计量经济学

第 二 章

简单线性回归模型

2007st.com 版图设计网

- 离散变量的期望值定义为**E(X)**

◆ 有关性质

1. b 为常数则E(b)=b
2. E(X+Y)= E(X)+ E(Y)
3. E(aX)= aE(X), if a=constant
4. E(aX+ b)= aE(X)+ b, if a,b=constant
5. E(X*Y)= E(X) E(Y) only X,Y 相互独立

2007st.com 版图设计网

• 方差的性质

• 方差 $Var(X) = E\left(X_i - E(X)\right)^2 = \sigma^2$

◆ 方差的有关性质

1. 常数的方差为0
2. var(X+b)= var(X)
3. var(aX)= a^2var(X), if a=constant
4. var(aX+ b)= a^2var (X), if a,b=constant
5. if X,Y 相互独立,

var(aX+bY)= a^2var(X) + b^2var(Y)

var(X+Y)= var(X) + var(Y); var(X-Y)= var(X) +var(Y)

2007st.com 版图设计网

- 协方差的性质

协方差　$Cov(XY) = E\big(X_i - E(X)\big)\big(Y_i - E(Y)\big)$

◆ 协方差的有关性质

1. 若随机变量相互独立，则协方差为0
2. cov(aX+b,c+dY)=ad cov(X,Y)
   where a,b,c,d=constant
3. cov(X,X)= var(X), if a=constant

2007st.com 版图设计网

第二章 简单线性回归模型

本章主要讨论：

- 回归分析与回归函数
- 简单线性回归模型参数的估计
- 拟合优度的度量
- 回归系数的区间估计和假设检验
- 回归模型预测

2007st.com 版权设计网

第一节 回归分析与回归方程

本节基本内容:

- **回归与相关**
- **总体回归函数**
- **随机扰动项**
- **样本回归函数**

2007st.com 版图设计网

一、回归与相关（对统计学的回顾）

1. 经济变量间的相互关系

◆确定性的函数关系

$$Y = f(X)$$

◆不确定性的统计关系—相关关系

$$Y = f(X) + \varepsilon$$ (ε为随机变量)

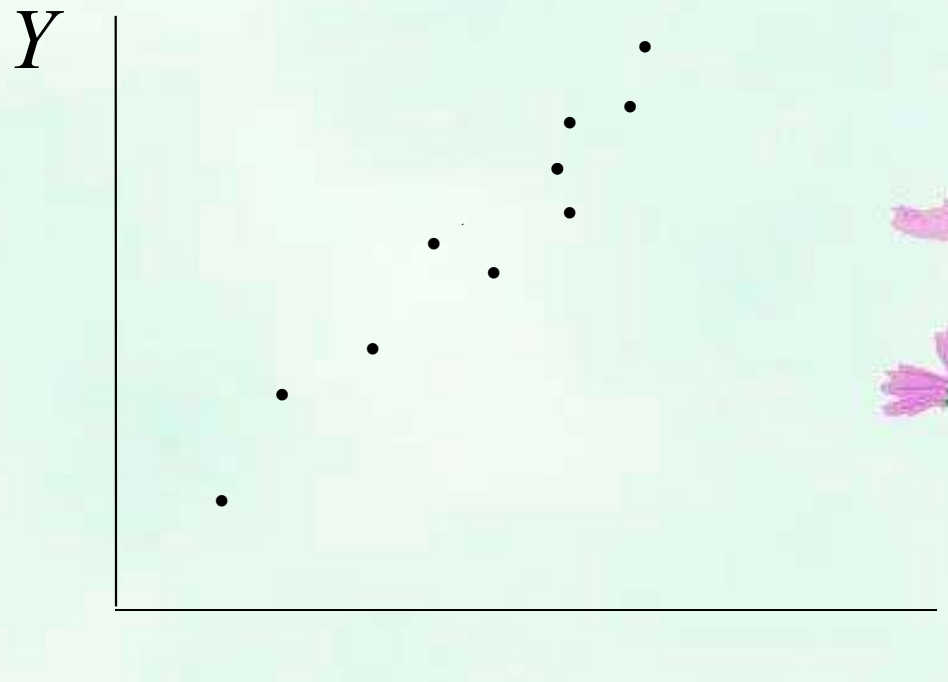
◆没有关系

2007st.com 版图设计网

2. 相关关系

◆ 相关关系的描述

相关关系最直观的描述方式——坐标图（散布图）



2007st.com 版图设计网

◆相关关系的类型

- 从涉及的变量数量看
  - 简单相关
  - 多重相关（复相关）
- 从变量相关关系的表现形式看
  - 线性相关——散布图接近一条直线
  - 非线性相关——散布图接近一条曲线
- 从变量相关关系变化的方向看
  - 正相关——变量同方向变化，同增同减
  - 负相关——变量反方向变化，一增一减
  - 不相关
- 从变量相关程度看
  - 完全相关；不相关；不完全相关

2007st.com 版权设计网

3. 相关程度的度量—相关系数

相关系数

十九世纪末——英国著名统计学家卡尔·皮尔逊（**Karl Pearson**）

——度量两个变量之间的线性相关程度的简单相关系数（简称相关系数）

两个变量X和Y的总体相关系数为

$$\rho_{XY}=\frac{Cov(X,\ Y)}{\sqrt{Var(X)}\cdot\sqrt{Var(Y)}}$$

其中，$Cov(X,\ Y)$是变量X、Y的协方差，

$Var(X)$、$Var(Y)$分别是变量X、Y的方差。

2007st.com 版权设计网

如果给定变量X、Y的一组样本$(X_i,\ Y_i)$,$i=1,2,\cdots n,$,

则总体相关系数的估计——样本相关系数为

$$r_{XY}=\frac{\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i-\bar{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i-\bar{Y})^2}}$$

或

$$r_{XY}=\frac{n\sum_{i=1}^{n}X_iY_i-\sum_{i=1}^{n}X_i\sum_{i=1}^{n}Y_i}{\sqrt{n\sum_{i=1}^{n}X_i^2-(\sum_{i=1}^{n}X_i)^2}\sqrt{n\sum_{i=1}^{n}Y_i^2-(\sum_{i=1}^{n}Y_i)^2}}$$

相关系数的取值介于–1—1之间，

取值为负表示两变量之间存在负相关关系；

取值为正表示两变量之间存在正相关关系；

取值为–1表示两变量之间存在完全负相关关系；

取值为0表示两变量不相关；

取值为1表示两变量之间存在完全正相关关系。

2007st.com 版权设计网

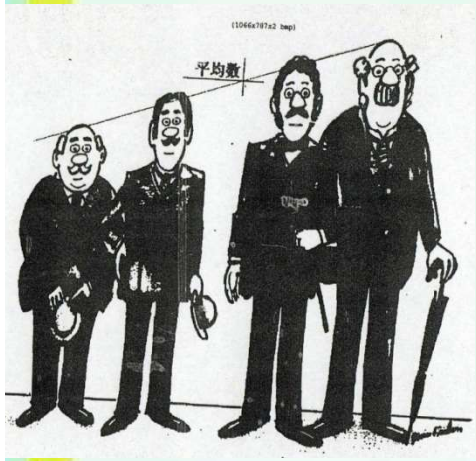
4. 回归分析

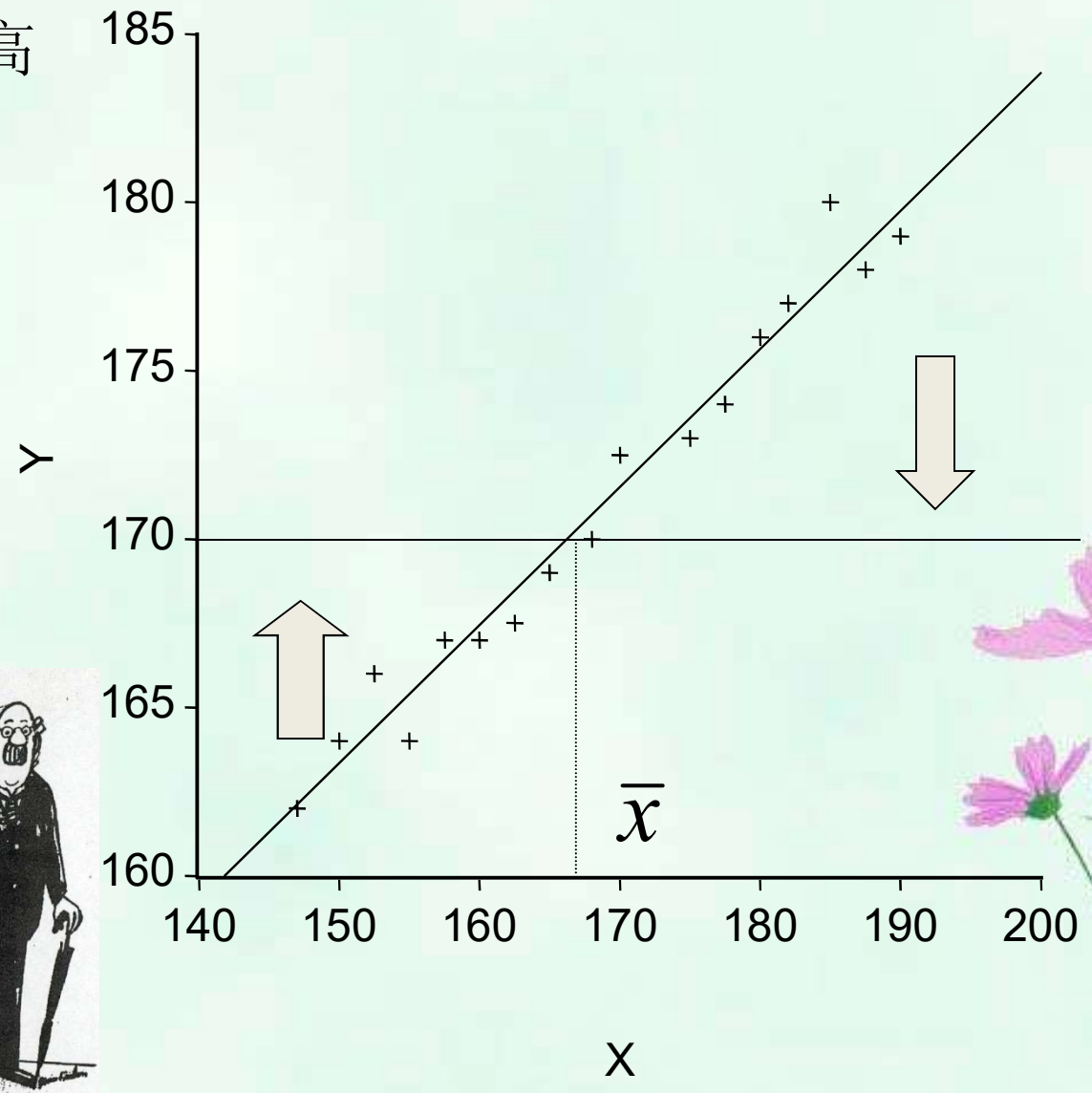
回归的由来：

- 1889年F.Gallton和他的朋友K.Pearson收集了上千个家庭的身高、臂长和腿长的记录
- 企图寻找出儿子们身高与父亲们身高之间关系的具体表现形式
- 下图是根据1078个家庭的调查所作的散点图（略图）

2007st.com 版图设计网

儿子们身高向着平均身高“回归”，以保持种族的稳定





2007st.com 版面设计网

“回归”一词的由来

- 从图上虽可看出，个子高的父亲确有生出个子高的儿子的倾向，同样地，个子低的父亲确有生出个子低的儿子的倾向。得到的具体规律如下：

$$y = a + bx + u$$

$$\hat{y} = 84.33 + 0.516x$$

- 如此以来，高的越来越高，矮的越来越矮。他百思不得其解，同时又发现某人种的平均身高是相当稳定的。最后得到结论：儿子们的身高回复于全体男子的平均身高，即“回归”——见1889年F.Gallton的论文《普用回归定律》。
- 后人将此种方法普遍用于寻找变量之间的规律

2007st.com 版图设计网

回归的现代意义：

一个因变量对若干解释变量依存关系 的研究

回归的目的（实质）：

由固定的解释变量去

估计因变量的平均值

2007st.com 版面设计网

回归分析与相关分析的异同

联系： 1）都是对存在相关关系的变量的统计相关关系的研究；
2）都能测度线性相关程度的大小；
3）都能判断线性相关关系是正相关还是负相关。

区别： 1）相关分析仅仅是从统计数据上测度变量之间的相关程度，不考虑两者之间是否存在因果关系，因而变量的地位在相关分析中是对等的；

回归分析是对变量之间的因果关系的分析，变量的地位是不对等的，有被解释变量和解释变量之分。

2） 相关分析假定所有变量均为随机变量；
回归分析通常假定解释变量是确定的，是非随机变量，被解释变量是随机变量

3）相关分析主要关注变量之间的相关程度和性质，不关注变量之间的具体依赖关系。

回归分析在关注变量之间的相关程度和性质的同时，更关注变量之间的具体依赖关系，因而可以深入分析变量间的依存关系，有可能达到掌握其内在规律的目的，具有更重要的实践意义。

2007st.com 版权设计网

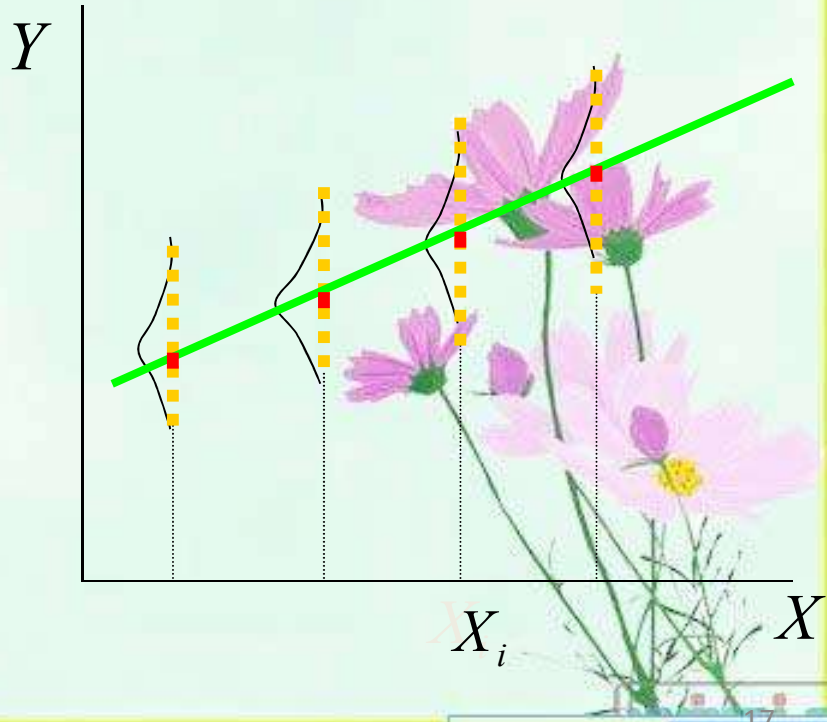
复习几个概念

- **Y**的条件分布

  当解释变量**X**取某固定值时（条件），**Y**的值不确定，**Y**的不同取值形成一定的分布，即**Y**的条件分布。

- **Y**的条件期望

  对于**X**的每一个取值，对**Y**所形成的分布确定其期望或均值，称为**Y**的条件期望或条件均值**E(Y|X)**

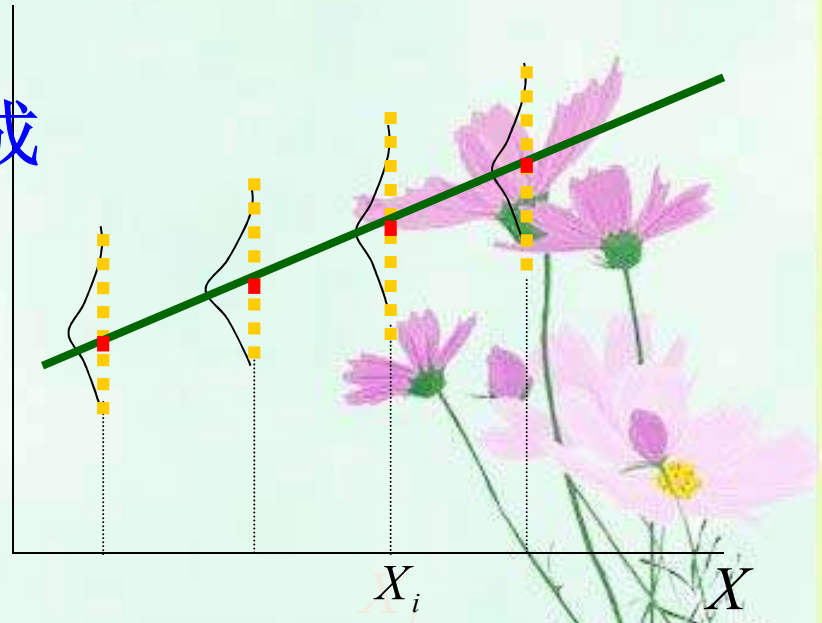


2007st.com 版图设计网

回归线与回归函数

● 回归线:

