

# 《传播统计学》

## 相关分析

教师：林志良

邮箱：[linzhl@nfu.edu.cn](mailto:linzhl@nfu.edu.cn)

个人网站：[www.zhilianglin.com](http://www.zhilianglin.com)

# 目录

- 相关分析介绍
- 协相关
- 相关系数：计算、显著性检验
- 实例与软件实操

# 相关分析介绍

- **作用：**确定两个数值型变量的关系
- 例如：
  - 对组织的满意程度与工作投入度的关系（不满是向上的车轮？）
  - 国民收入水平与幸福感的关系

## 相关概念

- 当一个变量增大，另一个也随之增大（或减少），我们称这种现象为**共变**，或相关（correlation）。两个变量有共变现象，称为**有相关关系**。
- 研究两个变量之间是否存在**线性**相关关系；如果存在**线性**相关关系，其相关的方向和密切程度如何——这个过程在统计学上称为**相关分析**。

## 相关分析介绍

### 相关关系≠因果关系

- 研究发现冰淇淋的销量增加时，溺水人数也增加。
- 结论——“冰淇淋销量→溺水人数”？
- 其实是“夏天的温度” → 冰淇淋销量增加  
    ↳ 溺水人数增加

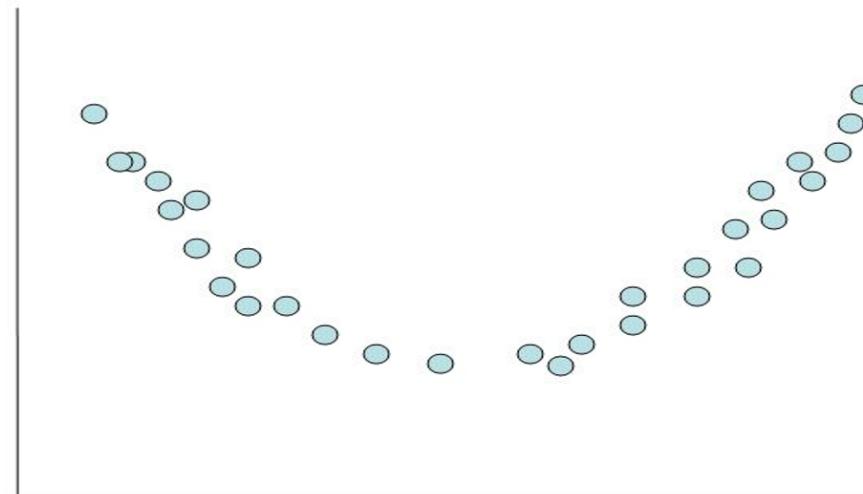
## 相关的分类

- **正相关**: 两个变量的变化方向一致的相关，即一个变量值变大时，另一个变量值也随着变大；一个变量值变小时，另一个变量也随着变小。
- **负相关**: 两个变量的变化方向相反的相关，即一个变量值变大时，另一个变量则变小；一个变量值变小时，另一个变量值则变大。
- **零相关**: 两个变量的变化方向无一定的规律，即一个变量值变化时，另一个变量值变大或变小的趋势均不明显。

## 相关分析介绍

### 零相关≠无相关

- 相关分析的是两个变量之间是否有**直线关系**。如下图两变量间是呈现曲线关系，但仍为零相关。



## 相关分析介绍

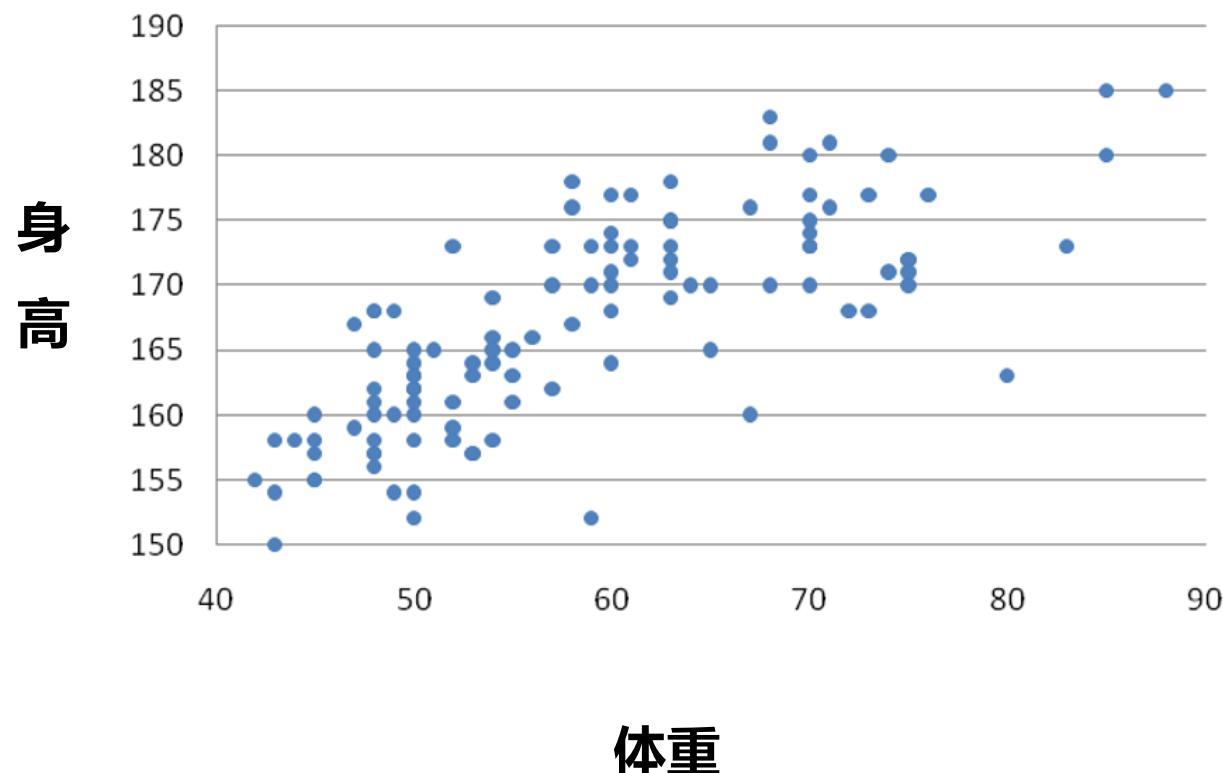
### 散点图 (scatter plots)

- 我们常常使用散点图来直观地观察两变量的相关关系。
- **散点图**: 用直角坐标系上, 以 $x$ 值为横轴,  $y$ 值为纵轴, 通过描点的方式来表示两个变量 ( $x$ 、 $y$ ) 之间的相关关系。

# 相关分析介绍

## 散点图 (scatter plots)

体重-身高散点图



# 协方差

## 协方差 (Covariance)

- 协方差衡量两个数值型变量(X & Y)是否存在线性关系。
- 协方差的计算公式：

$$\text{cov}(X, Y) = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{n - 1}$$

