

用考点来诠释回归分析的基本思想及其初步应用

◆程媛

(陕西省洋县中学)

回归分析的基本思想及其初步应用不仅是高考的考点要求,而是经济社会生活管理的需要,下面我们借助考点来认识一下它的魅力所在。

考点1: 线性回归方程

例1. 一台机器使用时间较长,但还可以使用. 它按不同的转速生产出来的某机械零件会有一些有缺点,每小时生产有缺点零件的多少,随机器运转的速度而变化,下表为抽样试验结果:

转速 x (转/秒)	16	14	12	8
每小时生产有缺点的零件数 y (件)	11	9	8	5

(1)对变量 y 与 x 进行相关性检验;

(2)如果 y 与 x 有线性相关关系,求回归直线方程.

解析: (1) $\bar{x} = 12.5$, $\bar{y} = 8.25$, $\sum_{i=1}^4 x_i y_i = 438.4$, $\bar{x} \bar{y} = 412.5$,
 $\sum_{i=1}^4 x_i^2 = 660$, $\sum_{i=1}^4 y_i^2 = 291$,

$$\text{所以 } r = \frac{\sum_{i=1}^4 x_i y_i - 4 \bar{x} \bar{y}}{\sqrt{\left(\sum_{i=1}^4 x_i^2 - 4 \bar{x}^2\right) \left(\sum_{i=1}^4 y_i^2 - 4 \bar{y}^2\right)}} = \frac{438 - 412.5}{\sqrt{(660 - 625) \times (291 - 272.25)}} = \frac{25.5}{\sqrt{656.25}} = \frac{25.5}{25.62} \approx 0.9953.$$

因为 $r > 0.75$, 所以 y 与 x 有很强的线性相关关系.

$$(2) \text{由(1)知: } b = \frac{\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y}}{\sum_{i=1}^n x_i^2 - n \bar{x}^2} = \frac{438 - 412.5}{660 - 4 \times 12.5^2} \approx 0.7286, a = \bar{y} - b \bar{x} = -0.8575. \therefore \text{回归直线方程为 } y = 0.7286x - 0.8575.$$

点拨: 这是一个回归分析实际应用问题,应先进行线性相关性检验或作散点图来判断 x 与 y 是否线性相关,如果线性相关,才可以求解回归直线方程才有意义.

考点2: 相关系数 r 的应用

例2. 有5名学生的数学和化学成绩如下表所示:

学生					
学科	A	B	C	D	E
数学成绩(x)	88	76	73	66	63
化学成绩(y)	78	65	71	64	61

(1)判断 y 与 x 是否具有相关关系; (2)如果 y 与 x 具有相关关系,求线性回归方程; (3)预测如果某学生数学成绩为79分,他的化学成绩为多少?

$$\text{解析: (1)} \because \bar{x} = \frac{1}{5} \times (88 + 76 + 73 + 66 + 63) = 73.2, \bar{y} = \frac{1}{5} \times (78 + 65 + 71 + 64 + 61) = 67.8, \sum_{i=1}^5 x_i^2 = 88^2 + 76^2 + 73^2 + 66^2 + 63^2 = 27174, \sum_{i=1}^5 y_i^2 = 78^2 + 65^2 + 71^2 + 64^2 + 61^2 = 23167,$$

$$\sum_{i=1}^5 x_i y_i = 88 \times 78 + 76 \times 65 + 73 \times 71 + 66 \times 64 + 63 \times 61 = 25054, \therefore \sum_{i=1}^5 x_i^2 - 5 \bar{x}^2 = 27174 - 5 \times 73.2^2 = 382.8, \sum_{i=1}^5 x_i y_i - 5 \bar{x} \bar{y}$$

$$= 25054 - 5 \times 73.2 \times 67.8 = 239.2, \sum_{i=1}^5 y_i^2 - 5 \bar{y}^2 = 23167 - 5 \times 67.8^2 = 182.8. \therefore r = \frac{239.2}{\sqrt{382.8 \times 182.8}} \approx 0.904. \therefore y \text{ 与 } x \text{ 之间具有线性相关关系.}$$

(2)设 y 与 x 的线性回归方程为 $y = bx + a$,

$$\therefore b = \frac{\sum_{i=1}^5 x_i y_i - 5 \bar{x} \bar{y}}{\sum_{i=1}^5 x_i^2 - 5 \bar{x}^2} = \frac{239.2}{382.8} \approx 0.625, a = \bar{y} - b \bar{x} = 67.8 - 0.625 \times 73.2 = 22.050, \therefore \text{线性回归方程为 } y = 22.050 + 0.625x.$$

(3)当 $x = 79$ 时, $y = 22.050 + 0.625 \times 79 = 71.425$,

即当某同学的数学成绩为79分时,他的化学成绩约为71分.

点拨: 在研究两个变量之间的关系时,应先进行相关性检验,若具备线性相关关系再求回归方程;如果本身两个变量不具备线性相关关系,即使求出线性回归方程也是毫无意义的,而且用其估计和预测的量也是不可信的.

考点3: 可线性化的回归分析

例3. 某地区的女性在不同年龄段的身高平均值 x (单位: cm) 和体重平均值 y (单位: kg) 的数据如下表:

身高 x /cm	60	70	80	90	100	110
体重 y /kg	6.13	7.90	9.99	12.15	15.02	17.50
身高 x /cm	120	130	140	150	160	170
体重 y /kg	20.92	26.85	31.11	38.85	47.25	55.05

(1)试建立 y 与 x 之间的回归方程; (2)若体重超过相同身高的女性体重平均值的1.2倍为肥胖,低于0.8倍为偏瘦,那么这个地区一名身高175 cm、体重82 kg的在校女生的体重是否正常?

解析: (1)根据上表中的数据

画出散点图如图(1)所示.

由图(1)可看出,样本点分布

在某条类似指数函数曲线 $y = e^{c_1 x} + c_2 x$ 的周围,其中 c_1 和 c_2 是待定参数,令 $z = \ln y$, 变换后的样本数据表如下:

x	60	70	80	90	100	110	120	130	140	150	160	170
z	1.81	2.07	2.30	2.50	2.71	2.86	3.04	3.29	3.44	3.66	3.86	4.01

作出散点图如图(2)所示.

设 z 与 x 之间的线性回归模型为 $z = a + bx$,

则由表中数据得 $b \approx 0.020$, $a = \bar{z} - b \bar{x} \approx 0.662$, 所以 z 与 x

之间的线性回归方程为 $z = 0.662 + 0.020x$, 所以 $y = e^{0.662 + 0.020x}$.

(2)当 $x = 175$ cm 时, 预测平均体重 $y = e^{0.662 + 0.020 \times 175} \approx 64.20$ (kg),

由于 $64.20 \times 1.2 = 77.04 < 82$, 所以这位女生肥胖.

点拨: 可线性化的回归分析问题, 画出已知数据的散点图, 选择跟散点拟合较好的函数模型, 如果散点图中的点分布在一个直线状带形区域, 可以选线性回归模型来建模; 如果散点图中的点分布在一个曲线状带形区域, 就需选择非线性回归模型来建模, 然后通过变量代换转化为新的线性回归方程, 最后再回代得到非线性回归方程.