



浙江财经大学
Zhejiang University of Finance & Economics

相关关系的含义与测度



陈玉娟

浙江财经大学 数据科学学院

2017.8.2



视频引入



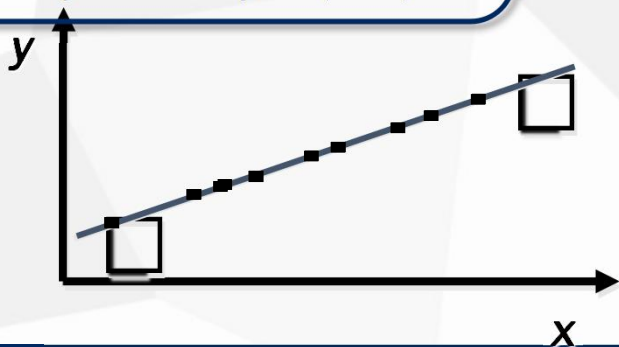
相关关系的含义与测度

相关关系的含义

变量之间的关系

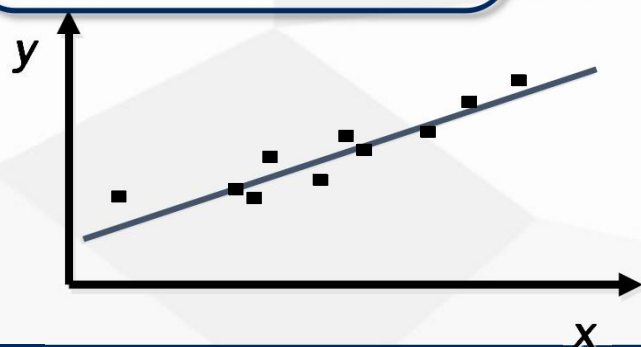
函数关系 (确定性关系)

$$y = f(x)$$



相关关系 (非确定性关系)

