

14. 【解】(1) 由题意得 $P(A) = \frac{81+9}{100} = \frac{9}{10}$, $P(B) = \frac{9}{100}$, $P(AB) = \frac{9}{100}$, $P(\overline{AB}) = \frac{81}{100}$,

$$\text{所以 } P(B|A) = \frac{P(AB)}{P(A)} = \frac{\frac{9}{100}}{\frac{9}{10}} = \frac{1}{10}, P(\overline{B}|A) = \frac{P(\overline{AB})}{P(A)} = \frac{\frac{81}{100}}{\frac{9}{10}} = \frac{9}{10},$$

$$\frac{\frac{81}{100}}{\frac{9}{10}} = \frac{9}{10},$$

$$\text{所以 } \frac{P(B|A)}{P(\overline{B}|A)} = \frac{\frac{1}{10}}{\frac{9}{10}} = \frac{1}{9}, \text{ 所以事件 } A \text{ 发生的条件下事件 } B$$

发生的似然比为 $\frac{1}{9}$.

(2) ①已知 $X \sim N(\mu, \sigma^2)$, 且 X 落在 $(0, 40)$ 和落在 $(60, 100)$ 内的概率相等, 根据正态曲线的对称性, 可得 $\mu = \frac{40+60}{2} = 50$.

②因为 $P(X \leq 10) = P(X \geq 90) = \frac{1}{10}$, 所以从一线工作者中

抽 1 人为轻压力工作者的概率为 $P(10 < X < 90) = \frac{8}{10} = \frac{4}{5}$, 所以从该区域该行业一线工作人员中随机地抽取 3 名, 设这 3 名工作人员中轻压力工作者人数为 Y , 则 $Y \sim B\left(3, \frac{4}{5}\right)$, 即 Y 的所有可能取值为 0, 1, 2, 3,

$$\text{则 } P(Y=0) = C_3^0 \times \left(\frac{1}{5}\right)^3 = \frac{1}{125}, P(Y=1) = C_3^1 \times \frac{4}{5} \times \left(\frac{1}{5}\right)^2 = \frac{12}{125},$$

$$P(Y=2) = C_3^2 \times \left(\frac{4}{5}\right)^2 \times \frac{1}{5} = \frac{48}{125}, P(Y=3) = C_3^3 \times \left(\frac{4}{5}\right)^3 = \frac{64}{125}.$$

所以 Y 的分布列为

Y	0	1	2	3
P	$\frac{1}{125}$	$\frac{12}{125}$	$\frac{48}{125}$	$\frac{64}{125}$

$$E(Y) = 3 \times \frac{4}{5} = \frac{12}{5}.$$

4.3 统计模型

4.3.1 一元线性回归模型

刷基础

1. C 【解析】考试号只是确定考生的位置, 与成绩无关, 则①错误; 勤能补拙具有相关关系, 水稻产量与气候具有相关关系,

②悟: 两个事物的关系可以是确定的函数关系, 也可以是不确定的相关关系. 相关关系描述的是因果关系或者是伴随关系

则②③正确; 正方形的边长与正方形的面积是函数关系, 则④错误. 故选 C.

规律方法 两个变量是否相关的两种判断方法

(1) 根据实际经验: 借助积累的经验进行分析判断.

(2) 利用散点图: 通过散点图, 观察它们的分布是否存在一定的规律, 直观地进行判断. 如果发现点的分布从整体上看大致在一条直线或曲线附近, 那么这两个变量就是相关的, 注意不要受个别点的位置的影响.

2. B 【解析】相关关系是一种非确定性关系.

对于 A, C, 两个变量具有函数关系, 是一种确定性关系, 故 A, C 错误;

对于 D, 图中的散点分布没有规律, 故两个变量之间不具有相关关系, 故 D 错误;

对于 B, 图中的散点分布在从左下到右上的一条直线附近, 两个变量具有相关关系, 故 B 正确. 故选 B.

名师点拨 散点图中的点从整体上看大致在一条直线附近

(并不是严格在一条直线上), 则称两变量之间具有线性相关关系. 若两变量之间具有线性相关关系且散点图在从左下角到右上角的区域内, 则两变量之间是正相关关系; 若两变量之间具有线性相关关系且散点图在从左上角到右下角的区域内, 则两变量之间是负相关关系.

3. A 【解析】由题意知 $\bar{x} = \frac{3+4+5+6+7}{5} = 5$, $\bar{y} =$

$$\frac{2.5+3+4+4.5+6}{5} = 4,$$

则将点 $(5, 4)$ 的坐标代入 $\hat{y} = \hat{b}x - 0.25$, 解得 $\hat{b} = 0.85$.

所以当 $x = 10$ 时, $\hat{y} = 0.85 \times 10 - 0.25 = 8.25$,

故选 A.

4. C 【解析】因为 x 与 y 正相关, 所以 A, D 不正确;

对于 B, 因为回归直线过点 (\bar{x}, \bar{y}) , 当 $x = 3$ 时, $\hat{y} = 2 \times 3 - 2.4 = 3.6$, 故 B 不正确;

对于 C, 当 $x = 3$ 时, $\hat{y} = 0.4 \times 3 + 2.3 = 3.5$, 故 C 正确. 故选 C.

5. B 【解析】由题意可知, $\bar{x} = \frac{1+2+3+4+5}{5} = 3$,

因为回归直线 $\hat{y} = -0.12x + 2.2$ 过样本点中心 (\bar{x}, \bar{y}) ,

所以 $\bar{y} = -0.12\bar{x} + 2.2 = -0.12 \times 3 + 2.2 = 1.84$,

所以 $\frac{1.7+2.4+2.0+1.6+t}{5} = 1.84$, 解得 $t = 1.5$. 故选 B.

②悟: 根据回归直线过样本点中心 (\bar{x}, \bar{y}) 求缺失数据

6. 【解】(1) 由题知 $\bar{x} = \frac{1}{5} \times (2+3+4+5+6) = 4$, $\bar{y} = \frac{1}{5} \times (1+1.1+1.5+1.8+2.1) = 1.5$,

$$\text{则 } \hat{b} = \frac{(-2) \times (-0.5) + (-1) \times (-0.4) + 1 \times 0.3 + 2 \times 0.6}{(-2)^2 + (-1)^2 + 1^2 + 2^2} = 0.29,$$

$$\text{则 } \hat{a} = 1.5 - 0.29 \times 4 = 0.34,$$

故 y 关于 x 的回归直线方程为 $\hat{y} = 0.29x + 0.34$.

(2) 由 (1) 知 y 关于 x 的回归直线方程为 $\hat{y} = 0.29x + 0.34$, 将 $x = 10$ 代入 $\hat{y} = 0.29x + 0.34$, 得 $\hat{y} = 0.29 \times 10 + 0.34 = 3.24$, 故当年的日营销费用为 1 000 元时, 日销售量约为 3.24 百件.