对于每一个**X**的取值，
都有**Y**的条件期望**E(Y|X)**
与之对应，代表这些**Y**的
条件期望的点的轨迹所形成
的直线或曲线，称为
回归线。



2007st.com 版图设计网

回归线与回归函数

回归函数：因变量 Y 的条件期望 $\mathrm{E}(Y|X_i)$ 随解释变量 X 的变化而有规律的变化，如果把 Y 的条件期望 $\mathrm{E}(Y|X_i)$ 表现为 X 的某种函数

$$E(Y|X_i) = f(X_i)$$

这个函数称为回归函数。

回归函数分为：总体回归函数和样本回归函数

举例：假如已知100个家庭构成的总体。

2007st.com 版权设计网

例:100个家庭构成的总体 (单位:元)

	每月家庭可支配收入 X									
	1000	1500	2000	2500	3000	3500	4000	4500	5000	5500
每月家庭消费支出 Y	820	962	1108	1329	1632	1842	2037	2275	2464	2824
	888	1024	1201	1365	1726	1874	2110	2388	2589	3038
	932	1121	1264	1410	1786	1906	2225	2426	2790	3150
	960	1210	1310	1432	1835	1068	2319	2488	2856	3201
		1259	1340	1520	1885	2066	2321	2587	2900	3288
		1324	1400	1615	1943	2185	2365	2650	3021	3399
			1448	1650	2037	2210	2398	2789	3064	
			1489	1712	2078	2289	2487	2853	3142	
			1538	1778	2179	2313	2513	2934	3274	
			1600	1841	2298	2398	2538	3110		
			1702	1886	2316	2423	2567			
				1900	2387	2453	2610			
				2012	2498	2487	2710			
					2589	2586				
$E(Y\|X_i)$	900	1150	1400	1650	1900	2150	2400	2650	2900	3150

二、总体回归函数（PRF）

1. 总体回归函数的概念

前提：假如已知所研究的经济现象的总体因变量 Y和解释变量 X 的每个观测值，可以计算出总体因变量 Y的条件均值 $\mathrm{E}(Y|X_i)$，并将其表现为解释变量 X 的某种函数

$$\mathrm{E}(Y|X_i)=f(X_i)$$

这个函数称为总体回归函数（**PRF**）

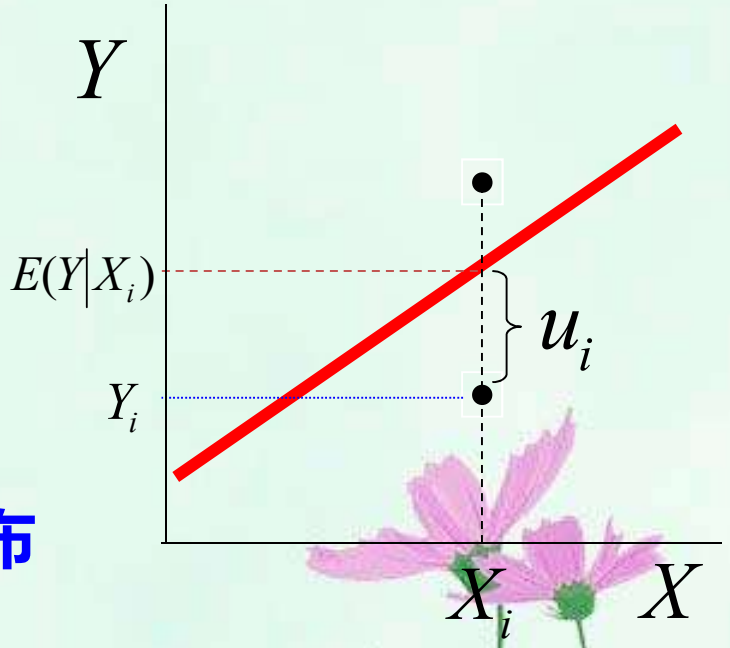
2007st.com 版权设计网

2. 总体回归函数的表现形式

（1）条件均值表现形式

假如 Y 的条件均值 $E(Y|X_i)$ 是解释变量 X 的线性函数，可表示为：

$$E(Y_i|X_i)=f(X_i)=\beta_1+\beta_2X_i$$



（2）个别值表现形式

对于一定的 X_i，Y 的各个别值 Y_i 分布在 $E(Y|X_i)$ 的周围，若令各个 Y_i 与条件均值 $E(Y|X_i)$ 的偏差为 u_i，显然 u_i 是随机变量,则有

$$u_i=Y_i-E(Y_i|X_i)=Y_i-\beta_1-\beta_2X_i$$ **或** $$Y_i=\beta_1+\beta_2X_i+u_i$$

结论：总体线性回归函数$E(Y|Xi)=\beta_1+\beta_2X_i$与$Y_i=\beta_1+\beta_2X_i+u_i$是等价形式。

3. 如何理解总体回归函数

- 实际的经济研究中总体回归函数通常是未知的，只能根据经济理论和实践经验去设定。“计量”的目的就是寻求PRF。
- 总体回归函数中Y与X的关系可是线性的，也可是非线性的。

对线性回归模型的“线性”有两种解释

就变量而言是线性的

——Y的条件均值是X的线性函数

就参数而言是线性的

——Y的条件均值是参数β的线性函数

2007st.com 版图设计网

“线性”的判断

$E(Y_i|X_i)=\beta_1+\beta_2 X_i$ 变量、参数均为“线性”

$E(Y_i|X_i)=\beta_1+\beta_2 X^2_i$ 参数“线性”，变量”非线性”

$E(Y_i|X_i)=\beta_1+\sqrt{\beta_2}X_i$ 变量“线性”，参数”非线性”

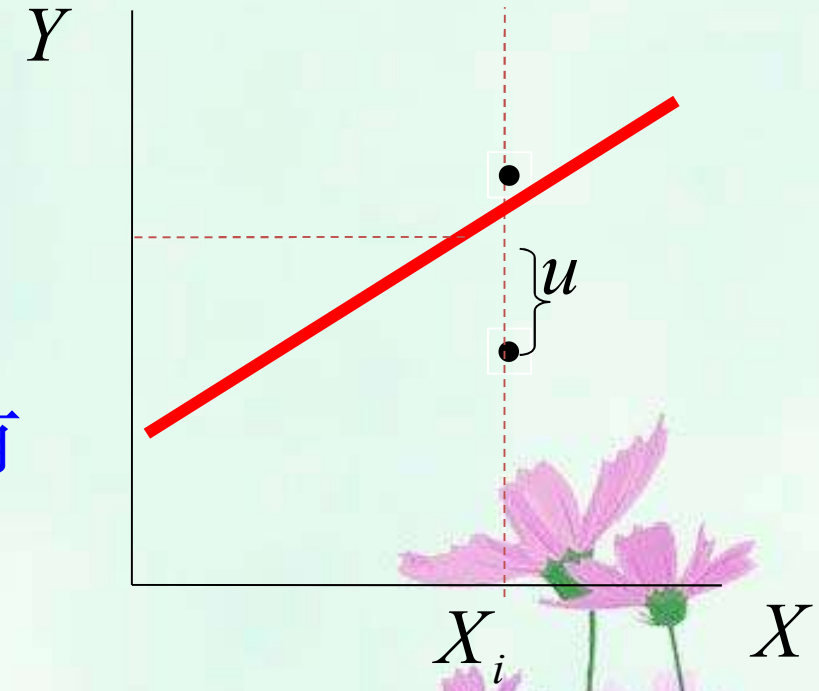
计量经济学中:线性回归模型主要指就参数而言是“线性”,因为只要对参数而言是线性的,都可以用类似的方法估计其参数。

2007st.com 版图设计网

三、随机扰动项

◆概念：

各个 Y_i 值与条件均值 $E(Y|X_i)$ 的偏差 u_i 代表排除在模型以外的所有因素对 Y 的影响。



◆性质：u_i 是期望为0有一定分布的随机变量

重要性：随机扰动项的性质决定着计量经济方法的选择

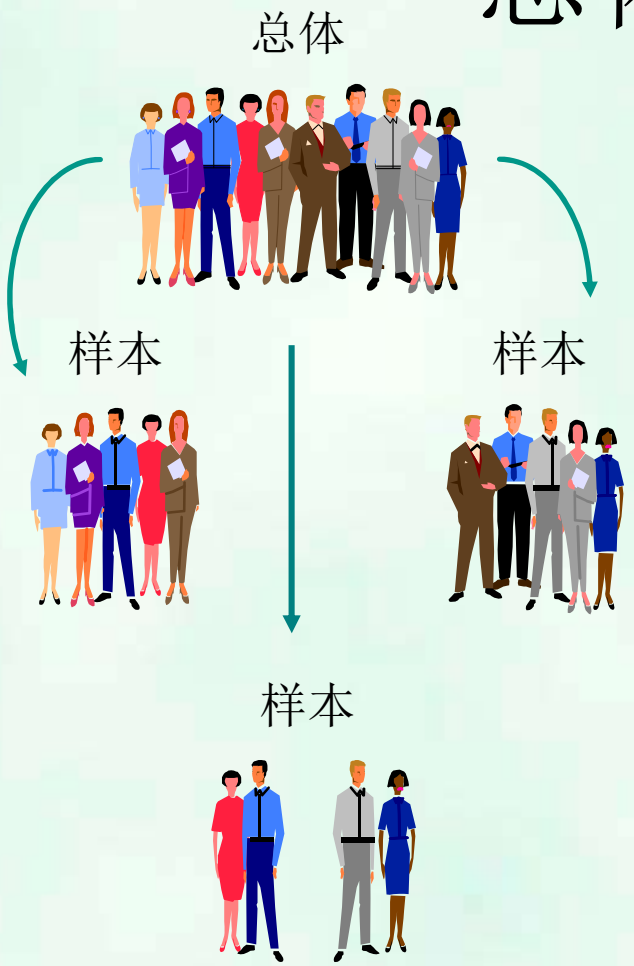
2007st.com 版权设计网

引入随机扰动项的原因

- 未知影响因素的代表
- 无法取得数据的已知影响因素的代表
- 众多细小影响因素的综合代表
- 模型的设定误差
- 变量的观测误差
- 变量内在随机性

2007st.com 版权设计网

总体与样本



- 总体是我们研究的目的，但是不能知道总体的全部数据
- 用总体中的一部分（样本）来推断总体的性质。

2007st.com 版图设计网

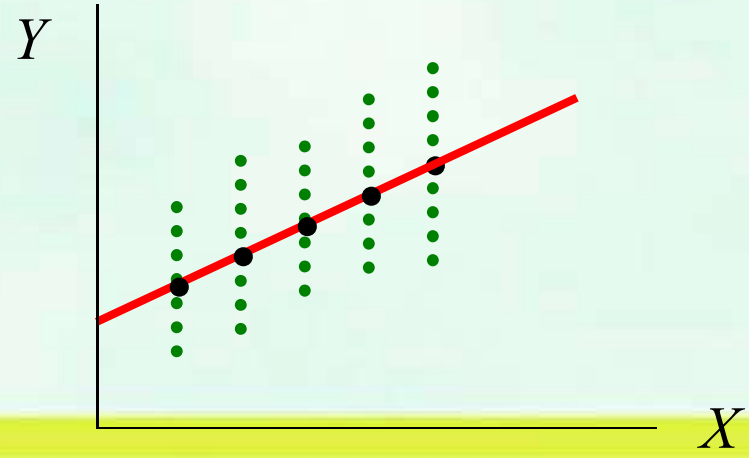
四、样本回归函数（SRF）

样本回归线：

对于 X 的一定值，取得 Y 的样本观测值，可计算其条件均值，样本观测值条件均值的轨迹称为样本回归线。

样本回归函数：

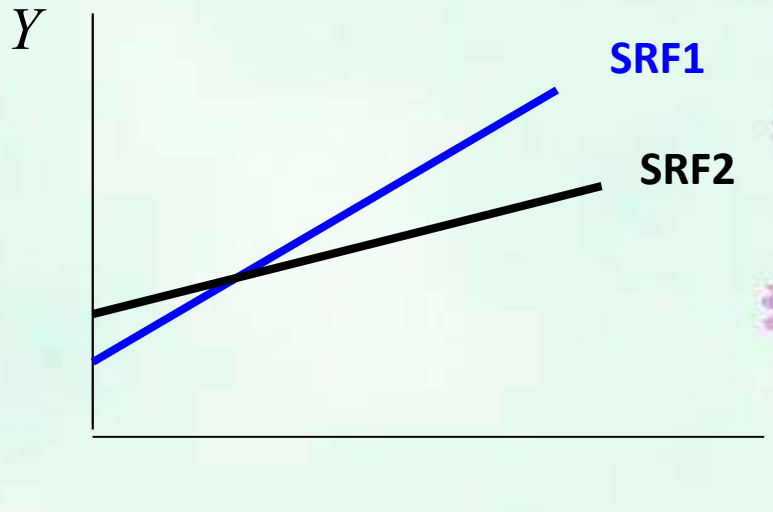
如果把因变量 Y 的样本条件均值表示为解释变量的某种函数，这个函数称为样本回归函数（SRF）。



2007st.com 版权设计网

SRF 的特点

- 每次抽样都能获得一个样本，就可以拟合一条样本回 归线，所以样本回归线随抽样波动而变化，可以有许多条（SRF不唯一）。



2007st.com 版图设计网

- 样本回归函数的函数形式应与设定的总体回归函数的函数形式一致。
- 样本回归线还不是总体回归线，至多只是未知总体回归线的近似表现。

2007st.com 版面设计网

样本回归函数的表现形式

样本回归函数如果为线性函数，可表示为

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i$$

其中：$\hat{Y}_i$是与X_i相对应的Y的样本条件均值

$\hat{\beta}_1$和$\hat{\beta}_2$分别是样本回归函数的参数

因变量Y的实际观测值Y_i不完全等于样本条件均值，二者之差用e_i表示：e_i

$e_i = Y_i - \hat{Y}_i$ 或者 $Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + e_i$

结论：样本的线性回归函数$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i$与$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + e_i$是等价形式其中，$e_i$称为样本的剩余项或残差。

