

二轮复习之概率与统计

◎ 福建厦门外国语学校 章少川

二轮复习

本专题是大学统计学的基础,起着承上启下的作用,也是每年高考命题的热点.在近年高考试题中,概率统计题通常是通过对课本原题的改编,通过对基础知识的重新组合、改编和拓展,从而加工为立意高、情境新、设问巧的问题.

随机事件的概率与古典概型

考情分析 古典概型是一种最基本的概率模型,在概率论中占有相当重要的地位,随机事件的概率多以古典概型出题.本部分试题通常以实际生活中的实例为背景,如以摸球、取数、掷骰子、分配等为模型来命制选择题或填空题.

破解技巧 解决有关古典概型问题的一般思路.

(1)要判断概率模型是否具备两大特点:①试验中所有可能出现的基本事件只有有限个;②每个基本事件出现的可能性相等.

(2)要准确理解基本事件的构成,正确求出基本事件总数和所求事件包含的基本事件数;一般用列举法,更多时候要用到计数原理与排列组合的相关知识.

(3)对于某些稍复杂的事件,通常有两种处理的方法:一是将所求事件的概率化成一些彼此互斥的事件的概率的和;二是转化为对立事件的概率求解.求解时一定要注意正确分类,并做到不重不漏.

例1 同时抛掷两枚骰子.

(I)求“点数之和为6”的概率;

(II)求“至少有一个5点或6点”的概率.

破解思路 本题是一道经典的古典概型问题,用列举法可求其概率.但第1问容易错解为:掷两枚骰子出现的点数之和的可能数值为 $\{2, 3, 4, \dots, 12\}$,“点数之和为6”的结果有 $1+5, 2+4, 3+3$ 三种情况,故 $P=\frac{3}{11}$.

实际上出现数值2和3不是等可能的,2只有 $(1, 1)$ 这一种情况,而3有 $(1, 2), (2, 1)$ 这两种情况,这种错误是对“等可能性”理解不足造成的.第2问是含“至多”“至少”型题目,用间接法,即先求“至少有一个5点或6点”的对立事件,再用公式计算.

完美答案 (I)点数之和为6的共有5个结果,所以点数之和为6的概率 $P=\frac{5}{36}$.

(II)至少有一个5点或6点的对立事件是既没有5点又没有6点,其结果共有16个,其概率 $P=\frac{16}{36}=\frac{4}{9}$,所以至少有一个5点或6点的概率为 $1-\frac{4}{9}=\frac{5}{9}$.

例2 盒中装着标有数字1, 2, 3, 4的卡片各2张,从盒中任意任取3张,每张卡片被抽出的可能性都

相等,求:

(I) 抽出的3张卡片上最大的数字是4的概率;

(II) 抽出的3张卡片上的数字互不相同的概率.

破解思路 利用古典概型的计算公式时要注意两点,(1)所有的基本事件必须是互斥的;(2)求基本事件总数和所求事件包含的基本事件数时,

要做到不重不漏.本例第1问可分为“3张卡片中只有一张为4”和“3张卡片中有两张为4”两类彼此互斥的事件,运用概率的加法公式求解.采用直接分类求第2问可能会造成重漏事件,因此考虑用间接法求解更好.

完美答案 (I)“抽出的3张卡片上最大的数字是4”的事件记为A,

$$\text{由题意得: } P(A) = \frac{C_2^1 C_6^2 + C_2^2 C_6^1}{C_8^3} = \frac{9}{14};$$

(II)“抽出的3张卡片上的数字互不相同”的事件记为C,“抽出的3张卡片中有两张数字相同”的事件记为D,由题意,C与D是对立事件,因为

$$P(D) = \frac{C_4^1 C_6^2}{C_8^3} = \frac{3}{7}, \text{ 所以 } P(C) = 1 - P(D) = \frac{4}{7}.$$

几何概型及随机模拟

考情分析 几何概型为新增加的概率内容,体现了数形结合的数学思想,是概率问题与几何问题的一种完美结合.近年的高考试题中,几何概型多与函数、方程、不等式等联系,常出现在客观题中.

破解技巧 当试验的结果构成的区域为长度、面积、体积、弧长、夹角等时,应考虑使用几何概型求解;利用几何概型求概率时,关键是寻找试验的全部结果构成的区域和事件发生的区域,有时还需要设出变量,在坐标系中表示所需的区域.

例3 假设王先生订了一份早报,送报人可能在早上6:30至7:30之间把报纸送到王先生家,王先生离开家去工作的时间在早上7:00至8:00之间,王先生在离开家前能得到报纸(称为事件A)的概率是多少?

破解思路 解答本题的关键是理解题意,由此将其归结为面积型几何概型,而不是长度型几何概型;认真审题,根据题意画出图形.将报纸送到家的时间与王先生离开家的时间作为坐标,并分别表示出来,由此

可直观地发现他们之间的联系,找到共同的区域,然后计算面积,代入公式求得结果.

完美答案 设“王先生离开家前能得到报纸”为事件A.在平面直角坐标系内,以x和y分别表示报纸送到和王先生离开家的时间,则王先生能得到报纸的充要条件是 $x \leq y$,而 (x, y) 的所有可能结果是边长为1的正方形,而能得到报纸的所有可能结果由图1中的阴影部分表示. $\mu_A = 1 - \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{7}{8}$, $\mu_\Omega = 1$,所以 $P(A) = \frac{\mu_A}{\mu_\Omega} = \frac{7}{8}$.

