | **18** | **19** | **20** |
| **数学成绩** | **67** | **93** | **64** | **78** | **77** | **90** | **57** | **83** | **72** | **83** |
| **物理成绩** | **77** | **82** | **48** | **85** | **69** | **91** | **61** | **84** | **78** | **86** |

**若单科成绩85以上(含85分)，则该科成绩优秀．**

**(1)根据上表完成下面的2×2列联表(单位：人).**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **数学成绩优秀** | **数学成绩不优秀** | **合计** |
| **物理成绩优秀** |  |  |  |
| **物理成绩不优秀** |  |  |  |
| **合计** |  |  |  |

**解析：(1)2×2列联表为(单位：人)：**

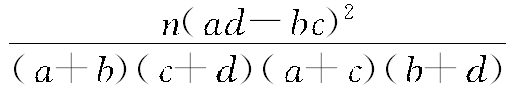
|  |  |  |  |
| --- | --- | --- | --- |
|  | **数学成绩优秀** | **数学成绩不优秀** | **合计** |
| **物理成绩优秀** | **5** | **2** | **7** |
| **物理成绩不优秀** | **1** | **12** | **13** |
| **合计** | **6** | **14** | **20** |

**(2)根据题(1)中表格的数据计算，能否在犯错误的概率不超过0.005的前提下认为学生的数学成绩与物理成绩之间有关系？**

**参数数据：**

1. **假设有两个分类变量*X*和*Y*，它们的值域分别为(*x*1，*x*2)和(*y*1，*y*2)，其样本频数列联表(称为2×2列联表)为：**

|  |  |  |  |
| --- | --- | --- | --- |
|  | ***y*1** | ***y*2** | **合计** |
| ***x*1** | ***a*** | ***b*** | ***a*＋*b*** |
| ***x*2** | ***c*** | ***d*** | ***c*＋*d*** |
| **合计** | ***a*＋*c*** | ***b*＋*d*** | ***a*＋*b*＋*c*＋*d*** |

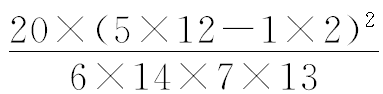
**则随机变量*K*2＝，其中*n*＝*a*＋*b*＋*c*＋*d*为样本容量；**

1. **独立检验随机变量*K*2的临界值参考表如下：**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| ***P*(*K*2≥*k*0)** | **0.50** | **0.40** | **0.25** | **0.15** | **0.10** |
| ***k*0** | **0.455** | **0.708** | **1.323** | **2.072** | **2.706** |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| ***P*(*K*2≥*k*0)** | **0.05** | **0.025** | **0.010** | **0.005** | **0.001** |
| ***k*0** | **3.841** | **5.024** | **6.635** | **7.879** | **10.828** |

**解析：根据列联表可以求得*K*2的观测值**

***k*＝≈8.802>7.879.**

**在犯错误的概率不超过0.005的前提下认为：学生的数学成绩与物理成绩之间有关系．**

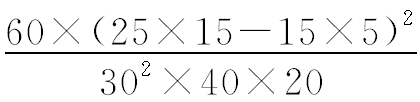
**7．(2013·深圳二模)2013年3月14日，CCTV财经频道报道了某地建筑市场存在违规使用未经淡化海砂的现象．为了研究使用淡化海砂与混凝土耐久性是否达标有关，某大学实验室随机抽取了60个样本，得到了相关数据如下表：**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **混凝土耐**  **久性达标** | **混凝土耐**  **久性不达标** | **总计** |
| **使用淡化海砂** | **25** | **5** | **30** |
| **使用未经淡化海砂** | **15** | **15** | **30** |
| **总计** | **40** | **20** | **60** |

**(1)根据表中数据，利用独立性检验的方法判断，能否在犯错误的概率不超过1%的前提下，认为使用淡化海砂与混凝土耐久性是否达标有关？**

**解析：提出假设*H*0：使用淡化海砂与混凝土耐久性是否达标无关．**

**根据表中数据，求得*K*2的观测值**

***k*＝＝7.5＞6.635.**

**查表得*P*(*K*2≥6.635)＝0.010.**

**∴能在犯错误的概率不超过1%的前提下，认为使用淡化海砂与混凝土耐久性是否达标有关．**

**(2)若用分层抽样的方法在使用淡化海砂的样本中抽取了6个，现从这6个样本中任取2个，则取出的2个样本混凝土耐久性都达标的概率是多少？**

**参考数据：**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| ***P*(*K*2≥*k*)** | **0.10** | **0.050** | **0.025** | **0.010** | **0.001** |
| ***k*** | **2.706** | **3.841** | **5.024** | **6.635** | **10.828** |

