



## 第9章 统计

### 9.1 线性回归分析

#### 9.1.1 变量的相关性

1. **D** 【解析】选项 A, B 中两个变量间是函数关系; 选项 C 中两个变量之间没有什么关系;

选项 D 中, 学习成绩与平均学习时间有关, 但不仅与时间有关, 还与其他变量有关, 如学习时的专注性, 个人的学习习惯等, 因此 D 中两个变量是相关关系. 故选 D.

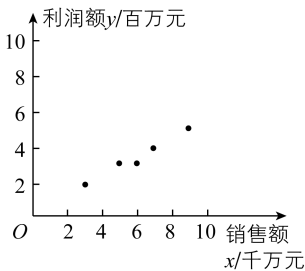
2. **C** 【解析】A 中的散点无规律可言, 看不出两个变量有什么相关性;

B 中两个变量具有正相关关系;

C 中两个变量具有负相关关系;

D 中两个变量具有相关性, 但既不是正相关, 也不是负相关. 故选 C.

3. 【解】根据该连锁经营公司的 5 个零售店某月的销售额和利润额资料画出散点图如图所示.



从图中可以看出, 这两个变量具有线性相关关系, 且是正相关.

4. **A** 【解析】由题设  $1 > |r_1| > |r_4| > |r_2| > |r_3| > 0$ , 则线性相关程度最强的是 A 组成对样本数据 (提示:  $|r|$  越接近 1, 相关性越强, 易直接错选成  $r$  较大组). 故选 A.

5. 【解】(1) 样本中 10 个这种零件的横

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

样本中 10 个这种零件的耗材量的平



$$\text{均值 } \bar{y} = \frac{3.90}{10} = 0.39,$$

据此可估计刘铭同学制作的这种零件平均每个的横截面面积为  $0.052 \text{ mm}^2$ , 平均一个零件的耗材量为  $0.39 \text{ mm}^3$ .

(2) 样本相关系数  $r =$

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

(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,

$$\text{可得 } \frac{0.052}{0.39} = \frac{182}{y}, \text{ 解得 } y = 1365 \text{ mm}^3,$$

故刘铭制作的零件的总耗材量的估计值为  $1365 \text{ mm}^3$ .

**6. 【解】**(1) 由题可知  $\bar{x} = \frac{1}{5}(12+12.5+13+13.5+14) = 13$ ,  $\bar{y} = \frac{1}{5}(14+13+11+9+8) = 11$ .

(2) 因为  $\sum_{i=1}^5 (x_i - \bar{x})^2 = (12-13)^2 + (12.5-13)^2 + (13-13)^2 + (13.5-13)^2 + (14-13)^2 = 2.5$ ,

$\sum_{i=1}^5 (y_i - \bar{y})^2 = (14-11)^2 + (13-11)^2 + (11-11)^2 + (9-11)^2 + (8-11)^2 = 26$ ,

$$\begin{aligned} \text{所以 } r &= \frac{\sum_{i=1}^5 (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^5 (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^5 (y_i - \bar{y})^2}} = \\ &= -\frac{8}{\sqrt{65}} \approx -0.992. \end{aligned}$$

因为  $|r| \approx 0.992 > 0.75$ , 所以可以推断  $y$  与  $x$  的线性相关性很强.

### 9.1.2 一元线性回归模型

**1. 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.

**2. D** 【解析】由题意可得  $\bar{x} = \frac{1}{5} \times (6+7+10+12+15) = 10$ .

$\therefore$  经验回归方程为  $\hat{y} = 0.7x - 6$ ,

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

$\therefore y_1, y_2, y_3, y_4, y_5$  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.

**3. 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, 经验回归直线恒过样本点的中心, 所以经验回归直线  $\hat{y} = 0.85x - 85.71$  必过点  $(\bar{x}, \bar{y})$ , 所以 B 正确. 对于 C, 由于经验回归方程为  $\hat{y} = 0.85x - 85.71$ , 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.

**4. 【解】**(1) 当  $x = 3$  时,  $\hat{y}_1 = 10.7 \times 3 + 3.4 = 35.5$ , 所以  $M = 43 - 35.5 = 7.5$ ;

当  $x = 4$  时,  $\hat{y}_2 = 35.5 \times \sqrt{4} - 22.8 = 48.2$ , 所以  $n = 45 - 48.2 = -3.2$ .

