



第4章 统计

4.1 成对数据的统计相关性

题型诀

1-1. D 【解析】当两个变量之间具有确定的关系时,两个变量之间是函数关系,而不是相关关系,故 A 错误;

球的体积与该球的半径之间是函数关系,故 B 错误;

农作物的产量与施肥量之间的关系是相关关系,是非确定性关系,故 C 错误;

学生的数学成绩与物理成绩之间的关系是相关关系,是非确定性关系,故 D 正确. 故选 D.

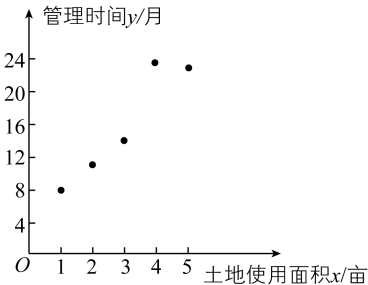
1-2. D 【解析】A 选项中的变量 x, y 具有确定的函数关系;

B 选项中的变量 x, y 有非线性相关关系;

C 选项中的变量 x, y 既没有函数关系,也没有相关关系;

D 中的点大致分布在一条直线附近,故 D 选项中的变量 x, y 有线性相关关系. 故选 D.

2-1. 【解】散点图如图所示.



由散点图可知,管理时间 y 与土地使用面积 x 线性相关.

依题意, $\bar{x} = \frac{1 + 2 + 3 + 4 + 5}{5} = 3$,

又 $\bar{y} = 16$,

$$\sum_{i=1}^5 (x_i - \bar{x})(y_i - \bar{y}) = 43,$$



$$\sum_{i=1}^5 (x_i - \bar{x})^2 = 10, \sum_{i=1}^5 (y_i - \bar{y})^2 = 206,$$

$$\text{则 } r_{xy} = \frac{\sum_{i=1}^5 (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^5 (x_i - \bar{x})^2 \sum_{i=1}^5 (y_i - \bar{y})^2}} =$$

$$\frac{43}{\sqrt{10 \times 206}} = \frac{43}{2\sqrt{515}} \approx \frac{43}{45.4} \approx 0.947.$$

因为 $0.947 > 0.8$,

所以管理时间 y 与土地使用面积 x 线性相关性较强.

巩固练

1. **C** 【解析】对于 A, 圆的面积与半径之间的关系是确定的关系, 是函数关系, 所以 A 错误;

对于 B, 吸烟与健康之间的关系不是函数关系, 是相关关系, 所以 B 错误;

对于 C, 一定范围内, 学生的成绩与学习时间是呈正相关关系的, 所以 C 正确;

对于 D, 人的体重与视力没有相关关系, 所以 D 错误. 故选 C.

2. **D** 【解析】对于题图①, 图中的点呈带状分布, 且从左到右上升, 是正相关关系, 选(1); 对于题图②, 图中的点没有明显的带状分布, 两变量不相关, 选(3); 对于题图③, 图中的点呈带状分布, 且从左到右是下降的, 是负相关关系, 选(2). 故选 D.

3. **D** 【解析】逐年比较, 2018 年该市二氧化硫年排放量最少, 故 A 正确;

2012 年比 2011 年二氧化硫年排放量明显减少, 故 2012 年该市治理二氧化硫排放显现成效, 故 B 正确;

2011 年以来每年该市二氧化硫年排放量除 2016 年外都在减少, 故 2011 年以来该市二氧化硫年排放量呈减少趋势, 故 C 正确;

2011 年以来该市二氧化硫年排放量随



年份逐渐减少,故与年份负相关,故 D 错误. 故选 D.

4. 【解】(1) 样本中 10 个这种零件的横

截面积的平均值 $\bar{x} = \frac{0.52}{10} = 0.052$,

样本中 10 个这种零件的耗材量的平

均值 $\bar{y} = \frac{3.9}{10} = 0.39$,

据此可估计该同学制作的这种零件平

均一个的横截面积为 0.052 mm^2 , 平均

一个零件的耗材量为 0.39 mm^3 .