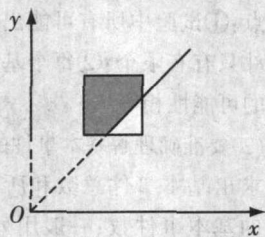


图1

例4 设关于x的一元二次方程 $x^2 + 2ax + b = 0$.

(I)若a是从0,1,2,3四个数中任取的一个数,b是从0,1,2三个数中任取的一个数,求上述方程有实根的概率;

概率;

(II)若a是从区间[0,3]任取的一个数,b是从区间[0,2]任取的一个数,求上述方程有实根的概率.

破解思路 古典概型与几何概型的区别在于所有可能出现的基本事件是有限个,还是无限个.本题第1问基本事件数为有限个,属于古典概型问题;第2问中a,b两个数都在连续的区间内取,基本事件数为无限个,属于几何概型问题.

完美答案 设事件A为“方程 $x^2 + 2ax + b = 0$ 有实根”.当 $a \geq 0, b \geq 0$ 时,方程 $x^2 + 2ax + b = 0$ 有实根的充要条件为 $a^2 \geq b$.

(I)事件A中包含9个基本事件,事件A发生的概率为 $P(A) = \frac{9}{12} = \frac{3}{4}$.

(II)试验的全部结果所构成的区域为 $\{(a, b) | 0 \leq a \leq 3, 0 \leq b \leq 2\}$,构成事件A的区域为 $\{(a, b) | 0 \leq a \leq 3, 0 \leq b \leq 2, a^2 \geq b\}$,而这个区域的面积为 $\int_0^{\sqrt{2}} a^2 da + 2 \times (3 - \sqrt{2}) = 6 - \frac{4\sqrt{2}}{3}$,所

以所求的概率为 $P(A) = 1 - \frac{2\sqrt{2}}{9}$.

离散型随机变量及其分布列、期望和方差

考情分析 离散型随机变量的分布列、期望与方差是高考的热点、

重点,题型以解答题为主,难度适中.它大多以实际问题为背景,涉及排

列、组合等知识,求互斥事件、相互独立事件的概率、条件概率等.

破解技巧 (1)一般方法:①理解

解 ξ 的意义,写出 ξ 的全部值;②求 ξ 取每个值的概率;③写出 ξ 的分布列;④由期望的定义求出 $E\xi$;⑤由方差的定义求 $D\xi$. (2)注意:真正弄清每一个随机变量 ξ 所对应的具体随机试验的结果;在具体解题中要充分利用期望和方差的性质解题;当断定随机变量服从两点分布或二项分布时,可不用列出分布列,直接用公式求出 $E\xi$ 与 $D\xi$.

例5 某班从6名班干部(其中男生4人,女生2人)中选3人参加学校学生会的干部竞选.

(I)设所选3人中女生人数为 ξ ,求 ξ 的分布列及数学期望;

(II)在男生甲被选中的情况下,求女生乙也被选中的概率.

破解思路 第1问的随机数分类较好把握;第2问考查条件概率问题,是新增加内容,对概念的理解要求较高,应引起足够的重视,可根据提示语“在……条件下”进行判断.

完美答案 (I) ξ 的所有可能取

值为0,1,2.依题意,得 $P(\xi=0)=\frac{C_4^3}{C_6^3}=\frac{1}{5}$, $P(\xi=1)=\frac{C_4^2 C_2^1}{C_6^3}=\frac{3}{5}$, $P(\xi=2)=\frac{C_4^1 C_2^2}{C_6^3}=\frac{1}{5}$.所以 ξ 的分布列为

ξ	0	1	2
P	$\frac{1}{5}$	$\frac{3}{5}$	$\frac{1}{5}$

$$E\xi=0 \times \frac{1}{5} + 1 \times \frac{3}{5} + 2 \times \frac{1}{5} = 1.$$

(II)设“男生甲被选中”为事件A,“女生乙被选中”为事件B,则 $P(A)=\frac{C_5^2}{C_6^3}=\frac{1}{2}$, $P(AB)=\frac{C_4^1}{C_6^3}=\frac{1}{5}$,所以 $P(B|A)=\frac{P(AB)}{P(A)}=\frac{2}{5}$.在男生甲被选中的情况下,女生乙也被选中的概率为 $\frac{2}{5}$.

例6 甲、乙两位学生参加数学竞赛培训.现分别从他们在培训期间参加的若干次预赛成绩中随机抽取8次,记录如下:

甲 82 81 79 78 95 88 93 84

乙 92 95 80 75 83 80 90 85

(I)用茎叶图表示这两组数据.

(II)现要从中选派一人参加数学竞赛,从统计学的角度考虑,你认为选派哪位学生参加合适?请说明理由.

(III)若将频率视为概率,对甲同学在今后的3次数学竞赛成绩进行预测,记这3次成绩中高于80分的次数为 ξ ,求 ξ 的分布列及数学期望 $E\xi$.