则模型①残差值的绝对值之和为

$1.1 + 2.8 + 7.5 + 1.2 + 1.9 + 0.4 = 14.9$ ,

模型②残差值的绝对值之和为  $0.3 +$

$5.4 + 4.3 + 3.2 + 1.6 + 3.8 = 18.6$ ,

因为  $14.9 < 18.6$ , 所以模型①的拟合效果较好, 应该选模型①.

(2) 由题意剔除异常数据即第 3 天的数据后,

得  $\bar{x} = \frac{1}{5} \times (3.5 \times 6 - 3) = 3.6$ ,

$\bar{y} = \frac{1}{5} \times (41 \times 6 - 43) = 40.6$ ,

$\sum_{i=1}^5 x_i y_i = 1\,049 - 3 \times 43 = 920$ ,



$$\sum_{i=1}^5 x_i^2 = 91 - 3^2 = 82,$$

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

$$= \frac{\sum_{i=1}^5 x_i y_i - 5\bar{x}\bar{y}}{\sum_{i=1}^5 x_i^2 - 5\bar{x}^2}$$

$$= \frac{920 - 5 \times 3.6 \times 40.6}{82 - 5 \times 3.6 \times 3.6} = 11,$$

$$\hat{a} = \bar{y} - \hat{b}\bar{x} = 40.6 - 11 \times 3.6 = 1,$$

故  $y$  关于  $x$  的经验回归方程为  $\hat{y} = 11x + 1$ .

5. 【解】(1)  $\bar{x} = \frac{2+5+8+9+11}{5} = 7, \bar{y} = \frac{12+10+8+8+7}{5} = 9.$

$$(2) \sum_{i=1}^5 (x_i - \bar{x})(y_i - \bar{y}) = (-5) \times 3 + (-2) \times 1 + 1 \times (-1) + 2 \times (-1) + 4 \times (-2) = -28,$$

$$\sqrt{\sum_{i=1}^5 (x_i - \bar{x})^2} = \sqrt{25+4+1+4+16} = 5\sqrt{2},$$

$$\sqrt{\sum_{i=1}^5 (y_i - \bar{y})^2} = \sqrt{9+1+1+1+4} = 4,$$

$$\text{代入公式得, } r = \frac{-28}{5\sqrt{2} \times 4} \approx \frac{-7}{7.07} \approx -0.99.$$

$$(3) \text{ 由于 } R^2 = 1 - \frac{\sum_{i=1}^5 (y_i - \hat{y}_i)^2}{\sum_{i=1}^5 (y_i - \bar{y})^2} = 1 -$$

$$\frac{0.32}{16} = 0.98, \text{ 故响应变量 } y \text{ 的差异有}$$

98% 由解释变量  $x$  引起.

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

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

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

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

$$\hat{d} = \frac{\sum_{i=1}^{20} w_i y_i - 20\bar{w}\bar{y}}{\sum_{i=1}^{20} w_i^2 - 20\bar{w}^2}$$



$$= \frac{272.1 - 20 \times 2.6 \times 5.12}{137 - 20 \times 2.6^2}$$

$$\approx 3.26,$$

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

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 吨.

## 9.2 独立性检验

**1. B** 【解析】当  $ad$  与  $bc$  差距越大时, 两个事件有关的可能性就越大. 检验四个选项中所给的  $ad$  与  $bc$  的差距.

$$A: ad - bc = 10 - 12 = -2;$$

$$B: ad - bc = 20 - 9 = 11;$$

$$C: ad - bc = 15 - 12 = 3;$$

$$D: ad - bc = 15 - 12 = 3.$$

显然 B 中  $|ad - bc|$  最大. 故选 B.

**2. D** 【解析】因为  $\chi^2 = 7.233 > 6.635$ ,

所以在犯错误的概率不超过 0.01 的前提下, 认为“该校高中生爱好数学与性别有关”. 故选 D.

**3. 【解】**(1) 甲学校竞赛成绩优秀的频率

$$\text{为 } \frac{60}{100} = \frac{3}{5},$$

乙学校竞赛成绩优秀的频率为

$$\frac{70}{100} = \frac{7}{10}.$$

(2) 由列联表可得  $\chi^2 =$

$$\frac{200 \times (60 \times 30 - 40 \times 70)^2}{100 \times 100 \times 130 \times 70} = \frac{200}{91} \approx 2.198 <$$

$$3.841,$$

故没有 95% 的把握认为甲校成绩优秀与乙校成绩优秀有差异.