

专题 16 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, $E(X) = 13$.

【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

$$\begin{aligned} P &= P(ABC) + P(\bar{A}BC) + P(A\bar{B}C) + P(AB\bar{C}) \\ &= 0.5 \times 0.4 \times 0.8 + 0.5 \times 0.4 \times 0.8 + 0.5 \times 0.6 \times 0.8 + 0.5 \times 0.4 \times 0.2 \\ &= 0.16 + 0.16 + 0.24 + 0.04 = 0.6. \end{aligned}$$

(2)

依题可知, X 的可能取值为 0,10,20,30, 所以,

$$P(X = 0) = 0.5 \times 0.4 \times 0.8 = 0.16,$$

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

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

$$P(X = 30) = 0.5 \times 0.6 \times 0.2 = 0.06.$$

即 X 的分布列为

X	0	10	20	30
P	0.16	0.44	0.34	0.06

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

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”， $\frac{P(B|A)}{P(B|\bar{A})}$ 与 $\frac{P(\bar{B}|\bar{A})}{P(\bar{B}|A)}$ 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： $R = \frac{P(A|B)}{P(\bar{A}|B)} \cdot \frac{P(\bar{A}|\bar{B})}{P(A|\bar{B})}$;

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

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

$P(K^2 \geq k)$	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患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)

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

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

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)

$$(i) \text{ 因为 } R = \frac{P(B|A)}{P(\bar{B}|A)} \cdot \frac{P(\bar{B}|\bar{A})}{P(B|\bar{A})} = \frac{P(AB)}{P(\bar{A}\bar{B})} \cdot \frac{P(A)}{P(\bar{A})} \cdot \frac{P(\bar{A}\bar{B})}{P(\bar{A})} \cdot \frac{P(\bar{A})}{P(AB)},$$

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

$$\text{所以 } R = \frac{P(A|B)}{P(\bar{A}|B)} \cdot \frac{P(\bar{A}|\bar{B})}{P(A|\bar{B})},$$

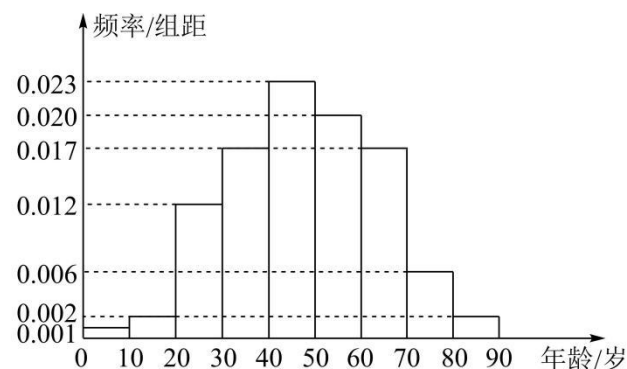
(ii)

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

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

$$\text{所以 } R = \frac{P(A|B)}{P(\bar{A}|B)} \cdot \frac{P(\bar{A}|\bar{B})}{P(A|\bar{B})} = 6$$

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

(3) 根据条件概率公式即可求出.

(1)

平均年龄 $\bar{x} = (5 \times 0.001 + 15 \times 0.002 + 25 \times 0.012 + 35 \times 0.017 + 45 \times 0.023 + 55 \times 0.020 + 65 \times 0.012 + 75 \times 0.006 + 85 \times 0.002) \times 10 = 44.65$ (岁).

(2)

设 $A = \{\text{一人患这种疾病的年龄在区间}[20, 70)\}$, 所以

$P(A) = 1 - P(\bar{A}) = 1 - (0.001 + 0.002 + 0.006 + 0.002) \times 10 = 1 - 0.11 = 0.89$.

(3)

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

则由条件概率公式可得

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

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 X 为小明的累计得分, 求 X 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) B 类.

【解析】

【分析】

(1) 通过题意分析出小明累计得分 X 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 B 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, X 的所有可能取值为 0, 20, 100.

$$P(X=0)=1-0.8=0.2;$$

$$P(X=20)=0.8(1-0.6)=0.32;$$

$$P(X=100)=0.8\times 0.6=0.48.$$

所以 X 的分布列为

X	0	20	100
P	0.2	0.32	0.48

$$(2) \text{ 由 (1) 知, } E(X)=0\times 0.2+20\times 0.32+100\times 0.48=54.4.$$

若小明先回答 B 问题, 记 Y 为小明的累计得分, 则 Y 的所有可能取值为 0, 80, 100.

$$P(Y=0)=1-0.6=0.4;$$

$$P(Y=80)=0.6(1-0.8)=0.12;$$

$$P(Y=100)=0.6\times 0.6=0.48.$$

$$\text{所以 } E(Y)=0\times 0.4+80\times 0.12+100\times 0.48=57.6.$$

因为 $54.4 < 57.6$, 所以小明应选择先回答 B 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,

$$P(X=i)=p_i(i=0,1,2,3).$$

(1) 已知 $p_0=0.4, p_1=0.3, p_2=0.2, p_3=0.1$, 求 $E(X)$;

(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:

$p_0+p_1x+p_2x^2+p_3x^3=x$ 的一个最小正实根, 求证: 当 $E(X)\leq 1$ 时, $p=1$, 当 $E(X)>1$ 时,

$p<1$;

(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 $E(X)$.

