

回归直线独立性检验综合

一、单选题

1. 下列说法中，错误的有（ ）

- ①回归直线 $\hat{y} = \hat{b}x + \hat{a}$ 恒过点 $(\bar{x}, \bar{y})$ ，且至少过一个样本点；
- ②根据 2×2 列联表中的数据计算得出 $\chi^2 \geq 6.635$ ，而 $P(\chi^2 \geq 6.635) \approx 0.01$ ，则有 99% 的把握认为两个分类变量有关系，即有 1% 的可能性使得“两个分类变量有关系”的推断出现错误；
- ③在做回归分析时，可以用决定系数 R^2 刻画模型的回归效果，若 R^2 越大，则说明模型拟合的效果越好；
- ④某项测量结果 ξ 服从正态分布 $N(1, \sigma^2)$ ，若 $P(\xi \leq 5) = 0.81$ ，则 $P(\xi \leq -3) = 0.19$

A. 1 个

B. 2 个

C. 3 个

D. 4 个

【答案】A

【分析】利用回归直线的特点可判断①；利用独立性检验可判断②；利用决定系数 R^2 与模型拟合效果的关系可判断③；利用正态分布可判断④，得出合适的选项即可.

【详解】对于①，回归直线 $\hat{y} = \hat{b}x + \hat{a}$ 恒过点 $(\bar{x}, \bar{y})$ ，不一定过样本点，故①错误；

对于②，根据 2×2 列联表中的数据计算得出 $\chi^2 \geq 6.635$ ，而 $P(\chi^2 \geq 6.635) \approx 0.01$ ，

则有 99% 的把握认为两个分类变量有关系，

即有 1% 的可能性使得“两个分类变量有关系”的推断出现错误，故②正确；

对于③，在做回归分析时，可以用决定系数 R^2 刻画模型的回归效果，

若 R^2 越大，则说明模型拟合的效果越好，故③正确；

对于④，某项测量结果 ξ 服从正态分布 $N(1, \sigma^2)$ ，若 $P(\xi \leq 5) = 0.81$ ，

则 $P(\xi \leq -3) = P(\xi \geq 5) = 1 - P(\xi \leq 5) = 0.19$ ，故④正确.

可得四个命题中，错误的有 1 个，故 A 正确.

故选：A

2. 以下四个命题中错误的是（ ）

- A. 在独立性检验中，由计算得 χ^2 的值，若 χ^2 的值越大，则两个变量相关的可能性就越大

- B. 在做回归分析时，可以用决定系数 R^2 刻画模型的回归效果，若 R^2 越大，则说明模型拟合的效果越好
- C. 在回归直线方程 $\hat{y} = 2 - 3x$ 中，变量 x 每增加 1 个单位时， y 平均增加 2 个单位
- D. 若变量 y 和 x 之间的相关系数为 $r = -0.99$ ，则变量 y 和 x 之间具有很强的线性相关性，而且是负相关

【答案】C

【分析】借助独立性检验的性质、决定系数定义、回归直线方程性质与相关系数定义逐项判定即可得.

【详解】对 A：由独立性检验的性质可知， χ^2 的值越大，则两个变量相关的可能性就越大，故 A 正确；

对 B：由决定系数定义可得， R^2 越大，则说明模型拟合的效果越好，故 B 正确；

对 C：在回归直线方程 $\hat{y} = 2 - 3x$ 中，变量 x 每增加 1 个单位时， y 平均减少 3 个单位，故 C 错误；

对 D：由 $r = -0.99 < 0$ ， $|r| = 0.99$ ，接近 1，故变量 y 和 x 之间具有很强的线性相关性，而且是负相关，故 D 正确.

3. 以下结论错误的是 ()

- A. 根据 2×2 列联表中的数据计算得出 $\chi^2 \geq 6.635$ ，而 $P(\chi^2 \geq 6.635) \approx 0.01$ ，则根据小概率值 $\alpha = 0.01$ 的独立性检验，认为两个分类变量有关系
- B. 在回归直线 $\hat{y} = 0.5x - 85$ 中，变量 $x = 200$ 时，变量 y 的值一定是 15
- C. χ^2 的值越大，两个事件的相关性的可能性就越大
- D. 在回归分析中，相关指数 R^2 越大，说明残差平方和越小，回归效果越好

【答案】B

【分析】对于 AC，根据独立性检验的定义和性质进行求解；对于 BD，根据回归直线的概念和性质进行求解.