(2) 相关系数

$$\begin{aligned} r_{xy} &= \frac{\sum_{i=1}^{10} x_i y_i - 10 \bar{x} \bar{y}}{\sqrt{(\sum_{i=1}^{10} x_i^2 - 10 \bar{x}^2)(\sum_{i=1}^{10} y_i^2 - 10 \bar{y}^2)}} \\ &= \frac{0.0115}{\sqrt{0.000149136}} \approx \frac{0.0115}{0.01221} \\ &\approx 0.94. \end{aligned}$$

(3) 设这种零件的总耗材量的估计值

为 $y \text{ mm}^3$.

又已知这种零件的耗材量及其横截面

积近似成正比,

可得 $\frac{0.052}{0.39} = \frac{182}{y}$, 解得 $y = 1365$,

故他制作的零件的总耗材量的估计值

为 1365 mm^3 .

5. AD 【解析】对于相关系数 r , 取值范

围是 $[-1, 1]$, $|r|$ 越大, 越接近于 1, 成

对样本数据的线性相关程度越强; $|r|$

越小, 越接近于 0, 成对样本数据的线

性相关程度越弱. 故选 AD.

4.2 一元线性回归模型

4.2.1 回归直线方程+4.2.2 一元线性回归模型的应用

题型诀

1-1. ABC 【解析】对于 A, 因为 $0.85 >$

0, 所以 y 与 x 是正相关的, 所以 A 正确.



对于 B, 回归直线恒过样本点中心, 所以回归直线 $\hat{y} = 0.85x - 85.71$ 必过点 (\bar{x}, \bar{y}) , 所以 B 正确. 对于 C, 由于回归直线方程为 $\hat{y} = 0.85x - 85.71$, 所以可知该中学某高中女生身高增加 1 cm, 则其体重约增加 0.85 kg, 所以 C 正确. 对于 D, 当 $x = 160$ 时, $\hat{y} = 0.85 \times 160 - 85.71 = 50.29$, 所以该中学某高中女生身高为 160 cm 时, 其体重约为 50.29 kg, 所以 D 错误. 故选 ABC.

1-2. 【解】(1) $r_{xy} =$

$$\frac{\sum_{i=1}^7 (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^7 (x_i - \bar{x})^2} \cdot \sqrt{\sum_{i=1}^7 (y_i - \bar{y})^2}} \approx \frac{395}{\sqrt{96 \times 1809}} = \frac{395}{36\sqrt{134}} \approx 0.95 > 0.8,$$

所以 y 与 x 两个变量相关性强, 可以用一元线性回归模型拟合.

$$(2) \text{ 因为 } \hat{b} = \frac{\sum_{i=1}^7 (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^7 (x_i - \bar{x})^2} \approx \frac{395}{96} \approx 4,$$

$$\bar{x} = \frac{1}{7} \times (2+3+4+6+8+10+13) = \frac{46}{7},$$

$$\bar{y} = \frac{1}{7} \times (13+22+31+42+50+56+58) = \frac{272}{7},$$

$$\text{所以 } \hat{a} = \bar{y} - \hat{b}\bar{x} = \frac{272}{7} - 4 \times \frac{46}{7} = \frac{88}{7} \approx 13,$$

故 y 关于 x 的一元线性回归方程为 $\hat{y} = 4x + 13$.

(3) 当 $x = 16$ 时, $\hat{y} = 4 \times 16 + 13 = 77$ (亿元),

故当研发投入为 16 亿元时, 公司的实际收益的预测值为 $77 + 4 = 81$ (亿元).

2-1. A 【解析】对于①, $z = \ln y = \ln(c \cdot e^{kx}) = kx + \ln c$, 又 $\hat{z} = 0.3x + 4$, 对比系数得 $k = 0.3$, $\ln c = 4$, 即 $c = e^4$, ①正确;
对于②, 回归直线过样本中心, 所以 $3 = \hat{a} + 2 \times 1 \Rightarrow \hat{a} = 1$, ②正确;



对于③,通过一元线性回归方程 $\hat{y} = \hat{b}x + \hat{a}$ 无法精确反映变量的取值,③错误.

所以错误的个数是 1. 故选 A.

2-2. 【解】(1) 根据散点图,开始的点在 某条直线旁,但后面的点会越来越偏离 这条直线,因此 $y = c + d \cdot \ln x$ 更适合作为 回归方程类型.