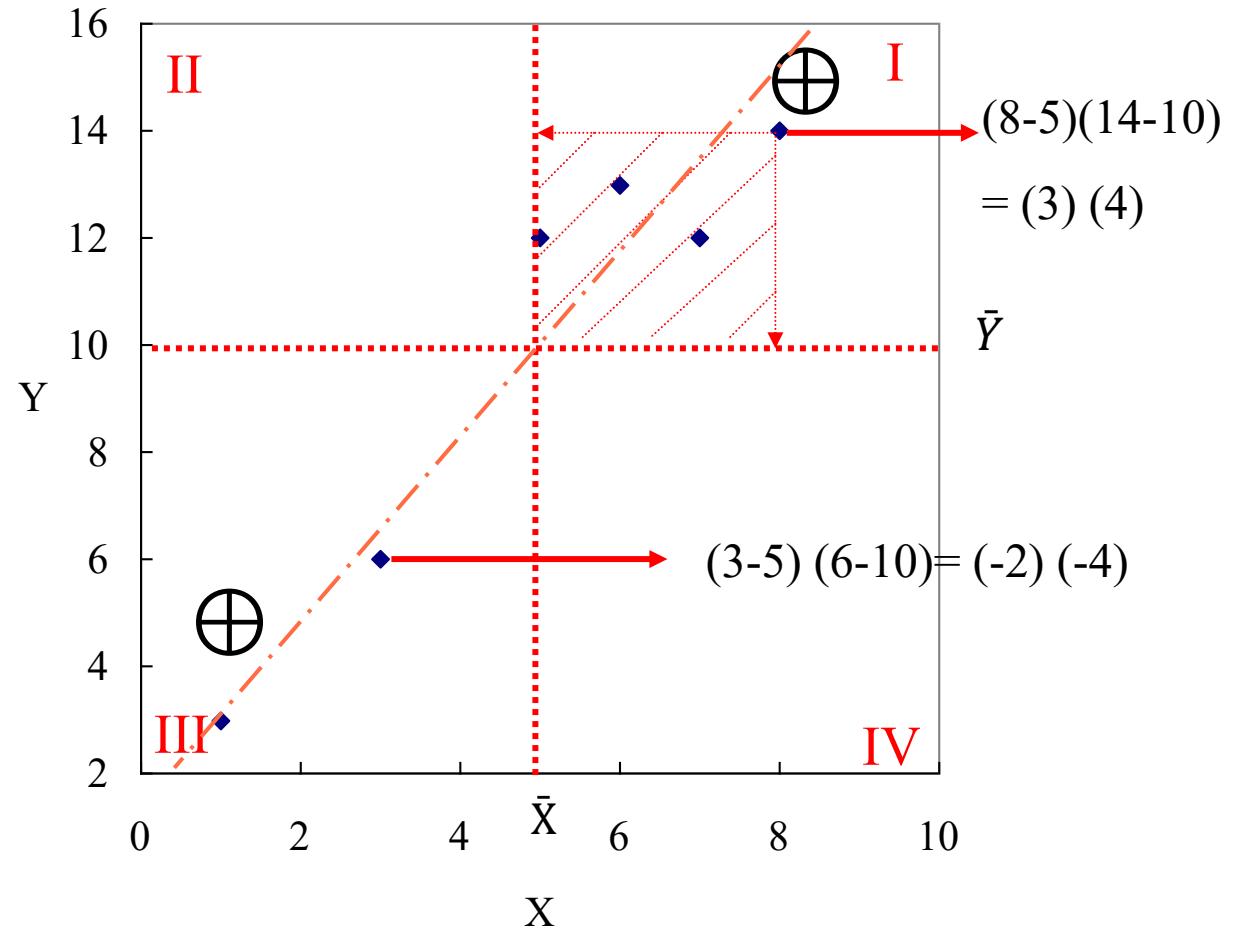
# 协方差

## 求X与Y的协方差

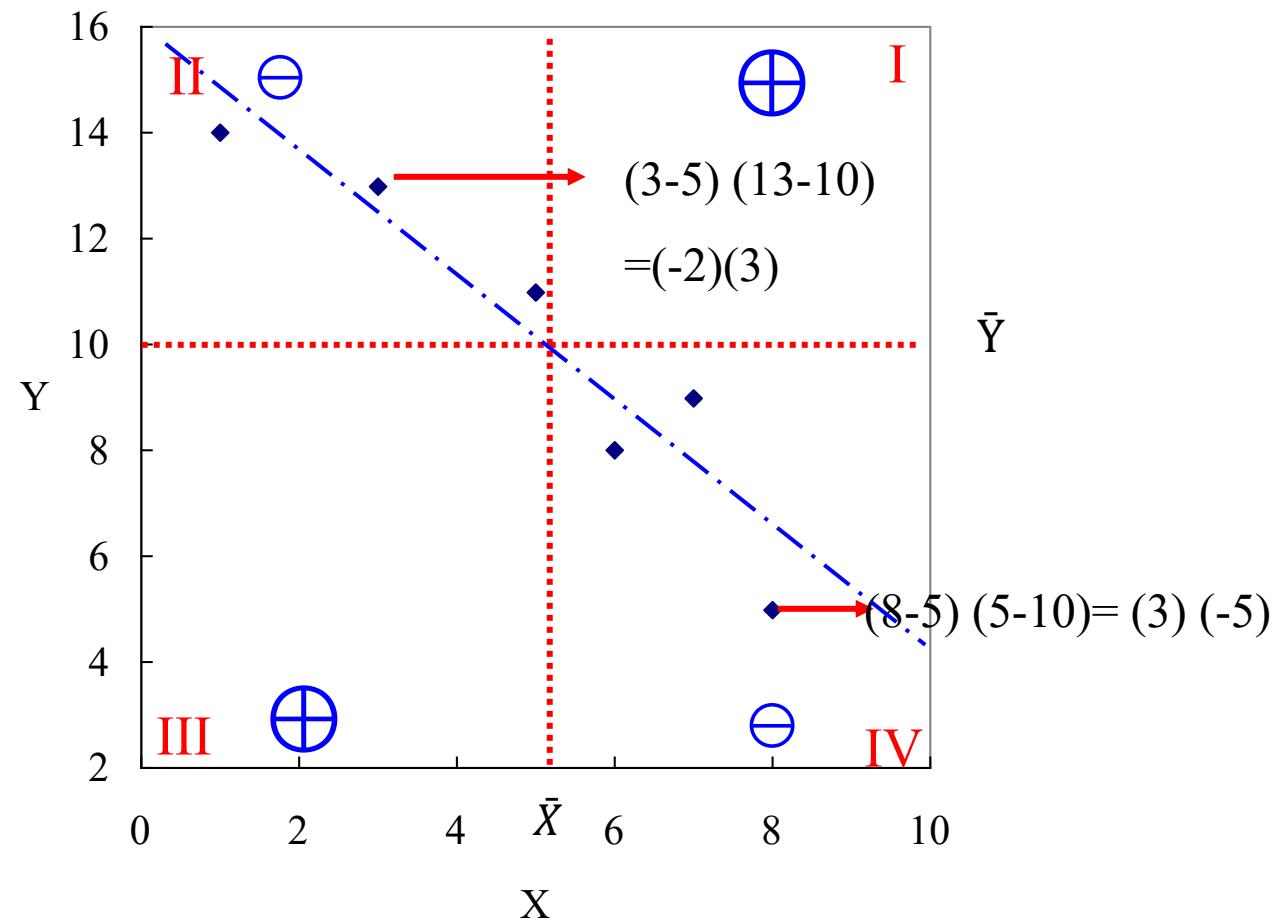
i	X <sub>i</sub>	Y <sub>i</sub>	X <sub>i</sub> - $\bar{X}$	(Y <sub>i</sub> - $\bar{Y}$ )	(X <sub>i</sub> - $\bar{X}$ )(Y <sub>i</sub> - $\bar{Y}$ )
1	3	6	-2	-4	8
2	5	12	0	2	0
3	1	3	-4	-7	28
4	6	13	1	3	3
5	8	14	3	4	12
6	7	12	2	2	4
	30	60		55	

$$\text{cov}(X, Y) = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{n - 1} = \frac{55}{6 - 1} = 11$$

