三年专题15 概率与统计（解答题）

（理科专用）

1．【2022年全国甲卷】甲、乙两个学校进行体育比赛，比赛共设三个项目，每个项目胜方得10分，负方得0分，没有平局．三个项目比赛结束后，总得分高的学校获得冠军．已知甲学校在三个项目中获胜的概率分别为0.5，0.4，0.8，各项目的比赛结果相互独立．

(1)求甲学校获得冠军的概率；

(2)用*X*表示乙学校的总得分，求*X*的分布列与期望．

【答案】(1)$0.6$；

(2)分布列见解析，$E\left(X\right)=13$.

【解析】

【分析】

（1）设甲在三个项目中获胜的事件依次记为$A,B,C$，再根据甲获得冠军则至少获胜两个项目，利用互斥事件的概率加法公式以及相互独立事件的乘法公式即可求出；

（2）依题可知，$X$的可能取值为$0,10,20,30$，再分别计算出对应的概率，列出分布列，即可求出期望．

(1)

设甲在三个项目中获胜的事件依次记为$A,B,C$，所以甲学校获得冠军的概率为

$$P=P\left(ABC\right)+P\left(\overbar{A}BC\right)+P\left(A\overbar{B}C\right)+P\left(AB\overbar{C}\right)$$

$$=0.5×0.4×0.8+0.5×0.4×0.8+0.5×0.6×0.8+0.5×0.4×0.2$$

$=0.16+0.16+0.24+0.04=0.6$．

(2)

依题可知，$X$的可能取值为$0,10,20,30$，所以，

$P\left(X=0\right)=0.5×0.4×0.8=0.16$,

$P\left(X=10\right)=0.5×0.4×0.8+0.5×0.6×0.8+0.5×0.4×0.2=0.44$，

$P\left(X=20\right)=0.5×0.6×0.8+0.5×0.4×0.2+0.5×0.6×0.2=0.34$，

$P\left(X=30\right)=0.5×0.6×0.2=0.06$.

即$X$的分布列为

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| $$X$$ | 0 | 10 | 20 | 30 |
| $$P$$ | 0.16 | 0.44 | 0.34 | 0.06 |

期望$E\left(X\right)=0×0.16+10×0.44+20×0.34+30×0.06=13$.

2．【2022年新高考1卷】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），得到如下数据：

|  |  |  |
| --- | --- | --- |
|  | 不够良好 | 良好 |
| 病例组 | 40 | 60 |
| 对照组 | 10 | 90 |

(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？

(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”．$\frac{P(B|A)}{P(\overbar{B}|A)}$与$\frac{P(B|\overbar{A})}{P(\overbar{B}|\overbar{A})}$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：$R=\frac{P(A|B)}{P(\overbar{A}|B)}⋅\frac{P(\overbar{A}|\overbar{B})}{P(A|\overbar{B})}$；

（ⅱ）利用该调查数据，给出$P(A|B),P(A|\overbar{B})$的估计值，并利用（ⅰ）的结果给出*R*的估计值．

附$K^{2}=\frac{n(ad−bc)^{2}}{(a+b)(c+d)(a+c)(b+d)}$，

|  |  |  |  |
| --- | --- | --- | --- |
| $$P\left(K^{2}\geq k\right)$$ | 0.050 | 0.010 | 0.001 |
| *k* | 3.841 | 6.635 | 10.828 |

【答案】(1)答案见解析

(2)（i）证明见解析；(ii)$R=6$；

【解析】

【分析】

(1)由所给数据结合公式求出$K^{2}$的值，将其与临界值比较大小，由此确定是否有99%的把握认为患该疾病群体与未黄该疾病群体的卫生习惯有差异；(2)(i) 根据定义结合条件概率公式即可完成证明；(ii)根据（i）结合已知数据求$R$.

(1)

由已知$K^{2}=\frac{n(ad−bc)^{2}}{(a+b)(c+d)(a+c)(b+d)}=\frac{200(40×90−60×10)^{2}}{50×150×100×100}=24$，

又$P(K^{2}\geq 6.635)=0.01$，$24>6.635$，

所以有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.