规律方法 求回归直线方程的一般步骤

- (1) 收集样本数据, 设为 $(x_i, y_i) (i=1, 2, \dots, n)$.
- (2) 作出散点图, 确定 x, y 具有线性相关关系.
- (3) 计算 $\bar{x}, \bar{y}, \sum_{i=1}^n x_i^2, \sum_{i=1}^n x_i y_i$.

$$(4) \text{ 代入公式计算 } \hat{b}, \hat{a}, \text{ 公式为 } \begin{cases} \hat{b} = \frac{\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y}}{\sum_{i=1}^n x_i^2 - n \bar{x}^2}, \\ \hat{a} = \bar{y} - \hat{b} \bar{x}. \end{cases}$$

- (5) 写出回归直线方程 $\hat{y} = \hat{b}x + \hat{a}$.

7. A 【解析】因为 y 与 x 的回归直线方程为 $\hat{y} = 6.5x + 17.5$,

所以当 $x=5$ 时, $\hat{y} = 6.5 \times 5 + 17.5 = 50$.

由表格知当广告支出为 5 万元时, 销售额为 60 万元, 所以误差为 $60 - 50 = 10$ 万元. 故选 A.

8. ABD 【解析】对于选项 A, 相关系数 r 的计算公式为 $r =$

$$\frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}, \text{ 回归系数 } \hat{b} \text{ 的计算公式为 } \hat{b} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2},$$

在这两个公式中, 分子均为 $\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$, 分母均为正数, 所以 r 与 \hat{b} 的符号由 $\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$ 决定, 二者符号相同, 选项 A 正确.

对于选项 B, 对于回归直线方程 $\hat{y} = \hat{b}x + \hat{a}$, 回归直线一定经过样本点中心 (\bar{x}, \bar{y}) , 选项 B 正确.

对于选项 C, 回归直线方程中的 \hat{b} 表示回归直线的斜率, 它反映的是一个变量 x 每变化一个单位时, 另一个变量 y 的平均变化量; 而变量的相关性强弱是由相关系数 r 来衡量的, $|r|$ 越接近 1, 两个变量的相关性越强, 因此, \hat{b} 的大小与两个变量的相关性强弱无关, 选项 C 错误.

对于选项 D, 相关系数 r 的绝对值 $|r|$ 越接近 1, 表明两个变量的线性相关性越强, 当 $|r| = 0.95$ 时, $|r|$ 非常接近 1, 说明两个变量的线性相关性足够强, 此时适合用回归直线方程进行预测, 选项 D 正确. 故选 ABD.

9. 【解】(1) 由题意得, $\bar{x} = \frac{1}{5} \times (40 + 45 + 50 + 55 + 60) = 50, \bar{y} = \frac{1}{5} \times (50 + 56 + 64 + 72 + 83) = 65,$

$$\text{所以 } y \text{ 关于 } x \text{ 的相关系数 } r = \frac{16\,660 - 5 \times 50 \times 65}{\sqrt{250 \times 680}} = \frac{410}{412} \approx 0.995,$$

所以 y 与 x 具有较强的线性相关关系.

$$(2) \text{ 由 (1) 得 } \hat{b} = \frac{410}{250} = \frac{41}{25}, \text{ 则 } \hat{a} = 65 - \frac{41}{25} \times 50 = -17,$$

所以 y 关于 x 的回归直线方程为 $\hat{y} = 1.64x - 17$,

当 $x=70$ 时, $\hat{y} = 1.64 \times 70 - 17 = 97.8$,

所以当土壤的湿度为 70% 时, 种子的发芽率 $y\%$ 的预测值为 97.8%.

10. A 【解析】 $y = c_1 e^{e_2 x} \Rightarrow \ln y = z = e_2 x + \ln c_1$, 由题可得 $e_2 = 0.2$,

$$\text{由题可得 } \bar{x} = \frac{20+23+25+27+30}{5} = 25, \bar{z} = \frac{2+2.4+3+3+4.6}{5} =$$

3, 因为回归直线过样本点中心 (\bar{x}, \bar{z}) ,

所以 $3 = 0.2 \times 25 + \ln c_1 \Rightarrow \ln c_1 = -2 \Rightarrow c_1 = e^{-2}$. 故选 A.

11. C 【解析】对于 A, 根据题中散点图知, 7:00~7:30 内, 每分钟的进校人数 y 与相应时间 x 呈正相关, 故 A 正确;

对于 B, 由题图知, 曲线 $\hat{y} = 0.82e^{0.16x}$ 的拟合效果更好, 故乙同学的回归方程拟合效果更好, 故 B 正确;

对于 C, 表格中并未给出 7:09~7:10 这一分钟内的进校人数对应的值, 而由甲的回归直线方程得到的只能是估计值, 不一定是实际值, 故 C 错误;

对于 D, 全校学生近 600 人, 从表格中的数据知, 7:26~7:30 进校的人数超过 300, 故 D 正确, 故选 C.

12. 【解】(1) 由散点图可知, $y = c \ln x + d$ 更适合作为每年 1 月份来哈尔滨的游客数量 y 关于年份代码 x 的回归方程类型.

因为 $u = \ln x$, 所以 $y = cu + d$.

$$\text{因为 } \bar{u} = 1.3, \bar{y} = 165, \sum_{i=1}^8 u_i^2 = 17.02, \sum_{i=1}^8 u_i y_i = 1\,901.5,$$

$$\text{所以 } \hat{c} = \frac{\sum_{i=1}^8 u_i y_i - 8 \bar{u} \bar{y}}{\sum_{i=1}^8 u_i^2 - 8 \bar{u}^2} = \frac{1\,901.5 - 8 \times 1.3 \times 165}{17.02 - 8 \times 1.3^2} =$$

$$\frac{185.5}{3.5} = 53,$$

所以 $\hat{d} = \bar{y} - \hat{c} \bar{u} = 165 - 53 \times 1.3 = 96.1$, 所以 $\hat{y} = 53u + 96.1$, 即 $\hat{y} = 53 \ln x + 96.1$.

所以每年 1 月份来哈尔滨的游客数量 y 关于年份代码 x 的回归方程为 $\hat{y} = 53 \ln x + 96.1$.

(2) 当 $x=10$ 时, $\hat{y} \approx 53 \times 2.3 + 96.1 = 218$,

所以预测 2026 年 1 月份来哈尔滨的游客数量为 218 万.