【详解】对于选项 A： $P(\chi^2 \geq 6.635) \approx 0.01$ ，故根据小概率值 $\alpha = 0.01$ 的独立性检验，认为两个分类变量有关系，即 A 正确；

对于选项 B，回归直线方程中，当变量等于 200 时， y 的值平均是 15，不能说一定是 15，故 B 错误；

对于选项 C: χ^2 越大, “ X 与 Y 有关系”可信程度越大, 即相关性的可能性就越大, 即 C 正确;

对于选项 D: 在回归分析中, 相关指数 R^2 越大, 说明残差平方和越小, 回归效果越好, 即 D 正确.

4. 下列命题错误的是 ()

- A. 设 $\xi \sim B(n, p)$, 若 $E(\xi) = 30$, $D(\xi) = 20$, 则 $n = 90$
- B. 独立性检验是对两个变量是否具有线性相关关系的一种检验
- C. 两个随机变量的线性相关性越强, 相关系数的绝对值越接近于 1
- D. 一个袋子中有 100 个大小相同的球, 其中有 40 个黄球、60 个白球, 从中不放回地随机摸出 20 个球作为样本, 用随机变量 X 表示样本中的黄球个数, 则 X 服从超几何分布, 且 $E(X) = 8$

【答案】B

【分析】根据二项分布的期望和方差公式、独立性检验的定义, 结合相关系数的性质和超几何分布的定义和期望公式逐一判断即可.

【详解】A: 因为 $\xi \sim B(n, p)$,

所以由
$$\begin{cases} E(\xi) = 30 \\ D(\xi) = 20 \end{cases} \Rightarrow \begin{cases} np = 30 \\ np(1-p) = 20 \end{cases} \Rightarrow \begin{cases} p = \frac{1}{3} \\ n = 90 \end{cases}, \quad \text{所以本选项说法正确;}$$

B: 因为独立性检验是利用卡方的取值推断两个分类变量是否独立的方法, 所以本选项说法不正确;

C: 因为两个随机变量的线性相关性越强, 相关系数的绝对值越接近于 1, 所以本选项说法正确;

D: 一个袋子中有 100 个大小相同的球, 其中有 40 个黄球、60 个白球, 从中随机摸出 1 个球是黄球的概率为 $\frac{40}{100} = 0.4$.

因为从中不放回地随机摸出 20 个球作为样本, 用随机变量 X 表示样本中的黄球个数, 所以 X 服从超几何分布, 且 $E(x) = 0.4 \times 20 = 8$, 所以本选项说法正确.

故选: B

5. 在性别与吃零食这两个分类变量的计算中, 下列说法正确的是 ()

①若 χ^2 的观测值为 $k = 6.635$, 我们有 99% 的把握认为吃零食与性别有关系, 那么在 100 个

吃零食的人中必有 99 人是女性；

②从独立性检验可知有 99% 的把握认为吃零食与性别有关系时，我们说某人吃零食，那么此人是女性的可能性为 99%；

③若从统计量中求出有 99% 的把握认为吃零食与性别有关系，是指有 1% 的可能性使得出的判断出现错误.

A. ①②

B. ①③

C. ②③

D. ③

【答案】D

【分析】由独立性检验相关概念可得答案.

【详解】①若 χ^2 的观测值为 $k > 6.635$ ，我们有 99% 的把握认为吃零食与性别有关系，那么在 100 个吃零食的人中必有 99 人是女性，故①不正确；

②独立性检验是用来考察两个分类变量是否具有关联性，并且能较精确地给出这种判断的可靠程度，

而不是给出事件的概率，故②不正确；

③若从统计量中求出有 99% 的把握认为吃零食与性别有关系，是指有 1% 的可能性使得出的判断出现错误，③正确。

故选：D

6. 为了验证牛的毛色（黑色、红色）和角（有角、无角）这两对相对性状是否相关，某学院进行了一次数据统计，根据形成的 2×2 列联表，计算得到 $\chi^2 \approx 2.786$ ，根据小概率值

$\alpha = 0.05$ 的独立性检验（已知 χ^2 独立性检验中 $x_{0.05} = 3.841$ ），下列结论正确的是（ ）

A. 牛的毛色与角无关

B. 牛的毛色与角无关，此推断犯错误的概率不超过 0.05

C. 牛的毛色与角有关

D. 牛的毛色与角有关，此推断犯错误的概率不超过 0.05

【答案】A

【分析】根据卡方独立性检验规则，比较 $\chi^2 \approx 2.786$ 与临界值 3.841 即可得出结论.