**解析：用分层抽样的方法在使用淡化海砂的样本中抽取6个，其中应抽取“混凝土耐久性达标”的为×6＝5，“混凝土耐久性不达标”的为6－5＝1，**

**“混凝土耐久性达标记”为*A*1，*A*2，*A*3，*A*4，*A*5”；“混凝土耐久性不达标”的记为*B*.**

**在这6个样本中任取2个，有以下几种可能：(*A*1，*A*2)，(*A*1，*A*3)，(*A*1，*A*4)，(*A*1，*A*5)，(*A*1，*B*)，(*A*2，*A*3)，(*A*2，*A*4)，(*A*2，*A*5)，(*A*2，*B*)，(*A*3，*A*4)，(*A*3，*A*5)，(*A*3，*B*)，(*A*4，*A*5)，(*A*4，*B*)(*A*5，*B*)，共15种．**

**设“取出的2个样本混凝土耐久性都达标”为事件*A*，它的对立事件为“取出的2个样本至少有1个混凝土耐久性不达标”，包含(*A*1，*B*)，(*A*2，*B*)，(*A*3，*B*)，(*A*4，*B*)，(*A*5，*B*)，共5种可能．**

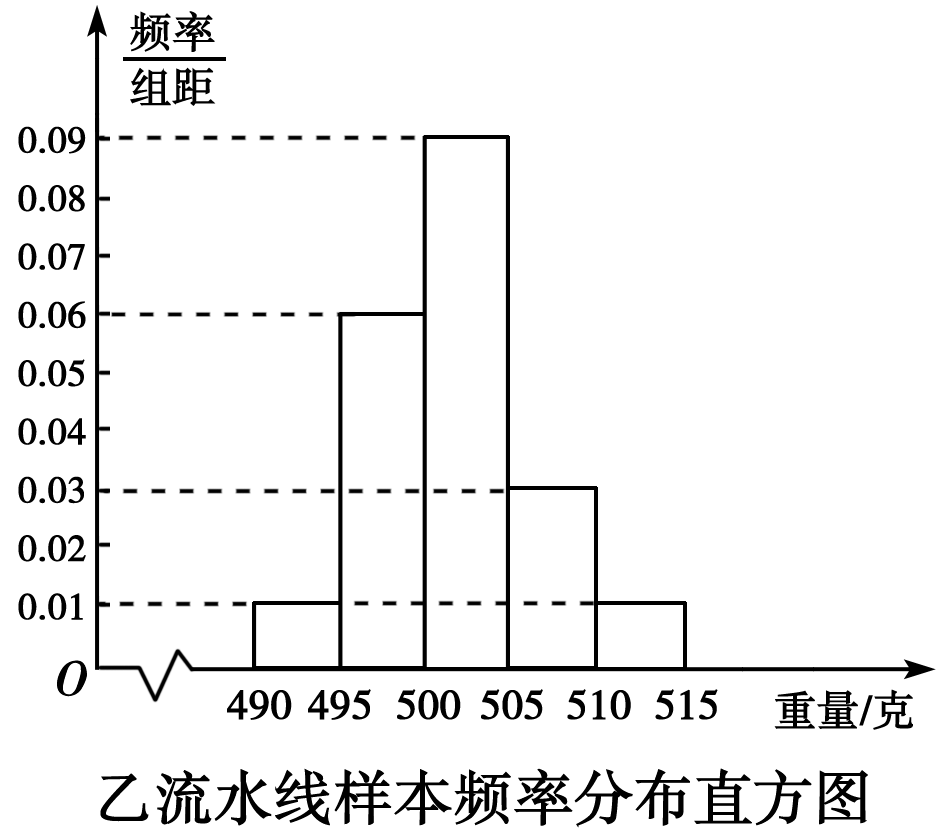
**∴*P*(*A*)＝1－*P*()＝1－＝.**

**即取出的2个样本混凝土耐久性都达标的概率是.**

**8．某食品厂为了检查甲、乙两条自动包装流水线的生产情况，随机在这两条流水线上各抽取40件产品作为样本称出它们的重量(单位：克)，重量值落在(495,510]的产品为合格品，否则为不合格品．左下表是甲流水线样本频数分布表，右下图是乙流水线样本的频率分布直方图．**