(2) 利用导数讨论函数的单调性, 结合 $f(1)=0$ 及极值点的范围可得 $f(x)$ 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) $E(X) = 0 \times 0.4 + 1 \times 0.3 + 2 \times 0.2 + 3 \times 0.1 = 1$.

(2) 设 $f(x) = p_3x^3 + p_2x^2 + (p_1 - 1)x + p_0$,

因为 $p_3 + p_2 + p_1 + p_0 = 1$, 故 $f(x) = p_3x^3 + p_2x^2 - (p_2 + p_0 + p_3)x + p_0$,

若 $E(X) \leq 1$, 则 $p_1 + 2p_2 + 3p_3 \leq 1$, 故 $p_2 + 2p_3 \leq p_0$.

$f'(x) = 3p_3x^2 + 2p_2x - (p_2 + p_0 + p_3)$,

因为 $f'(0) = -(p_2 + p_0 + p_3) < 0$, $f'(1) = p_2 + 2p_3 - p_0 \leq 0$,

故 $f'(x)$ 有两个不同零点 x_1, x_2 , 且 $x_1 < 0 < 1 \leq x_2$,

且 $x \in (-\infty, x_1) \cup (x_2, +\infty)$ 时, $f'(x) > 0$; $x \in (x_1, x_2)$ 时, $f'(x) < 0$;

故 $f(x)$ 在 $(-\infty, x_1)$, $(x_2, +\infty)$ 上为增函数, 在 (x_1, x_2) 上为减函数,

若 $x_2 = 1$, 因为 $f(x)$ 在 $(x_2, +\infty)$ 为增函数且 $f(1) = 0$,

而当 $x \in (0, x_2)$ 时, 因为 $f(x)$ 在 (x_1, x_2) 上为减函数, 故 $f(x) > f(x_2) = f(1) = 0$,

故 1 为 $p_0 + p_1x + p_2x^2 + p_3x^3 = x$ 的一个最小正实根,

若 $x_2 > 1$, 因为 $f(1) = 0$ 且在 $(0, x_2)$ 上为减函数, 故 1 为 $p_0 + p_1x + p_2x^2 + p_3x^3 = x$ 的一个最小正实根,

综上, 若 $E(X) \leq 1$, 则 $p = 1$.

若 $E(X) > 1$, 则 $p_1 + 2p_2 + 3p_3 > 1$, 故 $p_2 + 2p_3 > p_0$.

此时 $f'(0) = -(p_2 + p_0 + p_3) < 0$, $f'(1) = p_2 + 2p_3 - p_0 > 0$,

故 $f'(x)$ 有两个不同零点 x_3, x_4 , 且 $x_3 < 0 < x_4 < 1$,

且 $x \in (-\infty, x_3) \cup (x_4, +\infty)$ 时, $f'(x) > 0$; $x \in (x_3, x_4)$ 时, $f'(x) < 0$;

故 $f(x)$ 在 $(-\infty, x_3)$, $(x_4, +\infty)$ 上为增函数, 在 (x_3, x_4) 上为减函数,

而 $f(1) = 0$, 故 $f(x_4) < 0$,

又 $f(0) = p_0 > 0$, 故 $f(x)$ 在 $(0, x_4)$ 存在一个零点 p , 且 $p < 1$.

所以 p 为 $p_0 + p_1x + p_2x^2 + p_3x^3 = x$ 的一个最小正实根, 此时 $p < 1$,

故当 $E(X) > 1$ 时, $p < 1$.

(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 $\frac{1}{2}$,

(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) $\frac{1}{16}$; (2) $\frac{3}{4}$; (3) $\frac{7}{16}$.