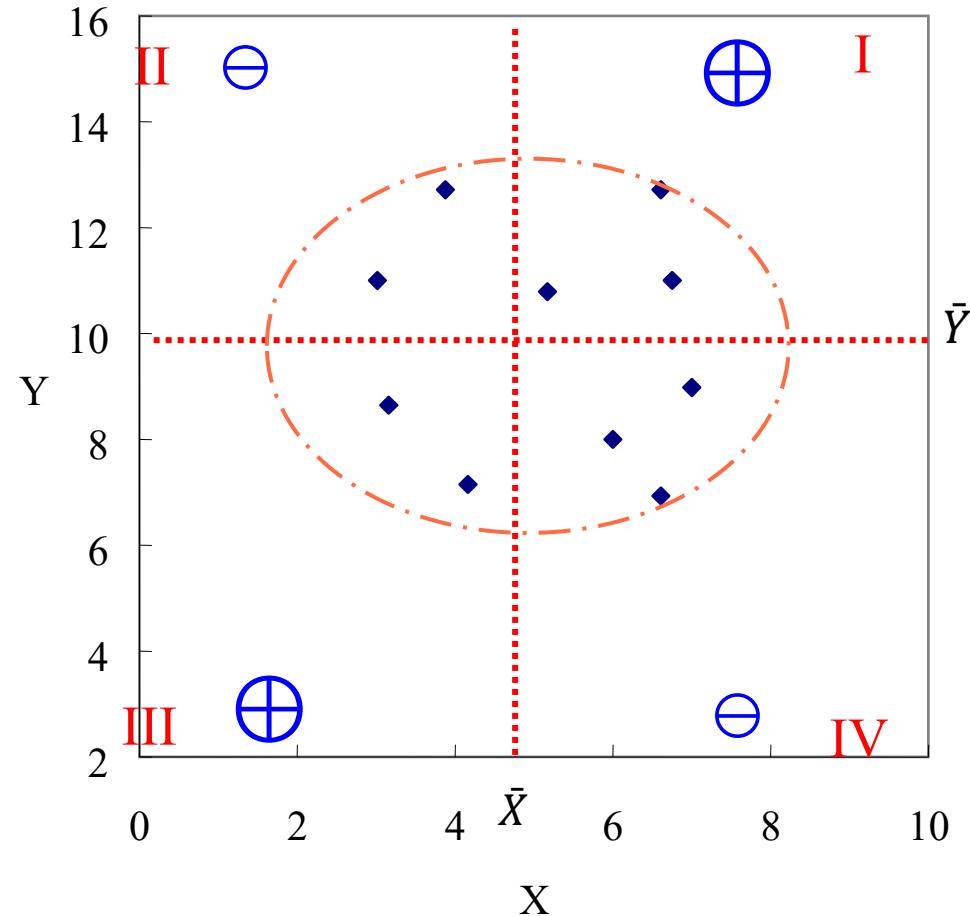
# 协方差



# 协方差



# 协方差



# 协方差

## 理解协方差

$\text{cov}(X,Y) > 0 \rightarrow X \text{与} Y \text{共变方向一致}$

$\text{cov}(X,Y) < 0 \rightarrow X \text{与} Y \text{共变方向相反}$

$\text{cov}(X,Y) = 0 \rightarrow X \text{与} Y \text{相互独立}$

# 协方差

## 理解协方差

- 用协方差来衡量两变量的关系有一个缺点，即其值受变量值大小的影响大。

$X_i$	$Y_i$
3	6
5	12
1	3
6	13
8	14
7	12

$$\text{cov}(X, Y) = 11$$

$X_i$	$Y_i$
30	60
50	120
10	30
60	130
80	140
70	120

$$\text{cov}(X, Y) = 1100$$



# 相关系数

## 相关系数的计算

- 相关系数(Coefficient of Correlation) 的含义：测量两个数值型变量线性关系强度的指标。

$$r = \frac{\text{cov}(X, Y)}{S_x S_y}$$

其中，

$$\text{cov}(X, Y) = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{n-1}$$

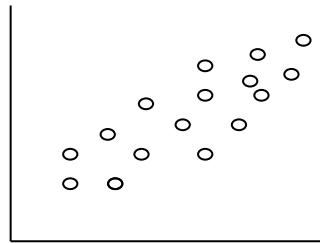
$$S_x = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}}$$

$$S_y = \sqrt{\frac{\sum_{i=1}^n (Y_i - \bar{Y})^2}{n-1}}$$

# 相关系数

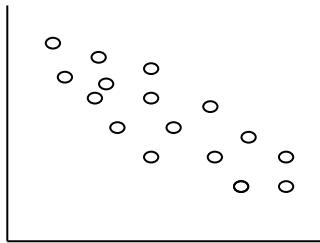
## 相关系数与散点图

正相关



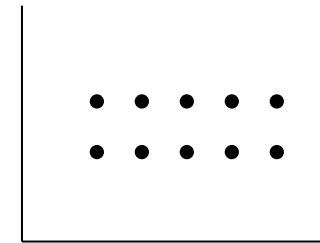
$$0 < r < 1$$

负相关

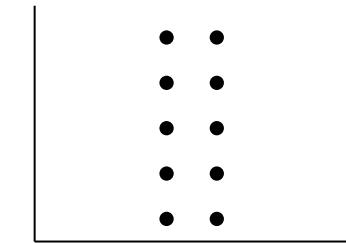


$$-1 < r < 0$$

零相关

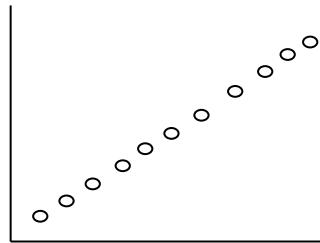


零相关



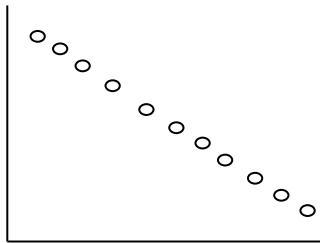
$$r=0$$

完全正相关



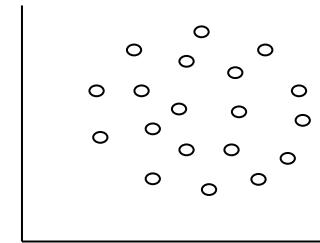
$$r=1$$

完全负相关



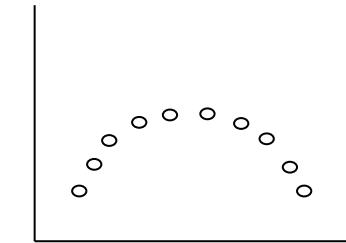
$$r=-1$$

零相关



$$r=0$$

零相关



$$r=0$$



# 相关系数

## 相关系数大小的含义

- 取值范围为-1到1(绝对值最大为1， 最小为0)
  - 相关系数绝对值大小的含义：
    - 低于0.1：基本**没有**相关关系
    - 0.1-0.3：**较小**的相关关系
    - 0.3-0.5：**中等**的相关关系
    - 0.5以上：**较强**的相关关系
- 不过也许更重要的是看p  
值有没有小于0.05**