破解思路 本题是一道概率与统计结合的经典好题,涉及的知识点很多,需要灵活运用各种知识分析问题.第1小题,考查用茎叶图处理数据;对于第2小题,“从统计学的角度考虑”有多种分析法,答案不唯一;对于第3小题,求分布列,首先要确定随机变量的取值,其次求其取某个值的概率,本题可通过 n 次独立重复试验有 k 次发生的概率求解.

完美答案: (I)

甲				乙			
9	8	7	5				
8	4	2	1	8	0	0	3
5	3	9		0	2	5	

(II)派甲参赛比较合适.理由如下

$$\bar{x}_甲=85, \bar{x}_乙=85, s_甲^2=\frac{1}{8}[(78-85)^2+(79-85)^2+(81-85)^2+(82-85)^2+(84-85)^2+(88-85)^2+(93-85)^2+(95-85)^2]=35.5,$$

$$s_乙^2=\frac{1}{8}[(75-85)^2+(80-85)^2+(80-85)^2+(83-85)^2+(85-85)^2+(90-85)^2+(92-85)^2+(95-85)^2]=41.$$

$$\text{因为 } \bar{x}_甲=\bar{x}_乙, s_甲^2 < s_乙^2, \text{ 所以甲的成绩较稳定, 派甲参赛比较合适.}$$

派乙参赛比较合适.理由如下,从统计的角度看,甲获得85分以上(含85分)的概率 $P_1=\frac{3}{8}$,乙获得85分以上(含85分)的概率 $P_2=\frac{4}{8}=\frac{1}{2}$.因为 $P_2 > P_1$,所以派乙参赛比较合适.

(III)记“甲同学在一次数学竞赛中成绩高于80分”为事件A,则 $P(A)=\frac{6}{8}=\frac{3}{4}$.随机变量 ξ 的可能取值为0,1,2,3,且 $\xi \in \left(3, \frac{3}{4}\right)$,所以 $P(\xi=k)=C_3^k \left(\frac{3}{4}\right)^k \left(\frac{1}{4}\right)^{3-k}, k=0,1,2,3$.所以变量 ξ 的分布列为

以上(含85分)的概率 $P_2=\frac{4}{8}=\frac{1}{2}$.因为 $P_2 > P_1$,所以派乙参赛比较合适.

(III)记“甲同学在一次数学竞赛中成绩高于80分”为事件A,则 $P(A)=\frac{6}{8}=\frac{3}{4}$.随机变量 ξ 的可能取值为0,1,2,3,且 $\xi \in \left(3, \frac{3}{4}\right)$,所以 $P(\xi=k)=C_3^k \left(\frac{3}{4}\right)^k \left(\frac{1}{4}\right)^{3-k}, k=0,1,2,3$.所以变量 ξ 的分布列为

以上(含85分)的概率 $P_2=\frac{4}{8}=\frac{1}{2}$.因为 $P_2 > P_1$,所以派乙参赛比较合适.

(III)记“甲同学在一次数学竞赛中成绩高于80分”为事件A,则 $P(A)=\frac{6}{8}=\frac{3}{4}$.随机变量 ξ 的可能取值为0,1,2,3,且 $\xi \in \left(3, \frac{3}{4}\right)$,所以 $P(\xi=k)=C_3^k \left(\frac{3}{4}\right)^k \left(\frac{1}{4}\right)^{3-k}, k=0,1,2,3$.所以变量 ξ 的分布列为

以上(含85分)的概率 $P_2=\frac{4}{8}=\frac{1}{2}$.因为 $P_2 > P_1$,所以派乙参赛比较合适.

(III)记“甲同学在一次数学竞赛中成绩高于80分”为事件A,则 $P(A)=\frac{6}{8}=\frac{3}{4}$.随机变量 ξ 的可能取值为0,1,2,3,且 $\xi \in \left(3, \frac{3}{4}\right)$,所以 $P(\xi=k)=C_3^k \left(\frac{3}{4}\right)^k \left(\frac{1}{4}\right)^{3-k}, k=0,1,2,3$.所以变量 ξ 的分布列为

以上(含85分)的概率 $P_2=\frac{4}{8}=\frac{1}{2}$.因为 $P_2 > P_1$,所以派乙参赛比较合适.

(III)记“甲同学在一次数学竞赛中成绩高于80分”为事件A,则 $P(A)=\frac{6}{8}=\frac{3}{4}$.随机变量 ξ 的可能取值为0,1,2,3,且 $\xi \in \left(3, \frac{3}{4}\right)$,所以 $P(\xi=k)=C_3^k \left(\frac{3}{4}\right)^k \left(\frac{1}{4}\right)^{3-k}, k=0,1,2,3$.所以变量 ξ 的分布列为

以上(含85分)的概率 $P_2=\frac{4}{8}=\frac{1}{2}$.因为 $P_2 > P_1$,所以派乙参赛比较合适.

(III)记“甲同学在一次数学竞赛中成绩高于80分”为事件A,则 $P(A)=\frac{6}{8}=\frac{3}{4}$.随机变量 ξ 的可能取值为0,1,2,3,且 $\xi \in \left(3, \frac{3}{4}\right)$,所以 $P(\xi=k)=C_3^k \left(\frac{3}{4}\right)^k \left(\frac{1}{4}\right)^{3-k}, k=0,1,2,3$.所以变量 ξ 的分布列为

以上(含85分)的概率 $P_2=\frac{4}{8}=\frac{1}{2}$.因为 $P_2 > P_1$,所以派乙参赛比较合适.