(2)

(i)因为$R=\frac{P(B|A)}{P(\overbar{B}|A)}⋅\frac{P(\overbar{B}|\overbar{A})}{P(B|\overbar{A})}=\frac{P(AB)}{P(A)}⋅\frac{P(A)}{P(A\overbar{B})}⋅\frac{P(\overbar{A}\overbar{B})}{P(\overbar{A})}⋅\frac{P(\overbar{A})}{P(\overbar{A}B)}$，

所以$R=\frac{P(AB)}{P(B)}⋅\frac{P(B)}{P(\overbar{A}B)}⋅\frac{P(\overbar{A}\overbar{B})}{P(\overbar{B})}⋅\frac{P(\overbar{B})}{P(A\overbar{B})}$

所以$R=\frac{P(A|B)}{P(\overbar{A}|B)}⋅\frac{P(\overbar{A}|\overbar{B})}{P(A|\overbar{B})}$，

(ii)

由已知$P(A|B)=\frac{40}{100}$，$P(A|\overbar{B})=\frac{10}{100}$，

又$P(\overbar{A}|B)=\frac{60}{100}$，$P(\overbar{A}|\overbar{B})=\frac{90}{100}$，

所以$R=\frac{P(A|B)}{P(\overbar{A}|B)}⋅\frac{P(\overbar{A}|\overbar{B})}{P(A|\overbar{B})}=6$

3．【2022年新高考2卷】在某地区进行流行病学调查，随机调查了100位某种疾病患者的年龄，得到如下的样本数据的频率分布直方图：



(1)估计该地区这种疾病患者的平均年龄（同一组中的数据用该组区间的中点值为代表）；

(2)估计该地区一位这种疾病患者的年龄位于区间$[20,70)$的概率；

(3)已知该地区这种疾病的患病率为$0.1\%$，该地区年龄位于区间$[40,50)$的人口占该地区总人口的$16\%$.从该地区中任选一人，若此人的年龄位于区间$[40,50)$，求此人患这种疾病的概率．（以样本数据中患者的年龄位于各区间的频率作为患者的年龄位于该区间的概率，精确到0.0001）.

【答案】(1)$44.65$岁；

(2)$0.89$；

(3)$0.0014$．

【解析】

【分析】

（1）根据平均值等于各矩形的面积乘以对应区间的中点值的和即可求出；

（2）设$A=${一人患这种疾病的年龄在区间$[20,70)$}，根据对立事件的概率公式$P(A)=1−P(\overbar{A})$即可解出；

（3）根据条件概率公式即可求出．

(1)

平均年龄$\overbar{x}=(5×0.001+15×0.002+25×0.012+35×0.017+45×0.023$

       $+55×0.020+65×0.012+75×0.006+85×0.002)×10=44.65$（岁）．

(2)

设$A=${一人患这种疾病的年龄在区间$[20,70)$}，所以

$P(A)=1−P(\overbar{A})=1−(0.001+0.002+0.006+0.002)×10=1−0.11=0.89$．

(3)

设$B=\{$任选一人年龄位于区间$\left.[40,50)\right\}$，$C=\{$任选一人患这种疾病$\}$，

则由条件概率公式可得

$P(C|B)=\frac{P(BC)}{P(B)}=\frac{0.1\%×0.023×10}{16\%}=\frac{0.001×0.23}{0.16}=0.0014375≈0.0014$．

4．【2021年新高考1卷】某学校组织“一带一路”知识竞赛，有*A*，*B*两类问题，每位参加比赛的同学先在两类问题中选择一类并从中随机抽取一个问题回答，若回答错误则该同学比赛结束；若回答正确则从另一类问题中再随机抽取一个问题回答，无论回答正确与否，该同学比赛结束.*A*类问题中的每个问题回答正确得20分，否则得0分；*B*类问题中的每个问题回答正确得80分，否则得0分，已知小明能正确回答*A*类问题的概率为0.8，能正确回答*B*类问题的概率为0.6，且能正确回答问题的概率与回答次序无关.

（1）若小明先回答*A*类问题，记为小明的累计得分，求的分布列；

（2）为使累计得分的期望最大，小明应选择先回答哪类问题？并说明理由.

【答案】（1）见解析；（2）类．

【解析】

【分析】

（1）通过题意分析出小明累计得分的所有可能取值，逐一求概率列分布列即可．（2）与（1）类似，找出先回答类问题的数学期望，比较两个期望的大小即可．

【详解】

（1）由题可知，的所有可能取值为，，．

；

；

．

所以的分布列为

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
|  |  |  |  |

（2）由（1）知，．

若小明先回答问题，记为小明的累计得分，则的所有可能取值为，，．

；

；

．

所以．

因为，所以小明应选择先回答类问题．

5．【2021年新高考2卷】一种微生物群体可以经过自身繁殖不断生存下来，设一个这种微生物为第0代，经过一次繁殖后为第1代，再经过一次繁殖后为第2代……，该微生物每代繁殖的个数是相互独立的且有相同的分布列，设*X*表示1个微生物个体繁殖下一代的个数，．

（1）已知，求；

（2）设*p*表示该种微生物经过多代繁殖后临近灭绝的概率，*p*是关于*x*的方程：的一个最小正实根，求证：当时，，当时，；

（3）根据你的理解说明（2）问结论的实际含义．

【答案】（1）1；（2）见解析；（3）见解析.