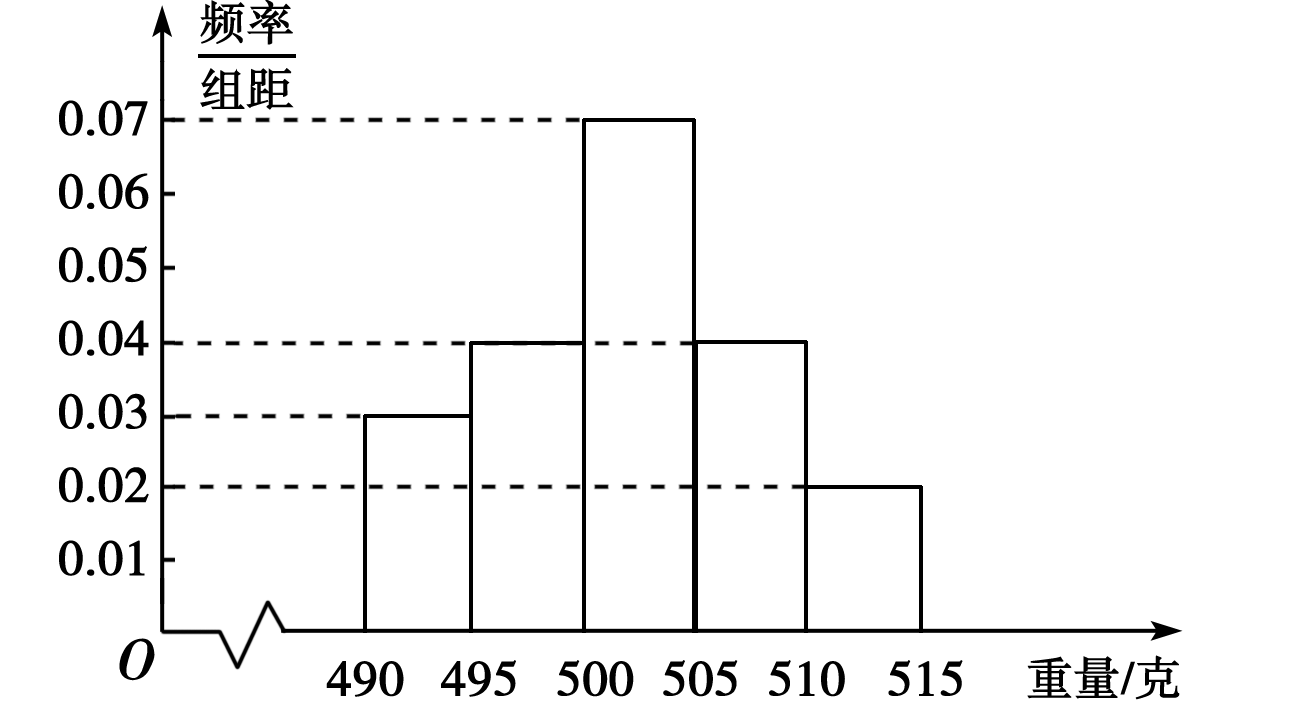
|  |  |
| --- | --- |
| **产品重量/克** | **频数** |
| **(490,495]** | **6** |
| **(495,500]** | **8** |
| **(500,505]** | **14** |
| **(505,510]** | **8** |
| **(510,515]** | **4** |

**甲流水线样本频数分布表**

****

**(1)根据上表数据作出甲流水线样本的频率分布直方图；**

**解析：甲流水线样本的频率分布直方图如下：**

****

**(2)若以频率作为概率，试估计从两条流水线分别任取1件产品，该产品恰好是合格品的概率；**

**解析：由题表知甲样本中合格品数为8＋14＋8＝30，由题图知乙样本中合格品数为(0.06＋0.09＋0.03)×5×40＝36，故甲样本合格品的频率为＝0.75，乙样本合格品的频率为＝0.9.**

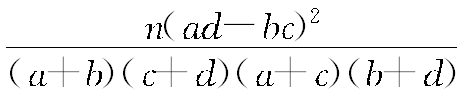
**据此可估计从甲流水线任取1件产品，该产品恰好是合格品的概率为0.75.从乙流水线任取1件产品，该产品恰好是合格品的概率为0.9.**

**(3)由以上统计数据完成下面2×2列联表，能否在犯错误的概率不超过0.1的前提下认为产品的包装质量与两条自动包装流水线的选择有关？**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **甲流水线** | **乙流水线** | **合计** |
| **合格品** | ***a*＝** | ***b*＝** |  |
| **不合格品** | ***c*＝** | ***d*＝** |  |
| **合计** |  |  | ***n*＝** |

**附表：**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| ***P*(*K*2≥*k*0)** | **0.15** | **0.10** | **0.05** | **0.025** | **0.010** | **0.005** | **0.001** |
| ***k*0** | **2.072** | **2.706** | **3.841** | **5.024** | **6.635** | **7.879** | **10.828** |