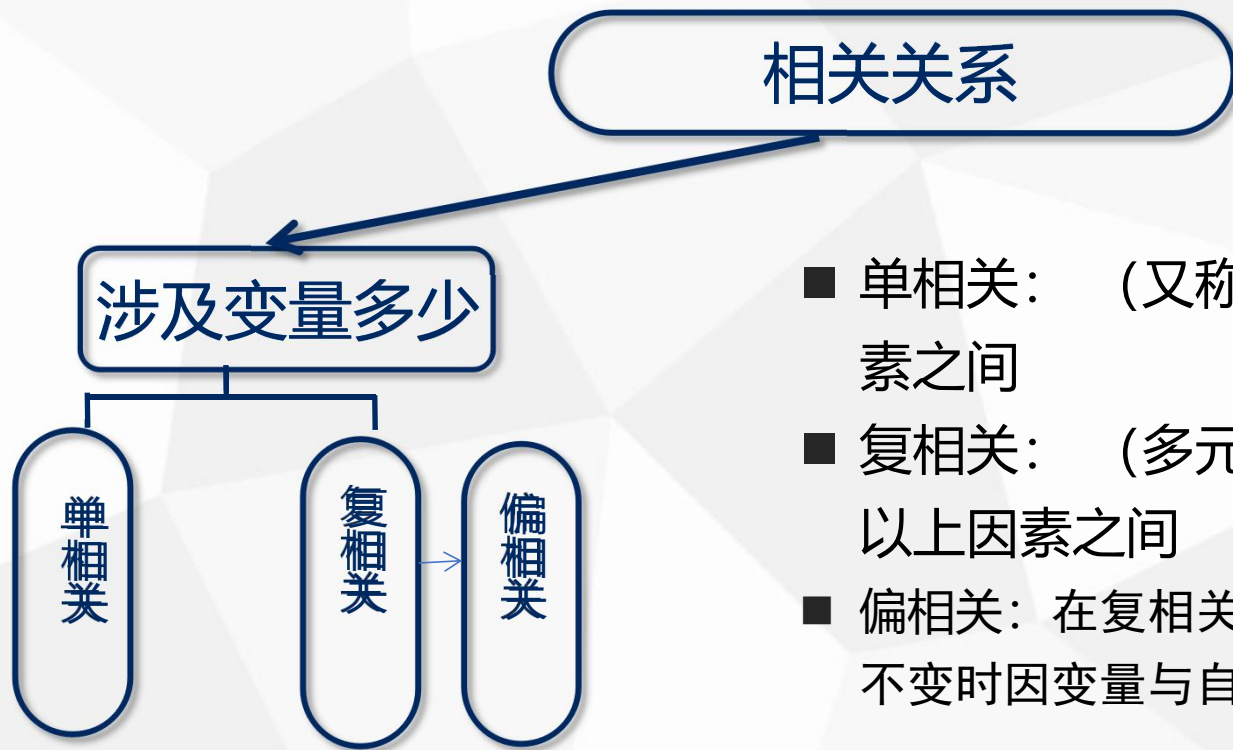
$$y = f(x) + \varepsilon$$





相关关系的含义与测度

相关关系的种类

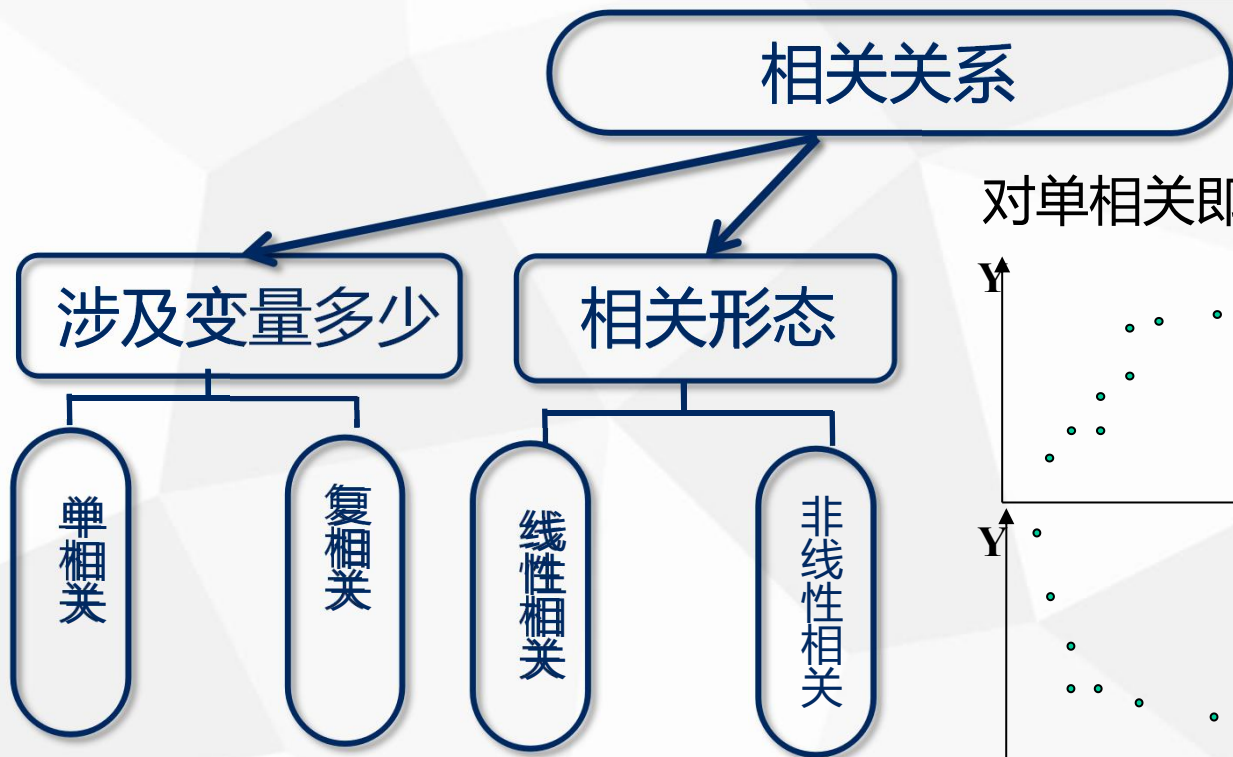


- 单相关：（又称一元相关）两个因素之间
- 复相关：（多元相关）三个或三个以上因素之间
- 偏相关：在复相关中，其他自变量固定不变时因变量与自变量之间的相关关系