【详解】因为 $\chi^2 \approx 2.786 < 3.841$ ，所以牛的毛色与角无关.

故选：A.

7. 下列说法正确的个数为（ ）

①根据具有线性相关关系的两个变量的统计数据所得的回归直线方程为 $\hat{y} = \hat{b}x + \hat{a}$ ，若 $\hat{b} = 2$ ， $\bar{x} = 1$ ， $\bar{y} = 3$ ，则 $\hat{a} = 1$ ；

②分类变量 A 与 B 的统计量 χ^2 越大，说明“ A 与 B 有关系”的可信度越小；

③一元线性回归模型中，如果相关系数 $r = 0.98$ ，表明两个变量的相关程度很强.

A. 0 B. 1 C. 2 D. 3

【答案】C

【分析】根据回归直线的性质判断①，根据 χ^2 的意义判断②，根据相关系数的概念判断③.

【详解】对于①，根据回归直线必过 $(\bar{x}, \bar{y})$ 得 $3 = 2 \times 1 + \hat{a}$ ，解得 $\hat{a} = 1$ ，故①正确；

对于②，根据统计量 χ^2 的意义可知， χ^2 越大，说明“ A 与 B 有关系”的可信度越大，故②错误；

对于③，一元线性回归模型中，相关系数的绝对值越接近 1，表明两个变量的相关性越强，所以如果相关系数 $r = 0.98$ ，表明两个变量的相关程度很强，故③正确.

故选：C.

8. 下列结论中，正确的选项个数是 ()

(1)对具有线性相关关系的变量 x, y ，且回归方程为 $y = 0.3x - m$ ，若样本点的中心为 $(m, 2.8)$ ，则实数 m 的值是 -4

(2) 若随机变量 $\xi \sim N(1, \sigma^2)$ ， $P(\xi \leq -2) = 0.21$ ，则 $P(\xi \leq 4) = 0.79$

(3) 若随机变量 ξ, η ，满足 $\eta = 3\xi - 2$ ，则 $E(\eta) = 3E(\xi) - 2$ ， $D(\eta) = 9D(\xi) - 4$

(4)根据分类变量 X 与 Y 成对样本数据，计算得到 $\chi^2 = 9.632$ ，依据小概率值 $\alpha = 0.001$ 的 χ^2 独立性检验 ($x_{0.001} = 10.828$)，可判断 X 与 Y 有关联，此推断犯错误的概率不大于 0.001

A. 1 个 B. 2 个 C. 3 个 D. 4 个

【答案】B

【分析】(1) 运用回归方程性质，样本点中心在回归直线上，利用此性质列方程求解参数；

(2) 运用正态分布特点，正态曲线有对称轴，对称区间概率相等，据此计算概率；

(3) 运用随机变量期望和方差的运算性质，根据公式计算随机变量线性变换后的期望和方差；

(4) 运用独立性检验知识, 通过比较 χ^2 值与临界值判断两个变量是否有关联。

【详解】(1) 已知回归方程 $y = 0.3x - m$, 样本点的中心 $(m, 2.8)$ 一定在回归直线上, 将样本点中心代入回归方程可得: $2.8 = 0.3m - m$,
解得: $m = -4$;

(2) 因为随机变量 $\xi \sim N(1, \sigma^2)$, 所以正态曲线关于 $x = 1$ 对称
 -2 与 4 关于 $x = 1$ 对称, 所以 $P(\xi \leq -2) = P(\xi \geq 4) = 0.21$

那么 $P(\xi \leq 4) = 1 - P(\xi \geq 4) = 0.79$;

(3) 根据期望与方差的性质 $E(a\xi + b) = aE(\xi) + b$, $D(a\xi + b) = a^2D(\xi)$ (a, b 为常数)
可得: $E(\eta) = 3E(\xi) - 2$, $D(\eta) = 9D(\xi)$;

(4) 因为 $\chi^2 = 9.632 < x_{0.001} = 10.828$, 所以不能判断 X 与 Y 有关系.