【解析】

【分析】

（1）利用公式计算可得.

（2）利用导数讨论函数的单调性，结合及极值点的范围可得的最小正零点.

（3）利用期望的意义及根的范围可得相应的理解说明.

【详解】

（1）.

（2）设，

因为，故，

若，则，故.

，

因为，，

故有两个不同零点，且，

且时，；时，；

故在，上为增函数，在上为减函数，

若，因为在为增函数且，

而当时，因为在上为减函数，故，

故为的一个最小正实根，

若，因为且在上为减函数，故1为的一个最小正实根，

综上，若，则.

若，则，故.

此时，，

故有两个不同零点，且，

且时，；时，；

故在，上为增函数，在上为减函数，

而，故，

又，故在存在一个零点，且.

所以为的一个最小正实根，此时，

故当时，.

（3）意义：每一个该种微生物繁殖后代的平均数不超过1，则若干代必然灭绝，若繁殖后代的平均数超过1，则若干代后被灭绝的概率小于1.

6．【2020年新课标1卷理科】甲、乙、丙三位同学进行羽毛球比赛，约定赛制如下：累计负两场者被淘汰；比赛前抽签决定首先比赛的两人，另一人轮空；每场比赛的胜者与轮空者进行下一场比赛，负者下一场轮空，直至有一人被淘汰；当一人被淘汰后，剩余的两人继续比赛，直至其中一人被淘汰，另一人最终获胜，比赛结束.经抽签，甲、乙首先比赛，丙轮空.设每场比赛双方获胜的概率都为，

（1）求甲连胜四场的概率；

（2）求需要进行第五场比赛的概率；

（3）求丙最终获胜的概率.

【答案】（1）；（2）；（3）.

【解析】

【分析】

（1）根据独立事件的概率乘法公式可求得事件“甲连胜四场”的概率；

（2）计算出四局以内结束比赛的概率，然后利用对立事件的概率公式可求得所求事件的概率；

（3）列举出甲赢的基本事件，结合独立事件的概率乘法公式计算出甲赢的概率，由对称性可知乙赢的概率和甲赢的概率相等，再利用对立事件的概率可求得丙赢的概率.

【详解】

（1）记事件甲连胜四场，则；

（2）记事件为甲输，事件为乙输，事件为丙输，

则四局内结束比赛的概率为

，

所以，需要进行第五场比赛的概率为；

（3）记事件为甲输，事件为乙输，事件为丙输，

记事件甲赢，记事件丙赢，

则甲赢的基本事件包括：、、、

、、、、，

所以，甲赢的概率为.

由对称性可知，乙赢的概率和甲赢的概率相等，

所以丙赢的概率为.

【点睛】

本题考查独立事件概率的计算，解答的关键就是列举出符合条件的基本事件，考查计算能力，属于中等题.

7．【2020年新课标2卷理科】某沙漠地区经过治理，生态系统得到很大改善，野生动物数量有所增加.为调查该地区某种野生动物的数量，将其分成面积相近的200个地块，从这些地块中用简单随机抽样的方法抽取20个作为样区，调查得到样本数据(*xi*，*yi*)(*i*=1，2，…，20)，其中*xi*和*yi*分别表示第*i*个样区的植物覆盖面积(单位：公顷)和这种野生动物的数量，并计算得，，，，.

（1）求该地区这种野生动物数量的估计值（这种野生动物数量的估计值等于样区这种野生动物数量的平均数乘以地块数）；

（2）求样本(*xi*，*yi*)(*i*=1，2，…，20)的相关系数（精确到0.01）；

（3）根据现有统计资料，各地块间植物覆盖面积差异很大.为提高样本的代表性以获得该地区这种野生动物数量更准确的估计，请给出一种你认为更合理的抽样方法，并说明理由.

附：相关系数*r*=，≈1.414.

【答案】（1）；（2）；（3）详见解析

【解析】

【分析】

（1）利用野生动物数量的估计值等于样区野生动物平均数乘以地块数，代入数据即可；

（2）利用公式计算即可；

（3）各地块间植物覆盖面积差异较大，为提高样本数据的代表性，应采用分层抽样.