对样本回归的理解

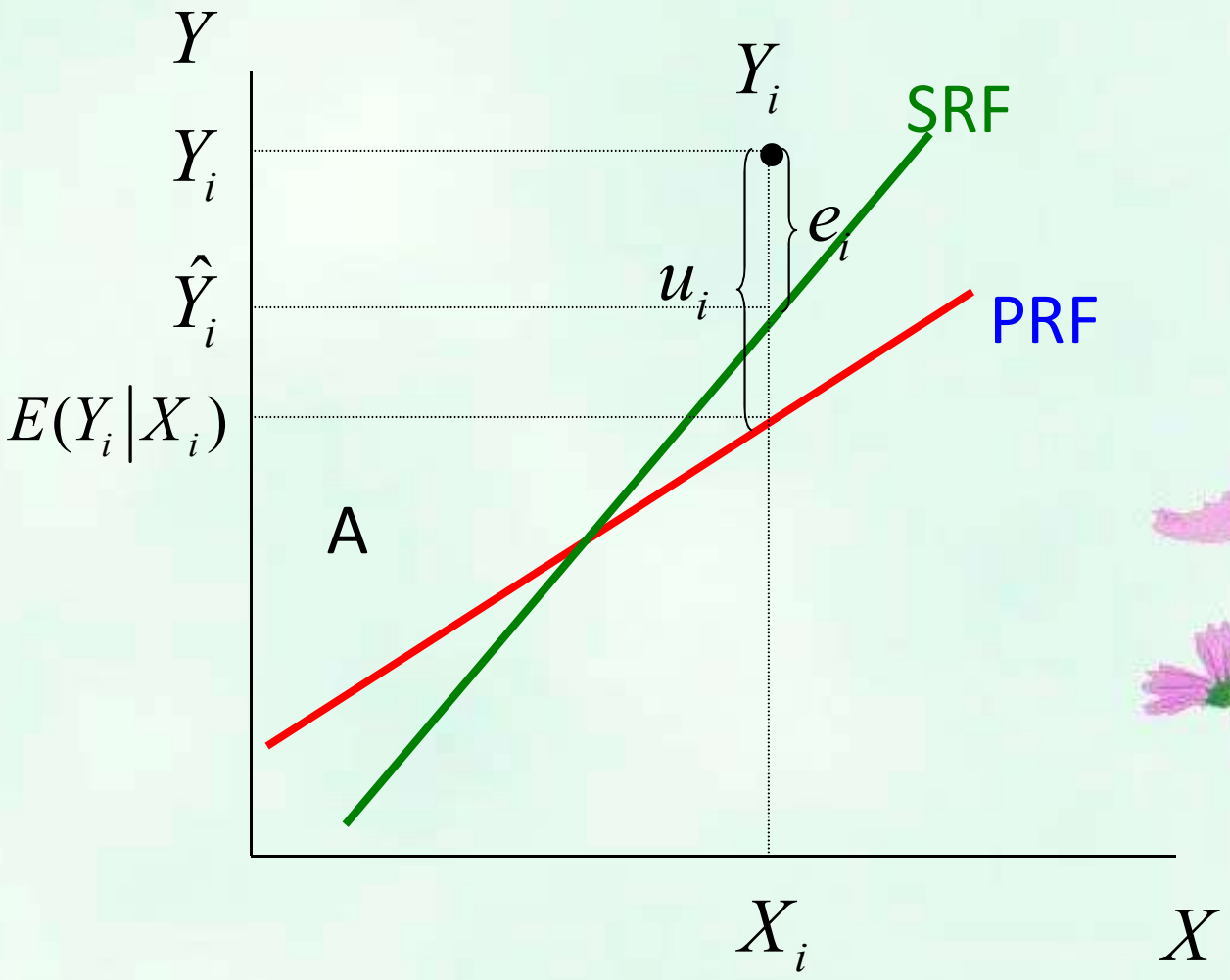
$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + e_i$$

如果能够获得 $\hat{\beta}_1$ 和 $\hat{\beta}_2$的数值，显然：

- $\hat{\beta}_1$ 和$\hat{\beta}_2$是对总体回归函数参数 β_1和β_2的估计
- $\hat{Y}_i$是对总体条件期望 $\mathrm{E}(Y|X_i)$的估计
- e_i在概念上类似总体回归函数中的u_i，可视为对 u_i 的估计。

2007st.com 版面设计网

样本回归函数与总体回归函数的关系



2007st.com 版面设计网

样本回归函数与总体回归函数的区别

首先，总体回归函数虽然未知，但它是确定的；样本回归线却是随抽样波动而变化的，可以有许多条。样本回归线至多只是未知的总体回归线的近似反映。

其次，总体回归函数的参数β_1和β_2是确定的常数；而样本回归函数的参数$\hat{\beta}_1$和$\hat{\beta}_2$是随抽样而变化的随机变量。

此外，总体回归函数中的u_i是不可直接观测的；而样本回归函数中的e_i是只要估计出样本回归的参数就可以计算的数值。

2007st.com 版图设计网

第二节
简单线性回归模型的最小二乘估计

本节基本内容:

- 简单线性回归的基本假定
- 普通最小二乘法
- OLS回归线的性质
- 参数估计式的统计性质

2007st.com 版图设计网

一、简单线性回归的基本假定

1. 为什么要作基本假定？

- 模型中有随机扰动，估计的参数是随机变量，只有对随机扰动的分布作出假定，才能确定所估计参数的分布性质，也才可能进行假设检验和区间估计
- 只有具备一定的假定条件，所作出的估计才具有较好的统计性质。

2007st.com 版面设计网

2、基本假定的内容

（1）对模型和变量的假定

如 $Y_i = \beta_1 + \beta_2 X_i + u_i$

假定解释变量X是非随机的，或者虽然是随机的，但与扰动项u是不相关的

假定解释变量X在重复抽样中为固定值

假定模型中的变量没有测量误差

假定变量和模型无设定误差

2007st.com 版图设计网

（2）对随机扰动项u 的假定

又称高斯假定、古典假定

假定1：零均值假定

在给定X_i的条件下，u_i 的条件期望为零

$$\mathrm{E}(u_i \mid X_i) = 0$$

假定2：同方差假定

在给定X_i的条件下，u_i的条件方差为某个常数σ^2

$$\mathrm{Var}(u_i \mid X_i) = \mathrm{E}[u_i - \mathrm{E}(u_i \mid X_i)]^2 = \sigma^2$$

2007st.com 版图设计网

假定3：无自相关假定

随机扰动项 u_i 的逐次值互不相关

$$Cov(u_i,u_j)=E[u_i-E(u_i)][u_j-E(u_j)]$$

$$=E(u_iu_j)=0 \qquad (i\neq j)$$

假定4：随机扰动 u_i 与解释变量 X_i 不相关

$$Cov(u_i,X_i)=E[u_i-E(u_i)][X_i-E(X_i)]=0$$

假定4也被称为　外生性假定

2007st.com 版图设计网

假定5：对随机扰动项分布的正态性假定

即假定 u_i 服从均值为零、方差为 σ^2 的正态分布

$$u_i \sim N(0, \sigma^2)$$

满足以上古典假定的线性回归模型，也称为古典线性回归模型（**Classical Linear Regression Model,** 简称**CLRM**）

2007st.com 版权设计网

Y的分布性质

由于　　$Y_i = \beta_1 + \beta_2 X_i + u_i$

u_i的分布性质决定了Y_i的分布性质。

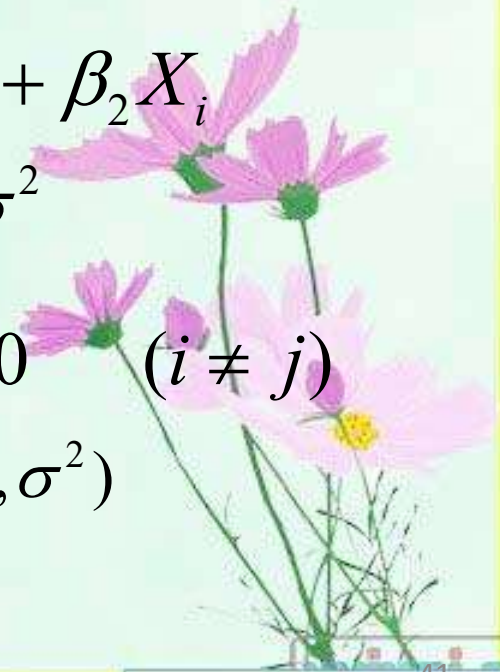
对u_i 的一些假定可以等价地表示为对Y_i的假定：

假定1：零均值假定　　$\mathrm{E}(Y_i \mid X_i) = \beta_1 + \beta_2 X_i$

假定2：同方差假定　　$\mathrm{Var}(Y \mid X_i) = \sigma^2$

假定3：无自相关假定　　$\mathrm{Cov}(Y_i, Y_j) = 0 \quad (i \neq j)$

假定5：正态性假定　　$Y_i \sim N(\beta_1 + \beta_2 X_i, \sigma^2)$



2007st.com 版图设计网

二、普通最小二乘法

（Ordinary Least Squares）

◆OLS的基本思想

- 不同的估计方法可得到不同的样本回归参数$\hat{\beta}_1$和$\hat{\beta}_2$，所估计的$\hat{Y}_i$也不同。
- 理想的估计方法应使Y_i与$\hat{Y}_i$的差即剩余e_i越小越好
- 因e_i可正可负，所以可以取$\sum e_i^2$最小

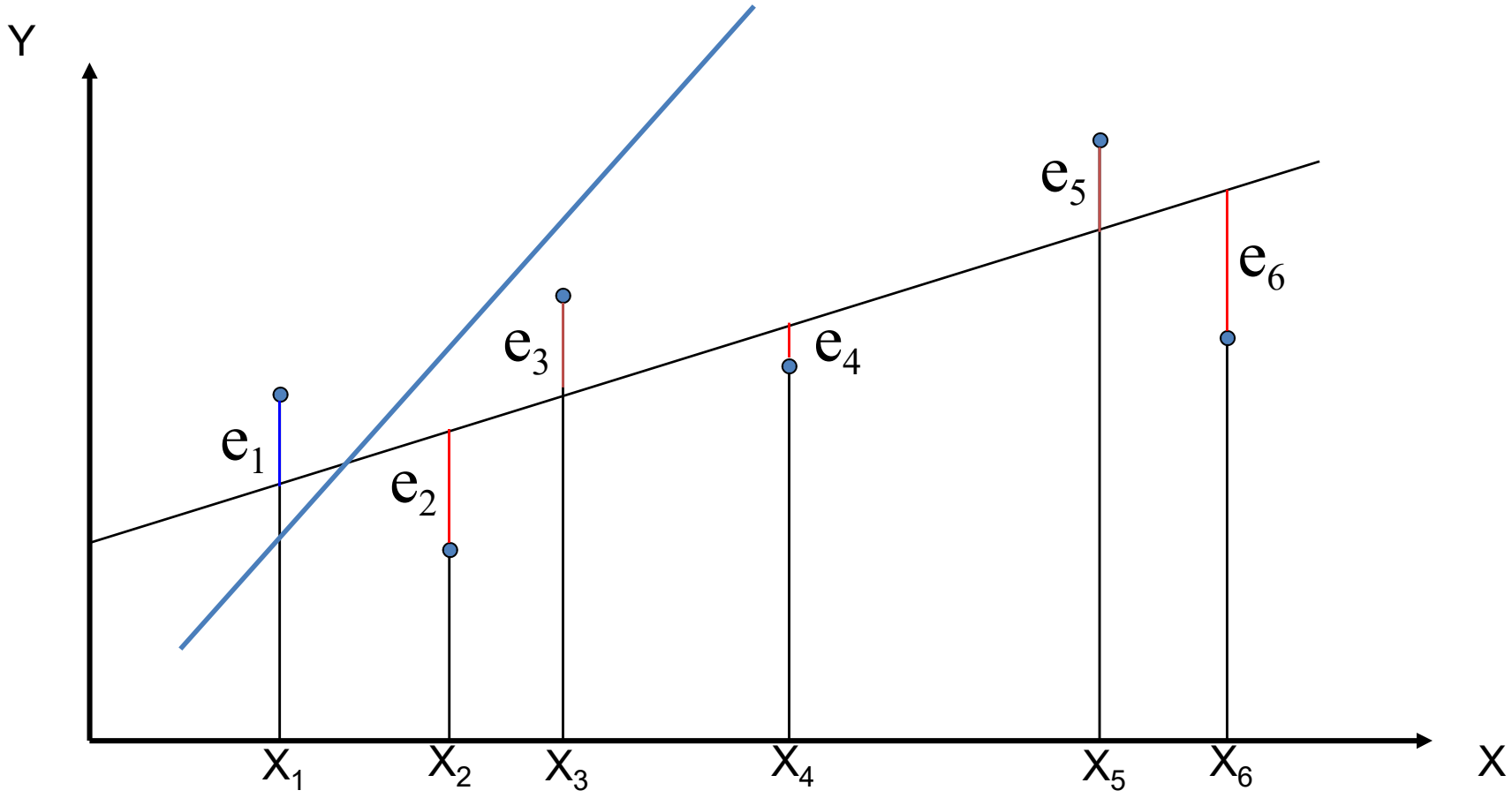
即

$$\min \sum e_i^2 = \min \sum (Y_i - \hat{Y}_i)^2 = \min \sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i)^2$$

2007st.com 版图设计网

普通最小二乘法（OLS）

普通最小二乘法是一种参数估计方法，确定估计参数的准则是使全部观察值的残差平方和最小，即 $\Sigma e_i^2 \to \min$，由此得出选择回归参数 $\hat{\beta}_1$、$\hat{\beta}_2$ 的最小二乘估计式。



正规方程和估计式

取偏导数为0，得正规方程

$$-2\sum(Y_i-\hat{\beta}_1-\hat{\beta}_2 X_i)=0$$

$$-2\sum(Y_i-\hat{\beta}_1-\hat{\beta}_2 X_i)X_i=0$$

或

$$\sum Y_i=n\hat{\beta}_1+\hat{\beta}_2\sum X_i$$

$$\sum X_iY_i=\hat{\beta}_1\sum X_i+\hat{\beta}_2\sum X_i^2$$

用克莱姆法则求解得观测值形式的OLS估计式：

$$\hat{\beta}_2=\frac{n\sum X_iY_i-\sum X_i\sum Y_i}{n\sum X_i^2-(\sum X_i)^2}$$

$$\hat{\beta}_1=\frac{\sum X_i^2\sum Y_i-\sum X_i\sum X_iY_i}{n\sum X_i^2-(\sum X_i)^2}$$

2007st.com 版图设计网

用离差表现的OLS估计式

为表达得更简洁，或者用离差形式OLS估计式：

$$\hat{\beta}_2 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} = \frac{\sum x_i y_i}{\sum x_i^2}$$

$$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X}$$

注意其中：$x_i = X_i - \bar{X}$

$y_i = Y_i - \bar{Y}$

而且样本回归函数可写为 $\hat{y}_i = \hat{\beta}_2 x_i$

【例2.2】见P32

2007st.com 版图设计网

三、OLS回归线的性质

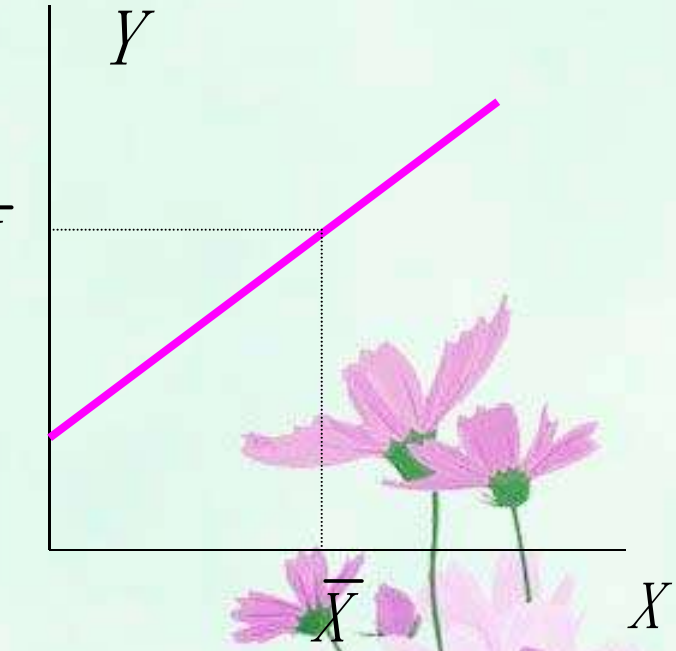
可以证明：

- 回归线通过样本均值

$$\bar{Y} = \hat{\beta}_1 + \hat{\beta}_2 \bar{X}$$

- 估计值 $\hat{Y}_i$ 的均值等于实际观测值 Y_i 的均值

$$\frac{\sum \hat{Y}_i}{n} = \bar{Y}$$



2007st.com 版图设计网

证明 OLS 回归线的第 2 条性质 $\bar{\hat{Y}} = \bar{Y}$

证明：由 $\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i = (\bar{Y} - \hat{\beta}_2 \bar{X}) + \hat{\beta}_2 X_i = \bar{Y} + \hat{\beta}_2 (X_i - \bar{X})$

所以有 $\dfrac{\sum \hat{Y}_i}{n} = \dfrac{\sum [\bar{Y} + \hat{\beta}_2 (X_i - \bar{X})]}{n} = \bar{Y} + \hat{\beta}_2 \dfrac{\sum (X_i - \bar{X})}{n}$

而 $\dfrac{\sum (X_i - \bar{X})}{n} = \dfrac{\sum X_i}{n} - \bar{X} = \bar{X} - \bar{X} = 0$

所以有 $\bar{\hat{Y}} = \bar{Y}$

2007st.com 版权设计网

证明OLS回归线的第3条性质：$\bar{e}=0$

法1：　由定义　$e_i = Y_i - \hat{Y}_i = Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i$