**(参考公式：*K*2＝，**

**其中*n*＝*a*＋*b*＋*c*＋*d*)**

**解析：2×2列联表如下：**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **甲流水线** | **乙流水线** | **合计** |
| **合格品** | ***a*＝30** | ***b*＝36** | **66** |
| **不合格品** | ***c*＝10** | ***d*＝4** | **14** |
| **合计** | **40** | **40** | ***n*＝80** |

**∵*K*2＝**

**＝≈3.117>2.706.**

**∴在犯错误的概率不超过0.1的前提下认为产品的包装质量与两条自动包装流水线的选择有关．**

**►品味高考**

**1．为调查某地区老年人是否需要志愿者提供帮助，用简单随机抽样方法从该地区调查了500位老人，结果如下：**

|  |  |  |
| --- | --- | --- |
| **性别**  **是否需要志愿者** | **男** | **女** |
| **需要** | **40** | **30** |
| **不需要** | **160** | **270** |

**(1)估计该地区老年人中需要志愿者提供帮助的老年人的比例．**

**解析：调查的500位老年人中有70位需要志愿者提供帮助，因此该地区老年人中需要帮助的老年人的比例的估计值为＝14%.**

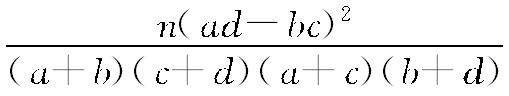
**(2)能否在犯错误的概率不超过0.01的前提下认为该地区的老年人是否需要志愿者提供帮助与性别有关？**

**解析：*K*2的观测值*k*＝≈9.967，**

**由于9.967>6.635所以在犯错误的概率不超过0.01的前提下认为该地区的老年人是否需要帮助与性别有关．**

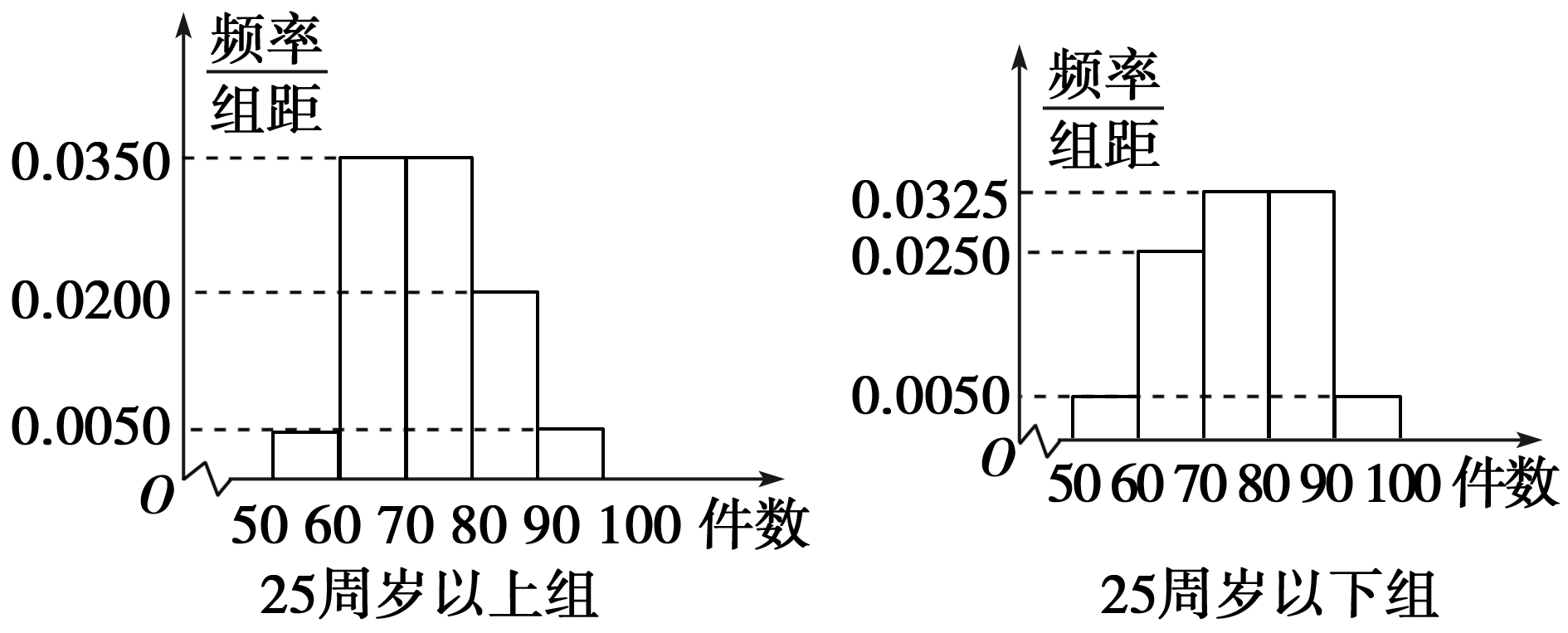
**(3)根据(2)的结论，能否提出更好的调查办法来估计该地区的老年人中需要志愿者提供帮助的老年人的比例？说明理由．**