综上所述: (1)、(2) 正确,

故选: B

9. 下列说法中, 正确的个数是 ()

(1) 两个随机变量的线性相关程度越强, 相关系数的绝对值越接近于 1

(2) 两个变量的 2×2 列联表中, 对角线上数据的乘积相差越大, 说明两个变量有关系成立的可能性就越大

(3) 从独立性检验可知: 有 99% 的把握认为吸烟与患肺癌有关, 是指在犯错误的概率不超过 1% 的前提下认为吸烟与患肺癌有关

A. 0 个 B. 1 个 C. 2 个 D. 3 个

【答案】D

【分析】根据相关系数的概念, 可知 (1) 正确; χ^2 观测值越大, 两个变量有关系的可能性越大, 所以 (2) 正确; 根据独立性检验的解释, 可得 (3) 正确.

【详解】对于 (1), 根据相关系数的概念, 可知两个随机变量的线性相关程度越强, 相关系数的绝对值越接近于 1, 故 (1) 正确;

对于 (2), 两个变量的 2×2 列联表中, 对角线上数据的乘积相差越大,

则 χ^2 观测值越大, 两个变量有关系的可能性越大, 所以 (2) 正确;

对于 (3), 从独立性检验可知: 有 99% 的把握认为吸烟与患肺癌有关,

是指在犯错误的概率不超过1%的前提下认为吸烟与患肺癌有关，是独立性检验的解释，所以（3）正确.

故选：D.

10. 下列说法中，不正确的是（ ）

- A. 在 1, 3, 6, 7, 9, 10, 12, 15 这组数据中，第 50 百分位数为 8
- B. 分类变量 A 与 B 的统计量 χ^2 越大，说明“ A 与 B 有关系”的可信度越大
- C. 已知经验回归方程为 $y = \hat{b}x + 1.8$ 且 $\bar{x} = 2, \bar{y} = 20$ ，则 $\hat{b} = 9.1$
- D. 两个模型中，残差平方和越大的模型拟合的效果越好

【答案】D

【分析】求出数据的第 50 百分位数，判断 A；根据 χ^2 的意义，判断 B；根据线性回归方程必过 $(\bar{x}, \bar{y})$ ，判断 C；根据残差平方和的意义判断 D.

【详解】对于 A，这组数据的第 50 百分位数为 $\frac{7+9}{2} = 8$ ，故 A 正确；

对于 B，根据统计量 χ^2 的意义可知 B 正确，故 B 正确；

对于 C，根据线性回归方程必过 $(\bar{x}, \bar{y})$ 得： $20 = \hat{b} \cdot 2 + 1.8 \Rightarrow \hat{b} = 9.1$ ，故 C 正确；

对于 D，由残差平方和越小，模型拟合效果越好，故 D 错误；

故选：D.

11. 下列命题中错误的是（ ）

- A. 在回归分析中，相关系数 r 的绝对值越大，两个变量的线性相关性越强
- B. 若变量 y 与 x 之间存在线性相关关系，且根据最小二乘法得到的经验回归方程为 $\hat{y} = 2x + \hat{a}$ ，样本点中心为 $(3, 6.5)$ ，则样本点 $(2.5, 7)$ 的残差为 1.5
- C. 在回归分析中，残差平方和越小，模型的拟合效果越好
- D. 对分类变量 X 与 Y ，它们的随机变量 K^2 的观测值 k 越小，说明“ X 与 Y 有关系”的把握越大

【答案】D

【分析】根据相关变量的回归分析，以及独立性检验的相关知识，即可判断选项.