刷易错

★易错点 1 混淆相关关系与函数关系致误

13. C 【解析】圆的半径与周长是函数关系, 故 A 错误;

吸烟越多, 越不健康, 所以吸烟与健康具有负相关关系, 故 B 错误;

汽车越重, 每消耗 1 L 汽油所行驶的平均路程越短, 所以汽车的质量与汽车每消耗 1 L 汽油所行驶的平均路程具有负相关关系, 故 D 错误;

一般来说, 数学成绩越好, 那么物理成绩越好, 所以数学成绩与物理成绩具有正相关关系, 故 C 正确. 故选 C.

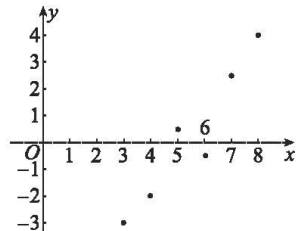
易错警示 函数关系是一种确定性关系, 相关关系是一种非确定性关系, 判断两个变量间的关系是不是相关关系的关键是看这个关系是不是具有不确定性.

★易错点 2 对回归直线方程的理解不到位致误

14. C 【解析】根据表格中样本数据的样本点作出散点图如图所示.

悟: 已知一组数据(数据较少)求回归直线的斜率与截距的正负时, 无需利用最小二乘法计算出具体数值, 直接画散点图分析即可.

从整体上看, 这些点大致分布在一条直线的周围, 且该回归直线的斜率为正, 在 y 轴上的截距为负, 则 $\hat{a} < 0, \hat{b} > 0$, 故选 C.



易错警示 分析随 x 的增加 y 的变化趋势可知 \hat{b} 的正负, 根据回归直线的纵截距可知 \hat{a} 的正负, 若算出具体数值则易出错.

★易错点 3 求回归直线方程计算错误

15.D 【解析】由题表得 $\bar{x} = \frac{7}{2}$, $\bar{y} = \frac{13}{6}$, 根据公式求得 $\hat{b} =$

$$\frac{58 - 6 \times \frac{7}{2} \times \frac{13}{6}}{91 - 6 \times \left(\frac{7}{2}\right)^2} = \frac{5}{7}, \hat{a} = \bar{y} - \hat{b}\bar{x} = \frac{13}{6} - \frac{5}{7} \times \frac{7}{2} = -\frac{1}{3}.$$

又根据题意可求得 $b' = 2, a' = -2$, 所以 $\hat{b} < b', \hat{a} > a'$, 故选 D.

→ **避坑:** 根据两点求直线方程时, 变量 x 与 y 是函数关系

易错警示 给出一组具体数据求其回归直线方程时, 由于

回归直线方程是 $\hat{y} = \hat{b}x + \hat{a}$, 其中 $\hat{b} = \frac{\sum_{i=1}^n x_i y_i - n\bar{x}\bar{y}}{\sum_{i=1}^n x_i^2 - n\bar{x}^2}$, $\hat{a} = \bar{y} - \hat{b}\bar{x}$, 计

算量较大, 因此计算要细心, 公式使用要准确, 要特别注意回归直线一定过样本点的中心 (\bar{x}, \bar{y}) 的应用.

★易错点 4 相关系数理解不正确致误

16.C 【解析】由变量 x 与 y 相对应的一组数据为 $(10, 1), (11.3, 2), (11.8, 3), (12.5, 4), (13, 5)$, 可得变量 y 与 x 正相关, 所以 $r_1 > 0$.

而由变量 u 与 v 相对应的一组数据为 $(10, 5), (11.3, 4), (11.8, 3), (12.5, 2), (13, 1)$, 可知变量 v 与 u 负相关, 所以 $r_2 < 0$, 所以 r_1 与 r_2 的大小关系是 $r_2 < 0 < r_1$. 故选 C.

易错警示 给出一组数据, 仅是判断该组数据的相关系数符号而非计算具体数值时, 可以先把第一维坐标按由小到大的顺序排列, 若此时对应的第二维坐标大致呈由小到大的顺序排列, 则相关系数为正; 若第二维坐标大致呈由大到小的顺序排列, 则相关系数为负. 切忌直接代入相关系数公式计算, 此法计算复杂, 耗时且易出错.

刷提升

1.A 【解析】根据相关系数的定义知, $|r|$ 越接近于 1 线性相关性越强, 散点图①③中变量之间为正相关关系, 且散点图①的线性相关性较强, 所以 $0 < r_3 < r_1$.

又因为散点图②④中变量之间为负相关关系, 且散点图②的线性相关性较强, 所以 $r_2 < r_4 < 0$. 故选 A.

2.ABC 【解析】对于 A, 回归直线方程 $\hat{y} = 0.8x + 0.5$ 中, 斜率 0.8 表示公共充电桩密度每增加 1, 新能源汽车通勤效率提升率大约增加 0.8%, 故 A 选项正确;

对于 B, 相关系数 $r = 0.92$, 绝对值接近 1 且为正, 所以公共充电桩密度与新能源汽车通勤效率提升率呈高度正线性相关关系, 故 B 选项正确;

对于 C, 将 $x = 5$ 代入回归直线方程, 得 $\hat{y} = 0.8 \times 5 + 0.5 = 4.5$, 即可预测其新能源汽车通勤效率提升率为 4.5%, 故 C 选项正确;

对于 D, 相关系数高说明相关性强, 但不能直接推断因果关系, 即不能推断新能源汽车通勤效率提升率由公共充电桩密度决定, 故 D 选项错误. 故选 ABC.

3.ABC 【解析】由新成对样本数据 $(x_1 - \bar{x}, y_1 - \bar{y}), (x_2 - \bar{x}, y_2 - \bar{y}), \dots, (x_n - \bar{x}, y_n - \bar{y})$, 可得其平均数为 $\bar{x}' = \frac{(x_1 - \bar{x}) + (x_2 - \bar{x}) + \dots + (x_n - \bar{x})}{n} = \frac{(x_1 + x_2 + \dots + x_n) - n\bar{x}}{n} = 0$,

$$\bar{y}' = \frac{(y_1 - \bar{y}) + (y_2 - \bar{y}) + \dots + (y_n - \bar{y})}{n} = \frac{(y_1 + y_2 + \dots + y_n) - n\bar{y}}{n} = 0,$$

$$\text{则 } (x_i - \bar{x}) - \bar{x}' = (x_i - \bar{x}) - 0 = x_i - \bar{x},$$

$$(y_i - \bar{y}) - \bar{y}' = (y_i - \bar{y}) - 0 = y_i - \bar{y},$$

$$\text{因为 } \hat{b} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}, r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}},$$

所以两条回归直线的斜率相同, 两组数据的相关系数相同, 故 A, C 正确;

因为回归直线经过样本点中心 (\bar{x}, \bar{y}) , 而新数据的样本点中心为 $(0, 0)$,

即原样本数据的回归直线方程为 $\hat{y} - \bar{y} = \hat{b}(x - \bar{x})$, 而新样本数据的回归直线方程为 $\hat{y}' = \hat{b}x$, 故两条回归直线的截距不相同, 故 D 错误;

原数据的残差为 $y_i - \hat{y}'_i = y_i - [\bar{y} + \hat{b}(x_i - \bar{x})]$, 新数据的残差为 $y_i - \bar{y} - \hat{y}'_i = y_i - \bar{y} - \hat{b}(x_i - \bar{x})$, 与原数据的残差相同, 因此残差平方和相同, 故 B 正确. 故选 ABC.