(III)记“甲同学在一次数学竞赛中成绩高于80分”为事件A,则 $P(A)=\frac{6}{8}=\frac{3}{4}$.随机变量 ξ 的可能取值为0,1,2,3,且 $\xi \in \left(3, \frac{3}{4}\right)$,所以 $P(\xi=k)=C_3^k \left(\frac{3}{4}\right)^k \left(\frac{1}{4}\right)^{3-k}, k=0,1,2,3$.所以变量 ξ 的分布列为

以上(含85分)的概率 $P_2=\frac{4}{8}=\frac{1}{2}$.因为 $P_2 > P_1$,所以派乙参赛比较合适.

(III)记“甲同学在一次数学竞赛中成绩高于80分”为事件A,则 $P(A)=\frac{6}{8}=\frac{3}{4}$.随机变量 ξ 的可能取值为0,1,2,3,且 $\xi \in \left(3, \frac{3}{4}\right)$,所以 $P(\xi=k)=C_3^k \left(\frac{3}{4}\right)^k \left(\frac{1}{4}\right)^{3-k}, k=0,1,2,3$.所以变量 ξ 的分布列为

以上(含85分)的概率 $P_2=\frac{4}{8}=\frac{1}{2}$.因为 $P_2 > P_1$,所以派乙参赛比较合适.

(III)记“甲同学在一次数学竞赛中成绩高于80分”为事件A,则 $P(A)=\frac{6}{8}=\frac{3}{4}$.随机变量 ξ 的可能取值为0,1,2,3,且 $\xi \in \left(3, \frac{3}{4}\right)$,所以 $P(\xi=k)=C_3^k \left(\frac{3}{4}\right)^k \left(\frac{1}{4}\right)^{3-k}, k=0,1,2,3$.所以变量 ξ 的分布列为

以上(含85分)的概率 $P_2=\frac{4}{8}=\frac{1}{2}$.因为 $P_2 > P_1$,所以派乙参赛比较合适.

(III)记“甲同学在一次数学竞赛中成绩高于80分”为事件A,则 $P(A)=\frac{6}{8}=\frac{3}{4}$.随机变量 ξ 的可能取值为0,1,2,3,且 $\xi \in \left(3, \frac{3}{4}\right)$,所以 $P(\xi=k)=C_3^k \left(\frac{3}{4}\right)^k \left(\frac{1}{4}\right)^{3-k}, k=0,1,2,3$.所以变量 ξ 的分布列为

独立重复事件与二项分布、正态分布

考情分析 (1)二项分布及其应

用主要以条件概率、相互独立事件同时发生的概率、独立重复试验的概率为载体,综合考查某一事件发生的概率,进而通过计算期望与方差考查总体取值的平均水平和稳定性;(2)正态

分布主要考查正态分布的意义和性质,通过把一般正态总体转化为标准正态,常以客观题的形式出现.

破解技巧 (1)准确判断某随机

变量是否服从二项分布,要看两点:①在每次试验中,试验的结果只有两

个,即发生与不发生;②在每次试验中,事件发生的概率相同.若满足,则在此独立重复试验中以事件发生的次数为随机变量,此时该随机变量服从二项分布.

(2)理解正态分布曲线的意义及

性质是解答此类问题的关键:如正态分布密度函数 $f(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$,图象关于直线 $x=\mu$ 对称,均值为 μ ,方差为 σ^2 等.

例7 在一个圆锥体的培养房内培养了40只蜜蜂,准备进行某种实验,过圆锥高的中点有一个不计厚度且平行于圆锥底面的平面把培养房分成两个实验区,其中小锥体叫第一实验区,圆台体叫第二实验区,且两个实验区是互通的.假设蜜蜂落入培养房内任何位置是等可能的,且蜜蜂落入哪个位置相互之间是不受影响的.

(I)求蜜蜂落入第二实验区的概率;

(II)若其中有10只蜜蜂被染上了红色,求恰有一只红色蜜蜂落入第二实验区的概率;

(III)记 X 为落入第一实验区的蜜蜂数,求随机变量 X 的数学期望 EX .

破解思路 恰当地回归到相应的概率模型中去,是解答概率与统计应用问题的突破口.只有找到合适的概率模型,我们才能迅速抓住问题的本质,进而设计相应的解题策略.第1小题考查几何概型的“三维”测度问题;第2小题实际上可转化为独立重复事件的概率;对于第3小题,“落入第一实验区的蜜蜂数”服从二项分布,不必通过列随机变量分布图求数学期望,直接代公式即可.

完美答案 (I)记“蜜蜂落入第一实验区”为事件 A ,“蜜蜂落入第二实验区”为事件 B .依题意, $P(A)=$

$$\frac{V_{\text{小锥体}}}{V_{\text{圆锥体}}} = \frac{\frac{1}{3} \cdot \frac{1}{4} \cdot S_{\text{圆锥底面}} \cdot \frac{1}{2} h_{\text{圆锥}}}{\frac{1}{3} \cdot S_{\text{圆锥底面}} \cdot h_{\text{圆锥}}} = \frac{1}{8},$$

所以 $P(B)=1-P(A)=\frac{7}{8}$,所以蜜蜂落入第二实验区的概率为 $\frac{7}{8}$.