【详解】A. 由相关系数 r 的意义可知，相关系数 r 的绝对值越大，越接近于 1，两个变量的线性相关性越强，故 A 说法正确；

B. 由条件可知， $6.5 = 2 \times 3 + \hat{a}$ ，得 $\hat{a} = 0.5$ ，即 $\hat{y} = 2x + 0.5$ ， $x = 2.5$ 时， $\hat{y} = 5.5$ ，

所以样本点(2.5,7)的残差为 $7-5.5=1.5$ ，故 B 说法正确；

C. 在回归分析中，残差平方和越小，模型的拟合效果越好，故 C 说法正确；

D. 对分类变量 X 与 Y ，它们的随机变量 K^2 的观测值 k 越小，说明“ X 与 Y 有关系”的把握越小，故 D 说法错误.

故选：D

12. 根据吸烟与患肺癌这两个分类变量的样本数据，计算得出 $\chi^2 \approx 6.816$ ，经查阅 χ^2 独立性检验的小概率值和相应的临界值 $x_{0.01} = 6.635$ ，则下列说法正确的是（ ）

- A. 在 100 个吸烟的人中就会有 99 人患肺癌
- B. 若某人吸烟，那么他有 99% 的可能患肺癌
- C. 若某人患肺癌，那么他有 99% 的可能为吸烟者
- D. 吸烟与患肺癌有关联，此推断犯错误的概率不大于 1%

【答案】D

【分析】根据给定条件，利用独立性检验的意义逐项判断即得.

【详解】由 $\chi^2 \approx 6.816 > 6.635 = x_{0.01}$ ，得吸烟与患肺癌有关联，此推断犯错误的概率不大于 1%，D 正确；

卡方检验仅说明吸烟与患肺癌两个变量间的关联性，无法量化个体情况，这两个变量间也无因果关系，ABC 错误.

故选：D

13. 下列实际问题不适合用独立性检验解决的是（ ）

- A. 不良的饮食习惯是否会导致肠胃疾病
- B. 某公司的营业额在过去 5 年逐年变化的情况
- C. 参加课外辅导能否提高学习成绩
- D. 男性和女性在职业选择偏好上是否有差异

【答案】B

【分析】根据独立性检验的基本思想，即可判断选项.

【详解】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 年的变化情况，不满足独立性检验的基本思想.

故选：B

14. 下列说法中, 错误的命题是 ()

- A. 在刻画回归模型的拟合效果时, R^2 的值越大, 说明拟合的效果越好
- B. 线性相关系数 r 越大, 两个变量的线性相关性越强, 反之, 线性相关性越弱
- C. 设随机变量 X 服从正态分布 $N(4, 2^2)$, 则 $P(X > 4) = \frac{1}{2}$
- D. 对分类变量 X 与 Y , 若计算出的 χ^2 越大, 则判断“ X 与 Y 有关系”的犯错误的概率越小

【答案】B

【分析】由 R^2 的值与回归模型的关系可判断 A 选项; 由相关系数的概念可判断 B, 利用正态分布密度曲线的对称性可判断 C 选项; 利用独立性检验可判断 D 选项.

【详解】对于 A: 在刻画回归模型的拟合效果时, R^2 的值越大, 说明拟合的效果越好, 正确;

对于 B: 线性相关系数 $|r|$ 越大, 两个变量的线性相关性越强, 反之, 线性相关性越弱, 错误;

对于 C: 由正太密度曲线的对称性可知: $P(X > 4) = \frac{1}{2}$, 正确;

对于 D: 对分类变量 X 与 Y , 若计算出的 χ^2 越大, 则判断“ X 与 Y 有关系”的犯错误的概率越小, 正确.

故选: B

15. 下列说法中, 正确的有 ()

- ①回归直线 $\hat{y} = \hat{b}x + \hat{a}$ 恒过点 $(\bar{x}, \bar{y})$, 且至少过一个样本点;
- ②根据 2×2 列联表中的数据计算得出 $\kappa^2 \geq 6.635$, 而 $P(\kappa^2 \geq 6.635) \approx 0.01$, 则有 99% 的把握认为两个分类变量有关系, 即有 1% 的可能性使得“两个分类变量有关系”的推断出现错误;
- ③ k^2 是用来判断两个分类变量是否相关的随机变量, 当 k^2 的值很小时可以推断两类变量不相关;
- ④某项测量结果 ξ 服从正态分布 $N(1, a^2)$, 若 $P(\xi \leq 5) = 0.81$, 则 $P(\xi \leq -3) = 0.19$.