4.(9, 5) 【解析】由 $\bar{x} = 8.2$, 可知 $\bar{y} = 8.2 - 3 = 5.2$, 即 $\sum_{i=1}^{10} x_i = 82$, $\sum_{i=1}^{10} y_i = 52$,

$$\text{剔除 } (1, 7) \text{ 后, 新数据的平均数 } \bar{x}' = \frac{82-1}{9} = 9, \bar{y}' = \frac{52-7}{9} = 5,$$

因为样本点中心 (\bar{x}', \bar{y}') 一定在回归直线上, 所以得到新的回归直线必过点 $(9, 5)$.

5. $e^{\frac{15}{2}}$ 【解析】依题意, $\bar{x} = \frac{1+2+3+4}{4} = 2.5, \bar{z} = \frac{1+3+4+6}{4} = 3.5$.

$$\text{由 } \bar{z} = \hat{b}\bar{x} - 0.5, \text{ 得 } 3.5 = 2.5\hat{b} - 0.5,$$

$$\text{解得 } \hat{b} = 1.6, \text{ 于是 } \hat{z} = 1.6x - 0.5, \text{ 则 } \hat{y} = e^{1.6x - 0.5}.$$

$$\text{所以当 } x = 5 \text{ 时, } \hat{y} = e^{1.6 \times 5 - 0.5} = e^{\frac{15}{2}}.$$

6.【解】(1) 模型一更合适.

模型一残差点比较均匀地落在水平的带状区域中, 且带状区域的宽度比模型二带状宽度窄,

所以模型一的拟合精度更高, 回归方程的预报精度相应就会越高, 故选模型一比较合适.

(2) 由(1)知模型一比较合适, 因为 $z = \ln y$, 所以 $z = bx + a$, 所以 z 与温度 x 可以用回归直线方程来拟合, 则 $\hat{z} = \hat{a} + \hat{b}x$.

$$\text{于是 } \hat{b} = \frac{\sum_{i=1}^8 (x_i - \bar{x})(z_i - \bar{z})}{\sum_{i=1}^8 (x_i - \bar{x})^2} = \frac{50.4}{168} = 0.3, \hat{a} = \bar{z} - \hat{b}\bar{x} = 2.9 -$$

$$0.3 \times 25 = -4.6,$$

因此 z 关于 x 的回归直线方程为 $\hat{z} = 0.3x - 4.6$, 即 $\ln \hat{y} = 0.3x - 4.6$,

所以产卵数 y 关于温度 x 的回归方程为 $\hat{y} = e^{0.3x - 4.6}$.

7. 【解】(1) 由题中数据得 $\bar{x} = 3$, 则 $\sum_{i=1}^5 (x_i - \bar{x})^2 = 10$,

因为回归直线方程为 $\hat{y} = -65.2x + 1210.4$,

$$\text{所以 } \hat{b} = \frac{\sum_{i=1}^5 (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^5 (x_i - \bar{x})^2} = \frac{\sum_{i=1}^5 (x_i - \bar{x})(y_i - \bar{y})}{10} =$$

$$-65.2,$$

$$\text{所以 } \sum_{i=1}^5 (x_i - \bar{x})(y_i - \bar{y}) = -652,$$

$$\text{所以 } r = \frac{\sum_{i=1}^5 (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^5 (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^5 (y_i - \bar{y})^2}} \approx \frac{-652}{3.16 \times 237.5} \approx$$

-0.87 . 因为 $|r| \approx 0.87 > 0.75$, 所以该回归直线方程有价值.

(2) 由题知 X 的所有可能取值为 $0, 1, 2$,

$$\text{所以 } P(X=0) = \frac{C_3^3}{C_5^3} = \frac{1}{10}, P(X=1) = \frac{C_2^1 C_3^2}{C_5^3} = \frac{3}{5}, P(X=2) =$$

$$\frac{C_2^2 C_3^1}{C_5^3} = \frac{3}{10},$$

所以 X 的分布列为

X	0	1	2
P	$\frac{1}{10}$	$\frac{3}{5}$	$\frac{3}{10}$

$$\text{所以 } E(X) = 0 \times \frac{1}{10} + 1 \times \frac{3}{5} + 2 \times \frac{3}{10} = \frac{6}{5}.$$

💡思: 也可以注意到 X 服从超几何分布 $H(5, 3, 2)$,

$$\text{从而求期望, 即 } E(X) = \frac{3 \times 2}{5} = \frac{6}{5}$$

4.3.2 独立性检验

刷基础

1. A 【解析】 $a = 35 - 8 = 27$, $b = a + 11 = 27 + 11 = 38$. 故选 A.

2. A 【解析】根据独立性检验的方法和 2×2 列联表可得, 若

$\frac{a}{a+10}$ 与 $\frac{c}{c+30}$ 相差越大, 则分类变量 X 和 Y 有关系的可能性越

大, 即 a, c 相差越大, $\frac{a}{a+10}$ 与 $\frac{c}{c+30}$ 相差越大.

由各选项可得 A 满足条件, 故选 A.

$$\text{名师点拨 } \chi^2 = \frac{n(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)} = n \left(\frac{a}{a+c} - \frac{b}{b+d} \right) \cdot$$

$$\left(\frac{a}{a+b} - \frac{c}{c+d} \right). \text{ 当 } b, d \text{ 一定时, } a, c \text{ 相差越大, } \frac{a}{a+10} \text{ 与 } \frac{c}{c+30} \text{ 相差越}$$

大, χ^2 就越大, 得出分类变量 X 和 Y 有关系的可能性越大.

3. B 【解析】独立性检验是通过统计学方法来检验两个分类变量之间是否存在关联性, ACD 满足独立性检验的基本思想, B

选项只是公司的营业额这一个变量在过去 5 年逐年变化的情况, 不满足独立性检验的基本思想. 故选 B.