(2) 记 $w = \ln x$, 则 $y = c + d \cdot \ln x$ 可化为 $y = c + d \cdot w$,

$$\bar{y} = \frac{\sum_{i=1}^{20} y_i}{20} = \frac{102.4}{20} = 5.12,$$

$$\bar{w} = \frac{\sum_{i=1}^{20} w_i}{20} = \frac{52}{20} = 2.6,$$

$$\begin{aligned} \hat{d} &= \frac{\sum_{i=1}^{20} w_i y_i - 20 \bar{w} \bar{y}}{\sum_{i=1}^{20} w_i^2 - 20 \bar{w}^2} \\ &= \frac{272.1 - 20 \times 2.6 \times 5.12}{137 - 20 \times 2.6^2} \\ &\approx 3.26, \end{aligned}$$

$$\hat{c} = \bar{y} - \hat{d} \bar{w} = 5.12 - 3.26 \times 2.6 \approx -3.36,$$

所以 $\hat{y} = 3.26w - 3.36$, 即 $\hat{y} = 3.26 \ln x - 3.36$.

(3) 当 $x = e^2$ 时, $\hat{y} = 3.26 \ln e^2 - 3.36 = 3.16$.

故当光照时长为 e^2 小时时,大棚蔬菜每公顷产量约为 3.16 千克.

巩固练

1. D 【解析】对两个变量 x, y 进行线性回归分析时,首先收集数据 (x_i, y_i) , $i = 1, 2, \dots, n$, 根据所收集的数据绘制散点图. 观察散点图的形状,求相关系数,判断线性相关关系的强弱,写出回归直线方程,最后对所求出的回归直线方程作出解释. 故正确顺序是②⑤④③①. 故选 D.

2. A 【解析】由题知 $\bar{x} = \frac{1+2+3+4}{4} = 2.5$,

$$\bar{y} = \frac{0.1+1.8+m+4}{4} = \frac{5.9+m}{4}. \text{ 又样本中}$$



心 (\bar{x}, \bar{y}) 在回归直线上, 所以 $\frac{5.9+m}{4} =$

$1.3 \times 2.5 - 1$, 解得 $m = 3.1$.

3. **D** 【解析】由题意可得, $\bar{x} = \frac{1}{5} \times (6 +$

$7 + 10 + 12 + 15) = 10$,

\therefore 回归直线方程为 $\hat{y} = 0.7x - 6$,

$\therefore \bar{y} = 0.7 \times 10 - 6 = 1$.

$\therefore y_1, y_2, y_3, y_4, y_5$ 成等差数列,

$\therefore \bar{y} = \frac{1}{5}(y_1 + y_2 + y_3 + y_4 + y_5) = y_3 = 1$.

故选 D.

4. **A** 【解析】根据题意, 适合用线性回

归模型拟合其中两个变量的散点图

中, 点的分布必须比较集中, 且大体接

近某一条直线, 分析选项可得 A 选项

的散点图杂乱无章, 最不符合条件. 故

选 A.

5. **D** 【解析】对 $y = ae^{bx+1}$ ($a > 0$) 两边同
时取自然对数, 得

$$\ln y = \ln(ae^{bx+1}) = \ln a + bx + 1,$$

令 $z = \ln y$,

则 $z = bx + \ln a + 1$, 所以 $\begin{cases} b = 2, \\ a = \ln a + 1, \end{cases}$

解得 $\begin{cases} b = 2, \\ a = 1, \end{cases}$

所以 $\frac{b}{a} = 2$. 故选 D.

6. **211** 【解析】因为 $\bar{x} = \frac{1}{4} \times (1 + 2 +$

$3 + 4) = 2.5$, $\bar{y} = \frac{1}{4} \times (197 + 193 +$

$201 + 209) = 200$, x 与 y 之间的回归直

线方程是 $\hat{y} = 4.4x + \hat{a}$,

所以 $200 = 4.4 \times 2.5 + \hat{a}$, 解得 $\hat{a} = 189$,

所以回归直线方程为 $\hat{y} = 4.4x + 189$,

当 $x = 5$ 时, $\hat{y} = 4.4 \times 5 + 189 = 211$,

所以预测第五代杂交水稻每穗的总粒

数约为 211.

7. **A** 【解析】令 $f(x) = 77.36 - 1.82x$,



因为 $f(x+1)-f(x)=77.36-1.82(x+1)-77.36+1.82x=-1.82$, 所以产量每增加 1 千件, 单位成本约下降 1.82 元, 故 A 正确, B 错误.