## | 相关系数的显著性检验

### 相关系数的显著性检验的含义

- 相关系数的显著性检验指的是由样本相关系数 $r$ 推断总体相关系数 $\rho$ 是否为零（总体是否相关）。相关系数符合自由度为 $n-2$ 的t分布。



# 相关系数的显著性检验

## 相关系数的显著性检验的步骤

- 步骤1：提出零假设与备择假设。
- 步骤2：计算检验统计量t值及自由度。

$$\begin{aligned} H_0: \rho &= 0 \\ H_1: \rho &\neq 0 \end{aligned}$$

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

$$df = n - 2$$

- 步骤3：查询t值表，进行推断统计。

## 实例

### 身高与体重

有五位学生，他们的身高分别是152、158、165、172、180厘米；体重分别是43、48、54、63、80公斤。求身高及体重的相关系数并做显著性检验。

<b>身高</b>	152	158	165	172	180
<b>体重</b>	43	48	54	63	80

# 实例

$$\bar{X} = 165.4$$

$$\bar{Y} = 57.6$$

$$S_X = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}} = \sqrt{\frac{(152 - 165.4)^2 + (158 - 165.4)^2 + (165 - 158)^2 + (172 - 165.4)^2 + (180 - 165.4)^2}{4}}$$
$$= 11.08$$

$$S_Y = \sqrt{\frac{\sum_{i=1}^n (Y_i - \bar{Y})^2}{n-1}} = \sqrt{\frac{(43 - 57.6)^2 + (48 - 57.6)^2 + (54 - 57.6)^2 + (63 - 57.6)^2 + (80 - 57.6)^2}{4}}$$
$$= 14.57$$

$$\text{cov}(X, Y) = \frac{\sum_{i=1}^{n-1} (X_i - \bar{X})(Y_i - \bar{Y})}{n-1} =$$
$$\frac{(152 - 165.4)(43 - 57.6) + (158 - 165.4)(48 - 57.6) + (165 - 158)(54 - 57.6) + (172 - 165.4)(63 - 57.6) + (180 - 165.4)(80 - 57.6)}{4}$$
$$= 157.7$$

$$r = \frac{\text{COV}(X, Y)}{S_X S_Y} = \frac{157.7}{11.08 \times 14.57} = 0.977$$

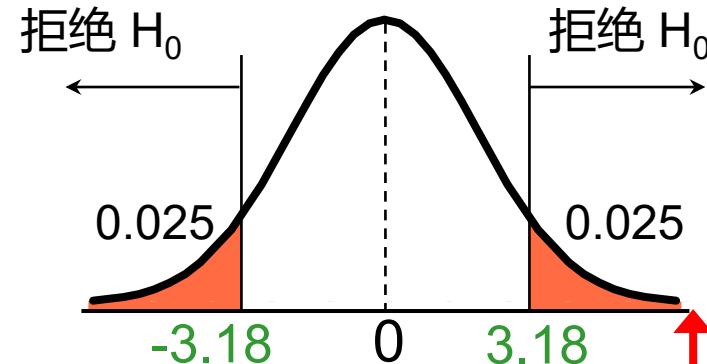
# 实例

$$\begin{aligned} H_0: \rho = 0 \\ H_1: \rho \neq 0 \end{aligned}$$

$$df = n - 2 = 5 - 2 = 3$$

临界值:  $t = \pm 3.18$

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{0.977*2}{\sqrt{1-0.977^2}} = 9.16$$



结论:

身高与体重显著相关，且相关系数为0.977

# 软件实操



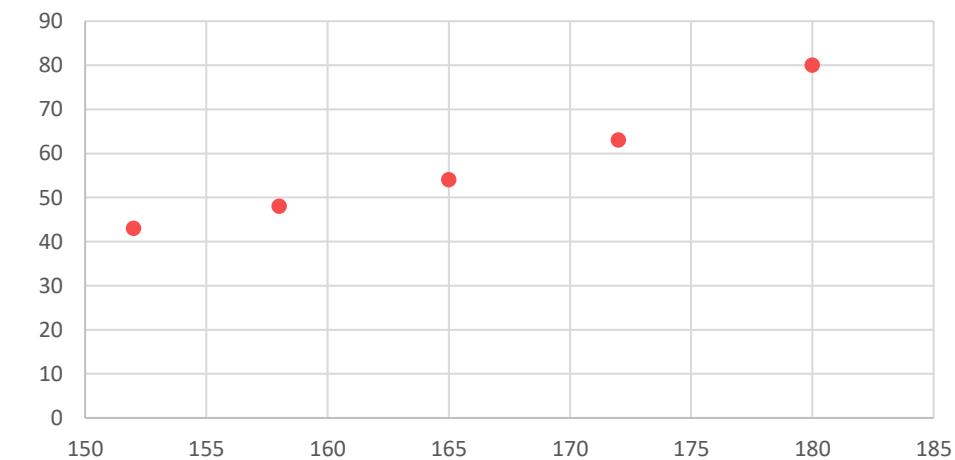
## 画散点图

选中数据，点击【插入】，选择散点图的图形

The screenshot shows the Microsoft Excel ribbon with the '插入' (Insert) tab highlighted by a red circle. Below the ribbon, there is a small data table:

152	158	165	172	180
43	48	54	63	80

身高与体重的散点图



# 软件实操



## 计算相关系数

注意：Excel不能对相关分析进行假设检验

- 方法一：使用函数**CORREL**(每次只能分析两个变量之间的相关关系)

	A	B	C	D	E	F	G
1	ranking	effort_overall					
2	3	3		0.28433			
3	2	3					
4	4	4					
5	3	3					
6	4	4					
7	3	3					
8	3	4					
9	3	2					
10	4	3					
11	5	4					
12	3	3					
13	4	4					
14	3	3					
15	3	3					
16	3	3					
17	2	4					
18	3	3					
19	3	3					
20	3	3					
21	3	4					
22	4	3					
23	3	3					
24	4	3					
25	2	4					
26	3	4					
27	5	4					