4. B 【解析】对于分类变量 X 与 Y 的统计量 χ^2 , χ^2 越大, “ X 与 Y 有关系”的可信程度越大; χ^2 越小, “ X 与 Y 有关系”的可信程度越小, 所以选项 B 正确. 故选 B.

5. D 【解析】根据所给的数据得到 2×2 列联表, 如下:

休闲方式	性别		
	男	女	合计
看电视	20	40	60
运动	35	25	60
合计	55	65	120

$$\text{计算 } \chi^2 = \frac{120 \times (20 \times 25 - 35 \times 40)^2}{55 \times 65 \times 60 \times 60} \approx 7.552 > 6.635, \text{ 所以有 } 99\%$$

的把握认为休闲方式与性别有关系. 故选 D.

6. C 【解析】因为 $\chi^2 \approx 6.816 > 6.635$,

所以数学成绩优秀与性别有关, 此推断犯错误的概率不大于 0.01 ,

即在犯错误的概率不超过 1% 的前提下认为数学成绩优秀与性别有关, 故 C 正确, D 错误;

若某人数学成绩优秀, 由已知数据不能判断他为男生的概率, 故 A 错误;

每 100 个数学成绩优秀的人中可能没有女生, 也有可能有多名女生, 由已知数据不能确定结论, 故 B 错误. 故选 C.

7. 【解】(1) 由列联表知 $s = 50 + 25 = 75$, $t = 25 + 225 = 250$.

$$(2) \text{ 由列联表得 } \chi^2 = \frac{500 \times (50 \times 225 - 25 \times 200)^2}{75 \times 425 \times 250 \times 250} \approx 9.804,$$

由于 $9.804 > 6.635$, 所以有 99% 的把握认为该地老年人是否需要帮助与性别有关.

(3) 采用分层抽样, 理由如下:

由(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抽取, 比采用简单随机抽样方法更好.

刷易错

★易错点 1 不理解独立性检验的基本思想

8. B 【解析】因为利用独立性检验时得到的结论与样本的选取有关, 所以样本不同得到的结论可能有差异, 故独立性检验得到的结论可能不正确, 因此 B 说法错误. 故选 B.

易错警示 独立性检验是利用随机变量 χ^2 来判断两个分类变量是否有关系的方法. 独立性检验得到的结论不一定可靠, 只能作为理论上的参考.

★易错点 2 对独立性检验的结果判断错误

9. D 【解析】因为 $3.841 > 2.974 > 2.706$, 所以变量 x 与 y 不独立, 这个结论犯错误的概率不超过 10% . 故选 D.

→即有 90% 的把握认为变量 x 与 y 不独立

易错警示 对独立性检验的结果进行描述时, 要注意“在犯错误的概率不超过 α 的前提下”与“有 $1-\alpha$ 以上的把握”的区别, 同时要注意判断两个变量是有关还是无关.

高中必刷题 数学

★易错点3 忽视数据的隐含条件而致误

10.9 【解析】由题意知 $\chi^2 \geq 6.635$,

$$\text{则 } \frac{65[a(30+a)-(20-a)(15-a)]^2}{15 \times 50 \times 20 \times 45} = \frac{13(13a-60)^2}{5400} \geq 6.635,$$

解得 $a \leq 0.577$ 或 $a \geq 8.654$. 因为 $a > 5$ 且 $15-a > 5, a \in \mathbb{Z}$, 所以 $8.654 \leq a < 10, a \in \mathbb{Z}$, 所以 $a = 9$.

易错警示 本题求解的易错之处: 一是忽视分类变量中的 $a \in \mathbb{Z}$; 二是忽视 a 和 $15-a$ 应满足大于 5 的条件.

刷提升

1.D 【解析】根据题意, 结合题目中的数据, 列出 2×2 列联表, 根据表中数据可得 χ^2 的值, 再与分位数 k 比较, 可得这些中学生假期里每天玩手机超过 1 小时是否与性别有关的结论, 故利用独立性检验的方法最有说服力. 故选 D.

2.ABD 【解析】对于 A, B, 由题中 2×2 列联表知, $m = 70 - 27 = 43, n = 110 - 58 = 52, b = 27 + n = 79, A, B$ 正确;

对于 C, D, 由 $\chi^2 \approx 1.315 < 2.706 < 3.841$ 知, C 错误, D 正确.

【易错黑板: χ^2 值越大, 说明相关性越强, 独立性越弱; χ^2 值越小, 说明相关性越弱, 独立性越强

故选 ABD.

3.D 【解析】由于 $\chi^2 = \frac{n(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)} =$

$$\frac{200 \times (70f - 30e)^2}{100 \times 100 \cdot m \cdot n} = \frac{8}{3} (*),$$

又 $m = 70 + e, n = 30 + f, 100 = e + f, 200 = m + n,$

则 $m = 200 - n, f = n - 30, e = 130 - n$, 代入 (*) 式可得 $38n^2 - 4600n + 135000 = 0$, 解得 $n = 50$ 或 $n = \frac{1350}{19}$ (舍). 故选 D.

4.A 【解析】设总人数为 $2n$, 则男生选学生物学的人数为 $\frac{4}{5}n$,

女生选学生物学的人数为 $\frac{3}{5}n$,

$$\text{则 } \chi^2 = \frac{2n \left(\frac{4n}{5} \times \frac{2n}{5} - \frac{3n}{5} \times \frac{n}{5} \right)^2}{n \times n \times \frac{7n}{5} \times \frac{3n}{5}} = \frac{2n}{21} \geq 2.706,$$

即 $n \geq \frac{2.706 \times 21}{2} = 28.413$, 又 n 为 5 的倍数, 结合选项可知男

生的人数可能为 30, 35, 40, 不可能为 20, 故选 A.

5.BD 【解析】男生有 35 人, 其中经常玩电子游戏的有 30 人,

频率为 $\frac{30}{35} = \frac{6}{7}$,

女生有 15 人, 其中经常玩电子游戏的有 10 人, 频率为 $\frac{10}{15} = \frac{2}{3}$,

$\frac{6}{7} > \frac{2}{3}$, 依据频率稳定于概率的原理, 可以认为性别对玩电子

游戏的经常性有影响, 故 A 错误, B 正确;

$\chi^2 = \frac{50 \times (5 \times 10 - 30 \times 5)^2}{35 \times 15 \times 10 \times 40} \approx 2.381 < 3.841$, 所以无法得到“在犯

错误的概率不超过 0.05 的前提下, 可以认为性别对玩电子游戏的经常性有影响”的结论, 又 $2.381 < 2.706$, 因此没有充分证据推断性别对玩电子游戏的经常性有影响, 故 C 错误, D 正确. 故选 BD.