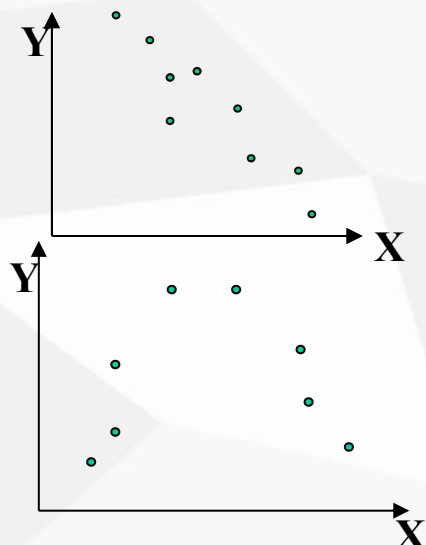
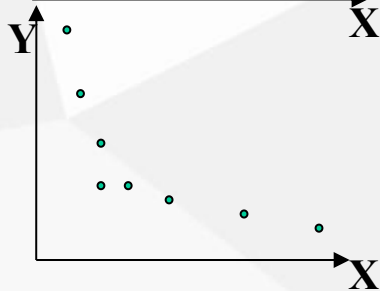
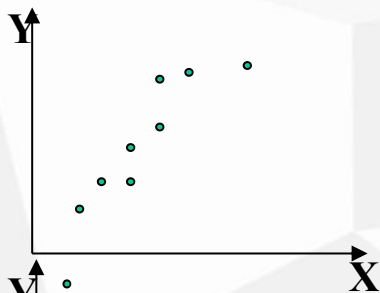


相关关系的含义与测度

相关关系的种类



对单相关即为直线相关和曲线相关





相关关系的含义与测度

相关关系的种类

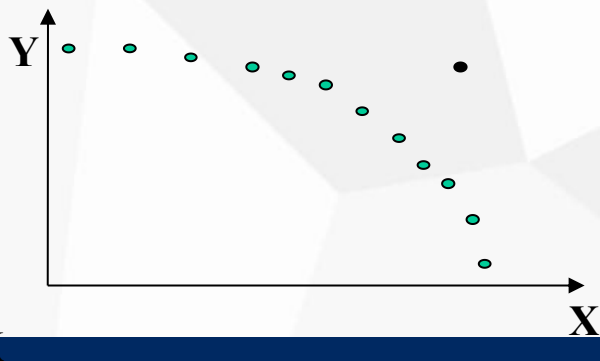
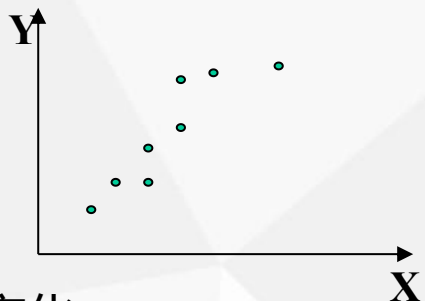
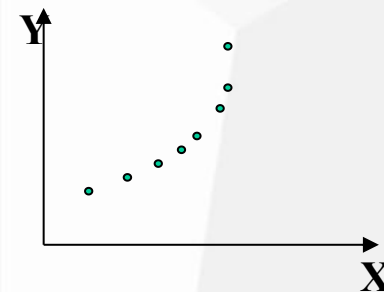
■ 正相关：同向变化