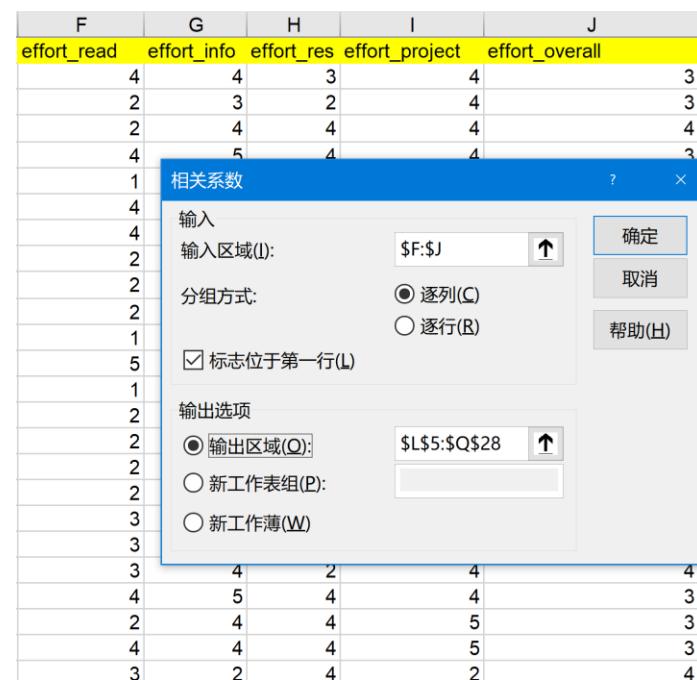
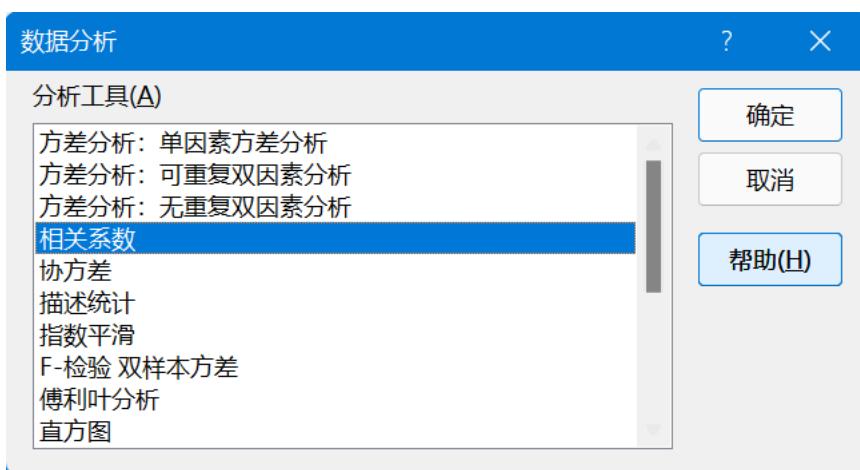
# 软件实操



## 计算相关系数

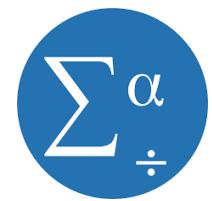
- 方法二：使用【分析工具】中的【相关系数】(可同时分析多个变量两两之间的相关系数)

这个表格叫：**相关系数表**



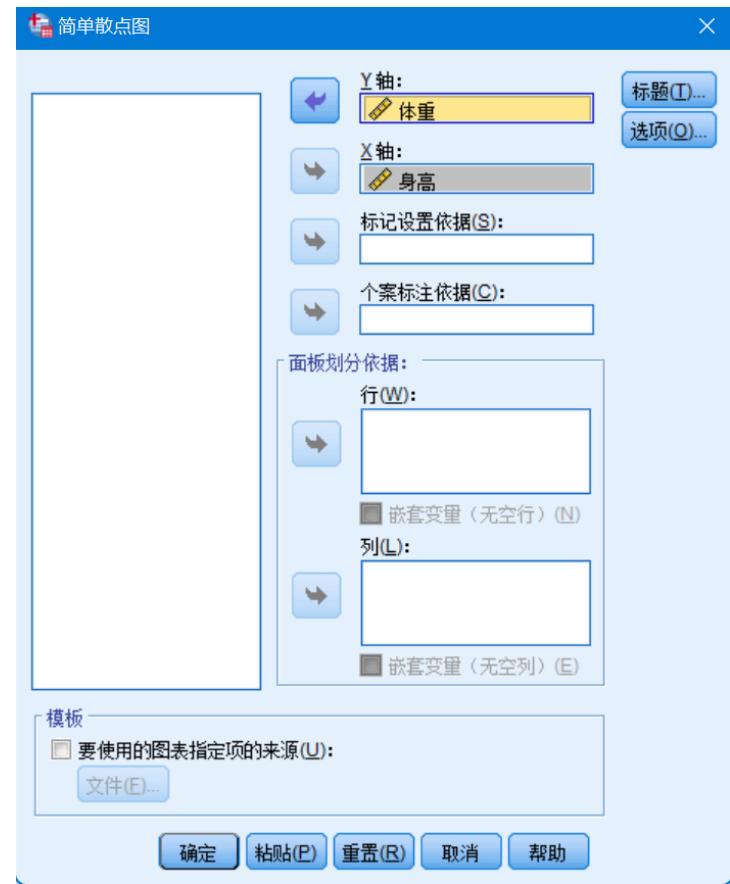
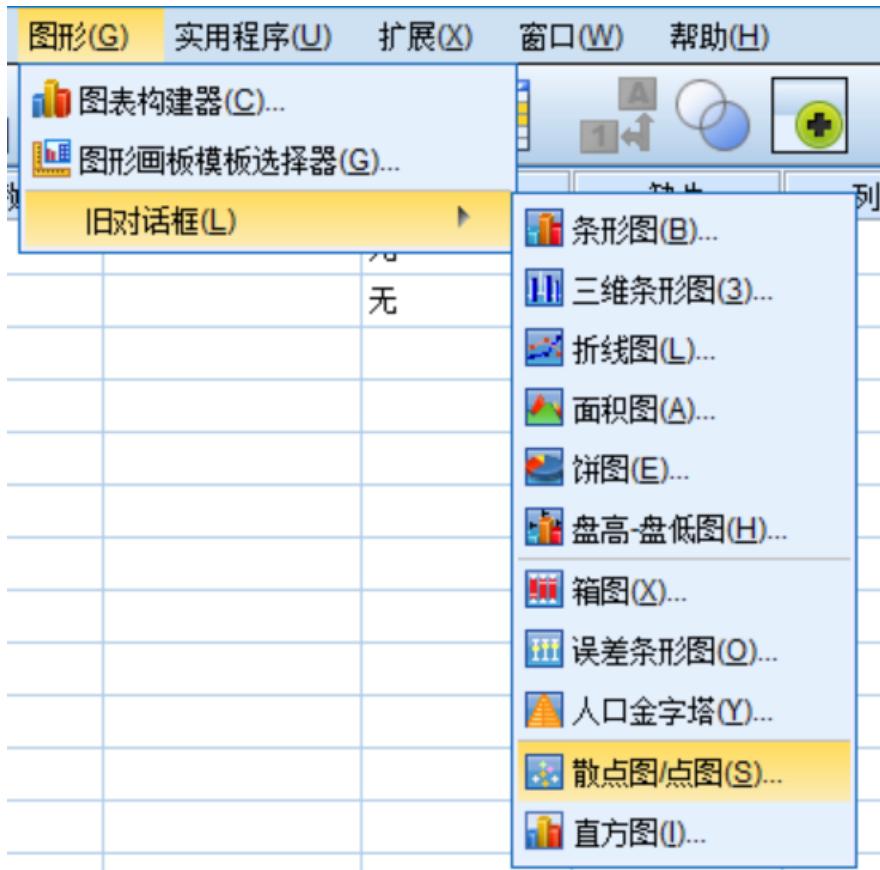
	effort_read	effort_info	effort_res	effort_project	effort_overall
effort_read	1				
effort_info	0.46734532	1			
effort_res	0.418507689	0.642064177	1		
effort_project	0.247580107	0.46872072	0.372938049	1	
effort_overall	0.239919349	0.252594288	0.324248546	0.32795651	1