当 $x=1$ 时, $\hat{y}=77.36-1.82\times 1=75.54$ 为单位成本的估计值, 并不一定是实际情况, 故 C 错误, 同理 D 错误. 故选 A.

8. **B** 【解析】因为重新求得的回归直线 l 的斜率为 1.2,

故变量 x 与 y 具有正相关关系, 故 A 正确.

因为回归直线过点 (\bar{x}, \bar{y}) ,

所以 $\bar{y}=1.5\times 3+0.5=5$.

设去除两个样本点后, 新的回归直线方程为 $\hat{y}=1.2x+\hat{a}_1$,

剩余样本点横坐标的平均值为 \bar{x}_1 , 纵坐标的平均值为 \bar{y}_1 ,

$$\text{则} \begin{cases} (n-2)\bar{x}_1=n\bar{x}-(1.2+4.8), \\ (n-2)\bar{y}_1=n\bar{y}-(2.2+7.8), \end{cases}$$

$$\text{即} \begin{cases} (n-2)\bar{x}_1=3(n-2), \\ (n-2)\bar{y}_1=5(n-2), \end{cases}$$

$$\text{解得} \begin{cases} \bar{x}_1=3, \\ \bar{y}_1=5. \end{cases} \text{ 将 } (3, 5) \text{ 代入 } \hat{y}=1.2x+\hat{a}_1,$$

得 $5=1.2\times 3+\hat{a}_1$, 解得 $\hat{a}_1=1.4$, 所

以新的回归直线方程为 $\hat{y}=1.2x+1.4$,

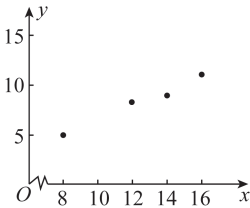
故 C 正确, B 错误.

去除两个样本点后, 样本点 $(2, 3.75)$

的随机误差为 $3.75-(1.2\times 2+1.4)=-0.05$, 故 D 正确.

故选 B.

9. 【解】(1) 根据题表中的数据作出散点图如图所示.





(2) 由散点图可得 y 与 x 线性相关, 由题中数据可得

$$\bar{x} = \frac{16 + 14 + 12 + 8}{4} = 12.5,$$

$$\bar{y} = \frac{11 + 9 + 8 + 5}{4} = 8.25,$$

$$\sum_{i=1}^4 x_i^2 = 16^2 + 14^2 + 12^2 + 8^2 = 660,$$

$$\sum_{i=1}^4 x_i y_i = 16 \times 11 + 14 \times 9 + 12 \times 8 + 8 \times 5 = 438,$$

$$\therefore \hat{b} = \frac{438 - 4 \times 12.5 \times 8.25}{660 - 4 \times 12.5^2} \approx 0.73,$$

$$\hat{a} = 8.25 - 0.73 \times 12.5 = -0.875.$$

$$\therefore \hat{y} = 0.73x - 0.875.$$

(3) 令 $0.73x - 0.875 = 10$, 解得 $x \approx 14.9$,

故机器的转速应控制在 14 转 / 秒内.

- 10. BCD** 【解析】回归直线是由点拟合而成, 可能不过任何一个样本点, 但必过数据的中心点, A 错误, B 正确; 由回归直线的定义知, 它是表示最接近 y 与 x 之间真实关系的一条直线, C 正确; 由相关系数的定义知 D 正确. 故选 BCD.

4.3 独立性检验

题型诀

- 1-1. B** 【解析】当 ad 与 bc 差距越大时, 两个变量有关系的可能性就越大. 检验四个选项中所给的 ad 与 bc 的差距.

A 中, $ad - bc = 10 - 12 = -2$;

B 中, $ad - bc = 20 - 9 = 11$;

C 中, $ad - bc = 15 - 12 = 3$;

D 中, $ad - bc = 15 - 12 = 3$.

显然 B 中 $|ad - bc|$ 最大. 故选 B.

- 1-2. 有** 【解析】因为在 20 至 40 岁的 58 名观众中有 18 名观众收看新闻节目, 而大于 40 岁的 42 名观众中有 27 名观众收看新闻节目,

$$\frac{b}{a+b} = \frac{18}{58} \approx 0.3103,$$



$$\frac{d}{c+d} = \frac{27}{42} \approx 0.6429, \text{ 两者相差较大, 所以}$$

可以在某种程度上认为收看新闻节目与年龄有关系.