【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 M : 甲连胜四场, 则 $P(M) = \left(\frac{1}{2}\right)^4 = \frac{1}{16}$;

(2) 记事件 A 为甲输, 事件 B 为乙输, 事件 C 为丙输,

则四局内结束比赛的概率为

$$P' = P(ABAB) + P(ACAC) + P(BCBC) + P(BABA) = 4 \times \left(\frac{1}{2}\right)^4 = \frac{1}{4},$$

所以, 需要进行第五场比赛的概率为 $P = 1 - P' = \frac{3}{4}$;

(3) 记事件 A 为甲输, 事件 B 为乙输, 事件 C 为丙输,

记事件 M : 甲赢, 记事件 N : 丙赢,

则甲赢的基本事件包括: $BCBC$ 、 $ABCBC$ 、 $ACBCB$ 、

$BABCC$ 、 $BACBC$ 、 $BCACB$ 、 $BCABC$ 、 $BCBAC$,

所以, 甲赢的概率为 $P(M) = \left(\frac{1}{2}\right)^4 + 7 \times \left(\frac{1}{2}\right)^5 = \frac{9}{32}$.

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 $P(N) = 1 - 2 \times \frac{9}{32} = \frac{7}{16}$.

【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 $(x_i, y_i) (i=1, 2, \dots, 20)$, 其中 x_i 和 y_i 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 $\sum_{i=1}^{20} x_i = 60$, $\sum_{i=1}^{20} y_i = 1200$, $\sum_{i=1}^{20} (x_i - \bar{x})^2 = 80$, $\sum_{i=1}^{20} (y_i - \bar{y})^2 = 9000$,

$\sum_{i=1}^{20} (x_i - \bar{x})(y_i - \bar{y}) = 800$.

(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 $(x_i, y_i) (i=1, 2, \dots, 20)$ 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 $r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}$, $\sqrt{2} \approx 1.414$.

【答案】(1) 12000; (2) 0.94; (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 $r = \frac{\sum_{i=1}^{20} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{20} (x_i - \bar{x})^2 \sum_{i=1}^{20} (y_i - \bar{y})^2}}$ 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 $\frac{1}{20} \sum_{i=1}^{20} y_i = \frac{1}{20} \times 1200 = 60$,

地块数为 200, 该地区这种野生动物的估计值为 $200 \times 60 = 12000$

(2) 样本 $(x_i, y_i) (i=1, 2, \dots, 20)$ 的相关系数为

$$r = \frac{\sum_{i=1}^{20} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{20} (x_i - \bar{x})^2 \sum_{i=1}^{20} (y_i - \bar{y})^2}} = \frac{800}{\sqrt{80 \times 9000}} = \frac{2\sqrt{2}}{3} \approx 0.94$$

(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.

【点睛】

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

$P(K^2 \geq k)$	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 1、2、3、4 的概率分别为 0.43、0.27、0.21、0.09；（2）350；（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 1、2、3、4 的概率；

(2) 利用每组的中点值乘以频数，相加后除以100可得结果；

(3) 根据表格中的数据完善 2×2 列联表，计算出 K^2 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1的概率为 $\frac{2+16+25}{100} = 0.43$ ，等级为2的概率为 $\frac{5+10+12}{100} = 0.27$ ，等级为3的概率为 $\frac{6+7+8}{100} = 0.21$ ，等级为4的概率为 $\frac{7+2+0}{100} = 0.09$ ；

(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

$$\frac{100 \times 20 + 300 \times 35 + 500 \times 45}{100} = 350$$

(3) 2×2 列联表如下：

	人次 ≤ 400	人次 > 400
空气质量好	33	37
空气质量不好	22	8

$$K^2 = \frac{100 \times (33 \times 8 - 37 \times 22)^2}{55 \times 45 \times 70 \times 30} \approx 5.820 > 3.841,$$

因此，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.

【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100天空气中的PM_{2.5}和SO₂浓度（单位： $\mu\text{g}/\text{m}^3$ ），得下表：

SO ₂ PM2.5	[0,50]	(50,150]	(150,475]
[0,35]	32	18	4
(35,75]	6	8	12
(75,115]	3	7	10

- (1) 估计事件“该市一天空气中 PM2.5 浓度不超过 75，且 SO₂ 浓度不超过 150”的概率；
- (2) 根据所给数据，完成下面的 2×2 列联表：

SO ₂ PM2.5	[0,150]	(150,475]
[0,75]		
(75,115]		

- (3) 根据 (2) 中的列联表，判断是否有 99% 的把握认为该市一天空气中 PM2.5 浓度与 SO₂ 浓度有关？

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

$P(K^2 \geq k)$	0.050	0.010	0.001
k	3.841	6.635	10.828

【答案】(1) 0.64；(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 2×2 列联表；
- (3) 计算出 K^2 ，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 $PM_{2.5}$ 浓度不超过 75, 且 SO_2 浓度不超过 150 的天数有 $32+6+18+8=64$ 天,

所以该市一天中, 空气中的 $PM_{2.5}$ 浓度不超过 75, 且 SO_2 浓度不超过 150 的概率为

$$\frac{64}{100} = 0.64;$$

(2) 由所给数据, 可得 2×2 列联表为:

SO_2 $PM_{2.5}$	$[0,150]$	$(150,475]$	合计
$[0,75]$	64	16	80
$(75,115]$	10	10	20
合计	74	26	100

(3) 根据 2×2 列联表中的数据可得

$$K^2 = \frac{n(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)} = \frac{100 \times (64 \times 10 - 16 \times 10)^2}{80 \times 20 \times 74 \times 26} = \frac{3600}{481} \approx 7.4844 > 6.635,$$

因为根据临界值表可知, 有 99% 的把握认为该市一天空气中 $PM_{2.5}$ 浓度与 SO_2 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 2×2 列联表, 考查了独立性检验, 属于中档题.