# 软件实操

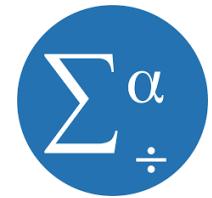


## 画散点图

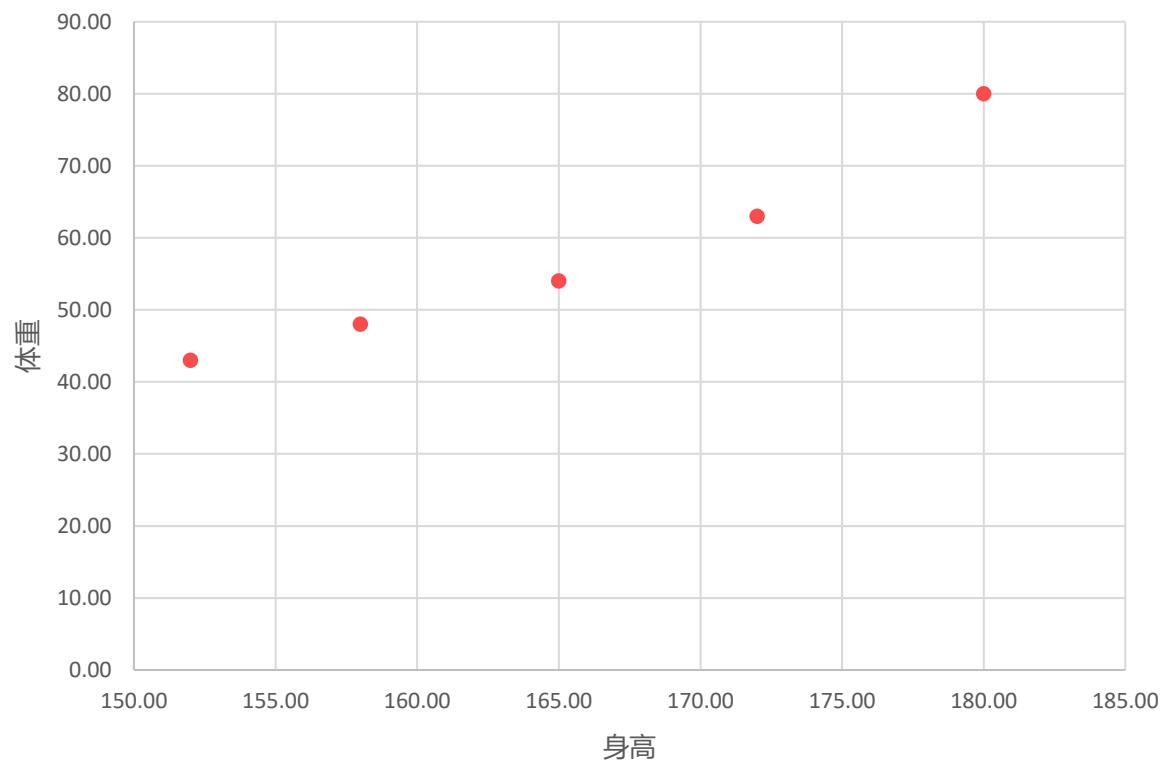
### 方法一



# 软件实操



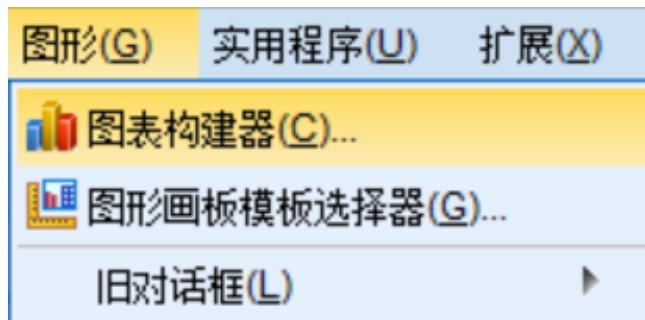
## 画散点图



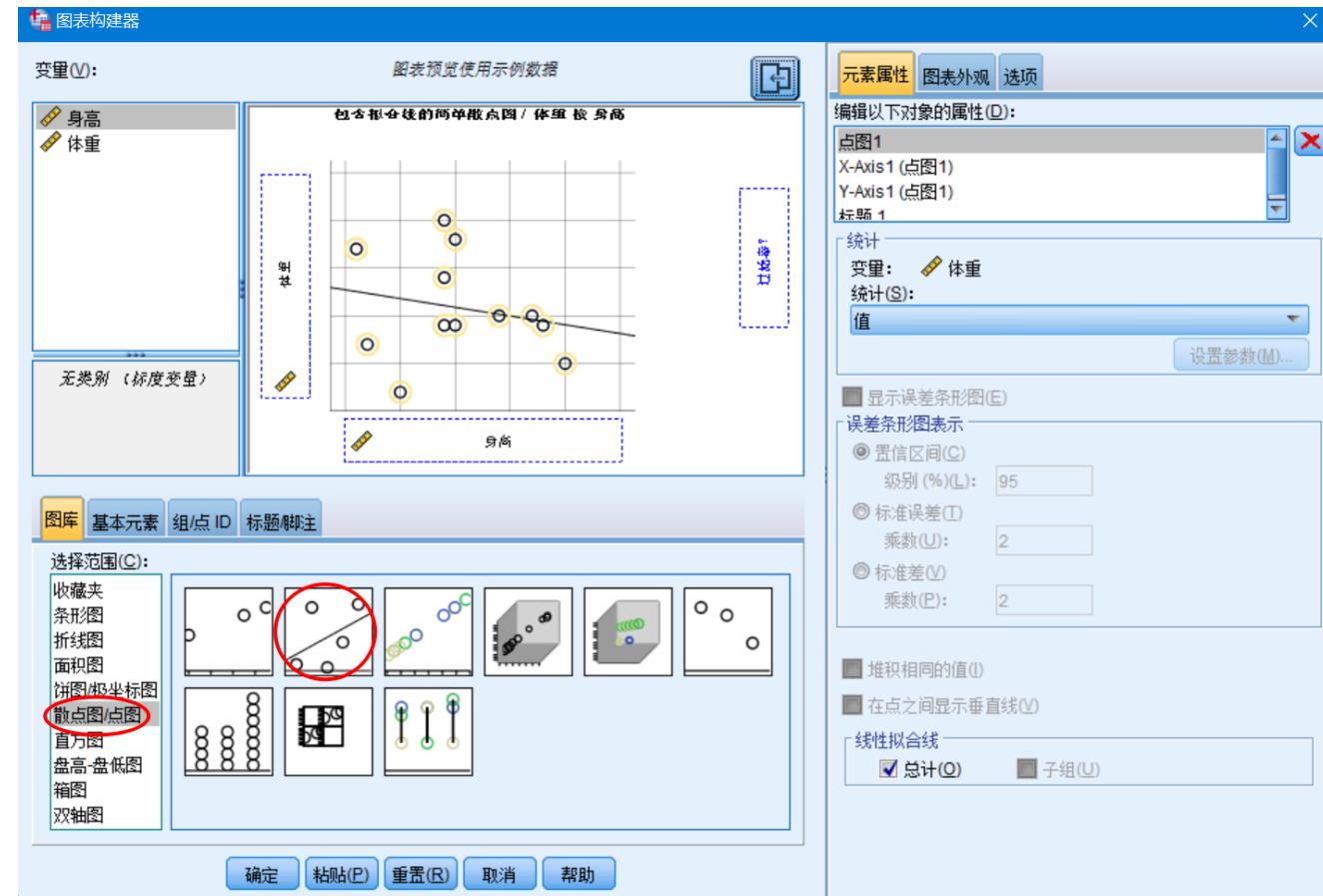
# 软件实操

$$\sum \alpha \div$$

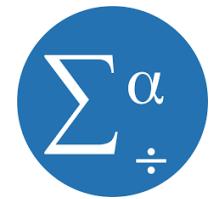