【详解】

（1）样区野生动物平均数为，

地块数为200，该地区这种野生动物的估计值为

（2）样本(*i*=1，2，…，20)的相关系数为



（3）由（2）知各样区的这种野生动物的数量与植物覆盖面积有很强的正相关性，

由于各地块间植物覆盖面积差异很大，从而各地块间这种野生动物的数量差异很大，

采用分层抽样的方法较好地保持了样本结构与总体结构的一致性，提高了样本的代表性，

从而可以获得该地区这种野生动物数量更准确的估计.

【点晴】

本题主要考查平均数的估计值、相关系数的计算以及抽样方法的选取，考查学生数学运算能力，是一道容易题.

8．【2020年新课标3卷理科】某学生兴趣小组随机调查了某市100天中每天的空气质量等级和当天到某公园锻炼的人次，整理数据得到下表（单位：天）：

|  |  |  |  |
| --- | --- | --- | --- |
| 锻炼人次空气质量等级 | [0，200] | (200，400] | (400，600] |
| 1（优） | 2 | 16 | 25 |
| 2（良） | 5 | 10 | 12 |
| 3（轻度污染） | 6 | 7 | 8 |
| 4（中度污染） | 7 | 2 | 0 |

（1）分别估计该市一天的空气质量等级为1，2，3，4的概率；

（2）求一天中到该公园锻炼的平均人次的估计值（同一组中的数据用该组区间的中点值为代表）；

（3）若某天的空气质量等级为1或2，则称这天“空气质量好”；若某天的空气质量等级为3或4，则称这天“空气质量不好”．根据所给数据，完成下面的2×2列联表，并根据列联表，判断是否有95%的把握认为一天中到该公园锻炼的人次与该市当天的空气质量有关？

|  |  |  |
| --- | --- | --- |
|  | 人次≤400 | 人次>400 |
| 空气质量好 |  |  |
| 空气质量不好 |  |  |

附：，

|  |  |  |  |
| --- | --- | --- | --- |
| *P*(*K2*≥*k*) | 0.050    | 0.010  | 0.001 |
| *k* | 3.841 | 6.635 | 10.828 |

【答案】（1）该市一天的空气质量等级分别为、、、的概率分别为、、、；（2）；（3）有，理由见解析.

【解析】

【分析】

（1）根据频数分布表可计算出该市一天的空气质量等级分别为、、、的概率；

（2）利用每组的中点值乘以频数，相加后除以可得结果；

（3）根据表格中的数据完善列联表，计算出的观测值，再结合临界值表可得结论.

【详解】

（1）由频数分布表可知，该市一天的空气质量等级为的概率为，等级为的概率为，等级为的概率为，等级为的概率为；

（2）由频数分布表可知，一天中到该公园锻炼的人次的平均数为

（3）列联表如下：

|  |  |  |
| --- | --- | --- |
|  | 人次 | 人次 |
| 空气质量好 |  |  |
| 空气质量不好 |  |  |

，

因此，有的把握认为一天中到该公园锻炼的人次与该市当天的空气质量有关.

【点睛】

本题考查利用频数分布表计算频率和平均数，同时也考查了独立性检验的应用，考查数据处理能力，属于基础题.

9．【2020年新高考1卷（山东卷）】为加强环境保护，治理空气污染，环境监测部门对某市空气质量进行调研，随机抽查了天空气中的和浓度（单位：），得下表：



（1）估计事件“该市一天空气中浓度不超过，且浓度不超过”的概率；

（2）根据所给数据，完成下面的列联表：



（3）根据（2）中的列联表，判断是否有的把握认为该市一天空气中浓度与浓度有关？

附：，



【答案】（1）；（2）答案见解析；（3）有.

【解析】

【分析】

（1）根据表格中数据以及古典概型的概率公式可求得结果；

（2）根据表格中数据可得列联表；

（3）计算出，结合临界值表可得结论.

【详解】

（1）由表格可知，该市100天中，空气中的浓度不超过75，且浓度不超过150的天数有天，

所以该市一天中，空气中的浓度不超过75，且浓度不超过150的概率为；

（2）由所给数据，可得列联表为：

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  | 合计 |
|  | 64 | 16 | 80 |
|  | 10 | 10 | 20 |
| 合计 | 74 | 26 | 100 |

（3）根据列联表中的数据可得

，

因为根据临界值表可知，有的把握认为该市一天空气中浓度与浓度有关.

【点睛】

本题考查了古典概型的概率公式，考查了完善列联表，考查了独立性检验，属于中档题.

