



## 第三节 分布形态的测度

### 一、偏态系数

偏态系数取决于离差三次方的平均数与标准差三次方的比值。计算公式为：

$$SK = \frac{n}{(n-1)(n-2)} \sum_{i=1}^n \left( \frac{X_i - \bar{X}}{s} \right)^3$$



### 第三节 分布形态的测度

1. 如果偏态系数等于0，说明数据的分布是对称的；

2. 如果偏态系数为正值，说明分布为右偏的，取值在0和0.5之间说明轻度右偏，取值在0.5和1之间说明中度右偏，取值大于1说明严重右偏；

3. 如果偏态系数为负值，说明分布为左偏，取值在0和-0.5之间说明轻度左偏，取值在-0.5和-1之间说明中度左偏，取值小于-1说明严重左偏。偏态系数的绝对值越大，说明数据分布的偏斜程度越大。



## 第三节 分布形态的测度

### 二、标准分数

标准分数也称为Z分数，是统计上常用的一种标准化方法，转变后的标准分数并没有改变数值在原分布中的位置，也没有改变原分布的偏度，但是标准分数的平均数为0，标准差为1。计算方法是用数值减去均值所得的差除以标准差，计算公式为：标准分数 $Z = (\text{原始分数} - \text{平均分数}) \div \text{标准差}s$

对于服从对称的钟形分布的标准分数，68%的标准分数在 $[-1, +1]$ 范围内，约有95%的标准分数在 $[-2, +2]$ 范围之内，约有99%的标准分数在 $[-3, +3]$ 范围之内。



## 典型真题

【真题·2023单选】偏态系数可用于描述数据分布的（ ）

。

- A. 集中趋势
- B. 对称程度
- C. 陡峭程度
- D. 偏斜方向
- E. 离散程度



## 典型真题

答案：BD

解析：偏度是指数据分布的偏斜向和程度，描述的是数据分布对称程度。测度数据分布偏度的统计量称为偏态系数。



## 第四节 变量间的相关分析

### 一、变量间的相关关系

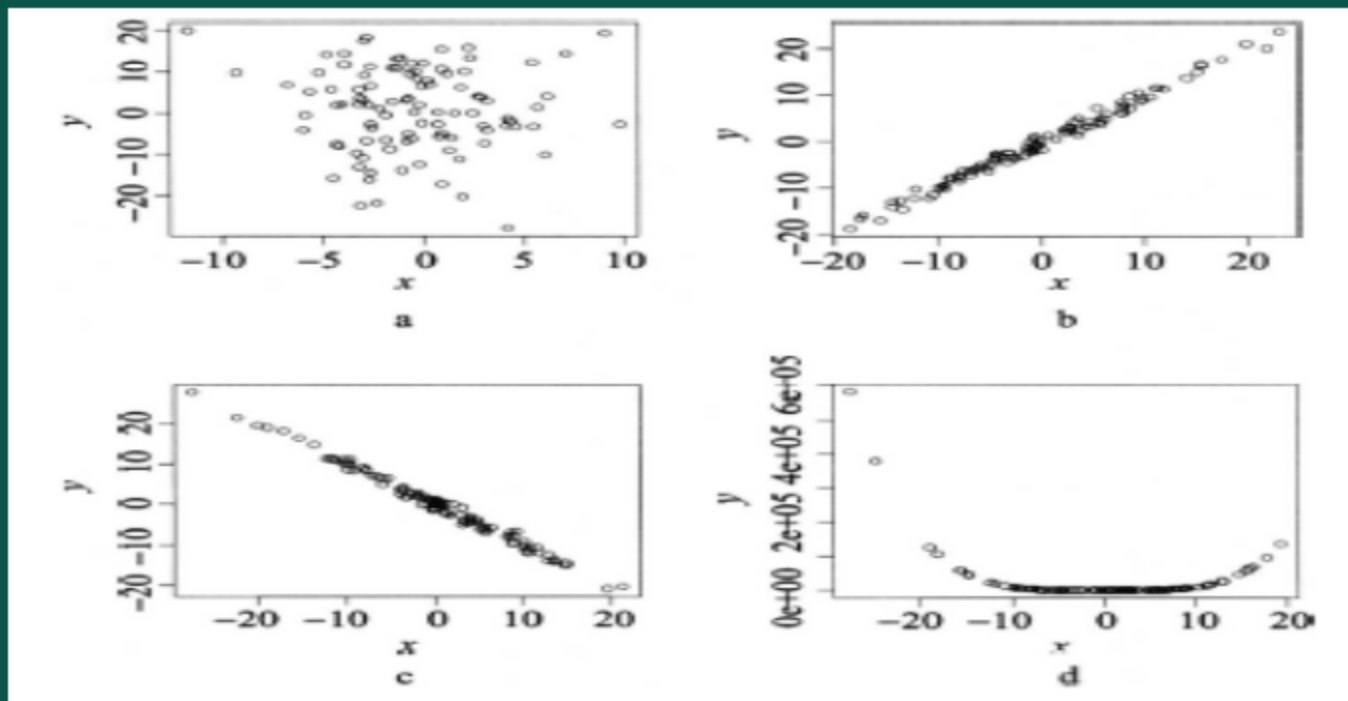
按相关的程度	完全相关	当一个变量的取值变化完全由另一个变量的取值变化所确定时
	不相关	当两个变量的取值变化彼此互不影响时
	不完全相关	当两个变量之间的关系介于完全相关和不相关之间
按相关的方向	正相关	当一个变量的取值由小变大，另一个变量的取值也相应由小变大
	负相关	当一个变量的取值由小变大，另一个变量的取值相反地由大变小
按相关的形式	线性相关	当两个相关变量之间的关系大致呈现为线性关系时
	非线性相关	如果两个相关变量之间，并不表现为直线的关系，而是近似于某种曲线方程的关系