相关关系

相关方向

正相关

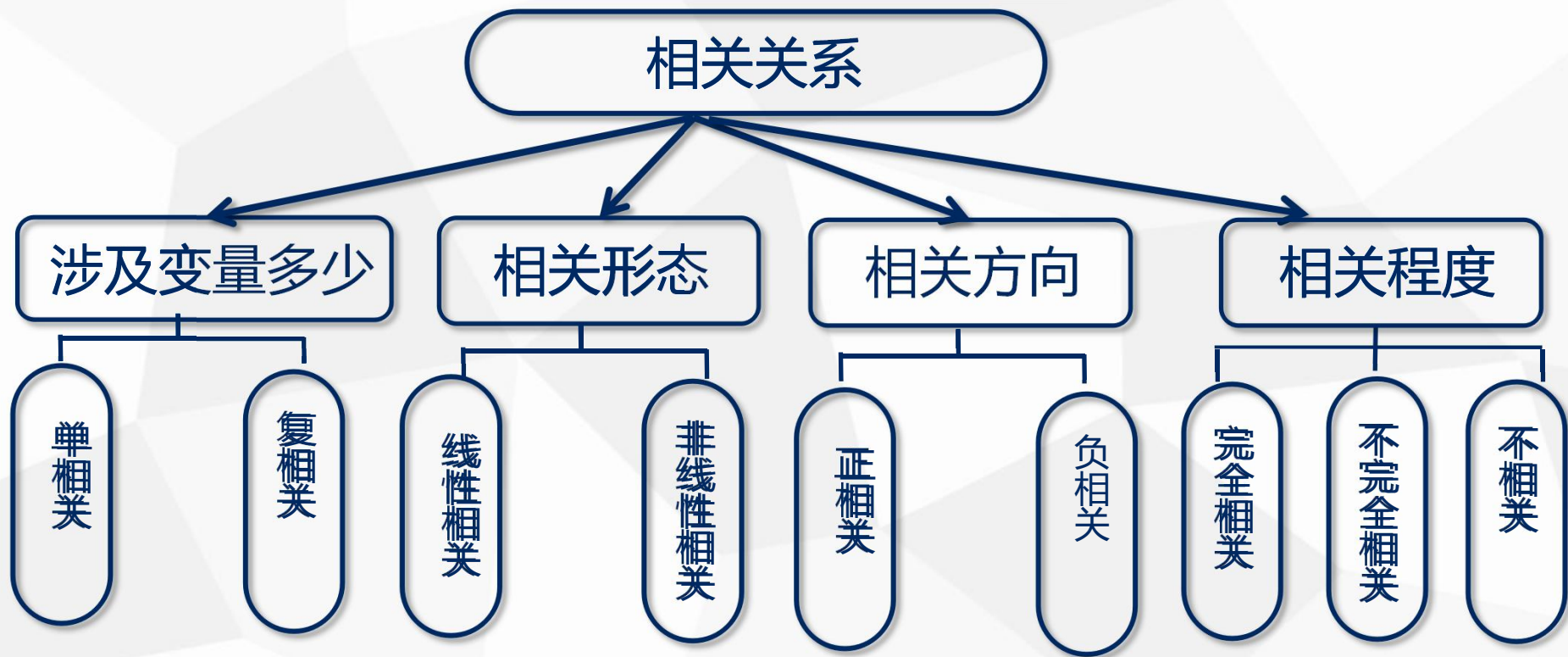
负相关





相关关系的含义与测度

相关关系的种类

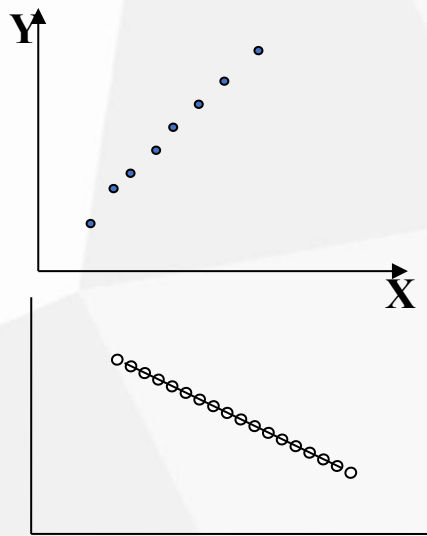




相关关系的含义与测度

相关关系的种类

- 完全相关：即确定性的函数关系



相关关系

相关程度

完全相关

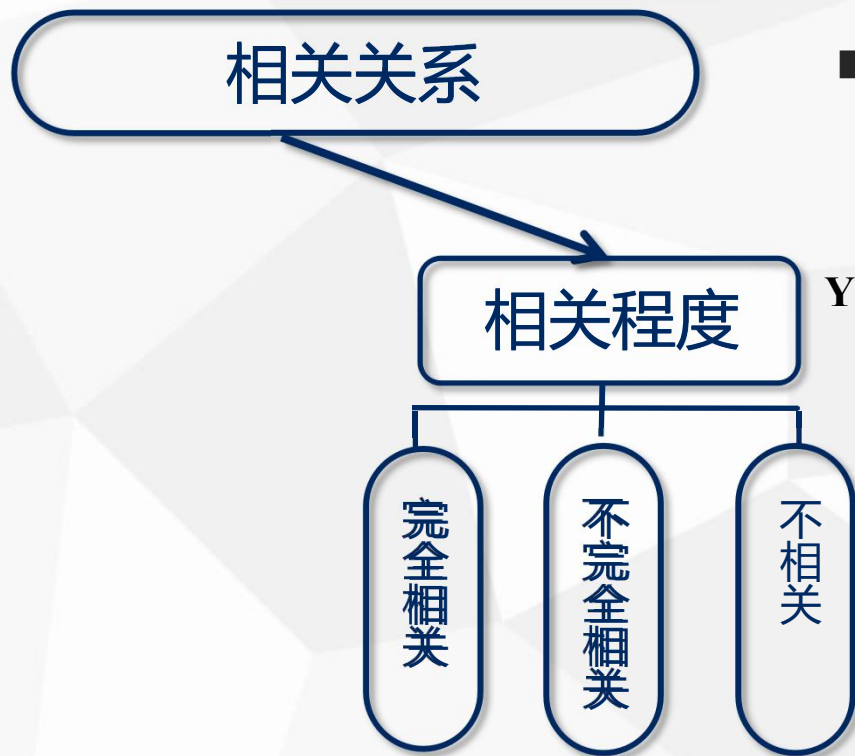
不完全相关

不相关

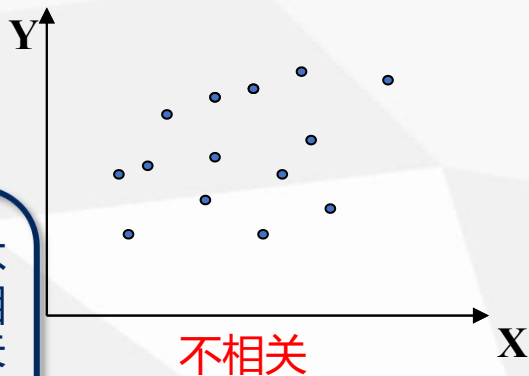


相关关系的含义与测度

相关关系的种类



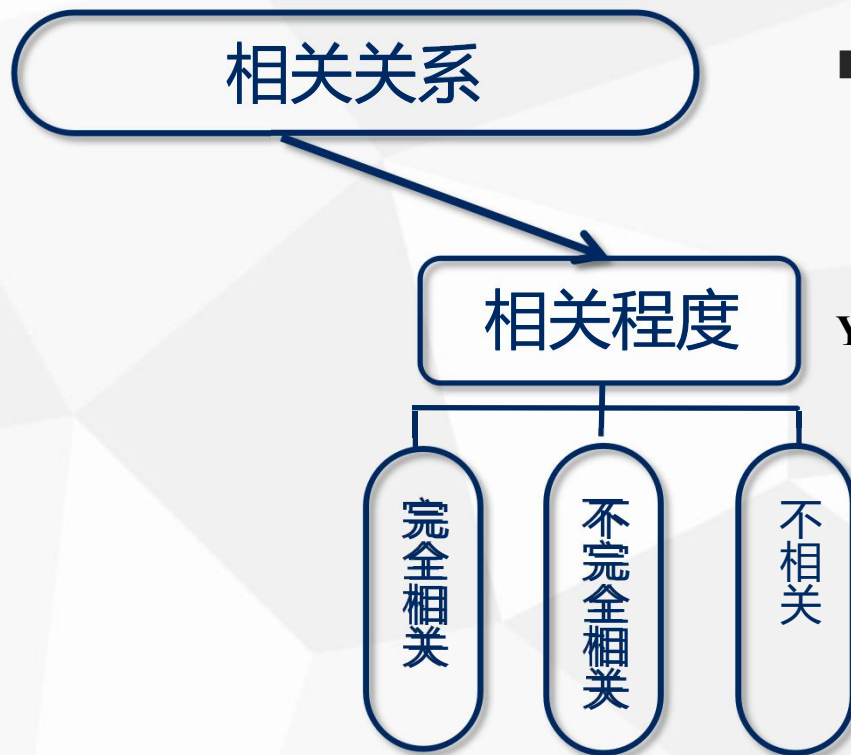
■ 无相关：变量各自独立，互不影响



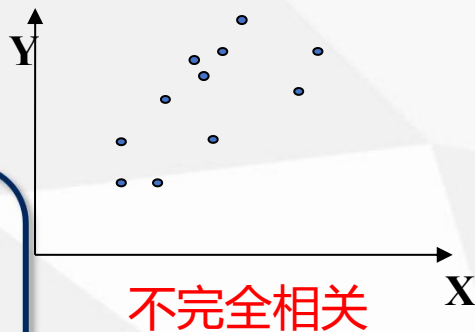


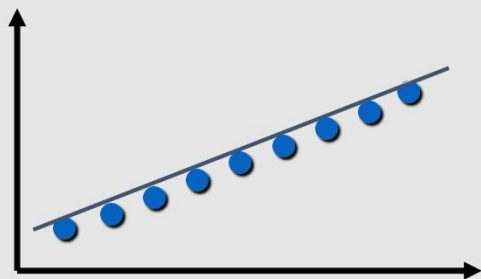