10. 【2019 年新课标 1 卷理科】为了治疗某种疾病, 研制了甲、乙两种新药, 希望知道哪种新药更有效, 为此进行动物试验. 试验方案如下: 每一轮选取两只白鼠对药效进行对比试验. 对于两只白鼠, 随机选一只施以甲药, 另一只施以乙药. 一轮的治疗结果得出后, 再安排下一轮试验. 当其中一种药治愈的白鼠比另一种药治愈的白鼠多 4 只时, 就停止试验, 并认为治愈只数多的药更有效. 为了方便描述问题, 约定: 对于每轮试验, 若施以甲药的白鼠治愈且施以乙药的白鼠未治愈则甲药得 1 分, 乙药得 -1 分; 若施以乙药的白鼠治愈且施以甲药的白鼠未治愈则乙药得 1 分, 甲药得 -1 分; 若都治愈或都未治愈则两种药均得 0 分. 甲、乙两种药的治愈率分别记为 α 和 β , 一轮试验中甲药的得分记为 X .

(1) 求 X 的分布列;

(2) 若甲药、乙药在试验开始时都赋予 4 分， $p_i (i=0,1,\cdots,8)$ 表示“甲药的累计得分为 i 时，最终认为甲药比乙药更有效”的概率，则 $p_0=0$ ， $p_8=1$ ， $p_i=ap_{i-1}+bp_i+cp_{i+1} (i=1,2,\cdots,7)$ ，其中 $a=P(X=-1)$ ， $b=P(X=0)$ ， $c=P(X=1)$ 。假设 $\alpha=0.5$ ， $\beta=0.8$ 。

(i) 证明： $\{p_{i+1}-p_i\} (i=0,1,2,\cdots,7)$ 为等比数列；

(ii) 求 p_4 ，并根据 p_4 的值解释这种试验方案的合理性。

【答案】(1) 见解析；(2) (i) 见解析；(ii) $p_4 = \frac{1}{257}$ 。

【解析】

【分析】

(1) 首先确定 X 所有可能的取值，再来计算出每个取值对应的概率，从而可得分布列；(2)

(i) 求解出 a, b, c 的取值，可得 $p_i = 0.4p_{i-1} + 0.5p_i + 0.1p_{i+1} (i=1,2,\cdots,7)$ ，从而整理出符合等比数列定义的形式，问题得证；(ii) 列出证得的等比数列的通项公式，采用累加的方式，结合 p_8 和 p_0 的值可求得 p_1 ；再次利用累加法可求出 p_4 。

【详解】

(1) 由题意可知 X 所有可能的取值为：-1，0，1

$$\therefore P(X=-1)=(1-\alpha)\beta; P(X=0)=\alpha\beta+(1-\alpha)(1-\beta); P(X=1)=\alpha(1-\beta)$$

则 X 的分布列如下：

X	-1	0	1
P	$(1-\alpha)\beta$	$\alpha\beta+(1-\alpha)(1-\beta)$	$\alpha(1-\beta)$

(2) $\because \alpha=0.5, \beta=0.8$

$$\therefore a=0.5\times 0.8=0.4, b=0.5\times 0.8+0.5\times 0.2=0.5, c=0.5\times 0.2=0.1$$

(i) $\because p_i=ap_{i-1}+bp_i+cp_{i+1} (i=1,2,\cdots,7)$

$$\text{即 } p_i=0.4p_{i-1}+0.5p_i+0.1p_{i+1} (i=1,2,\cdots,7)$$

整理可得： $5p_i=4p_{i-1}+p_{i+1} (i=1,2,\cdots,7) \therefore p_{i+1}-p_i=4(p_i-p_{i-1}) (i=1,2,\cdots,7)$

$\therefore \{p_{i+1}-p_i\} (i=0,1,2,\cdots,7)$ 是以 p_1-p_0 为首项，4 为公比的等比数列

(ii) 由 (i) 知: $p_{i+1} - p_i = (p_1 - p_0) \cdot 4^i = p_1 \cdot 4^i$

$$\therefore p_8 - p_7 = p_1 \cdot 4^7, \quad p_7 - p_6 = p_1 \cdot 4^6, \quad \dots, \quad p_1 - p_0 = p_1 \cdot 4^0$$