6.ABC 【解析】对于选项 A, 由列联表得 $\chi^2 = \frac{200 \times (40 \times 90 - 60 \times 10)^2}{100 \times 100 \times 50 \times 150} = 24 > 6.635$,

所以在犯错误概率不超过 0.01 的条件下, 认为患该疾病群体与未患该疾病群体的卫生习惯有差异, 故选项 A 正确;

对于选项 B, 由已知得 $P(A|B) = \frac{40}{100} = 0.4, P(A|\bar{B}) = \frac{10}{100} = 0.1$, 故选项 B 正确;

对于选项 C, 由题意得 $R = \frac{P(B|A)}{P(\bar{B}|A)} \cdot \frac{P(\bar{B}|\bar{A})}{P(B|\bar{A})} = \frac{P(AB)}{P(A)} \cdot$

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

$$\text{又 } \frac{P(A|B)}{P(\bar{A}|B)} \cdot \frac{P(\bar{A}|\bar{B})}{P(A|\bar{B})} = \frac{P(AB)}{P(B)} \cdot \frac{P(\bar{B})}{P(\bar{A})} \cdot \frac{P(\bar{A}|\bar{B})}{P(\bar{B})} \cdot$$

$$\frac{P(\bar{B})}{P(\bar{A})} = \frac{P(AB)P(\bar{A}|\bar{B})}{P(\bar{A})P(\bar{B})},$$

所以 $R = \frac{P(A|B)}{P(\bar{A}|B)} \cdot \frac{P(\bar{A}|\bar{B})}{P(A|\bar{B})}$, 故选项 C 正确;

对于选项 D, 由已知得 $P(A|B) = \frac{40}{100}, P(\bar{A}|B) = \frac{60}{100}, P(A|\bar{B}) = \frac{10}{100}, P(\bar{A}|\bar{B}) = \frac{90}{100}$,

所以 $R = \frac{P(A|B)}{P(\bar{A}|B)} \cdot \frac{P(\bar{A}|\bar{B})}{P(A|\bar{B})} = 6$, 故选项 D 错误. 故选 ABC.

7.【解】(1) 由题意知 X 的所有可能取值为 1, 2, 3,

$$\text{则 } P(X=1) = \frac{C_3^1 C_2^2}{C_5^3} = \frac{3}{10}, P(X=2) = \frac{C_3^2 C_2^1}{C_5^3} = \frac{3}{5}, P(X=3) =$$

$$\frac{C_3^3}{C_5^3} = \frac{1}{10},$$

则 X 的分布列为

X	1	2	3
P	$\frac{3}{10}$	$\frac{3}{5}$	$\frac{1}{10}$

$$\text{所以 } E(X) = 1 \times \frac{3}{10} + 2 \times \frac{3}{5} + 3 \times \frac{1}{10} = \frac{9}{5}.$$

(2) 由题中数据可得列联表如下:

成绩	日均阅读时长		
	[1, 2)	其他	合计
优秀	45	50	95
不优秀	180	300	480
合计	225	350	575

$$\text{则 } \chi^2 = \frac{575 \times (45 \times 300 - 50 \times 180)^2}{225 \times 350 \times 95 \times 480} \approx 3.242 < 3.841,$$

所以在犯错误概率不超过 0.05 的条件下, 不能认为学业成绩优秀与日均阅读时长不小于 1 小时且小于 2 小时有关.

第 4.3 节综合训练

刷能力

1.D 【解析】由题表中数据可得 y 随 x 的增大而增大, 故 y 与 x

正相关. 又 $\bar{x} = \frac{1}{5} \times (10 + 20 + 30 + 40 + 50) = 30, \bar{y} = \frac{1}{5} \times (64 + 69 +$

$75+82+90=76$, 所以样本点的中心为 $(30, 76)$, 而回归直线过样本点的中心, 因此其回归直线经过点 $(30, 76)$, 故选 D.

2. D 【解析】对于 A, 由散点图易知 y 与 x 不是负相关关系, 所以 A 错误;

对于 B, C, 由散点图可知变量 y 与 x 的变化趋向于一条曲线, 所以模型二能更好地拟合 GDP 值随年份代码的变化情况, 所以 B 错误, C 错误;

对于 D, 若选择模型二: $y = ke^x + b$ ($k > 0, x > 0$), 令 $t = e^x$, 则 $y = kt + b$ 的图象一定过点 (\bar{t}, \bar{y}) , 所以 $y = ke^x + b$ 的图象不一定过点 (\bar{x}, \bar{y}) , 故 D 正确.

故选 D.

3. BD 【解析】回归直线是通过最小二乘法拟合数据得到的直线, 它的目的是使样本数据点到该直线的距离的平方和最小, 而不是经过样本数据点最多, 所以 A 选项错误.

对于回归直线方程 $\hat{y} = 8x - m$, 因为样本点中心 $(m, 14)$ 一定在回归直线上, 所以将 $x = m, \hat{y} = 14$ 代入回归直线方程可得 $14 = 8m - m$, 即 $7m = 14$, 解得 $m = 2$, 所以 B 选项正确.

由简单随机抽样得到的成对样本数据的相关系数只是对变量之间相关关系的一个估计, 它受到样本随机性的影响, 不一定能确切地反映变量之间的相关关系, 所以 C 选项错误.

在独立性检验中, 随机变量 χ^2 的观测值越小, 说明两个变量之间越可能没有关系, 那么“认为两个变量有关”这种判断犯错误的概率就越大, 所以 D 选项正确. 故选 BD.

4. AD 【解析】对于 A, 因为 $-0.3 < 0$, 所以变量 y 与 x 负相关, 故 A 正确;

对于 B, 在分类变量 X, Y 的 2×2 列联表中, $|ad - bc|$ 越小, 说明两个变量有关的可能性越小, $|ad - bc|$ 越大, 说明两个变量有关的可能性越大, 故 B 不正确;

对于 C, 回归直线经过样本点中心, 不一定经过样本点, 故

❶ 黑板: 回归直线过样本点的中心

C 不正确;

对于 D, 以 $\hat{y} = ae^{bx}$ ($a > 0$) 拟合一组数据, 设 $z = \ln y$, 则 $z = \ln(ae^{bx}) = bx + \ln a$, 若 z 关于 x 的回归直线方程为 $\hat{z} = -x + 2$, 则 $b = -1, \ln a = 2$, 所以 $a = e^2$, 则 $ab = -e^2$, 故 D 正确. 故选 AD.