所以　$\bar{e} = \dfrac{\sum e_i}{n} = \dfrac{\sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i)}{n} = \bar{Y} - \hat{\beta}_1 - \hat{\beta}_2 \bar{X} = 0$

又因为　$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X}$　　(2.29)

所以　$\bar{e}_i = \bar{Y} - \bar{Y} + \hat{\beta}_2 \bar{X} - \hat{\beta}_2 \bar{X} = 0$

法2：　$e_i = Y_i - \hat{Y}_i$

所以　$\bar{e}_i = \bar{Y} - \bar{\hat{Y}} = 0$

2007st.com 版权设计网

证明OLS回归线的第4条性质：$Cov(\hat{Y},e)=0$

$$Cov\left(\hat{Y}_i,e_i\right)=E\left\{\left(\hat{Y}_i-\bar{\hat{Y}}\right)\left(e_i-\bar{e}\right)\right\}=E\left(\hat{Y}_ie_i-\bar{\hat{Y}}e_i-\hat{Y}_i\bar{e}+\bar{\hat{Y}}\bar{e}\right)$$

$\because \bar{e}_i=0$（性质3）

$$\therefore Cov\left(\hat{Y}_i,e_i\right)=E\left(\hat{Y}_ie_i-\bar{\hat{Y}}e_i-\hat{Y}_i\bar{e}_i+\bar{\hat{Y}}\bar{e}_i\right)=E\left[\left(\hat{Y}_i-\bar{\hat{Y}}\right)e_i\right]$$

根据离差公式$\hat{y}_i=\hat{Y}_i-\bar{\hat{Y}}$

$$\therefore Cov\left(\hat{Y}_i,e_i\right)=E\left(\hat{y}_i\cdot e_i\right)$$

Continued

2007st.com 版图设计网

Continued

$$\because e_i = Y_i - \hat{Y}_i \text{(2.11式)和} \hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X} \text{ (2.29式)}$$

$$\therefore e_i = Y_i - \hat{Y}_i = Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i = Y_i - \left(\bar{Y} - \hat{\beta}_2 \bar{X}\right) - \hat{\beta}_2 X_i$$

$$= \left(Y_i - \bar{Y}\right) - \left(\hat{\beta}_2 X_i - \hat{\beta}_2 \bar{X}\right) = \mathrm{y}_i - \hat{\beta}_2 x_i \text{（}\mathrm{y}_i\text{、}x_i\text{都是离差）}$$

$$\therefore \sum e_i \hat{\mathrm{y}}_i = \sum \hat{\mathrm{y}}_i \left(y_i - \hat{\beta}_2 x_i\right)$$

$$\text{又} \because \hat{\mathrm{y}}_i = \hat{\beta}_2 x_i$$

$$\therefore \sum e_i \hat{\mathrm{y}}_i = \sum \hat{\mathrm{y}}_i \left(y_i - \hat{\beta}_2 x_i\right) = \sum \left[\left(\hat{\beta}_2 x_i\right)\left(y_i - \hat{\beta}_2 x_i\right)\right]$$

$$= \hat{\beta}_2 \sum x_i y_i - \hat{\beta}_2^{\,2} \sum x_i^{\,2}$$

2007st.com 版图设计网

$$\because \sum x_i y_i = \hat{\beta}_2 \sum x_i^2 \text{（2.28式）}$$

$$\therefore \hat{\beta}_2 \sum x_i y_i = \sum x_i^2$$

$$\therefore \sum e_i \hat{y}_i = \hat{\beta}_2 \sum x_i y_i - \hat{\beta}_2^2 \sum x_i^2$$

$$= \hat{\beta}_2^2 \sum x_i^2 - \hat{\beta}_2^2 \sum x_i^2 = 0$$

$$\therefore Cov\left(\hat{Y}_i, e_i\right) = Ee_i \cdot \hat{y}_i = 0$$

2007st.com 版图设计网

证明 OLS 回归线第 5 条性质：$Cov(X,e)=0$

$$Cov(X_i,e_i)=E\left[(X_i-\bar{X})(e_i-\bar{e})\right]$$

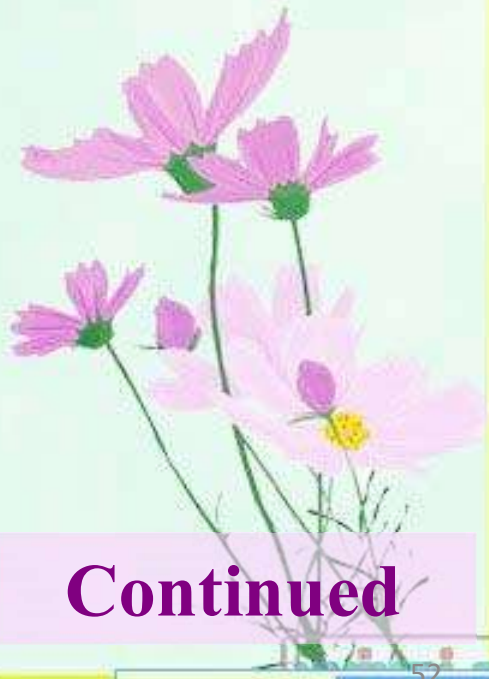
$$=\frac{1}{n}\sum\left(e_iX_i-\bar{e}X_i-e_i\bar{X}+\bar{e}\bar{X}\right)$$

$$=\frac{1}{n}\sum e_iX_i-\bar{e}\frac{\sum X_i}{n}-\bar{X}\frac{\sum e_i}{n}+\bar{e}\bar{X}$$

$$=\frac{1}{n}\sum e_iX_i-\bar{e}\frac{\sum X_i}{n}-\bar{X}\bar{e}+\bar{e}\bar{X}$$

由$\bar{e}$=0（性质3）得：

$$Cov(X_i,e_i)=\frac{1}{n}\sum e_iX_i \quad (*)$$



Continued

2007st.com 版图设计网

Continued

又∵求$\sum e_i^2$极值式的偏导式第二式

得证$Cov(X_i, e_i) = \frac{1}{n}\sum e_i X_i$

$$= \frac{1}{n}\sum\left(Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i\right)X_i = 0$$

2007st.com 版图设计网

- 剩余项 e_i 的均值为零

$$\bar{e} = \frac{\sum e_i}{n} = 0$$

- 因变量估计值 $\hat{Y}_i$ 与剩余项 e_i 不相关

$$\mathrm{Cov}(\hat{Y}_i, e_i) = 0$$

- 解释变量 X_i 与剩余项 e_i 不相关

$$\mathrm{Cov}(X_i, e_i) = 0$$

2007st.com 版图设计网

四、参数估计式的统计性质

(一)参数估计式的评价标准

1. 无偏性

前提：重复抽样中估计方法固定、样本数不变、经重复抽样的观测值，可得一系列参数估计值

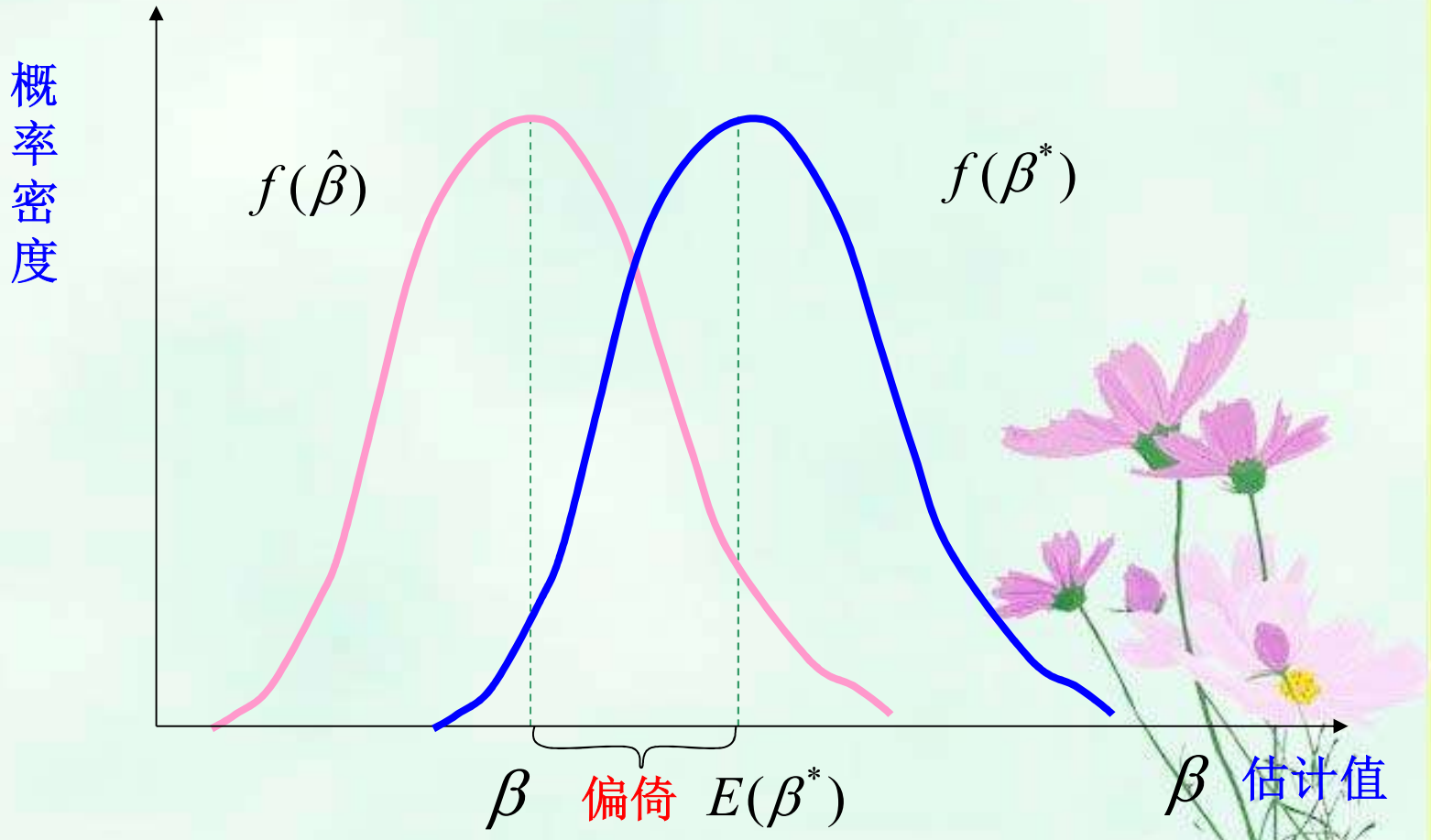
参数估计值 $\hat{\beta}$ 的分布称为 $\hat{\beta}$ 的抽样分布，密度函数记为 $f(\hat{\beta})$

如果 $\mathrm{E}(\hat{\beta})=\beta$，称 $\hat{\beta}$ 是参数 β 的无偏估计式，否则称 $\hat{\beta}$ 是有偏的，其偏倚为 $\mathrm{E}(\hat{\beta})-\beta$