相关关系的含义与测度

相关关系的种类

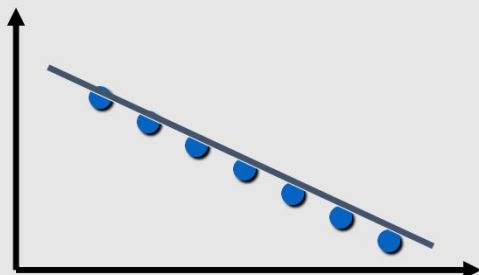


- 不完全相关：介于完全相关与无相关之间

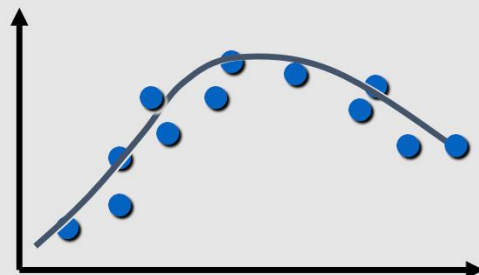




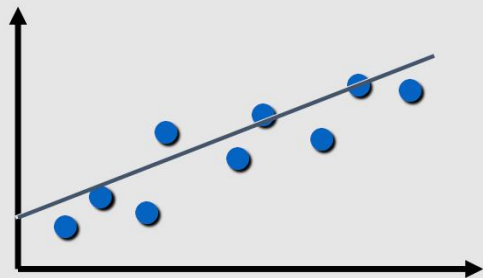
完全正线性相关



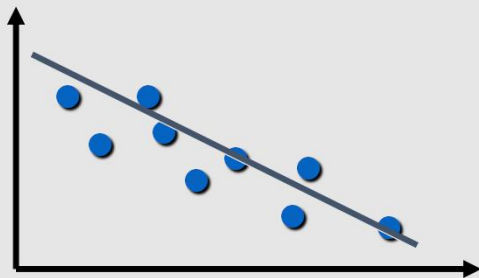
完全负线性相关



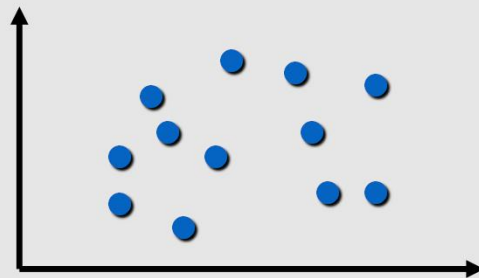
非线性相关



正线性相关



负线性相关



不相关



相关关系的含义与测度

相关关系的测度

相关关系的一般判断



定性分析



相关表



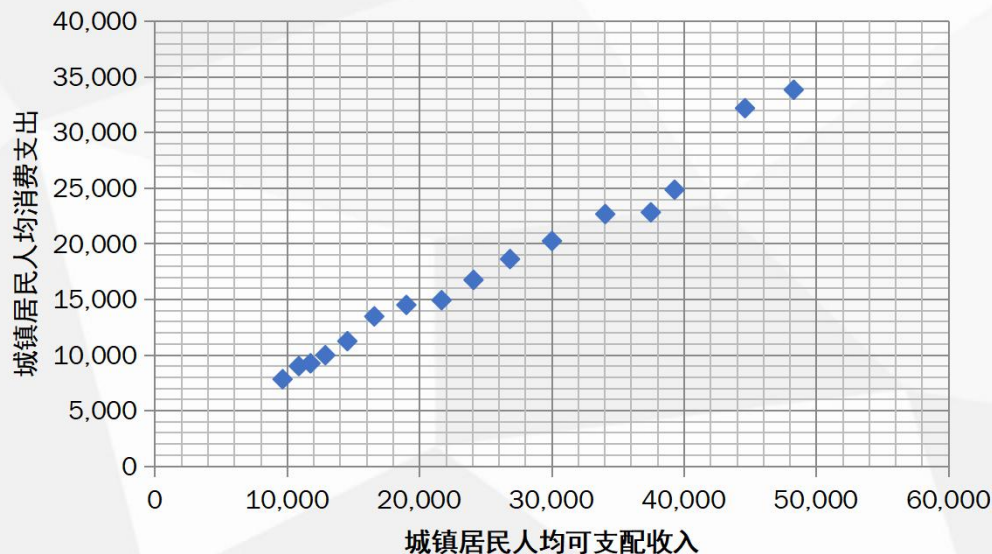
相关图



相关关系的含义与测度

引例的相关表和散点图