5. 【解】(1) 根据题意完成 2×2 列联表如下:

是否喜欢担任的情况	性别		
	男性	女性	合计
喜欢担任	10	15	25
不喜欢担任	20	5	25
合计	30	20	50

$$\text{则 } \chi^2 = \frac{50 \times (10 \times 5 - 20 \times 15)^2}{25 \times 25 \times 30 \times 20} = \frac{25}{3} \approx 8.333 > 7.879,$$

故在犯错误概率不超过 0.005 的条件下, 可以认为居民喜欢担任垃圾分类志愿者与性别有关.

$$(2) \text{ 由题中表格数据可知, } \bar{x} = \frac{1}{5} \times (2+3+4+5+6) = 4, \bar{y} = \frac{1}{5} \times (24+29+41+46+60) = 40,$$

$$\sum_{i=1}^5 x_i^2 = 2^2 + 3^2 + 4^2 + 5^2 + 6^2 = 90, \text{ 又 } \sum_{i=1}^5 x_i y_i = 889,$$

$$\text{则 } \hat{b} = \frac{\sum_{i=1}^5 x_i y_i - 5 \bar{x} \bar{y}}{\sum_{i=1}^5 x_i^2 - 5 \bar{x}^2} = \frac{889 - 5 \times 4 \times 40}{90 - 5 \times 4^2} = \frac{89}{10} = 8.9, \hat{a} = \bar{y} - \hat{b} \bar{x} = 40 -$$

$$8.9 \times 4 = 4.4,$$

所以回归直线方程为 $\hat{y} = 8.9x + 4.4$.

$$\text{当 } x = 10 \text{ 时, } \hat{y} = 8.9 \times 10 + 4.4 = 93.4,$$

所以当志愿者人数为 10 时, 该垃圾站的日垃圾分拣量大约为 93.4 千克.

$$\mathbf{6. 【解】} (1) \text{ 因为 } r = \frac{\sum_{i=1}^{10} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{10} (x_i - \bar{x})^2 \sum_{i=1}^{10} (y_i - \bar{y})^2}} =$$

$$\frac{\sum_{i=1}^{10} x_i y_i - 10 \bar{x} \bar{y}}{\sqrt{(\sum_{i=1}^{10} x_i^2 - 10 \bar{x}^2) \times (\sum_{i=1}^{10} y_i^2 - 10 \bar{y}^2)}}, \text{ 代入已知数据, 得}$$

$$r \approx \frac{357.3 - 10 \times 34.02}{\sqrt{(353.6 - 10 \times 33.62) \times (361.7 - 10 \times 34.42)}} = \frac{17.1}{\sqrt{304.5}} \approx$$

$$0.98.$$

所以汛期遥测雨量 y 与人工测雨量 x 有很强线性相关关系.

(2) 依题意, “I 类误差”有 5 组, “II 类误差”有 3 组, “III 类误差”有 2 组.

若从 “I 类误差”和 “II 类误差”数据中抽取 3 组, 抽到 “I 类误差”的组数 X 的所有可能取值为 0, 1, 2, 3.

$$\text{则 } P(X=0) = \frac{C_3^3}{C_8^3} = \frac{1}{56}, P(X=1) = \frac{C_5^1 C_3^2}{C_8^3} = \frac{15}{56},$$

$$P(X=2) = \frac{C_5^2 C_3^1}{C_8^3} = \frac{30}{56} = \frac{15}{28}, P(X=3) = \frac{C_5^3 C_3^0}{C_8^3} = \frac{10}{56} = \frac{5}{28}.$$

所以 X 的分布列为

X	0	1	2	3
P	$\frac{1}{56}$	$\frac{15}{56}$	$\frac{15}{28}$	$\frac{5}{28}$

$$\text{所以 } X \text{ 的数学期望 } E(X) = 1 \times \frac{15}{56} + 2 \times \frac{15}{28} + 3 \times \frac{5}{28} = \frac{15}{8}.$$

专题 2 概率、统计与其他知识的综合

刷难关

1. D 【解析】依题意知, $P(Y=1) = p_1 + p_{2m}, P(Y=2) = p_2 + p_{2m-1}, P(Y=3) = p_3 + p_{2m-2}, \dots, P(Y=m) = p_m + p_{m+1}, \therefore H(Y) = -[(p_1 + p_{2m}) \log_2(p_1 + p_{2m}) + (p_2 + p_{2m-1}) \log_2(p_2 + p_{2m-1}) + \dots + (p_m + p_{m+1}) \log_2(p_m + p_{m+1})],$

$$\text{又 } H(X) = -(p_1 \log_2 p_1 + p_2 \log_2 p_2 + \dots + p_m \log_2 p_m + \dots + p_{2m} \log_2 p_{2m}),$$

$$\therefore H(Y) - H(X) = p_1 \log_2 \frac{p_1}{p_1 + p_{2m}} + p_2 \log_2 \frac{p_2}{p_2 + p_{2m-1}} + \dots + p_{2m} \log_2 \frac{p_{2m}}{p_1 + p_{2m}},$$

$$\text{又 } \frac{p_1}{p_1 + p_{2m}} < 1, \frac{p_2}{p_2 + p_{2m-1}} < 1, \dots, \frac{p_{2m}}{p_1 + p_{2m}} < 1, \therefore H(Y) - H(X) < 0,$$

$\therefore H(X) > H(Y)$. 故选 D.

2. 【解】(1) 设甲同学成功晋级为事件 A, 事件 A 发生有以下三

高中必刷题 数学

种情况:前三题全对;前三题对两题,后两题答对一题;前三题答对一题,后两题全对.

$$\text{所以 } P(A) = \left(\frac{1}{2}\right)^3 + C_3^2 \times \left(\frac{1}{2}\right)^2 \times \left[\frac{1}{3} + \left(1 - \frac{1}{3}\right) \times \frac{1}{3}\right] + C_3^1 \times \left(\frac{1}{2}\right)^2 \times \left(\frac{1}{3}\right)^2 = \frac{3}{8}.$$

(2) 设选择方式一、二的班级团队挑战成功的概率分别为 P_1, P_2 .

当选择方式一时,因为两人都回答错误的概率为 $(1-p)^2$,则两人中至少有一人回答正确的概率为 $1 - (1-p)^2$,所以 $P_1 = [1 - (1-p)^2]^n = p^n(2-p)^n$.

当选择方式二时,因为一个小组闯关成功的概率为 p^n ,则一个小组闯关不成功的概率为 $1 - p^n$,所以 $P_2 = 1 - (1 - p^n)^2 = p^n(2 - p^n)$,

$$\text{所以 } P_1 - P_2 = p^n(2-p)^n - p^n(2-p^n) = p^n[(2-p)^n + p^n - 2],$$

$$\text{构造 } f(n) = (2-p)^n + p^n - 2, \text{ 则 } f(n+1) - f(n) = (2-p)^{n+1} + p^{n+1} - (2-p)^n - p^n = (2-p)^n(1-p) + p^n(p-1) = (1-p)[(2-p)^n - p^n],$$

因为 $0 < p < 1$, 则 $1-p > 0, 2-p > 1$, 可得 $(2-p)^n > 1, p^n < 1$,

所以 $f(n+1) - f(n) > 0$, 即 $f(n+1) > f(n)$, 所以 $f(n)$ 单调递增,

$$\text{又因为 } f(2) = (2-p)^2 + p^2 - 2 = 2p^2 - 4p + 2 = 2(p-1)^2 > 0,$$

且 $n \geq 10$, 所以 $f(n) > 0$, 从而 $P_1 - P_2 > 0$, 即 $P_1 > P_2$,

所以为使本班团队挑战成功的可能性更大,应选择方式一.