(II)记“恰有一只红色蜜蜂落入第二实验区”为事件 C ,则 $P(C)=C_{10}^1 \times \frac{7}{8} \times \left(\frac{1}{8}\right)^9 = \frac{70}{8^{10}} = \frac{70}{2^{30}}$,所以恰有一只红色蜜蜂落入第二实验区的概率 $\frac{70}{2^{30}}$.

(III)因为蜜蜂落入培养房内任何位置是等可能的,且蜜蜂落入哪个位置相互之间是不受影响的,所以变量 X 满足二项分布,即 $X \sim \left(40, \frac{1}{8}\right)$,
 $EX=40 \times \frac{1}{8}=5$

例8 在某校举行的数学竞赛中,全体参赛学生的竞赛成绩近似服从正态分布 $N(70,100)$.已知成绩表1

x_0	0	1	2	3	4	5	6	7	8	9
1.2	0.8849	0.8869	0.8880	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9278	0.9292	0.9306	0.9319
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9762	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857

在90分以上(含90分)的学生有12名.

(I)此次参赛学生总数约为多少人?

(II)若该校计划奖励竞赛成绩排在第50名的学生,设奖的分数线约为多少分?可查阅的(部分)标准正态分布表 $\Phi(x_0)=P(x < x_0)$ 见表1.

破解思路 本小题主要考查正态分布,考查运用概率统计知识解决实际问题的能力.

完美答案 (I)设参赛学生的分数为 ξ ,因为 $\xi \sim N(70,100)$,由条件知, $P(\xi \geq 90)=1-P(\xi < 90)=1-F(90)=1-\Phi\left(\frac{90-70}{10}\right)=1-\Phi(2)=1-0.9772=0.0228$.这说明成绩在90分以上(含90分)的学生人数约占全体参赛人数的2.28%,因此,参赛总人数约为 $\frac{12}{0.0228} \approx 526$ (人).

(II)假定设奖的分数线为 x 分,则 $P(\xi \geq x)=1-P(\xi < x)=1-F(x)=1-\Phi\left(\frac{x-70}{10}\right)=\frac{50}{526}=0.0951$,即 $\Phi\left(\frac{x-70}{10}\right)=0.9049$,查表得 $\frac{x-70}{10} \approx 1.31$,解得 $x=83.1$,故设奖的分数线约为83.1分.

抽样方法与总体分布的估计

考情分析 本部分内容是对初中数学统计初步的深化和扩展,以基础题(中、低档题)为主;热点问题是频率分布直方图和用样本的数值特征估计总体的数值特征.

破解技巧 (1)解决有关随机抽

样问题首先要深刻理解各种抽样方法的特点和实施步骤,其次要熟练掌握系统抽样中被抽个体号码的确定方法及分层抽样中各层人数的计算方法.

(2)注意频率分布直方图中小正

方形的面积=组距 $\times \frac{\text{频率}}{\text{组距}}=\text{频率}$,所有直方图的面积和应为1.

例9 随机抽取某中学甲、乙两班各10名同学,测量他们的身高(单位:cm),获得身高数据的茎叶图

如图2.

(I) 根据茎叶图判断哪个班的平均身高较高;

(II) 计算甲班的样本方差

(III) 现从乙班这10名同学中随机抽取两名身高不低于173 cm的同学,求身高为176 cm的同学被抽中的概率.

甲班					乙班				
		2	18	1					
9	9	1	0		17	0	3	6	8
8	8	3	2		16	2	5	8	
		8	15	9					

图2

破解思路 本题考查茎叶图的识图问题和平均数的计算方法,其中从茎叶图中读出数据是关键,为此,我们首先要弄清“茎”和“叶”分别代表什么;熟练掌握众数、中位数、平均数、方差、标准差的计算方法.

完美答案 (I) 乙班平均身高高.

(II) $\bar{x}=170$, 甲班的样本方差为57.2

(III) 设身高为176 cm的同学被抽中的事件为A,从乙班10名同学中抽中两名身高不低于173 cm的同学有:(181,173),(181,176),(181,178),

(181,179),(179,173),(179,176),(179,178),(178,173),(178,176),(176,173)共10个基本事件,而事件A含有4个基本事件, $P(A)=\frac{4}{10}=\frac{2}{5}$.

例10 某高校在2010年的

自主招生考试成绩中随机抽取100名学生的笔试成绩,按成绩分组,得到的频率分布表如表2所示.

表2

组号	分组	频数	频率
第1组	[160,165)	5	0.050
第2组	[165,170)	①	0.350
第3组	[170,175)	30	②
第4组	[175,180)	20	0.200
第5组	[180,185)	10	0.100
合计		100	1.00

(I) 请先求出频率分布表中①②位置相应的数据,再完成频率分布直方图;

(II) 为了能选拔出最优秀的学生,高校决定在笔试成绩高的第3、4、5组中用分层抽样抽取6名学生进入第二轮面试,求第3、4、5组每组各抽取多少名学生进入第二轮面试;

(III) 在(II)的前提下,学校决定在6名学生中随机抽取2名学生接受A

考官的面试,求第4组至少有一名学

生被A考官面试的概率.