**解析：由于(2)的结论知，该地区的老年人是否需要帮助与性别有关，并且从样本数据能看出该地区男性老年人与女性老年人中需要帮助的比例有明显差异，因此在调查时，先确定该地区老年人中男、女的比例，再把老年人分成男、女两层并采用分层抽样方法比采用简单随机抽样方法更好．**

**附：*K*2＝**

|  |  |  |  |
| --- | --- | --- | --- |
| ***P*(*K*2≥*k*0)** | **0.050** | **0.010** | **0.001** |
| ***k*0** | **3.841** | **6.635** | **10.828** |

**2．(2013·福建卷)某工厂有25周岁以上(含25周岁)工人300名，25周岁以下工人200名．为研究工人的日平均生产量是否与年龄有关，现采用分层抽样的方法，从中抽取了100名工人，先统计了他们某月的日平均生产件数，然后按工人年龄在“25周岁以上(含25周岁)”和“25周岁以下”分为两组，再将两组工人的日平均生产件数分为5组：[50,60)，[60,70)，[70,80)，[80,90)，[90,100)分别加以统计，得到如图所示的频率分布直方图．**

****

**(1)从样本中日平均生产件数不足60件的工人中随机抽取2人，求至少抽到一名“25周岁以下组”工人的概率；**

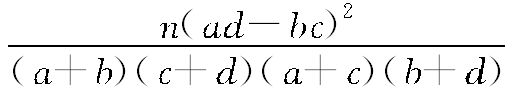
**解析：由已知得，样本中有25周岁以上组工人60名，25周岁以下组工人40名．**

**所以，样本中日平均生产件数不足60件的工人中，25周岁以上组工人有60×0.05＝3(人)，**

**记为*A*1，*A*2，*A*3；25周岁以下组工人有40×0.05＝2(人)，记为*B*1，*B*2.**

**从中随机抽取2名工人，所有的可能结果共有10种，它们是：(*A*1，*A*2)，(*A*1，*A*3)，(*A*2，*A*3)，(*A*1，*B*1)，(*A*1，*B*2)，(*A*2，*B*1)，(*A*2，*B*2)，(*A*3，*B*1)，(*A*3，*B*2)，(*B*1，*B*2)．其中至少有1名“25岁以下组”工人的可能结果共有7种，它们是：(*A*1，*B*1)，(*A*1，*B*2)，(*A*2，*B*1)，(*A*2，*B*2)，(*A*3，*B*1)，(*A*3，*B*2)，(*B*1，*B*2)．故所求的概率*P*＝.**

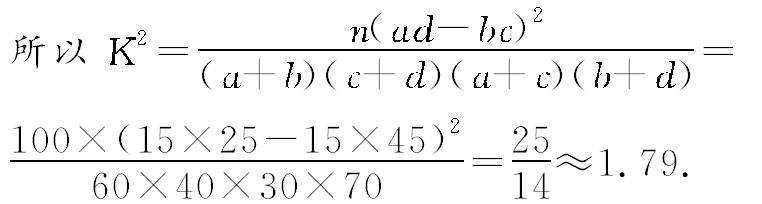
**(2)规定日平均生产件数不少于80件者为“生产能手”，请你根据已知条件完成列联表，并判断是否有90%的把握认为“生产能手与工人所在的年龄组有关”？**

**附：*K*2＝**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ***P*(*K*2≥*k*0)** | **0.100** | **0.050** | **0.010** | **0.001** |
| ***k*0** | **2.706** | **3.841** | **6.635** | **10.828** |

**解析：由频率分布直方图可知，在抽取的100名工人中，“25周岁以上组”中的生产能手60×0.25＝15(人)，“25周岁以下组”中的生产能手40×0.375＝15(人)，据此可得2×2列联表如下：**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **生产能手** | **非生产能手** | **合计** |
| **25周岁以上组** | **15** | **45** | **60** |
| **25周岁以下组** | **15** | **25** | **40** |
| **合计** | **30** | **70** | **100** |

****

**因为1.79＜2.706，所以没有90%的把握认为“生产能手与工人所在年龄组有关”．**