作和可得: $p_8 - p_0 = p_1 \cdot (4^0 + 4^1 + \dots + 4^7) = \frac{1-4^8}{1-4} p_1 = \frac{4^8-1}{3} p_1 = 1$

$$\therefore p_1 = \frac{3}{4^8-1}$$

$$\therefore p_4 = p_4 - p_0 = p_1 \cdot (4^0 + 4^1 + 4^2 + 4^3) = \frac{1-4^4}{1-4} p_1 = \frac{4^4-1}{3} \times \frac{3}{4^8-1} = \frac{1}{4^4+1} = \frac{1}{257}$$

p_4 表示最终认为甲药更有效的. 由计算结果可以看出, 在甲药治愈率为 0.5, 乙药治愈率为 0.

8 时, 认为甲药更有效的概率为 $p_4 = \frac{1}{257} \approx 0.0039$, 此时得出错误结论的概率非常小, 说明这种实验方案合理.

【点睛】

本题考查离散型随机变量分布列的求解、利用递推关系式证明等比数列、累加法求解数列通项公式和数列中的项的问题. 本题综合性较强, 要求学生能够熟练掌握数列通项求解、概率求解的相关知识, 对学生分析和解决问题能力要求较高.

11. 【2019 年新课标 2 卷理科】11 分制乒乓球比赛, 每赢一球得 1 分, 当某局打成 10:10 平后, 每球交换发球权, 先多得 2 分的一方获胜, 该局比赛结束. 甲、乙两位同学进行单打比赛, 假设甲发球时甲得分的概率为 0.5, 乙发球时甲得分的概率为 0.4, 各球的结果相互独立. 在某局双方 10:10 平后, 甲先发球, 两人又打了 X 个球该局比赛结束.

(1) 求 $P(X=2)$;

(2) 求事件“ $X=4$ 且甲获胜”的概率.

【答案】(1) 0.5; (2) 0.1

【解析】

【分析】

(1) 本题首先可以通过题意推导出 $P(X=2)$ 所包含的事件为“甲连赢两球或乙连赢两球”, 然后计算出每种事件的概率并求和即可得出结果;

(2) 本题首先可以通过题意推导出 $P(X=4)$ 所包含的事件为“前两球甲乙各得 1 分, 后两球均为甲得分”, 然后计算出每种事件的概率并求和即可得出结果.

【详解】

(1)由题意可知, $P(X=2)$ 所包含的事件为“甲连赢两球或乙连赢两球”

$$\text{所以 } P(X=2) = 0.5 \times 0.4 + 0.5 \times 0.6 = 0.5$$

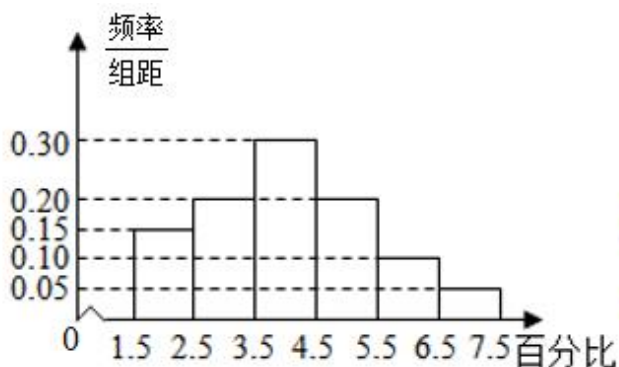
(2)由题意可知, $P(X=4)$ 包含的事件为“前两球甲乙各得1分, 后两球均为甲得分”

$$\text{所以 } P(X=4) = 0.5 \times 0.6 \times 0.5 \times 0.4 + 0.5 \times 0.4 \times 0.5 \times 0.4 = 0.1$$

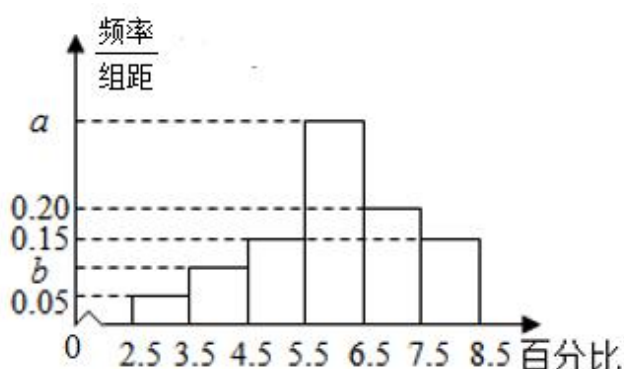
【点睛】

本题考查古典概型的相关性质, 能否通过题意得出 $P(X=2)$ 以及 $P(X=4)$ 所包含的事件是解决本题的关键, 考查推理能力, 考查学生从题目中获取所需信息的能力, 是中档题.