## 方法二（含拟合直线）



## 画散点图

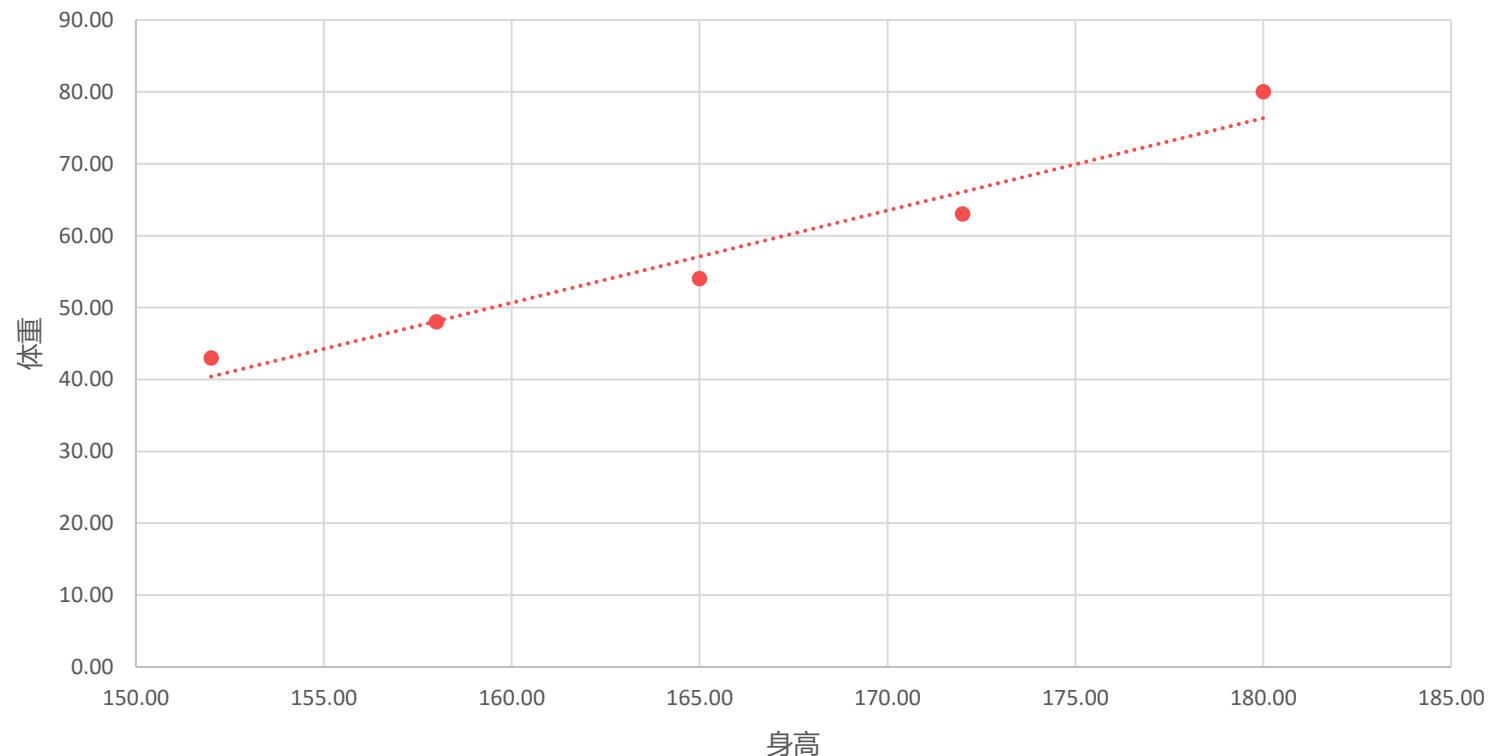


# 软件实操



## 画散点图

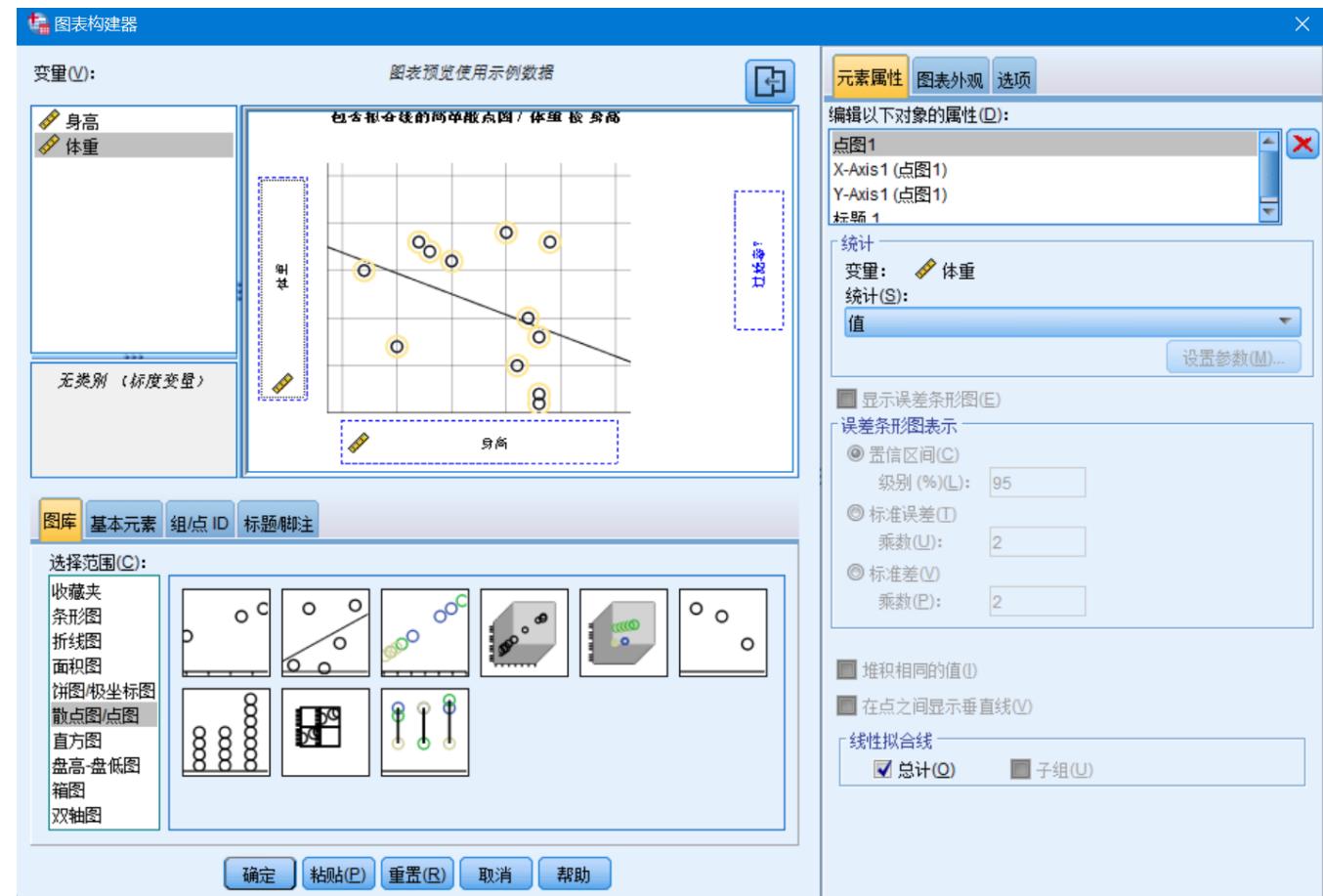
包含拟合线的简单散点图 / 体重按身高



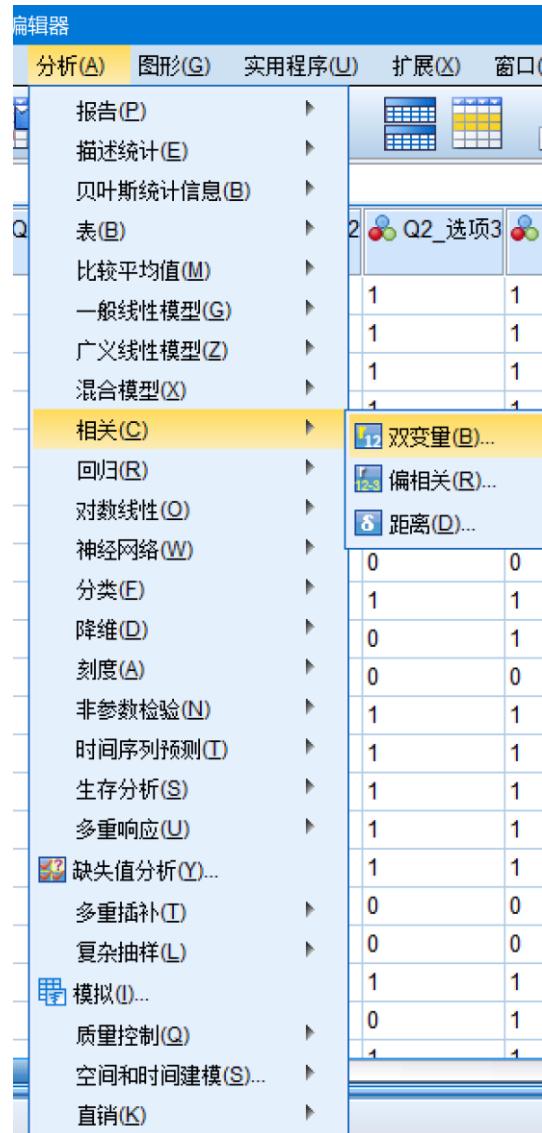
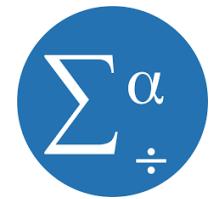
# 软件实操

$$\sum \alpha \div$$