破解思路 首先读懂表格的意义,利用概念求频数、频率等,再作出直方图;最后抽样分析,并得到结论.

完美答案 (I) 第2组的频数为 $0.35 \times 100 = 35$ 人,第3组的频率为 $\frac{30}{100} = 0.300$,频率分布直方图略.

(II) 因为第3、4、5组共有60名学生,所以第3组为 $\frac{30}{60} \times 6 = 3$ 人,第4组为 $\frac{20}{60} \times 6 = 2$ 人,第5组为 $\frac{10}{60} \times 6 = 1$ 人.

(III) 设第3组的3位同学为 A_1, A_2, A_3 ,第4组的2位同学为 B_1, B_2 ,第5组的1位同学为 C_1 ,则从6名同学中抽2名同学有15种可能,分别为 $(A_1, A_2), (A_1, A_3), (A_1, B_1), (A_1, B_2), (A_1, C_1), (A_2, A_3), (A_2, B_1), (A_2, B_2), (A_2, C_1), (A_3, B_1), (A_3, B_2), (A_3, C_1), (B_1, B_2), (B_1, C_1), (B_2, C_1)$.其中第4组至少有一位同学入选的有 $(A_1, B_1), (A_1, B_2), (A_2, B_1), (A_2, B_2), (A_3, B_1), (A_3, B_2), (B_1, B_2), (B_1, C_1), (B_2, C_1)$,共9种可能,其概率为 $\frac{9}{15} = \frac{3}{5}$.

回归分析与独立性检验

考情分析 本部分内容是新课标数学的新增内容,在目前高考中出现的次数较少,主要考查线性回归分析和独立性检验的统计方法.此部分的命题紧扣课本,通常考查单一知识点,大多是基础题.

破解技巧 一般情况下,在尚未断定两个变量之间是否具有线性相关关系的情况下,应先进行相关性检验;在确认其具有线性相关关系后,再求其回归直线方程;由部分数据得到的回归直线,可以对两个变量间的线性相关关系进行估计,其实质是

非确定性的相关关系问题转化成确定性的函数关系问题.

独立性检验的基本思想类似于反证法.要确认“两个分类变量有关系”这一结论成立的可信程度,首先假设该结论不成立,在该假设下构造的随机变量 K^2 应该很小,若由观测数据计算得到的 K^2 观测值 k 很大,则在一定程度上说明假设不合理.

例11 为了解某班学生喜爱

打篮球是否与性别有关,对本班50人进行了问卷调查得到了列联表.

	喜爱打篮球	不喜爱打篮球	合计
男生		5	
女生	10		
合计			50

已知在全部50人中随机抽取1人抽到喜爱打篮球的学生的概率为 $\frac{3}{5}$.

(I) 请将列联表补充完整.

(II) 是否有99.5%的把握认为喜爱打篮球与性别有关?说明你的理由.

(III) 已知喜爱打篮球的10位女生中, A_1, A_2, A_3, A_4, A_5 还喜欢打羽毛球, B_1, B_2, B_3 还喜欢打乒乓球, C_1, C_2

还喜欢踢足球. 现再从喜欢打羽毛球、打乒乓球、踢足球的女生中各选出1名进行其他方面的调查, 求 B_1 和 C_1 不全被选中的概率.

临界值表供参考

$P(K^2 \geq k)$	0.15	0.10	0.05	0.025	0.010	0.005	0.001
k	2.072	2.706	3.841	5.024	6.635	7.879	10.828

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

其中 $n=a+b+c+d$)

破解思路 根据“二联表”准确计算 K^2 , 并与临界表的数进行比较判断.

完美答案 (I) 列联表是

	喜爱打篮球	不喜爱打篮球	合计
男生	20	5	25
女生	10	15	25
合计	30	20	50

(II) 因为 $K^2 = \frac{50 \times (20 \times 15 - 10 \times 5)^2}{30 \times 20 \times 25 \times 25} \approx 8.333 > 7.879$, 所以有99.5%的把握认为喜爱打篮球与性别有关.

(III) 从10位女生中选出喜欢打羽毛球、打乒乓球、踢足球的各1名, 其一切可能的结果组成的基本事件如下: $(A_1, B_1, C_1), (A_1, B_1, C_2), (A_1, B_2, C_1), (A_1, B_2, C_2), (A_1, B_3, C_1), (A_1, B_3, C_2), (A_2, B_1, C_1), (A_2, B_1, C_2), (A_2, B_2, C_1), (A_2, B_2, C_2), (A_2, B_3, C_1), (A_2, B_3, C_2)$.