年份	人均可支配收入	人均消费支出
2000	9668	7790
2001	10896	8968
2002	11778	9215
2003	12898	9950
2004	14565	11213
2005	16601	13438
2006	19027	14472
2007	21689	14896
2008	24104	16719
2009	26864	18595
2010	30035	20219
2011	34065	22642
2012	37511	22800
2013	39310	24833
2014	44632	32165
2015	48316	33818





相关关系的含义与测度

相关系数的测定

1 线性相关系数的定义

相关系数(coefficient of correlation)是描述两个变量之间线性相关密切程度和相关方向的统计分析指标。



根据变量的不同类型采用不同相关系数：

- (1) 皮尔逊相关系数----定距变量或定比变量；**
- (2) 斯皮尔曼等级相关系数或肯特尔等级相关系数----定序变量。**



相关关系的含义与测度

相关系数的测定

2 计算公式

若根据总体全部数据计算的，称为总体相关系数，记为 ρ ；

若根据样本数据计算的，称为样本相关系数，记为 r 。

总体相关系数

$$\rho = \frac{Cov(X, Y)}{\sqrt{Var(X)Var(Y)}}$$



相关关系的含义与测度

相关系数的测定

2 计算公式

样本相关系数

$$r = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{\sum (x - \bar{x})^2 \cdot \sum (y - \bar{y})^2}}$$

简捷计算公式：

$$r = \frac{n \sum xy - \sum x \sum y}{\sqrt{n \sum x^2 - (\sum x)^2} \sqrt{n \sum y^2 - (\sum y)^2}}$$