12. 【2019 年新课标 3 卷理科】为了解甲、乙两种离子在小鼠体内的残留程度, 进行如下试验: 将 200 只小鼠随机分成 A, B 两组, 每组 100 只, 其中 A 组小鼠给服甲离子溶液, B 组小鼠给服乙离子溶液. 每只小鼠给服的溶液体积相同、摩尔浓度相同. 经过一段时间后用某种科学方法测算出残留在小鼠体内离子的百分比. 根据试验数据分别得到如下直方图:



甲离子残留百分比直方图



乙离子残留百分比直方图

记 C 为事件: “乙离子残留在体内的百分比不低于 5.5”, 根据直方图得到 $P(C)$ 的估计值为 0.70.

(1) 求乙离子残留百分比直方图中 a, b 的值;

(2) 分别估计甲、乙离子残留百分比的平均值 (同一组中的数据用该组区间的中点值为代表).

【答案】(1) $a = 0.35$, $b = 0.10$; (2) 4.05, 6.

【解析】

【分析】

(1)由 $P(C) = 0.70$ 及频率和为 1 可解得 a 和 b 的值; (2)根据公式求平均数.

【详解】

(1)由题得 $a+0.20+0.15=0.70$ ，解得 $a=0.35$ ，由 $0.05+b+0.15=1-P(C)=1-0.70$ ，解得 $b=0.10$ 。

(2)由甲离子的直方图可得，甲离子残留百分比的平均值为

$$0.15 \times 2 + 0.20 \times 3 + 0.30 \times 4 + 0.20 \times 5 + 0.10 \times 6 + 0.05 \times 7 = 4.05,$$

$$\text{乙离子残留百分比的平均值为 } 0.05 \times 3 + 0.10 \times 4 + 0.15 \times 5 + 0.35 \times 6 + 0.20 \times 7 + 0.15 \times 8 = 6$$

【点睛】

本题考查频率分布直方图和平均数，属于基础题。

13. 【2018 年新课标 1 卷理科】某工厂的某种产品成箱包装，每箱 200 件，每一箱产品在交付用户之前要对产品作检验，如检验出不合格品，则更换为合格品。检验时，先从这箱产品中任取 20 件作检验，再根据检验结果决定是否对余下的所有产品作检验，设每件产品为不合格品的概率都为 $p(0 < p < 1)$ ，且各件产品是否为不合格品相互独立。

(1) 记 20 件产品中恰有 2 件不合格品的概率为 $f(p)$ ，求 $f(p)$ 的最大值点 p_0 ；

(2) 现对一箱产品检验了 20 件，结果恰有 2 件不合格品，以 (1) 中确定的 p_0 作为 p 的值。已知每件产品的检验费用为 2 元，若有不合格品进入用户手中，则工厂要对每件不合格品支付 25 元的赔偿费用。

(i) 若不对该箱余下的产品作检验，这一箱产品的检验费用与赔偿费用的和记为 X ，求 EX ；

(ii) 以检验费用与赔偿费用之和的期望值为决策依据，是否该对这箱余下的所有产品作检验？

【答案】(1) 0.1；(2) (i) 490；(ii) 应该对余下的产品作检验。

【解析】

【分析】

(1) 利用独立重复实验成功次数对应的概率，求得 $f(p) = C_{20}^2 p^2 (1-p)^{18}$ ，之后对其求导，利用导数在相应区间上的符号，确定其单调性，从而得到其最大值点，这里要注意 $0 < p < 1$ 的条件；

(2) 先根据第一问的条件，确定出 $p=0.1$ ，在解 (i) 的时候，先求件数对应的期望，之后应用变量之间的关系，求得赔偿费用的期望；在解 (ii) 的时候，就通过比较两个期望的大小，得到结果。

【详解】

(1) 20 件产品中恰有 2 件不合格品的概率为 $f(p) = C_{20}^2 p^2 (1-p)^{18}$.

因此 $f'(p) = C_{20}^2 [2p(1-p)^{18} - 18p^2(1-p)^{17}] = 2C_{20}^2 p(1-p)^{17} (1-10p)$.

令 $f'(p) = 0$, 得 $p = 0.1$. 当 $p \in (0, 0.1)$ 时, $f'(p) > 0$; 当 $p \in (0.1, 1)$ 时, $f'(p) < 0$.

所以 $f(p)$ 的最大值点为 $p_0 = 0.1$;

(2) 由 (1) 知, $p = 0.1$.