$(C_2), (A_3, B_1, C_1), (A_3, B_1, C_2), (A_3, B_2, C_1), (A_3, B_2, C_2), (A_3, B_3, C_1), (A_3, B_3, C_2), (A_4, B_1, C_1), (A_4, B_1, C_2), (A_4, B_2, C_1), (A_4, B_2, C_2), (A_4, B_3, C_1), (A_4, B_3, C_2), (A_5, B_1, C_1), (A_5, B_1, C_2), (A_5, B_2, C_1), (A_5, B_2, C_2), (A_5, B_3, C_1), (A_5, B_3, C_2)$, 基本事件的总数为30, 用 M 表示“ B_1, C_1 不全被选中”这一事件, 则其对立事件 $\bar{M}$ 表示“ B_1, C_1 全被选中”这一事件. 由于 $\bar{M}$ 由 $(A_1, B_1, C_1), (A_2, B_1, C_1), (A_3, B_1, C_1), (A_4, B_1, C_1), (A_5, B_1, C_1)$ 5个基本事件组成, 所以 $P(M) = 1 - P(\bar{M}) = 1 - \frac{1}{6} = \frac{5}{6}$.

综合测试

1. 从数字1, 2, 3, 4, 5中随机抽取3个数字(允许重复), 组成一个三位数, 其各位数字之和等于9的概率是 ()

- A. $\frac{1}{3}$ B. $\frac{16}{125}$ C. $\frac{18}{125}$ D. $\frac{19}{125}$

2. 三个元件 T_1, T_2, T_3 正常工作的概率分别为 $\frac{1}{2}, \frac{3}{4}, \frac{3}{4}$, 将它们中某两个元件并联后再和第三个元件串联接入电路, 在如图3的电路中, 电路不发生故障的概率是 ()

- A. $\frac{15}{32}$ B. $\frac{9}{32}$ C. $\frac{7}{32}$ D. $\frac{17}{32}$

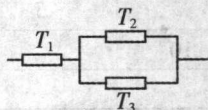


图3

3. 若在区间 $(-1, 1)$ 内任取一个实数 a , 在区间 $(0, 1)$ 内任取一个实数 b , 则直线 $ax - by = 0$ 与圆 $(x-1)^2 + (y-2)^2 = 1$ 相交的概率为 ()

- A. $\frac{3}{16}$ B. $\frac{5}{16}$ C. $\frac{3}{8}$ D. $\frac{5}{8}$

4. 口袋里放有大小相同的2个红球和1个白球, 有放回的每次摸取一

个球, 定义数列 $\{a_n\}: a_n = \begin{cases} -1, & \text{第}n\text{次摸取红球,} \\ 1, & \text{第}n\text{次摸取白球.} \end{cases}$ 如果 S_n 为数列 $\{a_n\}$ 的前 n 项之和, 那么 $S_7 = 3$ 的概率为 ()

- A. $\frac{224}{729}$ B. $\frac{28}{729}$ C. $\frac{35}{2387}$ D. $\frac{28}{75}$

5. 连续两次掷一颗质地均匀的骰子, 记出现向上的点数分别为 m, n . 设向量 $a = (m, n), b = (3, -3)$, 则 a 与 b 的夹角为锐角的概率是_____.

6. 在平面直角坐标系 xOy 中, 设不等式组 $\begin{cases} -1 \leq x \leq 2, \\ 0 \leq y \leq 2 \end{cases}$ 所表示的平面区域是 W , 从区域 W 中随机取点 $M(x, y)$.

(I) 若 $x, y \in \mathbf{Z}$, 求点 M 位于第一象限的概率;

(II) 若 $x, y \in \mathbf{R}$, 求 $|OM| \leq 2$ 的概率.

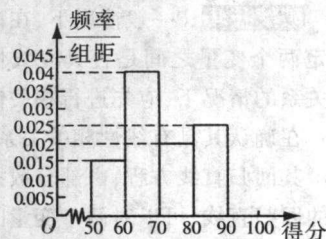
7. $f_1(x) = x, f_2(x) = x^2, f_3(x) = x^3, f_4(x) = \sin x, f_5(x) = \cos x, f_6(x) = \lg(|x| + 1)$. 分别写在六张卡片上, 放在一盒子里.

(I) 现从盒子中任取两张卡片, 将卡片上的函数相加得一个新函数, 求所得函数是奇函数的概率;

(II) 现从盒子中逐一抽取卡片, 且每次取出后均不放回, 若取到一张记有偶函数的卡片则停止抽取, 否则继续进行, 求抽取次数 ξ 的分布列和数学期望.

8. 为抗击金融风暴, 某系统决定对所属企业给予低息贷款扶持. 该系统制定了评分标准, 并根据标准对企业进行评估. 该系统依据评估得分将这些企业分别定为优秀、良好、合格、不合格四个等级, 并根据等级分配相应的低息贷款数额. 为了更好地掌握贷款总额, 该系统随机抽查了所属的部分企业. 以下图表给出了有关数据(将频率看做概率).

检查得分	[50, 60)	[60, 70)	[70, 80)	[80, 90]
评定类型	不合格	合格	良好	优秀
贷款金额(万元)	0	200	400	800



(I) 任抽一家所属企业, 求抽到

的企业等级是优秀或良好的概率.

(II)对照标准,部分企业进行了整改.整改后,优秀企业数量不变,不合格企业、合格企业、良好企业的数量成等差数列.要使所属企业获得贷款的平均值(即数学期望)不低于410万元,试求整改后不合格企业占企业总数百分比的最大值.