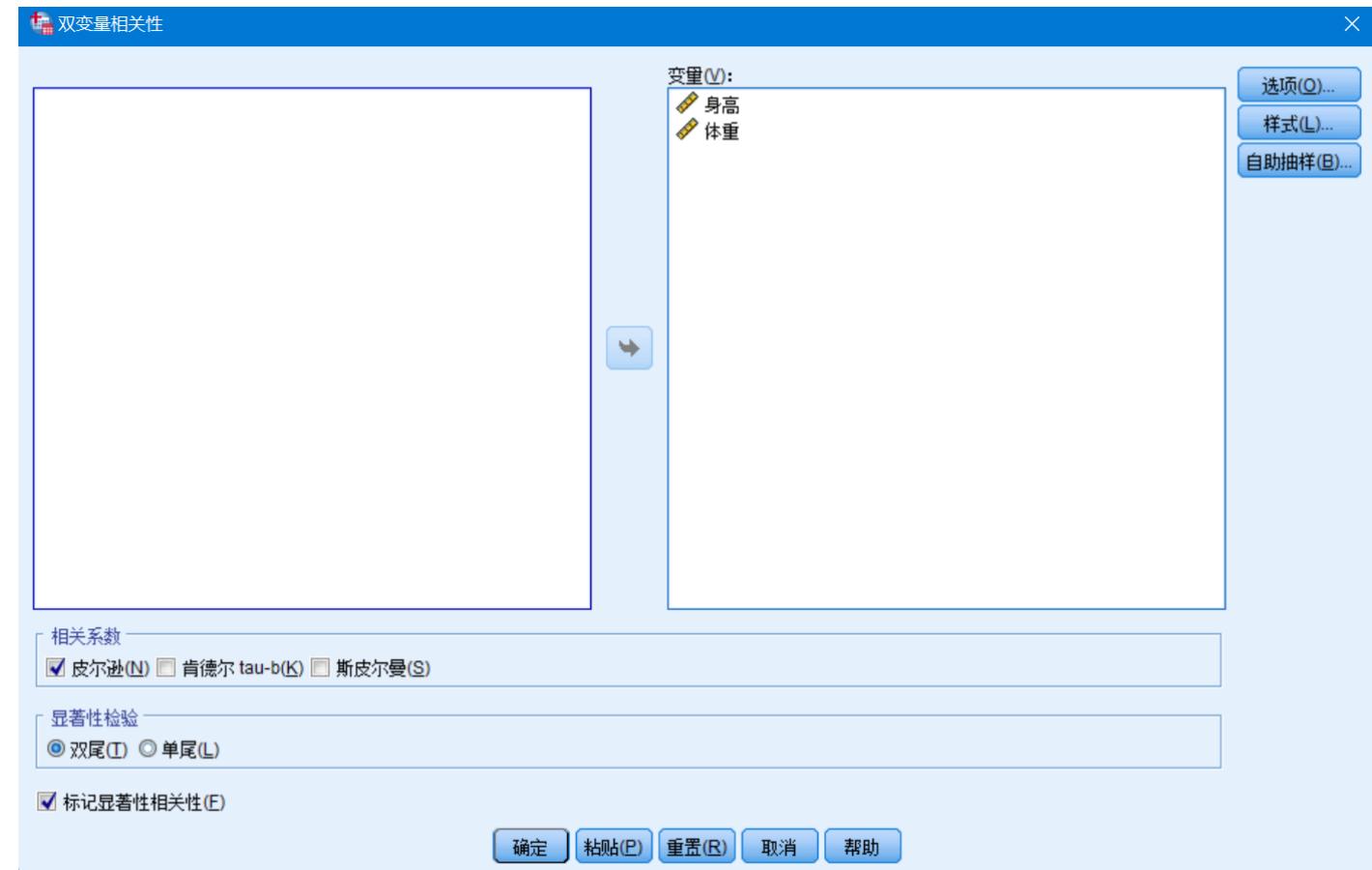
## 画散点图



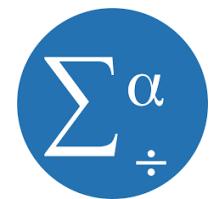
# 软件实操



## 计算相关系数



# 软件实操



## 计算相关系数

相关性

		身高	体重
身高	皮尔逊相关性	1	.977**
	Sig. (双尾)		.004
体重	个案数	5	5
	皮尔逊相关性	.977**	1
	Sig. (双尾)		.004
	个案数	5	5

p值

\*\*. 在 0.01 级别 (双尾)，相关性显著。

谢谢！