（见图1.2）

2007st.com 版图设计网

图 1.2



2007st.com 版图设计网

2. 有效性

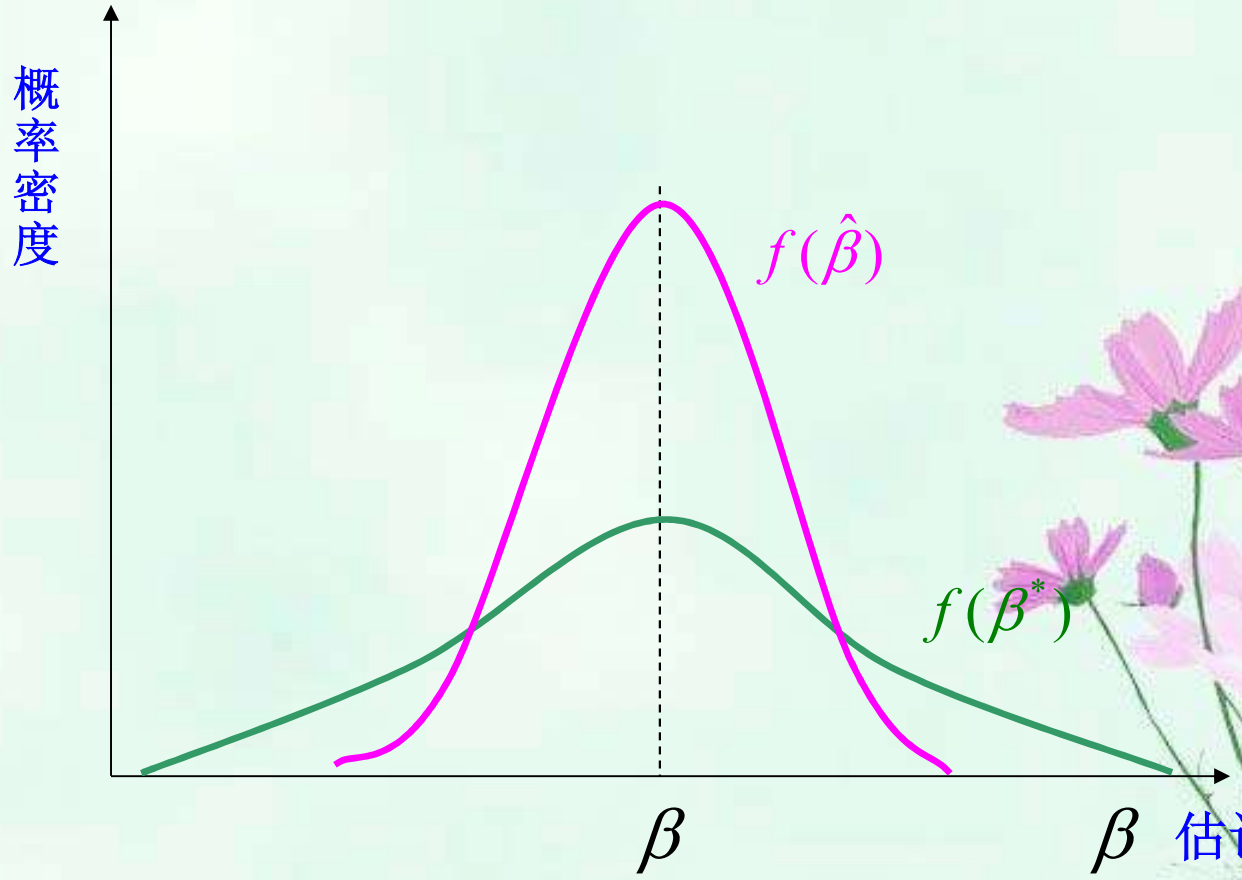
前提：样本相同、用不同的方法估计参数，
可以找到若干个不同的估计式

目标：努力寻求其抽样分布具有最小方差的
估计式。

既是无偏的同时又具有最小方差的估计式，称为最佳无偏估计式，或称为有效估计式。

2007st.com 版权设计网

图 1.3



2007st.com 版图设计网

3. 一致性

思想:当样本容量较小时，有时很难找到最佳无偏估计，需要考虑样本扩大后的性质

一致性：

当样本容量 n 趋于无穷大时，如果估计式$\hat{\beta}$依概率收敛于总体参数的真实值，就称这个估计式$\hat{\beta}$是β的一致估计量。

$$\lim_{n\to\infty} P(\left|\hat{\beta}-\beta\right|)<\varepsilon)=1$$

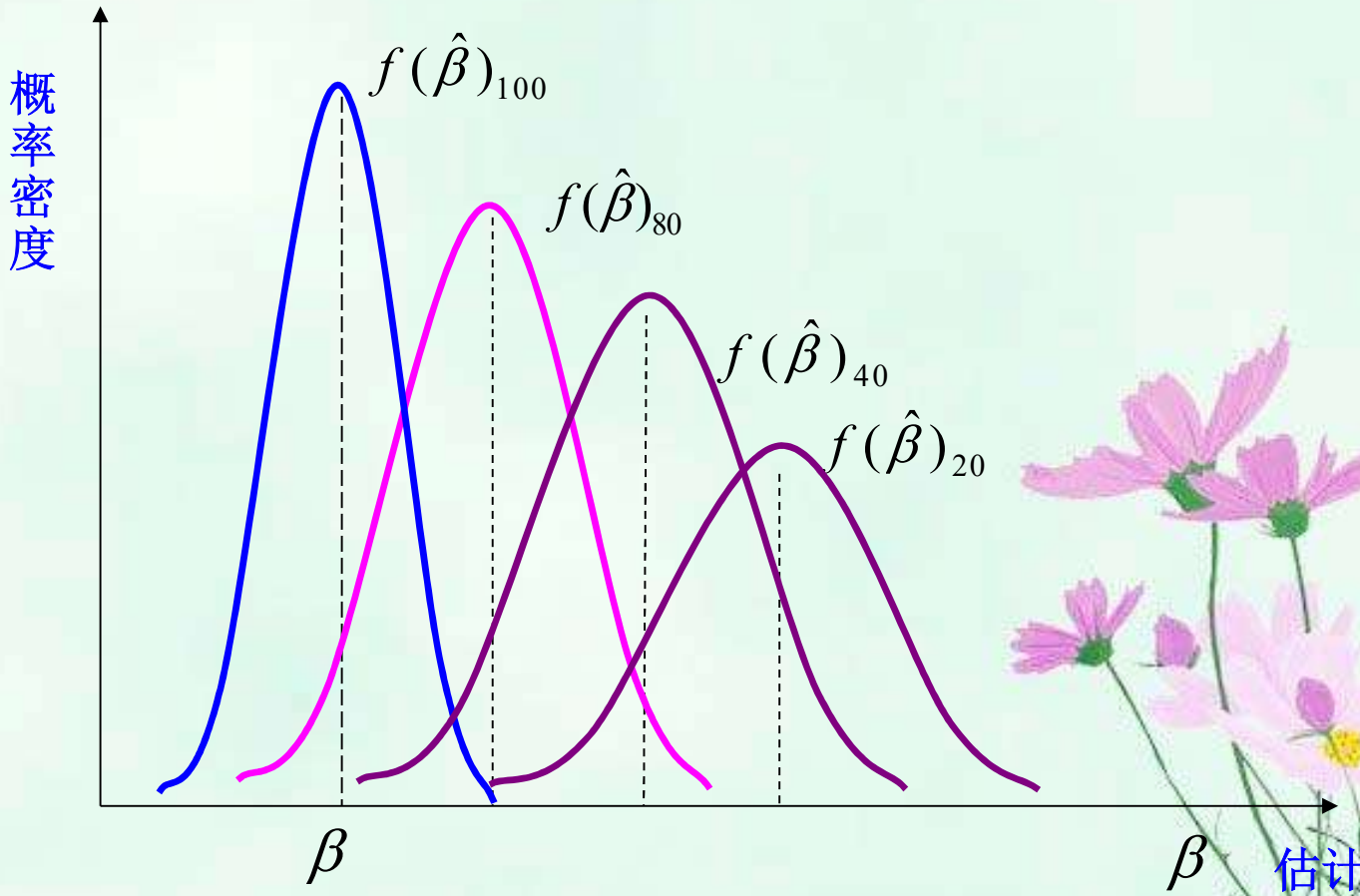
或

$$P\lim_{n\to\infty}(\hat{\beta})=\beta$$

渐近有效性：当样本容量 n 趋于无穷大时，在所有的一致估计式中，具有最小的渐近方差。(见图1.4)

2007st.com 版图设计网

图 1.4



2007st.com 版权设计网

4. 渐近无偏性

即样本容量趋于无穷大时，是否它的均值序列趋于总体真值。

5. 渐近有效性

即样本容量趋于无穷大时，是否它在所有的一致估计量中具有最小的渐近方差。

2007st.com 版面设计网

（二） OLS估计式的统计性质

- 由OLS估计式可以看出

$$\hat{\beta}_2 = \frac{n\sum X_i Y_i - \sum X_i \sum Y_i}{n\sum X_i^2 - (\sum X_i)^2} \qquad \hat{\beta}_1 = \frac{\sum X_i^2 \sum Y_i - \sum X_i \sum X_i Y_i}{n\sum X_i^2 - (\sum X_i)^2}$$

$\hat{\beta}_k$ 由可观测的样本值 X_i 和 Y_i 唯一表示。

- 因存在抽样波动，OLS估计 $\hat{\beta}_k$ 是随机变量
- OLS估计式是点估计式

2007st.com 版面设计网

OLS估计式的统计性质

1. 线性特征　$\hat{\beta}_k$ 是 Y 的线性函数

$$\hat{\beta}_2 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} = \frac{\sum x_i y_i}{\sum x_i^2} = \sum k_i y_i \qquad k_i = \frac{x_i}{\sum x_i^2}$$

2. 无偏特性

$$E(\hat{\beta}_k) = \beta_k$$ （证明见教材P37）

3. 最小方差特性

（证明见教材P68附录2·1）

$$Var(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_i^2} \qquad Var(\hat{\beta}_1) = \sigma^2 \frac{\sum X_i^2}{N \sum x_i^2}$$

在所有的线性无偏估计中，OLS估计 $\hat{\beta}_k$ 具有最小方差

高斯—马尔可夫定理(Gauss-Markov theorem):

在给定经典线性回归的假定下，最小二乘估计量是具有最小方差的线性无偏估计量（BLUE）。

2007st.com 版权设计网

1. 线性性 (linear)

$$\hat{\beta}_2 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} = \frac{\sum x_i y_i}{\sum x_i^2} = \sum \frac{x_i}{\sum x_i^2} y_i = \sum k_i \mathrm{Y}_i$$

其中，$k_i = \dfrac{x_i}{\sum x_i^2}$

$$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X} = \bar{Y} - \sum k_i Y_i \times \bar{X} = \frac{1}{n} \sum Y_i - \sum k_i Y_i \times \bar{X}$$

$$= \sum \left(\frac{1}{n} - k_i \bar{X} \right) Y_i = \sum l_i Y_i$$

其中，$li = \dfrac{1}{n} - k_i \bar{X}$

2007st.com 版面设计网

2. 无偏性(unbiased)

$$\hat{\beta}_2 = \sum k_i Y_i = \sum k_i (\beta_1 + \beta_2 X_i + u_i)$$

$$= \beta_1 \sum k_i + \beta_2 \sum k_i X_i + \sum k_i u_i$$

$$= \beta_2 + \sum k_i u_i$$

$$\therefore \quad E(\hat{\beta}_2) = \beta_2$$

$$\hat{\beta}_1 = \sum l_i Y_i = \sum l_i (\beta_1 + \beta_2 X_i + u_i)$$

$$= \beta_1 \sum l_i + \beta_2 \sum l_i X_i + \sum k_i u_i$$

$$= \beta_1 + \sum l_i u_i$$

$$\therefore \quad E(\hat{\beta}_1) = \beta_1$$

2007st.com 版权设计网

3. 有效性 (efficient)

$$Var\left(\hat{\beta}_1\right)=Var\left(\sum l_iY_i\right)$$

由假设$Cov\left(Y_i,Y_j\right)=0$

$$Var\left(\hat{\beta}_1\right)=\sum Var\left(l_iY_i\right)=\sum l_i^{\,2}\times Var\left(Y_i\right)$$

$$=\sum l_i^2\times\sigma^2=\sigma^2\sum l_i^2=\sigma^2\sum\left(\frac{1}{n}-k_i\overline{X}\right)^2$$

因为$\sum k_i^{\,2}=\dfrac{1}{\sum x_i^{\,2}},\sum k_i=0$

故 $\sum l_i^{\,2}=\dfrac{1}{n}+\overline{X}^2\sum k_i^{\,2}-\dfrac{2}{n}\times\overline{X}\sum k_i=\dfrac{1}{n}+\overline{X}^2\sum k_i^{\,2}=\dfrac{1}{n}+\dfrac{\overline{X}^2}{\sum x_i^{\,2}}$

$$Var\left(\hat{\beta}_1\right)=\sigma^2\times\sum l_i^{\,2}=\sigma^2\times\left(\frac{1}{n}+\frac{\overline{X}^2}{\sum x_i^{\,2}}\right)=\sigma^2\times\frac{\sum x_i^{\,2}+n\overline{X}^2}{n\sum x_i^{\,2}}$$

$$=\sigma^2\times\frac{\sum x_i^{\,2}+n\overline{X}^2}{n\sum x_i^{\,2}}=\sigma^2\times\frac{\sum X_i^{\,2}-n\overline{X}^2+n\overline{X}^2}{n\sum x_i^{\,2}}=\sigma^2\times\frac{\sum X_i^{\,2}}{n\sum x_i^{\,2}}$$

因为:$\sum x_i^{\,2}=\sum\left(X_i-\overline{X}\right)^2=\sum\left(X_i^{\,2}-2\overline{X}X_i+\overline{X}^2\right)$

$$=\sum X_i^{\,2}-2\overline{X}\sum X_i+n\overline{X}^2=\sum X_i^{\,2}-n\overline{X}^2$$

综上：　$Var\left(\hat{\beta}_1\right)=\sigma^2\times\frac{\sum X_i^{\,2}}{n\sum x_i^{\,2}}$

2007st.com 版图设计网

$$Var\left(\hat{\beta}_2\right)=Var\left(\sum k_iY_i\right)$$

由假设$Cov\left(Y_i,Y_j\right)=0$

$$Var\left(\hat{\beta}_2\right)=\sum Var\left(k_iY_i\right)=\sum k_i^{\ 2}\times Var\left(Y_i\right)$$

$$=\sum k_i^{\ 2}\times\sigma^2=\sigma^2\sum k_i^{\ 2}=\frac{\sigma^2}{\sum x_i^{\ 2}}$$

因为$\sum k_i^{\ 2}=\frac{1}{\sum x_i^{\ 2}}$

综上，$Var\left(\hat{\beta}_2\right)=\frac{\sigma^2}{\sum x_i^{\ 2}}$

2007st.com 版图设计网

证明有效性，即最小方差性

假设 $\hat{\beta}_1^*$ 是其他估计方法得到的关于 β_1 的线性无偏估计量：

$$\hat{\beta}_1^* = \sum c_i Y_i$$

其中，$c_i=l_i+d_i$，d_i为不全为零的常数

则容易证明

$$\mathrm{var}(\hat{\beta}_1^*) \geq \mathrm{var}(\hat{\beta}_1)$$

同理，可以证明 $\widehat{\beta_2}$ 具有最小方差性。

普通最小二乘估计量（ordinary least Squares Estimators）称为**最佳线性无偏估计量（best linear unbiased estimator, BLUE）**

2007st.com 版权设计网

$\hat{\beta}_1$、$\hat{\beta}_2$的标准差： (Standard Deviation)

$$SD(\hat{\beta}_1) = \sigma\sqrt{\frac{\sum X_I^2}{n\sum x_i^2}}$$

$$SD(\hat{\beta}_2) = \sigma\frac{1}{\sqrt{\sum x_i^2}}$$

2007st.com 版权设计网

随机误差项μ的方差σ^2的估计

在参数估计量的方差和标准差公式中，包含随机扰动项u的方差σ^2．而σ^2通常是未知的。

由于随机项μ_i不可观测，只能从μ_i的估计——残差e_i出发，对总体方差进行估计。

可以证明，σ^2的**最小二乘估计量**为

$$\hat{\sigma}^2 = \frac{\sum e_i^2}{n-2}$$

它是关于σ^2的无偏估计量。

2007st.com 版权设计网

$\widehat{\beta_1}$和$\widehat{\beta_2}$的方差估计值和标准误（Standard Error）

$$\widehat{Var}(\widehat{\beta_1}) = \frac{\sum e_i^2}{n-2} * \frac{\sum X_i^2}{n\sum x_i^2}$$

$$SE(\widehat{\beta_1}) = \sqrt{\frac{\sum e_i^2}{n-2} * \frac{\sum X_i^2}{n\sum x_i^2}} = \hat{\sigma}\sqrt{\frac{\sum X_i^2}{n\sum x_i^2}}$$

$$\widehat{Var}(\widehat{\beta_2}) = \frac{\sum e_i^2}{n-2} * \frac{1}{\sum x_i^2}$$

$$SE(\widehat{\beta_2}) = \sqrt{\frac{\sum e_i^2}{n-2} * \frac{1}{\sum x_i^2}} = \hat{\sigma}\frac{1}{\sqrt{\sum x_i^2}}$$

2007st.com 版权设计网

例:从100个家庭构成的总体中抽出10个家庭 (单位:元)

	每月家庭可支配收入 X									
	1000	1500	2000	2500	3000	3500	4000	4500	5000	5500
每月家庭消费支出 Y	820	1024	1310	1886	2589	2487	2321	2789	3274	3399

2007st.com 版面设计网

Exercise

- 令kids表示一名妇女生育孩子的数目，educ表示该妇女接受过教育的年数。生育率对教育年数的简单线性回归模型为

$$kids = \beta_0 + \beta_1 educ + u$$

（1）随机扰动项包含什么样的因素？它们可能与教育水平相关吗？

（2）上述简单回归分析能够揭示教育对生育率在其他条件不变下的影响吗？请解释。

2007st.com 版图设计网

（1）随机扰动项包含什么样的因素？它们可能与教育水平相关吗？

解答：收入、年龄、家庭状况、政府的相关政策等也是影响生育率的重要的因素，在上述简单回归模型中，它们被包含在了随机扰动项之中。有些因素可能与教育水平相关，如收入水平与教育水平往往呈正相关、年龄大小与教育水平呈负相关等。

2007st.com 版图设计网

（2）上述简单回归分析能够揭示教育对生育率在其他条件不变下的影响吗？请解释。

解答：当归结在随机扰动项中的重要影响因素与模型中的教育水平**educ**相关时，上述回归模型不能够揭示教育对生育率在其他条件不变下的影响，因为这时出现解释变量与随机扰动项相关的情形，基本假设**4**不满足。

2007st.com 版图设计网

第三节　拟合优度的度量

本节基本内容:

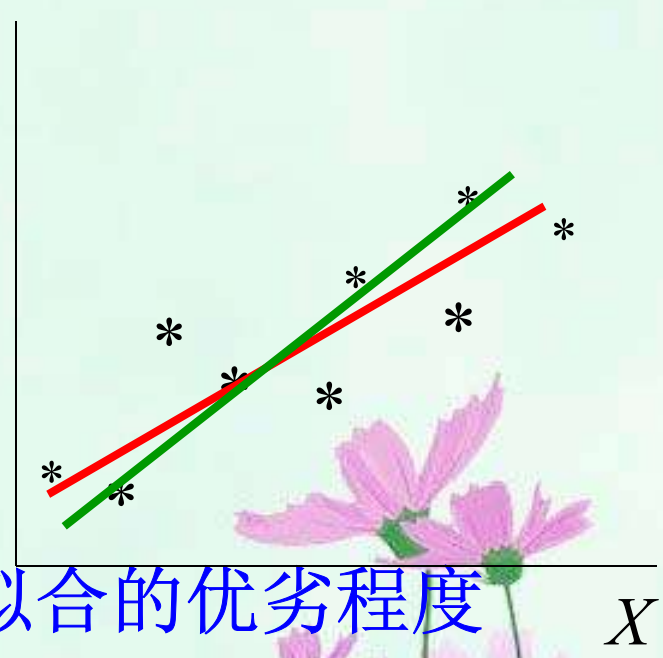
- 什么是拟合优度
- 总变差的分解
- 可决系数

2007st.com 版权设计网

一、什么是拟合优度?