9. 第16届亚运会于2010年11月12日至27日在中国广州进行,为了搞

好接待工作,组委会招募了16名男志愿者和14名女志愿者,调查发现,男、女志愿者中分别有10人和6人喜爱运动,其余不喜爱.

(I)根据以上数据完成以下 2×2 列联表.

	喜爱运动	不喜爱运动	总计
男	10		16
女	6		14
总计			30

(II)根据列联表的独立性检验,能否在犯错误的概率不超过0.10的前提下认为性别与喜爱运动有关?

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

其中 $n=a+b+c+d$.

参考数据:

$P(K^2 \geq k_0)$	0.40	0.25	0.10	0.010
k_0	0.708	1.323	2.706	6.635

完美答案

1. $P(A) = \frac{6+6+3+3+1}{125} = \frac{19}{125}$, 选D

2. $P = \frac{1}{2} \cdot \left(1 - \frac{1}{4} \cdot \frac{1}{4}\right) = \frac{15}{32}$, 选A

3. 几何概型,由直线与圆相交得 $\frac{|a-2b|}{\sqrt{a^2+b^2}} \leq 1$, 即 $3b \leq 4a$, 又 $-1 < a < 1, 0 < b < 1$, 作出可行域, 求出面积得 $P = \frac{\frac{1}{2} \left(\frac{1}{4} + 1\right) \times 1}{2 \times 1} = \frac{5}{16}$, 选B

4. $S_7=3$ 意味着7次里有2次摸红球, 有5次摸白球, $C_7^2 \left(\frac{2}{3}\right)^2 \left(\frac{1}{3}\right)^5 = \frac{28}{729}$, 选B

5. $\frac{5}{12}$

6. (I)若 $x, y \in \mathbb{Z}$, 则点 M 的个数共有12个. 当点 M 的坐标为 $(1, 1), (1, 2), (2, 1), (2, 2)$ 时, 点 M 位于第一象限, 其概率为 $\frac{4}{12} = \frac{1}{3}$

(II)这是一个几何概型. 区域 W 的面积是 $3 \times 2 = 6$. 满足 $|OM| \leq 2$ 的点 M 构成的区域为 $\{(x, y) | -1 \leq x \leq 2, 0 \leq y \leq 2, x^2 + y^2 \leq 4\}$, 即图4中的阴影部分, 易知 $E(-1, \sqrt{3})$, $\angle EOA = 60^\circ$, 所以

扇形 BOE 的面积是 $\frac{4\pi}{3}$, $\triangle EOA$ 的面积是 $\frac{\sqrt{3}}{2}$, 所以 $|OM| \leq 2$ 的概率为 $\frac{8\pi + 3\sqrt{3}}{36}$

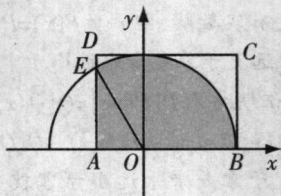


图4

7. (I) $P = \frac{C_3^2}{C_6^2} = \frac{1}{5}$

(II) ξ 可取1, 2, 3, 4. $P(\xi=1) = \frac{C_3^1}{C_6^1} =$

$\frac{1}{2}, P(\xi=2) = \frac{C_3^1}{C_6^1} \cdot \frac{C_3^1}{C_5^1} = \frac{3}{10}, P(\xi=3) = \frac{C_3^1}{C_6^1} \cdot$

$\frac{C_2^1}{C_5^1} \cdot \frac{C_3^1}{C_4^1} = \frac{3}{20}, P(\xi=4) = \frac{C_3^1}{C_6^1} \cdot \frac{C_2^1}{C_5^1} \cdot \frac{C_1^1}{C_4^1} = \frac{1}{20},$

故 ξ 的分布列为

ξ	1	2	3	4
P	$\frac{1}{2}$	$\frac{3}{10}$	$\frac{3}{20}$	$\frac{1}{20}$

$E\xi = \frac{7}{4}$

8. (I) 0.45

(II)设整改后,任意抽取一家企业,抽到不合格企业、合格企业、良好企业的概率分别为 a, b, c . 因为整改后,不合格企业、合格企业、良好企业的数量成等差数列,所以 a, b, c 也成等差数列,即 $2b = a + c$. 又 $a + b + c + 0.25 = 1$, 所以 $b = 0.25, a + c = 0.5$. 设整改后一家企业获得的低息贷款为 ξ 万元, 则 ξ 的分布列是

ξ	0	200	400	800
P	a	0.25	c	0.25

所以 $E\xi = 0 \times a + 200 \times 0.25 + 400 \times c + 800 \times 0.25 = 450 - 400a$. 由已知得 $E\xi \geq 410$, 所以 $450 - 400a \geq 410$, 解得 $a_{\max} = 10\%$

9. (I)

	喜爱运动	不喜爱运动	合计
男生	10	6	16
女生	6	8	14
合计	16	14	30

(II)假设是否喜爱运动与性别无关. 由已知数据可求得 $K^2 =$

$$\frac{30 \times (10 \times 8 - 6 \times 6)^2}{(10+6)(6+8)(10+6)(6+8)} \approx 1.1575 <$$

2.706, 因此,在犯错的概率不超过0.10的前提下不能判断喜爱运动与性别有关