(i) 令 Y 表示余下的 180 件产品中的不合格品件数, 依题意知 $Y \sim B(180, 0.1)$,

$X = 20 \times 2 + 25Y$, 即 $X = 40 + 25Y$.

所以 $EX = E(40 + 25Y) = 40 + 25EY = 490$.

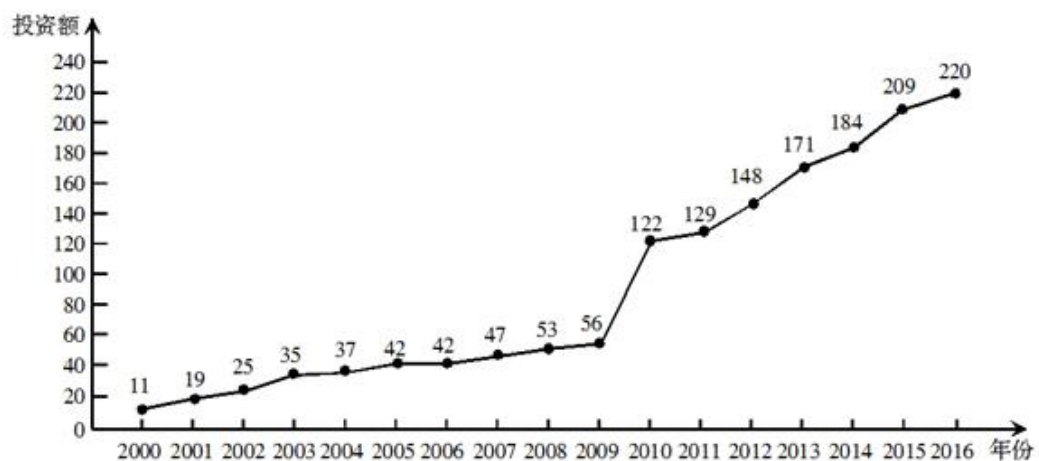
(ii) 如果对余下的产品作检验, 则这一箱产品所需要的检验费为 400 元.

由于 $EX > 400$, 故应该对余下的产品作检验.

【点睛】

该题考查的是有关随机变量的问题, 在解题的过程中, 一是需要明确独立重复试验成功次数对应的概率公式, 再者就是对其用函数的思想来研究, 应用导数求得其最小值点, 在做第二问的时候, 需要明确离散型随机变量的可取值以及对应的概率, 应用期望公式求得结果, 再有就是通过期望的大小关系得到结论.

14. 【2018 年新课标 2 卷理科】下图是某地区 2000 年至 2016 年环境基础设施投资额 y (单位: 亿元) 的折线图.



为了预测该地区 2018 年的环境基础设施投资额, 建立了 y 与时间变量 t 的两个线性回归模型. 根据 2000 年至 2016 年的数据 (时间变量 t 的值依次为 1, 2, ..., 17) 建立模型①:

$\hat{y} = -30.4 + 13.5t$; 根据 2010 年至 2016 年的数据 (时间变量 t 的值依次为 1, 2, ..., 7) 建立模

型②: $\hat{y}=99+17.5t$.

(1) 分别利用这两个模型, 求该地区 2018 年的环境基础设施投资额的预测值;

(2) 你认为用哪个模型得到的预测值更可靠? 并说明理由.

【答案】(1) 利用模型①预测值为 226.1, 利用模型②预测值为 256.5, (2) 利用模型②得到的预测值更可靠.

【解析】

【详解】

分析: (1) 两个回归直线方程中无参数, 所以分别求自变量为 2018 时所对应的函数值, 就得结果; (2) 根据折线图知 2000 到 2009, 与 2010 到 2016 是两个有明显区别的直线, 且 2010 到 2016 的增幅明显高于 2000 到 2009, 也高于模型 1 的增幅, 因此所以用模型 2 更能较好得到 2018 的预测.

详解: (1) 利用模型①, 该地区 2018 年的环境基础设施投资额的预测值为

$$\hat{y}=-30.4+13.5\times 19=226.1 \text{ (亿元)}.$$

利用模型②, 该地区 2018 年的环境基础设施投资额的预测值为

$$\hat{y}=99+17.5\times 9=256.5 \text{ (亿元)}.$$

(2) 利用模型②得到的预测值更可靠.