3. 【解】 (1) 设底面边长为 a , 则 $\frac{1}{3} \times a^2 \times \sqrt{2} = \frac{4\sqrt{2}}{3}$, 得 $a = 2$.

连接 AC 交 BD 于点 O , 连接 PO , 作 $OM \perp PD$ 交 PD 于点 M , 连接 AM , 如图所示.

因为 $PO \perp$ 平面 $ABCD, AC \subset$ 平面 $ABCD$, 所以 $PO \perp AC$,

又 $AC \perp BD, PO \cap BD = O, PO, BD \subset$ 平面 PBD ,

所以 $AC \perp$ 平面 PBD , 因为 $PD \subset$ 平面 PBD , 所以 $AC \perp PD$.

又 $OM \perp PD$, 且 $AC \cap OM = O, AC, OM \subset$ 平面 AOM ,

所以 $PD \perp$ 平面 AOM , 又 $AM \subset$ 平面 AOM ,

所以 $PD \perp AM$, 所以 $\angle AMO$ 为平面 PAD 与平面 PBD 的夹角.

$$\text{因为 } PO = \sqrt{2}, AO = \frac{1}{2}AC = \sqrt{2}, \text{ 所以 } PA = \sqrt{PO^2 + AO^2} = 2,$$

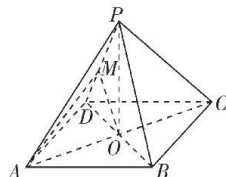
所以 $\triangle PAD$ 是等边三角形, $AM = \sqrt{3}$, 且 M 为 PD 的中点, 所以

$$OM = \frac{1}{2}PD = 1,$$

因为 $AC \perp$ 平面 $PBD, OM \subset$ 平面 PBD , 所以 $AC \perp OM$,

$$\text{所以 } \cos \angle AMO = \frac{MO}{AM} = \frac{1}{\sqrt{3}} = \frac{\sqrt{3}}{3},$$

所以平面 PAD 与平面 PBD 的夹角的余弦值为 $\frac{\sqrt{3}}{3}$.



(2) 由题意可知, 蚂蚁从点 P 处沿 PA, PB, PC, PD 移动的概率都是 $\frac{1}{4}$.

2 秒后蚂蚁移动了 2 个单位长度, 由 (1) 知侧棱长为 2, 所以若沿 PA 移动, 蚂蚁到达点 A , 若沿 PB 移动, 蚂蚁到达点 B , 若沿 PC 移动, 蚂蚁到达点 C , 若沿 PD 移动, 蚂蚁到达点 D , 因为 $AB = AD = 2, AC = 2\sqrt{2}$, 所以 X 的所有可能取值为 0, 2, $2\sqrt{2}$, 则 X 的分布列为

X	0	2	$2\sqrt{2}$
P	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

$$\text{则数学期望 } E(X) = 0 \times \frac{1}{4} + 2 \times \frac{1}{2} + 2\sqrt{2} \times \frac{1}{4} = 1 + \frac{\sqrt{2}}{2}.$$

第四章素养检测

刷速度

1. B 【解析】依题意可得 $P(X=1) = 0.8, P(X=0) = 0.2$, 所以 $E(X) = 0 \times 0.2 + 1 \times 0.8 = 0.8$. 故选 B.

2. D 【解析】由题可得 $E(X) = 1, E(Y) = 4 \times \frac{1}{2} = 2, D(X) = 3$,

$$D(Y) = 4 \times \frac{1}{2} \times \left(1 - \frac{1}{2}\right) = 1, \text{ 所以 } E(X) = D(Y). \text{ 故选 D.}$$

3. C 【解析】记事件 $A =$ “第一次正面朝上的点数为奇数”, 事件 $B =$ “第二次正面朝上的点数大于 4”.

投掷一枚正方体骰子两次, 所有的样本点 (x, y) (x 为第一次正面朝上的点数, y 为第二次正面朝上的点数) 共 36 个, 其中事件 A 包含的样本点有 18 个, 事件 AB 包含的样本点有 6 个, 所以 $P(AB) = \frac{6}{36} = \frac{1}{6}, P(A) = \frac{18}{36} = \frac{1}{2}$, 所以 $P(B|A) =$

$$\frac{P(AB)}{P(A)} = \frac{\frac{1}{6}}{\frac{1}{2}} = \frac{1}{3}. \text{ 故选 C.}$$

4. A 【解析】由题设知样本数据 (x_i, y_i) ($i = 1, 2, \dots, 9$) 满足

$$\bar{x} = \frac{33}{9} = \frac{11}{3}, \text{ 则 } \bar{y} = 2\bar{x} - 1 = 2 \times \frac{11}{3} - 1 = \frac{19}{3},$$

增加数据 $(-3, 3)$ 后, $\bar{x}_1 = \frac{33-3}{10} = 3, \bar{y}_1 = \frac{9 \times \frac{19}{3} + 3}{10} = 6$, 且回归直线方程为 $\hat{y} = 2.1x + \hat{a}$,

$$\text{所以 } 6 = 2.1 \times 3 + \hat{a} \Rightarrow \hat{a} = -0.3, \text{ 则 } \hat{y} = 2.1x - 0.3,$$

所以当 $x = 4$ 时, 有 $\hat{y} = 2.1 \times 4 - 0.3 = 8.1$, 故残差的绝对值为 $|8 - 8.1| = 0.1$. 故选 A.

5. A 【解析】由已知表格中的数据, 可知 $\bar{X} < \bar{Y}$, 即数学成绩分布曲线的对称轴在物理成绩分布曲线的对称轴的左侧.

又 $\sigma(X) = 94 > \sigma(Y) = 23$, 所以数学成绩分布曲线“矮胖”, 物理成绩分布曲线“瘦高”, 结合选项可知 A 正确. 故选 A.

6. D 【解析】设事件 A 表示“小孩诚实”, 事件 B 表示“小孩说谎”, 则 $P(B|A) = 0.1, P(B|\bar{A}) = 0.5, P(A) = 0.9, P(\bar{A}) = 0.1$, 所以 $P(B) = P(A)P(B|A) + P(\bar{A})P(B|\bar{A}) = 0.9 \times 0.1 + 0.1 \times 0.5 = 0.14$, 故选 D.