2-1. AC 【解析】 经计算 $\chi^2 =$

$$\frac{100 \times (64 \times 10 - 16 \times 10)^2}{80 \times 20 \times 74 \times 26} \approx 7.4844.$$

对于 A, 该市一天空气中 PM2.5 浓度不超过 $75 \mu\text{g}/\text{m}^3$, 且 SO_2 浓度不超过

$150 \mu\text{g}/\text{m}^3$ 的概率估计值是 $\frac{64}{100} = 0.64$,

故 A 正确;

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

$$\frac{1000 \times (640 \times 100 - 160 \times 100)^2}{800 \times 200 \times 740 \times 260} \approx 74.844 \neq$$

7.4844, 故 B 错误;

对于 C, $\because 7.4844 > 6.635$, \therefore 有超过 99% 的把握认为该市一天空气中 PM2.5 浓度与 SO_2 浓度有关, 故 C 正确;

对于 D, 在犯错的概率不超过 1% 的条件下, 认为该市一天空气中 PM2.5 浓度与 SO_2 浓度有关, 故 D 错误. 故选 AC.

2-2. ②③④ 【解析】 对于①, χ^2 的值越

大, 说明有更大的把握认为 X 与 Y 有关系, 但却不能判断其相关性大小, 故①错误,

对于②, 由临界值数据可知 χ^2 的值越小, “ X 与 Y 有关系” 的可信程度越小, 故②正确;

对于③, $\chi^2 = 4 > 3.841$, 所以有 95% 的把握认为 “ X 与 Y 有关系”, 即 “ X 与 Y 有关系” 的推断出现错误的概率不会超过 0.05, 故③正确;

对于④, 在 2×2 列联表中, 若每个数据变为原来的 3 倍, 则

$$\chi^2 = \frac{3n(3a \times 3d - 3b \times 3c)^2}{(3a+3b)(3c+3d)(3a+3c)(3b+3d)} =$$

$$3 \times \frac{n(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)},$$



变为原来的 3 倍,所以④正确.

故答案为②③④.

2-3. 【解】(1) 由已知可得调查中男生共有 80 人,女生有 80 人,其中喜欢阅读古典文学的女生有 60 人,故 2×2 列联表为

	喜欢	不喜欢	合计
男生	40	40	80
女生	60	20	80
合计	100	60	160

$$\therefore \chi^2 = \frac{160 \times (40 \times 20 - 60 \times 40)^2}{100 \times 60 \times 80 \times 80} \approx 10.667 >$$

7.879.

故能在犯错误概率不超过 0.005 的情况下认为学生喜欢阅读古典文学与性别有关系.

(2) 由题意可知, ξ 的所有可能取值为 2, 3, 4, 5, 6.

$$\text{则 } P(\xi=2) = \frac{C_6^2 \cdot C_4^4}{C_{10}^6} = \frac{15}{210} = \frac{1}{14},$$

$$P(\xi=3) = \frac{C_6^3 \cdot C_4^3}{C_{10}^6} = \frac{80}{210} = \frac{8}{21},$$

$$P(\xi=4) = \frac{C_6^4 \cdot C_4^2}{C_{10}^6} = \frac{90}{210} = \frac{3}{7},$$

$$P(\xi=5) = \frac{C_6^5 \cdot C_4^1}{C_{10}^6} = \frac{24}{210} = \frac{4}{35},$$

$$P(\xi=6) = \frac{C_6^6 \cdot C_4^0}{C_{10}^6} = \frac{1}{210}.$$

$\therefore \xi$ 的分布列为

ξ	2	3	4	5	6
P	$\frac{1}{14}$	$\frac{8}{21}$	$\frac{3}{7}$	$\frac{4}{35}$	$\frac{1}{210}$

$$\therefore E(\xi) = 2 \times \frac{1}{14} + 3 \times \frac{8}{21} + 4 \times \frac{3}{7} + 5 \times \frac{4}{35} + 6 \times$$

$$\frac{1}{210} = \frac{756}{210} = \frac{18}{5}.$$

巩固练

1. C 【解析】 本题探究性别对判断“人机大战是人类的胜利”是否有影响,符合独立性检验的基本思想,因此,本题



所选择的统计方法是独立性检验,故选 C.