相关关系的含义与测度

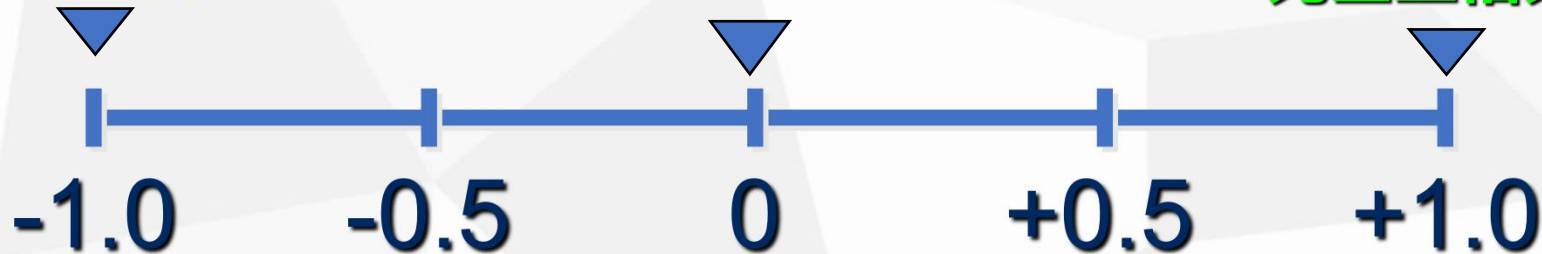
相关系数的测定

3 相关关系密切程度的判断

完全负相关

无线性相关

完全正相关



负相关程度增加


正相关程度增加



相关关系的含义与测度

相关系数的测定

3 相关关系密切程度的判断

- ⊕ r 的取值范围是 $[-1, 1]$
- ⊕ $r > 0$ 正相关; $r < 0$ 负相关
- ⊕ $|r| = 0$ 表示变量间**不存在线性相关**关系
- ⊕ $|r| = 1$ 完全线性相关
- ⊕ $0 < |r| < 1$ 不完全线性相关 

$0 \leq |r| < 0.3$ 微弱相关
 $0.3 \leq |r| < 0.5$ 低度相关
 $0.5 \leq |r| < 0.8$ 中度相关
 $0.8 \leq |r| < 1$ 高度相关



相关关系的含义与测度

引例相关系数的计算

年份	人均可支配收入X	人均消费支出Y	X ²	Y ²	XY
2000	9668	7790	93470224	60684100	75313720
2001	10896	8968	118722816	80425024	97715328
2002	11778	9215	138721284	84916225	108534270
2003	12898	9950	166358404	98997724	128332004
2004	14565	11213	212139225	125726435	163314141
2005	16601	13438	275593201	180579844	223084238
2006	19027	14472	362021401	209431259	275351771
2007	21689	14896	470428337	221883368	323079284
2008	24104	16719	580982569	279528305	402990164
2009	26864	18595	721670735	345764728	499528062
2010	30035	20219	902101225	408807961	607277665
2011	34065	22642	1160424225	512660164	771299730
2012	37511	22800	1407075121	519840000	855250800
2013	39310	24833	1545276100	616677889	976185230
2014	44632	32165	1992015424	1034587225	1435588280
2015	48316	33818	2334435856	1143657124	1633950488
合计	401959	281732	12481436148	5924167374	8576795176

$$r = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{\sqrt{\left[n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2 \right] \left[n \sum_{i=1}^n y_i^2 - \left(\sum_{i=1}^n y_i \right)^2 \right]}}$$
$$= \frac{16 \times 8576795176 - 401959 \times 281732}{\sqrt{3813215773} \sqrt{1541382578}}$$
$$= 0.989$$

表明杭州市人均可支配收入与人均消费支出高度正相关



相关关系的含义与测度

相关系数的测定

4 相关关系的显著性检验

相关系数的显著性检验过程就是由样本相关系数对总体相关系数的理论真值做出具有一定把握程度的推断。

随机样
本数据

样本相
关系数

总体相
关系数



相关关系的含义与测度

相关系数的测定

4 相关关系的显著性检验

1

提出假设: $H_0 : \rho = 0 \quad H_1 : \rho \neq 0$

2

构造计算检验统计量: $t = r \sqrt{n-2} / \sqrt{1-r^2} \sim t(n-2)$

3

根据给定的显著性水平 α , 确定临界值 $t_{\alpha/2}$



相关关系的含义与测度

相关系数的测定

4 相关关系的显著性检验

4

作出决策:

若 $|t| < t_{\alpha/2}(n-2)$, 则接受 H_0 , 表示总体两变量间线性相关性不显著

若 $|t| \geq t_{\alpha/2}(n-2)$, 则拒绝 H_0 , 表示总体两变量间线性相关性显著

对前例，检验杭州市城镇居民年人均消费支出和年人均可支配收入总体的相关系数是否显著（ $\alpha = 0.05$ ）

- 2000-2015年样本相关系数计算结果 $r = 0.989$

根据检验步骤：首先提出假设
然后计算检验统计量t：

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{0.989 \times \sqrt{16-2}}{\sqrt{1-0.989^2}} = 25.018$$

在给定的显著性水平 $\alpha=0.05$ ，查得临界值 $t_{\alpha/2}(14) = 2.145$ ，
由于 $|t| > t_{\alpha/2}$ ，故拒绝 H_0 ，认为两变量
总体相关系数显著。



浙江财经大学
Zhejiang University of Finance & Economics

谢谢

日期：2017/08/23