概念:

样本回归线是对样本数据的一种拟合，不同估计方法可拟合出不同的回归线，拟合的回归线与样本观测值总有偏离。



样本回归线对样本观测数据拟合的优劣程度

——拟合优度

拟合优度的度量建立在对总变差分解的基础上

2007st.com 版权设计网

二、总变差的分解

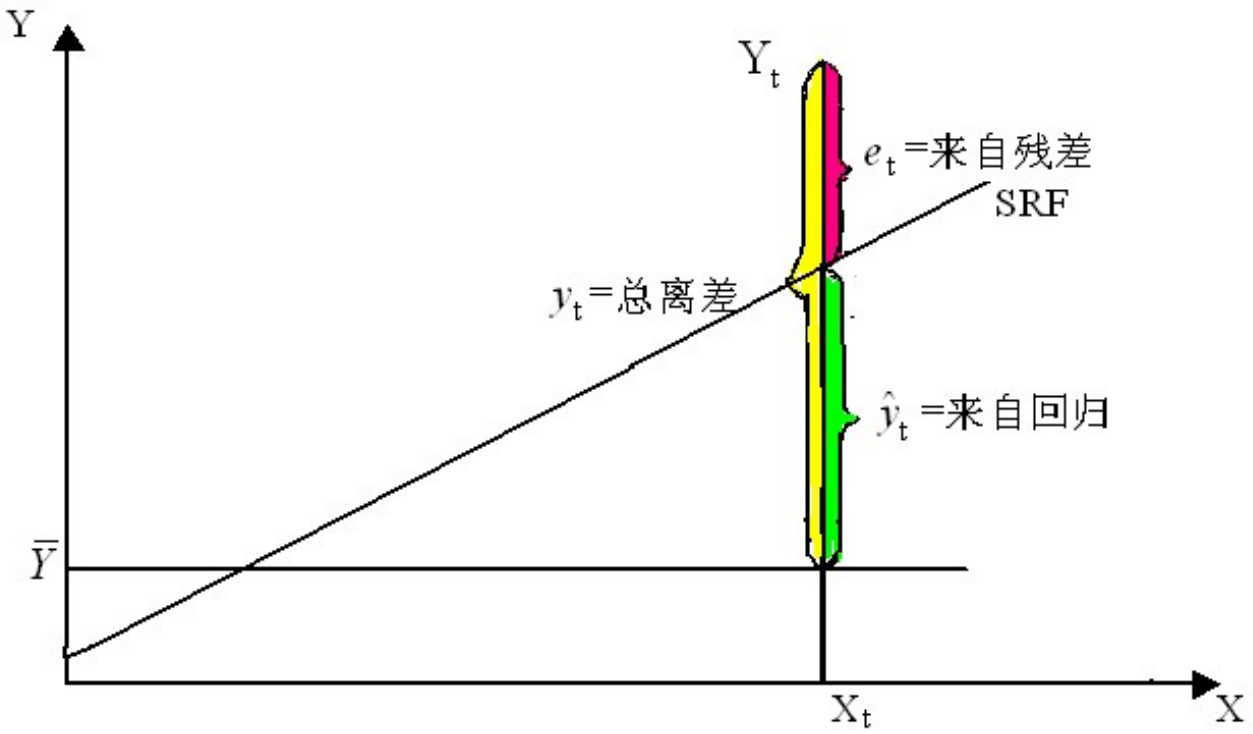
已知由一组样本观测值（X_i,Y_i），i=1,2…,n得到如下样本回归直线

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i$$

而Y的第i个观测值与样本均值的离差 $y_i = (Y_i - \bar{Y})$ 可分解为两部分之和

$$y_i = Y_i - \bar{Y} = (Y_i - \hat{Y}_i) + (\hat{Y}_i - \bar{Y}) = e_i + \hat{y}_i$$

2007st.com 版图设计网



$\hat{y}_i = (\hat{Y}_i - \bar{Y})$是样本回归拟合值与观测值的平均值之差，可认为是**由回归直线解释的部分**；

$e_i = (Y_i - \hat{Y}_i)$是实际观测值与回归拟合值之差，是**回归直线不能解释的部分**。

2007st.com 版图设计网

对于所有样本点，则需考虑这些点与样本均值离差的平方和,可以证明：

$$\sum y_i^2 = \sum \hat{y}_i^2 + \sum e_i^2 + 2\sum \hat{y}_i e_i$$

$$= \sum \hat{y}_i^2 + \sum e_i^2$$

记 $TSS = \sum y_i^2 = \sum (Y_i - \bar{Y})^2$ 总离差平方和（**Total Sum of Squares**）

$ESS = \sum \hat{y}_i^2 = \sum (\hat{Y}_i - \bar{Y})^2$ 回归平方和（**Explained Sum of Squares**）

$RSS = \sum e_i^2 = \sum (Y_i - \hat{Y}_i)^2$ 残差平方和（**Residual Sum of Squares**）

$$\sum (Y_i - \bar{Y})^2 = \sum (\hat{Y}_i - \bar{Y})^2 + \sum (Y_i - \hat{Y}_i)^2$$

2007st.com 版权设计网

三、可决系数

以TSS同除总变差等式两边：

$$\frac{TSS}{TSS}=\frac{ESS}{TSS}+\frac{RSS}{TSS} \qquad \text{或} \qquad 1=\frac{\sum \hat{y}^2}{\sum y_i^2}+\frac{\sum e_i^2}{\sum y_i^2}$$

定义：回归平方和（解释了的变差ESS）$\sum \hat{y}_i^2$ 在总变差（TSS）$\sum y_i^2$ 中所占的比重称为可决系数，用 r^2 表示：

$$R^2=\frac{\sum \hat{y}^2}{\sum y_i^2} \qquad \text{或} \qquad R^2=1-\frac{\sum e_i^2}{\sum y_i^2}$$

2007st.com 版图设计网

可决系数的作用

作用：可决系数越大，说明在总变差中由模型作出了解释的部分占的比重越大，模型拟合优度越好。反之可决系数小，说明模型对样本观测值的拟合程度越差。

2007st.com 版面设计网

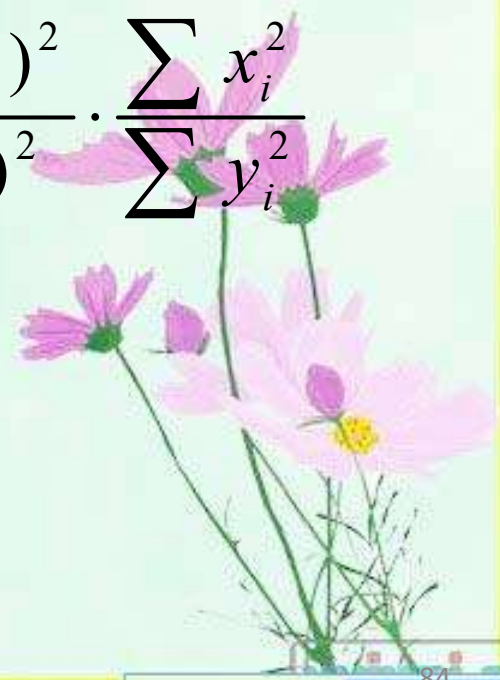
可决系数与相关系数的关系

（1）联系

数值上，可决系数等于因变量与解释变量之间简单相关系数的平方:

$$R^2 = \frac{\sum \hat{y}_i^2}{\sum y_i^2} = \frac{\hat{\beta}_2^2 \sum x_i^2}{\sum y_i^2} = \frac{(\sum x_i y_i)^2}{(\sum x_i^2)^2} \cdot \frac{\sum x_i^2}{\sum y_i^2}$$

$$= \frac{(\sum x_i y_i)^2}{(\sum x_i^2)(\sum y_i^2)} = r^2$$



2007st.com 版图设计网

可决系数与相关系数的关系

（2）区别

可决系数	相关系数
就模型而言	就两个变量而言
说明解释变量对因变量的解释程度	度量两个变量线性依存程度。
度量不对称的因果关系	度量不含因果关系的对称相关关系
取值：**[0,1]**	取值：**[－1,1]**

2007st.com 版面设计网

第四节

回归系数的区间估计和假设检验

本节基本内容：

- OLS估计的分布性质
- 回归系数的区间估计
- 回归系数的假设检验

2007st.com 版图设计网

问题的提出

为什么要作区间估计？

OLS估计只是通过样本得到的点估计，不一定等于真实参数，还需要找到真实参数的可能范围，并说明其可靠性

为什么要作假设检验？

OLS 估计只是用样本估计的结果，是否可靠？是否抽样的偶然结果？还有待统计检验。

区间估计和假设检验都是建立在确定参数估计值概率分布性质的基础上。

2007st.com 版面设计网

一、OLS估计的分布性质

基本思想

$\hat{\beta}_k$ 是随机变量，必须确定其分布性质才可能进行区间估计和假设检验

u_i 是服从正态分布的随机变量，决定了 Y_i 也是服从正态分布的随机变量，$\hat{\beta}_k$ 是 Y_i 的线性函数，决定了 $\hat{\beta}_k$ 也是服从正态分布的随机变量，只要确定 $\hat{\beta}_k$ 的期望和方差，即可确定 $\hat{\beta}_k$ 的分布性质

2007st.com 版权设计网

$\hat{\beta}_k$ 的期望和方差

- $\hat{\beta}_k$ 的期望：$\mathrm{E}(\hat{\beta}_k) = \beta_k$ (无偏估计）

- $\hat{\beta}_k$ 的方差和标准差

  (标准误差是方差的算术平方根)

$$\mathrm{Var}(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_i^2} \qquad \mathrm{SD}(\hat{\beta}_2) = \frac{\sigma}{\sqrt{\sum x_i^2}}$$

$$\mathrm{Var}(\hat{\beta}_1) = \sigma^2 \frac{\sum X_i^2}{n\sum x_i^2} \qquad \mathrm{SD}(\hat{\beta}_1) = \sigma\sqrt{\frac{\sum X_i^2}{n\sum x_i^2}}$$

注意： 以上各式中σ^2未知，其余均是样本观测值

2007st.com 版图设计网

$\hat{\beta}_k$ 的分布

$$\hat{\beta}_2 \sim N(\beta_2, \frac{\sigma^2}{\sum x_i^2})$$

$$\hat{\beta}_1 \sim N(\beta_1, \sigma^2 \frac{\sum X_i^2}{n\sum x_i^2})$$

2007st.com 版图设计网

对随机扰动项方差 σ^2 的估计

可以证明（见教材**P70**附录**2.2)**

σ^2 的无偏估计为

$$\hat{\sigma}^2 = \frac{\sum e_i^2}{n-2}$$

(*n*-2为自由度,即可自由变化的样本观测值个数**)**

2007st.com 版图设计网

将 $\hat{\beta}_k$ 作标准化变换

$$z_1 = \frac{\hat{\beta}_1 - \beta_1}{SD(\hat{\beta}_1)} = \frac{\hat{\beta}_1 - \beta_1}{\sigma\sqrt{\frac{\sum X_i^2}{n\sum x_i^2}}} \sim N(0,1)$$

$$z_2 = \frac{\hat{\beta}_2 - \beta_2}{SD(\hat{\beta}_2)} = \frac{\hat{\beta}_2 - \beta_2}{\frac{\sigma}{\sqrt{\sum x_i^2}}} \sim N(0,1)$$

2007st.com 版图设计网

又，$\frac{RSS}{\sigma^2} \sim \chi^2(n-2)$　　$\frac{\sum e_i^2}{\sigma^2} \sim \chi^2(n-2)$

$$\frac{\frac{\hat{\beta}_1 - \beta_1}{\sigma\sqrt{\frac{\sum X_i^2}{n\sum x_i^2}}}}{\sqrt{\frac{\frac{\sum e_i^2}{\sigma^2}}{n-2}}} = \frac{\hat{\beta}_1 - \beta_1}{\sqrt{\frac{\sum e_i^2}{n-2} * \frac{\sum X_i^2}{n\sum x_i^2}}} \sim t(n-2)$$

综上，$t = \frac{\hat{\beta}_k - \beta_k}{SE(\hat{\beta}_k)} \sim t(n-2)$

2007st.com 版面设计网

二、回归系数的区间估计

概念：

对参数作出的点估计是随机变量，虽然是无偏估计，但还不能说明估计的可靠性和精确性，需要找到包含真实参数的一个范围，并确定这个范围包含参数真实值的可靠程度。

在确定参数估计式概率分布性质的基础上，可找到两个正数**δ**和**α**（$0 \le \alpha \le 1$），使得区间 $(\hat{\beta}_k - \delta, \hat{\beta}_k + \delta)$ 包含真实 β_k 的概率为$1-\alpha$，即

$$P(\hat{\beta}_k - \delta \le \beta_k \le \hat{\beta}_k + \delta) = 1 - \alpha$$

这样的区间称为所估计参数的置信区间。

2007st.com 版面设计网

回归系数区间估计的方法

对回归系数的区间估计，可分为三种情况，一般情况下，总体方差 σ^2 未知，用无偏估计 $\hat{\sigma}^2$ 去代替，由于样本容量较小，统计量 t 不再服从正态分布，而服从 t 分布。可用 t 分布去建立参数估计的置信区间。

$$t^* = \frac{\hat{\beta}_2 - \beta_2}{SE(\hat{\beta}_2)} \sim t(n-2)$$

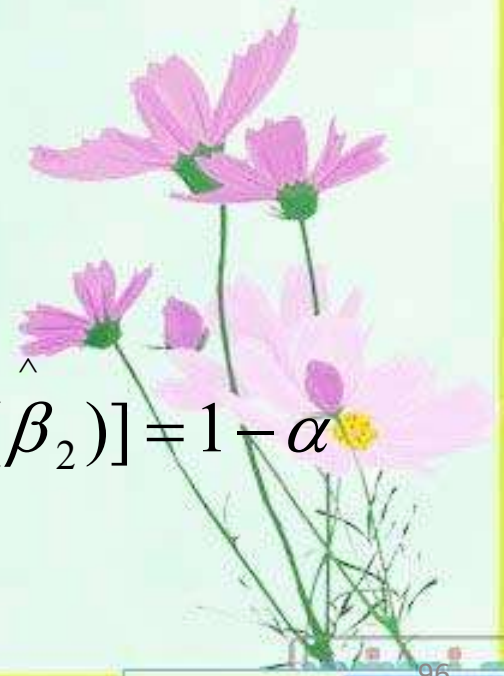
2007st.com 版权设计网

选定**α**，查 t 分布表得显著性水平为 $\alpha/2$，自由度为$n-2$ 的临界值，则有

$$P[-t_{\alpha/2} \le \frac{\hat{\beta}_2 - \beta_2}{SE(\hat{\beta}_2)} \le t_{\alpha/2}] = 1-\alpha$$

即

$$P[\hat{\beta}_2 - t_{\alpha/2} SE(\hat{\beta}_2) \le \beta_2 \le \hat{\beta}_2 + t_{\alpha/2} SE(\hat{\beta}_2)] = 1-\alpha$$



2007st.com 版面设计网

三、回归系数的假设检验

1. 假设检验的基本思想

为什么要作假设检验？

所估计的回归系数 $\hat{\beta}_1$ 、$\hat{\beta}_2$ 和方差 $\hat{\sigma}^2$ 都是通过样本估计的，都是随抽样而变动的随机变量，它们是否可靠？是否抽样的偶然结果呢？还需要加以检验。

2007st.com 版面设计网

对回归系数假设检验的方式

计量经济学中，主要是针对变量的参数真值是否为零来进行显著性检验的。

目的：对简单线性回归，判断解释变量 X 是否是被解释变量 Y 的显著影响因素。在一元线性模型中，就是要判断 X 是否对 Y 具有显著的线性影响。这就需要进行变量的显著性检验。

2007st.com 版图设计网

2. 回归系数的检验方法

一般情况下，总体方差 σ^2 未知，只能用 $\hat{\sigma}^2$ 去代替，可利用 t 分布作 t 检验

$$t^* = \frac{\hat{\beta}_2 - \beta_2}{SE(\hat{\beta}_2)} = \frac{\hat{\beta}_2}{SE(\hat{\beta}_2)} \sim t(n-2)$$

给定 α ，查 t 分布表得 $t_{\alpha/2}(n-2)$

▼如果 $t^* \le -t_{\alpha/2}(n-2)$ 或者 $t^* \ge -t_{\alpha/2}(n-2)$ 则拒绝原假设 $H_0 : \beta_2 = 0$ ，而接受备择假设 $H_1 : \beta_2 \neq 0$