## 第四节 变量间的相关分析

### 二、散点图

图a的点几乎无规律而言，表示这两个变量不相关，图b和图c中，观测点密集在一条直线周围，表现为较强的线性相关，但相关的方向不同。图b中的两个变量为正相关关系，图c的两个变量为负相关关系。图d中的观测点呈现出曲线模式，这表示两个变量为非线性相关。





## 第四节 变量间的相关分析

### 三、Pearson相关系数

#### 1. 计算公式

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}$$

Pearson相关系数只适用于线性相关关系的判断。





## 第四节 变量间的相关分析

### 2. 取值范围

若 $0 < r \leq 1$ ，表明变量X和Y之间存在正线性相关关系

若 $-1 \leq r < 0$ ，表明变量X和Y之间存在负线性相关关系

若 $r=1$ ，表明变量X和Y之间为完全正线性相关

若 $r=-1$ ，表明变量X和Y之间为完全负线性相关（ $|r|=1$ 时，变量Y的取值完全依赖于X）。

当 $r=0$ 时，说明X和Y之间不存在线性相关关系



## 第四节 变量间的相关分析

### 2. 取值范围

根据经验可将相关程度分为以下几种情况：

当 $|r| \geq 0.8$ 时，可视为高度相关；

$|r| < 0.8$ 时，可视为中度相关；

$0.3 \leq |r| < 0.5$ 时，视为低度相关；

$|r| < 0.3$ 时，说明两个变量之间的相关程度极弱，可视为无线性相关关系



## 典型真题

【真题·2023单选】一组数据的偏态系数为0.85，则该组数据的分布为（ ）。

- A. 中度左偏
- B. 中度右偏
- C. 严重左偏
- D. 严重右偏



## 典型真题

答案：B

解析：如果偏态系数为正值，说明分布为右偏的。（1）取值在0和0.5之间说明轻度右偏；（2）取值在0.5和1之间说明中度右偏；（3）取值大于1说明严重右偏。如果偏态系数为负值，说明分布为左偏。（1）取值在0和-0.5之间说明轻度左偏；（2）取值在-0.5和-1之间说明中度左偏；（3）取值小于-1说明严重左偏。



# 本章内容总结

## 第24章 描述统计

### 集中趋势

- 均值 • 适用于数值型数据、比较稳定、容易受到极端值的影响
- 中位数 • 顺序数据和数值型数据、不受极端值的影响、适用于收入
- 众数 • 分类数据和顺序数据、不受极端值的影响、缺乏稳定性

### 离散程度

- 方差
- 标准差 • 方差的平方根，不能直接用标准差比较不同变量的离散程度
- 离散系数 • 又称为：变异系数，标准差系数  
标准差与均值的比值，用于不同类别数据离散程度的比较

### 分布形态

- 偏态系数 • 离差三次方的平均数与标准差三次方的比值。  
等于0对称，>0右偏，<0左偏  
绝对值0和0.5之间轻度；0.5和1之间中度；大于1严重
- 标准分数 • 标准分数=(数值-均值)÷标准差，“168、295和399”

### 相关分析

- 相关关系 • 相关程度（完全相关、不完全相关和不相关）；方向（正相关和负相关）；相关形式（线性相关和非线性相关）  
相关关系并不等同于因果关系
- Pearson • 取值范围： $-1 \leq r \leq 1$ ， $0 < r \leq 1$ 正线性相关； $-1 \leq r < 0$ 负线性相关； $r=1$ 完全正线性相关； $r=-1$ 完全负线性相关； $r=0$ 不存在线性相关关系

谢谢 观看  
THANK YOU