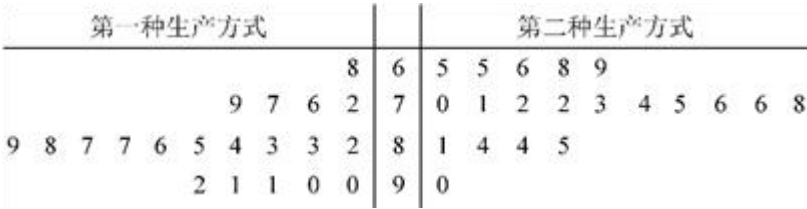
理由如下:

(i) 从折线图可以看出, 2000 年至 2016 年的数据对应的点没有随机散布在直线 $y=-30.4+13.5t$ 上下, 这说明利用 2000 年至 2016 年的数据建立的线性模型①不能很好地描述环境基础设施投资额的变化趋势. 2010 年相对 2009 年的环境基础设施投资额有明显增加, 2010 年至 2016 年的数据对应的点位于一条直线的附近, 这说明从 2010 年开始环境基础设施投资额的变化规律呈线性增长趋势, 利用 2010 年至 2016 年的数据建立的线性模型 $\hat{y}=99+17.5t$ 可以较好地描述 2010 年以后的环境基础设施投资额的变化趋势, 因此利用模型②得到的预测值更可靠.

(ii) 从计算结果看, 相对于 2016 年的环境基础设施投资额 220 亿元, 由模型①得到的预测值 226.1 亿元的增幅明显偏低, 而利用模型②得到的预测值的增幅比较合理, 说明利用模型②得到的预测值更可靠.

点睛: 若已知回归直线方程, 则可以直接将数值代入求得特定要求下的预测值; 若回归直线方程有待定参数, 则根据回归直线方程恒过点 $(\bar{x}, \bar{y})$ 求参数.

15. 【2018 年新课标 3 卷理科】某工厂为提高生产效率，开展技术创新活动，提出了完成某项生产任务的两种新的生产方式. 为比较两种生产方式的效率，选取 40 名工人，将他们随机分成两组，每组 20 人，第一组工人用第一种生产方式，第二组工人用第二种生产方式. 根据工人完成生产任务的工作时间（单位：min）绘制了如下茎叶图：



- (1) 根据茎叶图判断哪种生产方式的效率更高？并说明理由；
- (2) 求 40 名工人完成生产任务所需时间的中位数 m ，并将完成生产任务所需时间超过 m 和不超过 m 的工人数填入下面的列联表：

	超过 m	不超过 m
第一种生产方式		
第二种生产方式		

- (3) 根据 (2) 中的列联表，能否有 99% 的把握认为两种生产方式的效率有差异？

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

$P(K^2 \geq k)$	0.050	0.010	0.001
k	3.841	6.635	10.828

【答案】(1) 第二种生产方式的效率更高. 理由见解析

(2) 80

(3) 能

【解析】

【详解】

- 分析：(1) 计算两种生产方式的平均时间即可.
- (2) 计算出中位数，再由茎叶图数据完成列联表.
- (3) 由公式计算出 k^2 ，再与 6.635 比较可得结果.

详解：(1) 第二种生产方式的效率更高.

理由如下：

(i) 由茎叶图可知：用第一种生产方式的工人中，有 75% 的工人完成生产任务所需时间至少 80 分钟，用第二种生产方式的工人中，有 75% 的工人完成生产任务所需时间至多 79 分钟.因此第二种生产方式的效率更高.

(ii) 由茎叶图可知：用第一种生产方式的工人完成生产任务所需时间的中位数为 85.5 分钟，用第二种生产方式的工人完成生产任务所需时间的中位数为 73.5 分钟.因此第二种生产方式的效率更高.

(iii) 由茎叶图可知：用第一种生产方式的工人完成生产任务平均所需时间高于 80 分钟；用第二种生产方式的工人完成生产任务平均所需时间低于 80 分钟，因此第二种生产方式的效率更高.

(iv) 由茎叶图可知：用第一种生产方式的工人完成生产任务所需时间分布在茎 8 上的最多，关于茎 8 大致呈对称分布；用第二种生产方式的工人完成生产任务所需时间分布在茎 7 上的最多，关于茎 7 大致呈对称分布，又用两种生产方式的工人完成生产任务所需时间分布的区间相同，故可以认为用第二种生产方式完成生产任务所需的时间比用第一种生产方式完成生产任务所需的时间更少，因此第二种生产方式的效率更高.

以上给出了 4 种理由，考生答出其中任意一种或其他合理理由均可得分.

(2) 由茎叶图知 $m = \frac{79+81}{2} = 80$.

列联表如下：

	超过 m	不超过 m
第一种生产方式	15	5
第二种生产方式	5	15

(3) 由于 $K^2 = \frac{40(15 \times 15 - 5 \times 5)^2}{20 \times 20 \times 20 \times 20} = 10 > 6.635$ ，所以有 99% 的把握认为两种生产方式的效率

有差异.

点睛：本题主要考查了茎叶图和独立性检验，考察学生的计算能力和分析问题的能力，贴近生活.

