

第四节 变量间的相关关系、统计案例



变量间的相关关系、统计案例

1. 变量间的相关关系

- (1) 会作两个有关联变量的数据的散点图，会利用数点图认识变量间的相关关系。
- (2) 了解最小二乘法的思想，能根据给出的线性回归方程系数公式建立线性回归方程。

2. 统计案例

了解下列一些常见的统计方法，并能应用这些方法解决一些实际问题。

(1) 独立性检验

了解独立性检验(只要求 2×2 列联表)的基本思想、方法及其简单应用。

(2) 回归分析

了解回归分析的基本思想、方法及其简单应用。

【最新考纲】

— 备考新动向



1. 变量间的相关关系

(1)常见的两变量之间的关系有两类：一类是函数关系，另一类是相关关系；与函数关系不同，相关关系是一种非确定性关系。

(2)从散点图上看，点分布在从左下角到右上角的区域内，两个变量的这种相关关系称为正相关，点分布在左上角到右下角的区域内，两个变量的相关关系为负相关。



知识点一

知识点一

知识点二

2. 两个变量的线性相关

(1)从散点图上看，如果这些点从整体上看大致分布在通过散点图中心的一条直线附近，称两个变量之间具有线性相关关系，这条直线叫回归直线。

(2)回归方程为 $\hat{y} = \hat{b}x + \hat{a}$ ，其中 $\hat{b} = \frac{\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y}}{\sum_{i=1}^n x_i^2 - n \bar{x}^2}$ ， $\hat{a} = \bar{y} - \hat{b} \bar{x}$ 。

(3)通过求 $Q = \sum_{i=1}^n (y_i - bx_i - a)^2$ 的最小值而得出回归直线的方法，即求回归直线，使得样本数据的点到它的距离的平方和最小，这一方法叫作最小二乘法。

知识点一

知识点一

(4)相关系数:

当 $r > 0$ 时, 表明两个变量正相关;

当 $r < 0$ 时, 表明两个变量负相关.

知识点二

r 的绝对值越接近于 **1**, 表明两个变量的线性相关性越强. r 的绝对值越接近于 **0** 时, 表明两个变量之间几乎不存在线性相关关系. 通常 $|r|$ 大于 **0.75** 时, 认为两个变量有很强的线性相关性.



知识点一

知识点一

知识点二

易误提醒

1. 易混淆相关关系与函数关系，两者的区别是函数关系是一种确定的关系，而相关关系是一种非确定的关系，函数关系是一种因果关系，而相关关系不一定是因果关系，也可能是伴随关系。
2. 回归分析中易误认为样本数据必在回归直线上，实质上回归直线必过 (\bar{x}, \bar{y}) 点，可能所有的样本数据点都不在直线上。
3. 利用回归方程分析问题，所得的数据易误认为准确值，而实质上是预测值(期望值)。

知识点一

知识点一

知识点二

[自测练习]

1. 已知 x, y 的取值如下表, 从散点图可以看出 y 与 x 线性相关, 且回归方程为 $\hat{y} = 0.95x + \hat{a}$, 则 $\hat{a} =$ (**B**)

x	0	1	3	4
y	2.2	4.3	4.8	6.7

A. 3.25

B. 2.6

C. 2.2

D. 0

答案

试题

解析

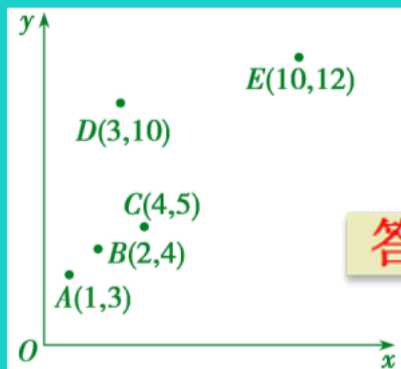
\because 回归直线必过样本点的中心 (\bar{x}, \bar{y}) , 又 $\bar{x} = 2$, $\bar{y} = 4.5$, 代入回归方程, 得 $\hat{a} = 2.6$.

知识点一

知识点一

知识点二

2. (2016·镇江模拟)如图所示,有 A, B, C, D, E 5 组 (x, y) 数据, 去掉 **D** 组数据后, 剩下的 4 组数据具有较强的线性相关关系.



答案

试题

解析

由散点图知呈带状区域时具有较强的线性相关关系, 故去掉 D .