- A. 1 个 B. 2 个 C. 3 个 D. 4 个

【答案】B

【分析】根据回归直线的特征即可判断①, 理解独立性检验的基本思想即可判断②, 正确把握卡方值的含义即可判断③, 利用正态曲线的对称性可判断④.

【详解】回归直线 $\hat{y} = \hat{b}x + \hat{a}$ 的性质是恒过样本点的中心 $(\bar{x}, \bar{y})$ ，但不一定会经过任何一个具体的样本点.所以说法①错误.

在独立性检验中，我们先提出一个假设 H_0 .当根据 2×2 列联表中的数据计算得出 $\kappa^2 \geq 6.635$ ，且 $P(\kappa^2 \geq 6.635) \approx 0.01$ 时，这意味着在假设 H_0 成立的条件下，出现这样的 κ^2 值是一个小概率事件.小概率事件在一次试验中几乎不可能发生，但现在却发生了，所以我们有理由拒绝假设 H_0 ，从而有 99% 的把握认为两个分类变量有关系，同时也就意味着有 1% 的可能性使得“两个分类变量有关系”的推断出现错误，所以说法②正确.

κ^2 是用于判断两个分类变量是否相关的随机变量.当 κ^2 的值很小时，只能说明我们较小的把握认为两类变量相关，但不能就此推断两类变量不相关.因为即使 κ^2 值小，也有可能是由于样本量等因素的影响，不能绝对地得出两类变量无关的结论，所以说法③错误.

已知某项测量结果 ξ 服从正态分布 $N(1, a^2)$ ，正态分布具有对称性，其对称轴为 $\mu = 1$.又因为 $\frac{5 + (-3)}{2} = 1$ ，这表明 $\xi \leq 5$ 与 $\xi \leq -3$ 关于对称轴 $\mu = 1$ 对称.根据正态分布的对称性可知， $P(\xi \leq 5)$ 与 $P(\xi \leq -3)$ 之和为 1，已知 $P(\xi \leq 5) = 0.81$ ，那么 $P(\xi \leq -3) = 1 - 0.81 = 0.19$ ，所以说法④正确.

故选：B.

16. 为了研究高中学生对乡村音乐的态度（喜欢和不喜欢两种态度）与性别的关系，运用 2×2 列联表进行独立性检验，经计算得 $\chi^2 = 7.01$ ，则认为“喜欢乡村音乐与性别有关系”的把握约为（ ）

- A. 0.1% B. 1% C. 99% D. 99.9%

【答案】C

【分析】由独立性检验相关知识可得答案.

【详解】易知 $\chi^2 = 7.01 > 6.635 = x_{0.01}$ ，对照临界值表知，有 99% 的把握认为“喜欢乡村音乐与性别有关系”.

故选：C

17. 在对吸烟与患肺病这两个分类变量的独立性检验中，下列说法正确的是（参考数据：

$$P(K^2 \geq 6.635) = 0.01) \quad ()$$

①若 K^2 的观测值满足 $K^2 \geq 6.635$ ，我们有 99% 的把握认为吸烟与患肺病有关系；

- ②若 K^2 的观测值满足 $K^2 \geq 6.635$ ，那么在 100 个吸烟的人中约有 99 人患有肺病；
- ③从独立性检验可知，如果有 99% 的把握认为吸烟与患肺病有关系时，那么我们就认为：每个吸烟的人有 99% 的可能性会患肺病；
- ④从统计量中得知有 99% 的把握认为吸烟与患肺病有关系时，是指有 1% 的可能性使推断出现错误。

A. ②③ B. ②③④ C. ①②④ D. ①④

【答案】D

【分析】由给出的数据，结合 K^2 观测值的意义判定即可。

【详解】若 K^2 的观测值满足 $K^2 \geq 6.635$ ，则我们有 99% 的把握认为吸烟与患肺病有关系，而得知有 99% 的把握认为吸烟与患肺病有关系时，仍有 1% 的可能性使推断出现错误，但不能说明 100 个吸烟的人中约有 99 人患有肺病，也不能说明每个吸烟的人有 99% 的可能性会患肺病。