2. **D** 【解析】对于 A, $P(\chi^2 \geq 6.635) \approx 0.01$, 故有 99% 的把握认为两个分类变量有关系, 即 A 正确;
- 对于 B, χ^2 越大, “X 与 Y 有关系”可信程度越大, 即 B 正确;
- 对于 C, 随机误差是各估计值与实际值之间的误差, 所以随机误差平方和越小, 回归效果越好, 即 C 正确;
- 对于 D, 在回归直线方程中, 当自变量 $x=200$ 时, 因变量 y 的平均值是 15, 不能说一定是 15, 即 D 不正确.
- 故选 D.

3. **30 0.050** 【解析】补全 2×2 列联表如表所示.

	被某病毒感染	未被某病毒感染	合计
注射疫苗	10	40	50
未注射疫苗	20	30	50
合计	30	70	100

所以 $a=30$.

因为 $\chi^2 = \frac{100 \times (10 \times 30 - 40 \times 20)^2}{50 \times 50 \times 30 \times 70} \approx$

4.762, 又 $4.762 > 3.841$, $4.762 <$

5.024, 所以在犯错误的概率最多不超过 0.050 的前提下, 可认为“给基因编辑小鼠注射该种疫苗能起到预防此病毒感染的效果”.

4. **A** 【解析】设被调查的学生中男生的人数为 x , 则由题可得列联表如下:

	男生	女生	合计
喜欢航天	$\frac{4}{5}x$	$\frac{3}{5}x$	$\frac{7}{5}x$
不喜欢航天	$\frac{1}{5}x$	$\frac{2}{5}x$	$\frac{3}{5}x$
合计	x	x	$2x$



$$\chi^2 = \frac{2x \cdot \left(\frac{4}{5}x \cdot \frac{2}{5}x - \frac{1}{5}x \cdot \frac{3}{5}x \right)^2}{x \cdot x \cdot \frac{7}{5}x \cdot \frac{3}{5}x} = \frac{2x}{21},$$

因为有 95% 的把握认为是否喜欢航天与学生性别有关,

所以 $\frac{2x}{21} \geq 3.841$, 解得 $x \geq 40.3305$, 所以

被调查的学生中男生的人数不可能为 25. 故选 A.

5. 【解】(1) 列联表补充完整如下:

性别	参与意愿		合计
	愿意参与	不愿意参与	
男性	48	12	60
女性	22	18	40
合计	70	30	100

假设 H_0 : 参与意愿与性别无关系, 根据

列联表的数据可得, $\chi^2 =$

$$\frac{100 \times (48 \times 18 - 22 \times 12)^2}{60 \times 40 \times 70 \times 30} = \frac{50}{7} \approx 7.143 >$$

6.635,

对照附表, 依据独立性检验, 我们推断

H_0 不成立, 所以认为参与意愿与性别

有关系, 此推断犯错的概率不大于

0.01.

(2) X 的可能取值为 0, 1, 2, 3,

$$P(X=0) = \frac{C_4^0 C_3^3}{C_7^3} = \frac{1}{35},$$

$$P(X=1) = \frac{C_4^1 C_3^2}{C_7^3} = \frac{12}{35},$$

$$P(X=2) = \frac{C_4^2 C_3^1}{C_7^3} = \frac{18}{35},$$

$$P(X=3) = \frac{C_4^3 C_3^0}{C_7^3} = \frac{4}{35}.$$

所以 X 的分布列为

X	0	1	2	3
P	$\frac{1}{35}$	$\frac{12}{35}$	$\frac{18}{35}$	$\frac{4}{35}$

根据超几何分布的数学期望有 $E(X) =$

$$\frac{4}{7} \times 3 = \frac{12}{7}.$$



6. AC 【解析】该地区后半夜下雨的概率

率约为 $\frac{60}{100} = \frac{3}{5}$, 则 A 正确.

该地区未出现“日落云里走”时, 后半夜

下雨的概率为 $\frac{15}{40} = \frac{3}{8}$, 则 B 不正确.

假设 H_0 : 该地区“日落云里走”是否出现与后半夜是否下雨无关. 因为 $\chi^2 \approx 14.06 > 10.828$, 所以根据小概率值 $P(\chi^2 \geq x_0) = 0.001$ 的独立性检验, 我们推断 H_0 不成立, 即认为该地区“日落云里走”是否出现与后半夜是否下雨有关, 此推断犯错误的概率不大于 0.001, 则 C 正确, D 不正确. 故选 AC.