▼如果 $-t_{\alpha/2}(n-2) \leqslant t^* \leqslant t_{\alpha/2}(n-2)$

则接受原假设 $H_0 : \beta_2 = 0$

2007st.com 版图设计网

用 P 值判断参数的显著性

假设检验的 p 值:

p 值是基于既定的样本数据所计算的统计量，是拒绝原假设的最低显著性水平。

统计分析软件中通常都给出了检验的 p 值

相对于显著性水平 α 的临界值: t_α 或 $t_{\alpha/2}$

由样本计算的统计量为: t^*

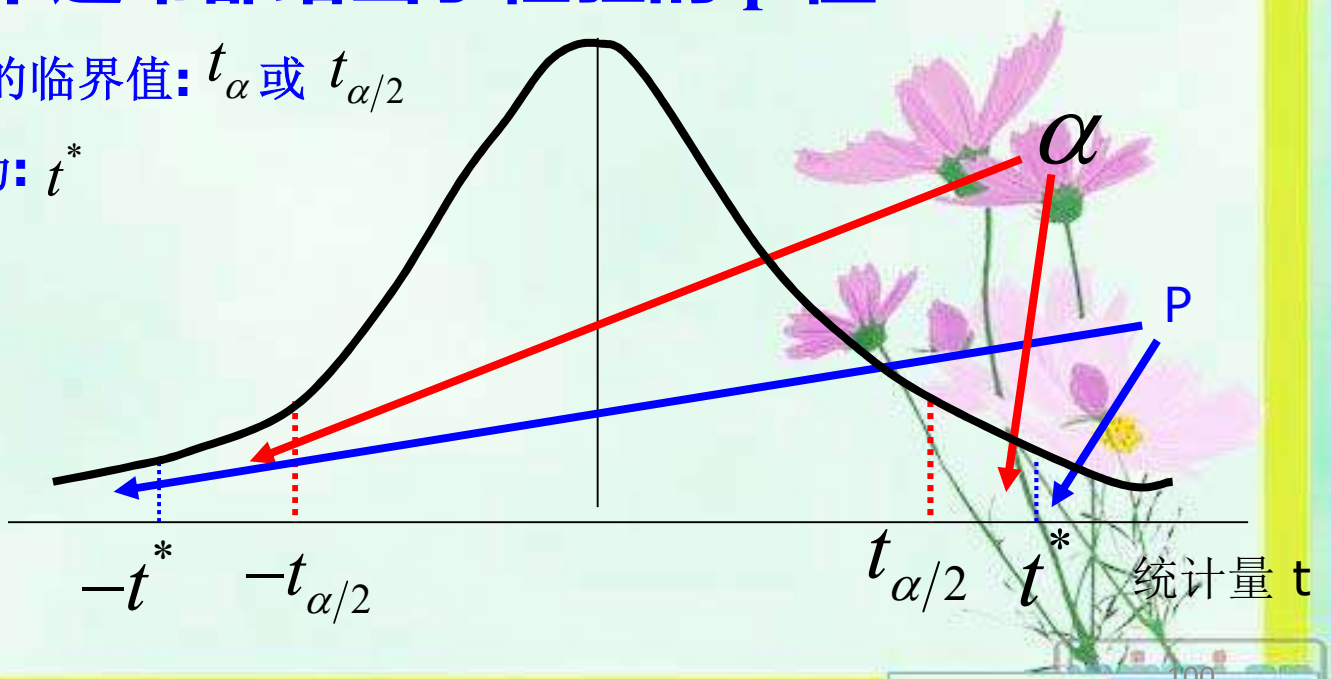
$t_{\alpha/2}$ 与 α 相对应

t^* 与 P 相对应

注意:

t检验是比较 t^* 和 $t_{\alpha/2}$

P值检验是比较 α 和 p



2007st.com 版图设计网

用 *P* 值判断参数的显著性

假设检验的 p 值：

p 值是根据既定的样本数据所计算的统计量，拒绝原假设的最小显著性水平。

统计分析软件中通常都给出了检验的 p 值。

2007st.com 版权设计网

用 P 值判断参数的显著性的方法

方法： 将给定的显著性水平 α 与 p 值比较：

- 若 $\alpha > p$ 值，则在显著性水平 α 下拒绝原假设 $H_0 : \beta_k = 0$，即认为 X 对 Y 有显著影响
- 若 $\alpha \leqslant p$ 值，则在显著性水平 α 下接受原假设 $H_0 : \beta_k = 0$，即认为 X 对 Y 没有显著影响

规则： 当 $p < \alpha$ 时，p 值越小，越能拒绝原假设

2007st.com 版面设计网

第五节　回归模型预测

本节主要内容：

- **回归分析结果的报告**
- **被解释变量平均值预测**
- **被解释变量个别值预测**

2007st.com 版图设计网

一、回归分析结果的报告

经过模型的估计、检验，得到一系列重要的数据，为了简明、清晰、规范地表述这些数据，计量经济学通常采用了以下规范化的方式：

例如：回归结果为

$$\hat{Y}_i = 352.00 + 0.5300 X_i$$

$$(76.5826) \quad (0.0216)$$

$$t = (4.5963) \quad (24.5902)$$

$$r^2 = 0.9869 \qquad df = 8$$

标准误差*SE*

t 统计量

可决系数和自由度

2007st.com 版面设计网

二、被解释变量平均值预测

1.基本思想

●运用计量经济模型作预测：指利用所估计的样本回归函数，用解释变量的已知值或预测值，对预测期或样本以外的被解释变量数值作出定量的估计。

●计量经济预测是一种条件预测：

条件：◆模型设定的关系式不变

◆所估计的参数不变

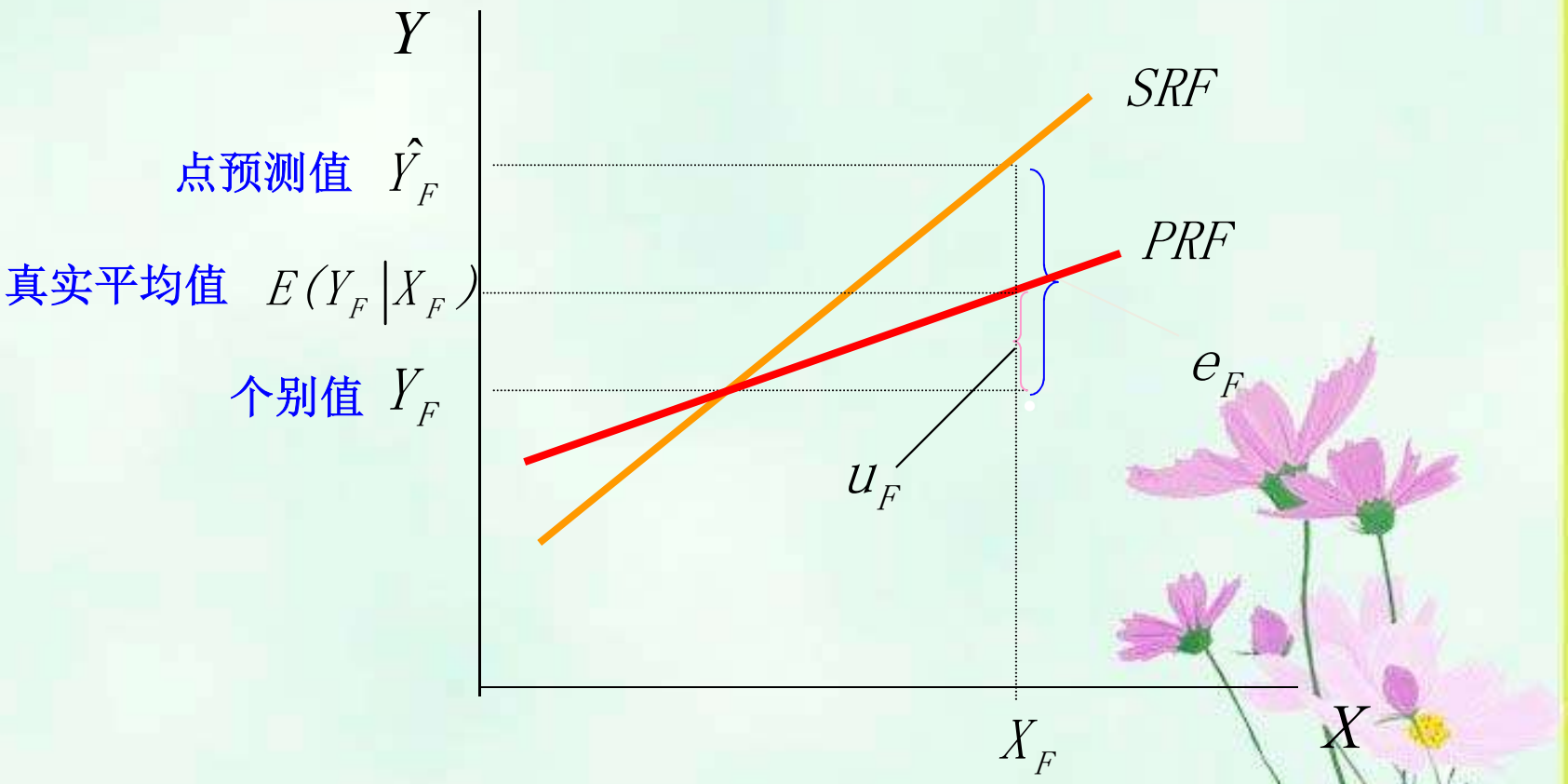
◆ 解释变量在预测期的取值已作出预测

对因变量的预测分为平均值预测和个别值预测

对因变量的预测又分为点预测和区间预测

2007st.com 版权设计网

预测值、平均值、个别值的相互关系



$\hat{Y}_F$是真实平均值的点估计,也是对个别值的点估计

2007st.com 版图设计网

2 .Y 平均值的点预测

将解释变量预测值直接代入估计的方程

$$\hat{Y}_F = \hat{\beta}_1 + \hat{\beta}_2 X_F$$

这样计算的 $\hat{Y}_F$ 是一个点估计值

2007st.com 版面设计网

3. Y平均值的区间预测

基本思想：

- 由于存在抽样波动，预测的平均值 $\hat{Y}_F$不一定等于真实平均值 $E(Y_F|X_F)$ 还需要对$E(Y_F|X_F)$作区间估计。
- 为对Y作区间预测，必须确定平均值预测值的抽样分布， 必须找出与 $\hat{Y}_F$ 和 $E(Y_F|X_F)$ 都有关的统计量

2007st.com 版图设计网

具体作法（从 $\hat{Y}_F$ 的分布分析）

已知

$$E(\hat{Y}_F) = E(Y_F \mid X_F) = \beta_1 + \beta_2 X_F$$

可以证明

$$Var(\hat{Y}_F) = \sigma^2 [\frac{1}{n} + \frac{(X_F - \overline{X})^2}{\sum x_i^2}]$$

$$SD(\hat{Y}_F) = \sigma \sqrt{\frac{1}{n} + \frac{(X_F - \overline{X})^2}{\sum x_i^2}}$$

$\hat{Y}_F$ 服从正态分布，将其标准化，

当σ^2未知时，只得用$\hat{\sigma}^2 = \sum e_i^2 / (n-2)$ 代替，这时有

$$t = \frac{\hat{Y}_F - E(Y_F \mid X_F)}{\hat{\sigma} \sqrt{\frac{1}{n} + \frac{(X_F - \overline{X})^2}{\sum x_i^2}}} \sim t(n-2)$$

2007st.com 版面设计网

构建平均值的预测区间

显然这样的 t 统计量与$\hat{Y}_F$和$E(Y_F|X_F)$都有关。

给定显著性水平α，查t分布表，得自由度n－2的临界值 $t_{\alpha/2}(n-2)$则有

$$P(-t_{\alpha/2} \le t = \frac{\hat{Y}_F - E(Y_F|X_F)}{SE(\hat{Y}_F)} \le t_{\alpha/2}) = 1-\alpha$$

$$p\{[\hat{Y}_F - t_{\alpha/2}SE(\hat{Y}_F)] \le E(Y_F|X_F) \le [\hat{Y}_F + t_{\alpha/2}SE(\hat{Y}_F)]\} = 1-\alpha$$

Y平均值的置信度为$1-\alpha$的预测区间为

$$[\hat{Y}_F - t_{\alpha/2}\hat{\sigma}\sqrt{\frac{1}{n}+\frac{(X_F-\overline{X})^2}{\sum x_i^2}}, \hat{Y}_F + t_{\alpha/2}\hat{\sigma}\sqrt{\frac{1}{n}+\frac{(X_F-\overline{X})^2}{\sum x_i^2}}]$$

2007st.com 版权设计网

三、因变量个别值预测

基本思想：

- $\hat{Y}_F$既是对 Y平均值的点预测，也是对 Y个别值的点预测
- 由于存在随机扰动 u_i的影响， Y的平均值并不等于Y 的个别值Y_F
- 为了对 Y的个别值 Y_F作区间预测，需要寻找与预测值$\hat{Y}_F$ 和个别值 Y_F有关的统计量，并要明确其概率分布

2007st.com 版面设计网

具体作法：

已知剩余项$e_F = Y_F - \hat{Y}_F$是与预测值$\hat{Y}_F$及个别值Y_F都有关的变量,并且已知e_F服从正态分布，且可证明

$$E(e_F) = 0$$

$$Var(e_F) = E(Y_F - \hat{Y}_F)^2 = \sigma^2[1 + \frac{1}{n} + \frac{(X_F - \overline{X})^2}{\sum x_i^2}]$$

当用$\hat{\sigma}^2 = \sum e_i^2 / (n-2)$代替$\sigma^2$时，对$e_F$标准化的变量 t 为

$$t = \frac{e_F - E(e_F)}{\hat{SE}(e_F)} = \frac{Y_F - \hat{Y}_F}{\hat{\sigma}\sqrt{1 + \frac{1}{n} + \frac{(X_F - \overline{X})^2}{\sum x_i^2}}} \sim t(n-2)$$

2007st.com 版面设计网

构建个别值的预测区间

给定显著性水平α，查 t 分布表得自由度为 $n-2$ 的临界值 $t_{\alpha/2}(n-2)$，则有

$$P\{[\hat{Y}_F - t_{\alpha/2}SE(e_F)] \le Y_F \le [\hat{Y}_F + t_{\alpha/2}SE(e_F)]\} = 1-\alpha$$

因此，一元回归时 Y的个别值的置信度为$1-\alpha$的预测区间上下限为

$$Y_F = \hat{Y}_F \mp t_{\alpha/2}\hat{\sigma}\sqrt{1+\frac{1}{n}+\frac{(X_F-\overline{X})^2}{\sum x_i^2}}$$

2007st.com 版面设计网

因变量Y区间预测的特点

1、Y平均值的预测值与真实平均值有误差，主要是受抽样波动影响

$$Y_F = \hat{Y}_F \mp t_{\alpha/2}\hat{\sigma}\sqrt{\frac{1}{n}+\frac{(X_F-\overline{X})^2}{\sum x_i^2}}$$

Y个别值的预测值与真实个别值的差异,不仅受抽样波动影响，而且还受随机扰动项的影响

$$Y_F = \hat{Y}_F \mp t_{\alpha/2}\hat{\sigma}\sqrt{1+\frac{1}{n}+\frac{(X_F-\overline{X})^2}{\sum x_i^2}}$$

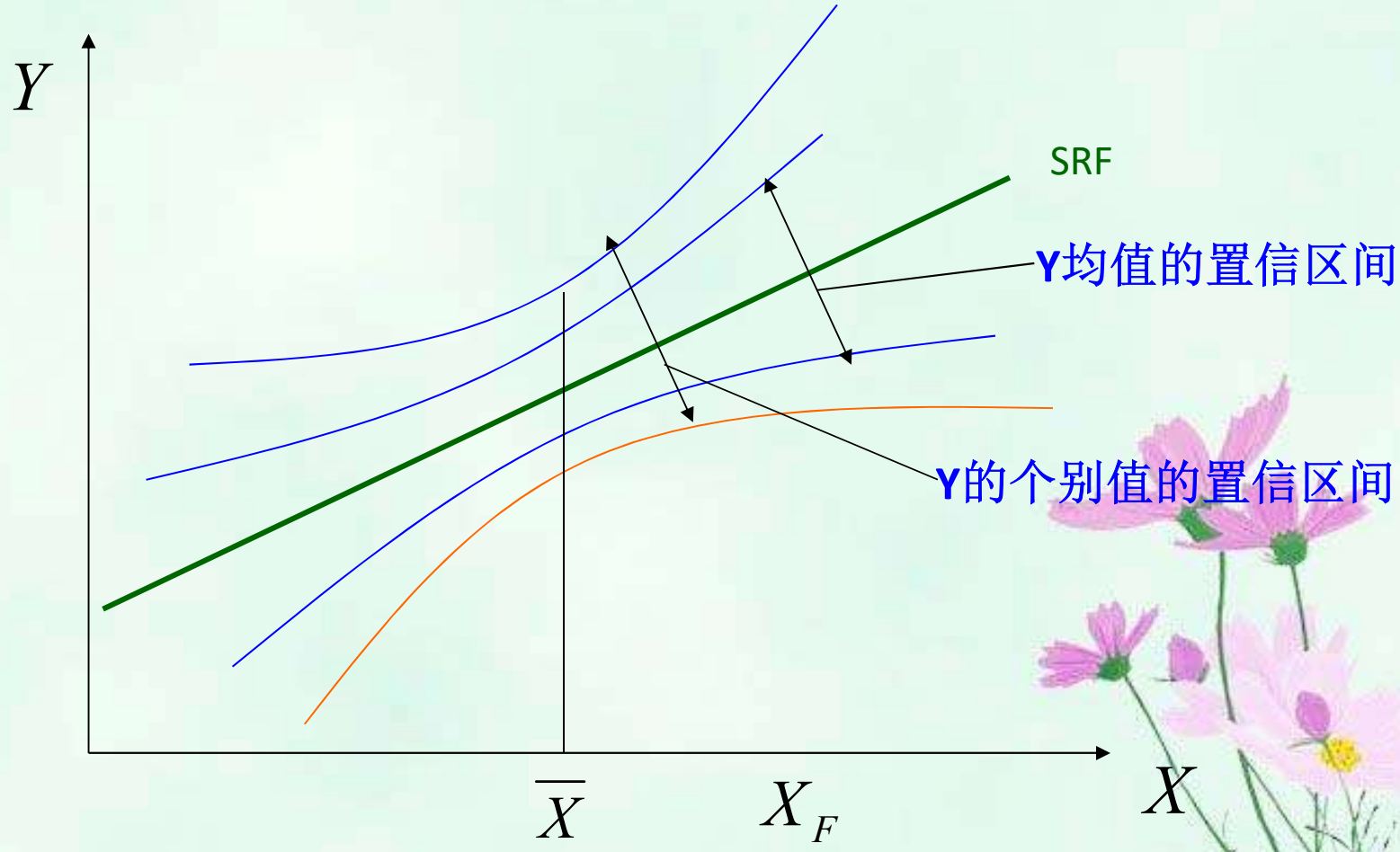
2007st.com 版图设计网

2、平均值和个别值预测区间都不是常数，是随 X_F 的变化而变化的

3、预测区间上下限与样本容量有关，当样本容量 $n \to \infty$ 时个别值的预测误差只决定于随机扰动的方差

2007st.com 版面设计网

各种预测值的关系



当$X_F = \overline{X}$时，置信区间最小

2007st.com 版面设计网

第八节　案例分析

案例：分析各地区城镇居民计算机拥有量与城镇居民收入水平的关系

提出问题：随着信息化程度和居民收入水平的提高，作为居民耐用消费品重要代表的计算机已为众多城镇居民家庭所拥有。研究中国各地区城镇居民计算机拥有量与居民收入水平的数量关系，对于探寻居民消费增长的规律性，分析各地区居民消费的差异，预测地区全体居民消费水平和结构的发展趋势，合理规划信息产业的发展，都有重要的意义。