故①④正确、②③错误。

故选：D

18. 下列说法错误的是（ ）

- A. 线性相关系数 $|r|$ 越接近 1，两个变量的线性相关程度越强；
- B. 独立性检验可以 100% 确定两个变量之间是否具有某种关系；
- C. 在回归分析中，可用残差图判断模型的拟合效果，残差点比较均匀地落在水平的带状区域中，说明这样的模型比较合适，带状区域的宽度越窄，说明模型的拟合精度越高；
- D. 甲、乙两个模型的决定系数 R^2 分别约为 0.88 和 0.80，则模型甲的拟合效果更好。

【答案】B

【分析】利用线性相关系数、独立性检验、残差图、决定系数等相关概念，逐一判断选项即可得出结论。

【详解】对于 A，根据线性相关系数的定义可判断 A 正确；

对于 B，独立性检验是存在某种程度的错误概率的，因此可得 B 错误；

对于 C，利用回归分析残差概念以及残差图可判断 C 正确；

对于 D，决定系数 R^2 的值越大，说明拟合效果越好，显然 $0.88 > 0.80$ ，即模型甲的拟合效果更好，可得 D 正确。

故选：B

19. 下列结论错误的是 ()

- A. 若随机变量 ξ, η 满足 $\eta = 2\xi + 1$, 则 $D(\eta) = 4D(\xi)$
- B. 数据 1, 3, 5, 7, 9, 11, 13 的第 60 百分位数为 9
- C. 用简单随机抽样的方法从 51 个个体中抽取 2 个个体, 则每个个体被抽到的概率都是 $\frac{1}{51}$
- D. 根据分类变量 X 与 Y 的成对样本数据, 计算得到 $\chi^2 = 4.712$. 依据 $\alpha = 0.05$ 的独立性检验 ($x_{0.05} = 3.841$), 可判断 X 与 Y 有关

【答案】C

【分析】A 选项由方差性质判断; B 选项应用百分位数求法判断; C 选项根据随机抽样的性质分析判断; D 选项由独立检验的基本思想判断.

【详解】若随机变量 ξ, η 满足 $\eta = 2\xi + 1$, 由方差的性质可知 $D(\eta) = 4D(\xi)$, A 选项正确;
 $7 \times 60\% = 4.2$, 数据 1, 3, 5, 7, 9, 11, 13 的第 60 百分位数为第 5 个数据 9, B 选项正确;
用简单随机抽样的方法从 51 个个体中抽取 2 个个体, 则每个个体被抽到的概率都是 $\frac{2}{51}$, C 选项错误;
由 $\chi^2 = 4.712 > x_{0.05} = 3.841$, 结合独立检验的基本思想, 在 $\alpha = 0.05$ 小概率情况下可判断 X 与 Y 有关, D 选项正确.
故选: C.

20. 下列说法正确的个数是 ()

- ①线性相关系数 $|r|$ 越接近 1, 两个变量的线性相关程度越强;
- ②独立性检验可以 100% 确定两个变量之间是否具有某种关系;
- ③在回归分析中, 可用残差图判断模型的拟合效果, 残差点比较均匀地落在水平的带状区域中, 说明这样的模型比较合适, 带状区域的宽度越窄, 说明模型的拟合精度越高;
- ④甲、乙两个模型的决定系数 R^2 分别约为 0.88 和 0.80, 则模型甲的拟合效果更好.

- A. 1 B. 2 C. 3 D. 4

【答案】C

【分析】根据线性相关系数, 独立性检验, 残差图及决定系数的概念分别判断即可.

【详解】线性相关系数 $|r|$ 越接近 1, 两个变量的线性相关程度越强, 故①正确;

独立性检验并不能 100%确定两个变量之间是否具有某种关系，故②错误；

回归分析中，可用残差图判断模型的拟合效果，残差点比较均匀地落在水平的带状区域中，说明这样的模型比较合适，带状区域的宽度越窄，说明模型的拟合精度越高，故③正确；

回归分析中，可用 R^2 判断模型的拟合效果， R^2 越大，模型的拟合效果越好，故④正确；

故选：C.