知识点二

独立性检验

独立性检验

假设有两个分类变量 X 和 Y ，它们的取值分别为 $\{x_1, x_2\}$ 和 $\{y_1, y_2\}$ ，其样本频数列联表(称为 2×2 列联表)为：

知识点一

	y_1	y_2	总计
x_1	a	b	$a+b$
x_2	c	d	$c+d$
总计	$a+c$	$b+d$	$a+b+c+d$

知识点二

$$K^2 = \frac{n(ad-bc)^2}{(a+b)(a+c)(b+d)(c+d)} \text{ (其中 } n=a+b+c+d \text{ 为样本容量).}$$

知识点二

知识点一

知识点二

易误提醒 (1)独立性检验是对两个变量有关系的可

信程度的判断，而不是对其是否有关系的判断。

(2)独立性检验得出的结论是带有概率性质的，只能说结论成立的概率有多大，而不能完全肯定一个结论，因此才出现了临界值表。在分析问题时一定要注意这点，不可对某个问题下确定性结论，否则就可能对统计计算的结果做出错误的解释。



知识点二

[自测练习]

试题

解析

3. 下面是 2×2 列联表:

	y_1	y_2	总计
x_1	a	21	73
x_2	22	25	47
总计	b	46	120

则表中 a, b 的值分别为(C)

A. 94,72

B. 52,50

C. 52,74

D. 74,52

答案

$$\because a+21=73, \therefore a=52,$$

$$\text{又 } a+22=b, \therefore b=74.$$

知识点一

知识点二

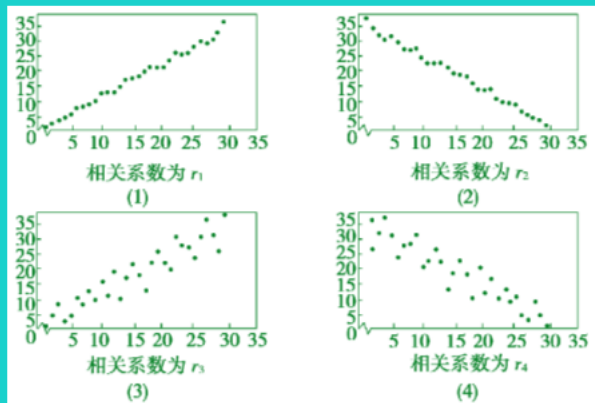
考点一

题组训练

相关关系判断

自主探究 ZIZHU TANJIU

1. 对四组数据进行统计, 获得如图所示的散点图, 关于其相关系数的比较, 正确的是(A)



A. $r_2 < r_4 < 0 < r_3 < r_1$

B. $r_4 < r_2 < 0 < r_1 < r_3$

C. $r_4 < r_2 < 0 < r_3 < r_1$

D. $r_2 < r_4 < 0 < r_1 < r_3$

答案

试题

解析

易知题中图(1)与图(3)是正相关, 图(2)与图(4)是负相关, 且图(1)与图(2)中的样本点集中分布在一条直线附近, 则 $r_2 < r_4 < 0 < r_3 < r_1$.

考点一

题组训练

2. (2015·高考湖北卷)已知变量 x 和 y 满足关系 $y = -0.1x + 1$, 变量 y 与 z 正相关. 下列结论中正确的是

(C)

- A. x 与 y 正相关, x 与 z 负相关
- B. x 与 y 正相关, x 与 z 正相关
- C. x 与 y 负相关, x 与 z 负相关
- D. x 与 y 负相关, x 与 z 正相关

答案

试题

解析

因为 $y = -0.1x + 1$, x 的系数为负, 故 x 与 y 负相关; 而 y 与 z 正相关, 故 x 与 z 负相关.

考点一

题组训练

»规律方法

相关关系的判断的两种方法

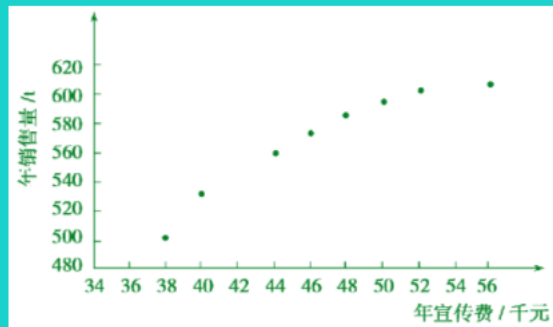
- (1)散点图法.
- (2)相关系数法：利用相关系数判定，当 $|r|$ 越趋近于1相关性越强.

考点二

回归分析 | 合作探究

HEZUO TANJIU

典例 1 (2015·高考全国卷 I) 某公司为确定下一年度投入某种产品的宣传费, 需了解年宣传费 x (单位: 千元) 对年销售量 y (单位: t) 和年利润 z (单位: 千元) 的影响. 对近 8 年的年宣传费 x_i 和年销售量 y_i ($i=1, 2, \dots, 8$) 数据作了初步处理, 得到下面的散点图及一些统计量的值.



\bar{x}	\bar{y}	\bar{w}	$\sum_{i=1}^8 (x_i - \bar{x})^2$	$\sum_{i=1}^8 (w_i - \bar{w})^2$	$\sum_{i=1}^8 (x_i - \bar{x})(y_i - \bar{y})$	$\sum_{i=1}^8 (w_i - \bar{w})(y_i - \bar{y})$
46.6	563	6.8	289.8	1.6	1 469	108.8

表中 $w_i = \sqrt{x_i}$, $\bar{w} = \frac{1}{8} \sum_{i=1}^8 w_i$.

典题悟法

演练冲关

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：<https://d.book118.com/75813105306006055>