理论分析：影响居民计算机拥有量的因素有多种，但从理论和经验分析，最主要的影响因素应是居民收入水平。从理论上说居民收入水平越高，居民计算机拥有量越多。

变量选择：被解释变量选择能代表城乡所有居民消费的“城镇居民家庭平均每百户计算机拥有量”(单位:台)；解释变量选择表现城镇居民收入水平的“城镇居民平均每人全年家庭总收入”（单位:元）

研究范围：全国各省市**2011**年底的城镇居民家庭平均每百户计算机拥有量和城镇居民平均每人全年家庭总收入数据。

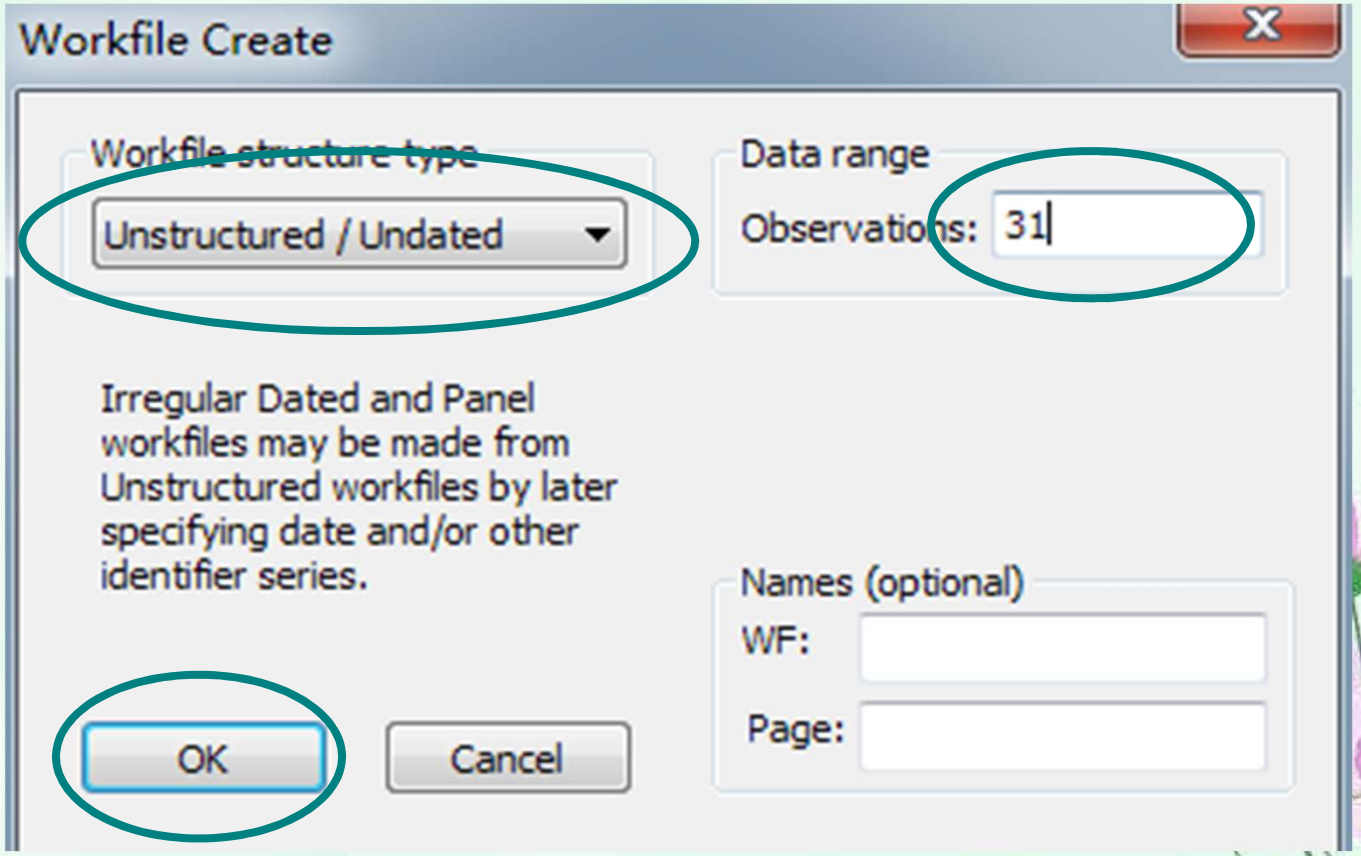
2007st.com 版图设计网

2011年中国各地区城镇居民每百户计算机拥有量和人均总收入

地区	2011年底城镇居民家庭平均每百户计算机拥有量(台)Y	城镇居民平均每人全年家庭总收入（元）X
北 京	103.51	37124.39
天 津	95.4	29916.04
河 北	74.74	19591.91
山 西	69.45	19666.1
内蒙古	60.83	21890.19
辽 宁	71.66	22879.77
吉 林	68.04	19211.71
黑龙江	55.36	17118.49
上 海	137.7	40532.29
江 苏	96.94	28971.98
浙 江	103.17	34264.38
安 徽	74.04	20751.11
福 建	103	27378.11
江 西	73.87	18656.52
山 东	85.88	24889.8

地区	2011年底城镇居民家庭平均每百户计算机拥有量(台)Y	城镇居民平均每人全年家庭总收入（元）X
河 南	71.41	19526.92
湖 北	75.49	20193.27
湖 南	66.36	20083.87
广 东	104.13	30218.76
广 西	91.72	20846.11
海 南	63.82	20094.18
重 庆	76.07	21794.27
四 川	68.86	19688.09
贵 州	63.89	17598.87
云 南	63.55	20255.13
西 藏	58.83	18115.76
陕 西	82.43	20069.87
甘 肃	56.14	16267.37
青 海	52.65	17794.98
宁 夏	59.39	19654.59
新 疆	61.2	17631.15

建立工作文件

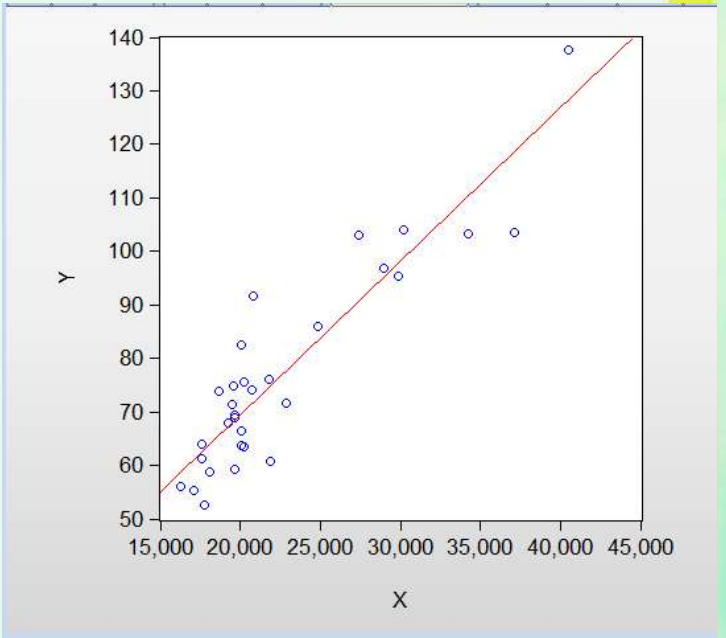


录入数据

2007st.com 版面设计网

模型设定：

为了初步分析城镇居民家庭平均每百户计算机拥有量**(Y)**与城镇居民平均每人全年家庭总收入**(X)**的关系，作以**X**为横坐标，以**Y**为纵坐标的散点图。



从散点图可以看出城镇居民家庭平均每百户计算机拥有量**(Y)**与城镇居民平均每人全年家庭总收入**(X)** 大体呈现线性关系。

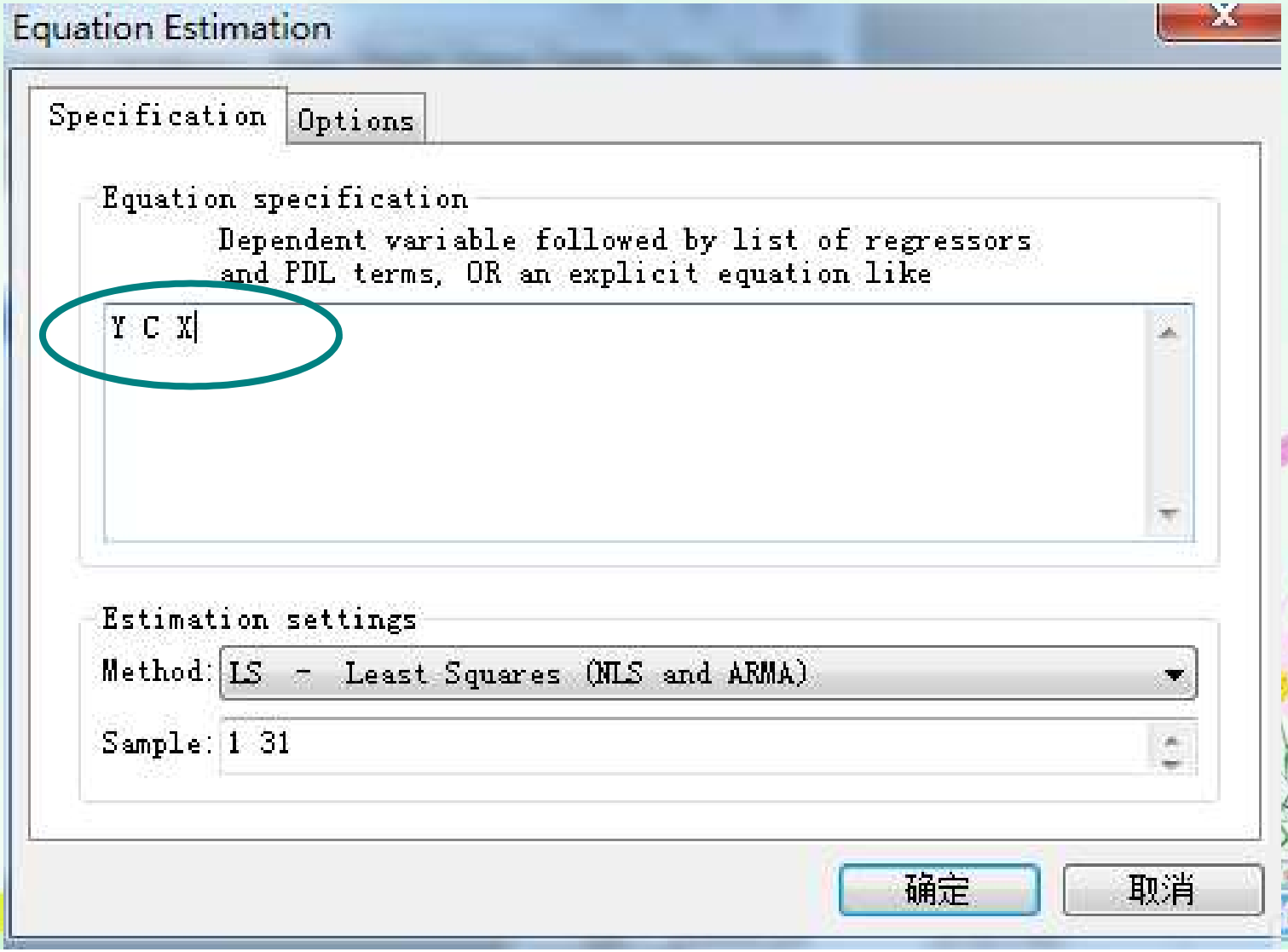
可以建立如下简单线性回归模型：

$$Y_t = \beta_1 + \beta_2 X_t + u_t$$

2007st.com 版权设计网

估计方程

假定模型中随机扰动满足基本假定，可用OLS法。



估计参数

假定模型中随机扰动满足基本

具体操作：使用***EViews*** 软件，

Dependent Variable: Y
Method: Least Squares
Date: 09/08/13 Time: 19:37
Sample: 1 31
Included observations: 31

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	11.95802	5.622841	2.126686	0.0421
X	0.002873	0.000240	11.98264	0.0000

R-squared	0.831966	Mean dependent var	77.08161
Adjusted R-squared	0.826171	S.D. dependent var	19.25503
S.E. of regression	8.027957	Akaike info criterion	7.066078
Sum squared resid	1868.995	Schwarz criterion	7.158593
Log likelihood	-107.5242	Hannan-Quinn criter.	7.096236
F-statistic	143.5836	Durbin-Watson stat	1.656123
Prob(F-statistic)	0.000000		

Dependent ——被解释变量
Variable ——解释变量
Coefficient ——回归系数
Std.Error ——系数标准差
t-Statistic——t 检验值
Prob.——零系数概率
R-squared ——可决系数
Adjusted R-squared ——调整后的可决系数
S.E.of regression——回归标准差
Sum squared resid ——残差平方和
Log likelihood ——对数似然估计值
Durbin-Watson stat——DW 检验值
Mean of dependent var——被解释变量均值
S.D.of dependent var——被解释变量方差
F-statistic——总体F 检验值

2007st.com 版图设计网

用规范的形式将参数估计和检验的结果写为：

$$\hat{Y}_t = 11.9580 + 0.002873 X_t$$

（5.6228） (0.00024)

t= (2.1267) (11.9826)

$R^2 = 0.8320$ F=143.5836 n=31

2007st.com 版图设计网

模型检验

1. 可决系数： $R^2 = 0.8320$ 模型整体上拟合较好。

2. 系数显著性检验：取 $\alpha = 0.05$，查t分布表得自由度为 $n-2=31-2=29$ 的临界值为 $t_{0.025}(29)=2.045$。

因为 $t(\hat{\beta}_1)=2.1267 > t_{0.025}(29)=2.045$ 应拒绝 $H_0:\beta_1=0$

$t(\hat{\beta}_2)=11.9826 > t_{0.025}(29)=2.045$ 应拒绝 $H_0:\beta_2=0$

3. 用P值检验 $\alpha = 0.05 >> p$=0.0000

表明，城镇居民人均总收入对城镇居民每百户计算机拥有量确有显著影响。

2007st.com 版权设计网

4. 经济意义检验：

所估计的参数 ，，说明城镇居民家庭人均总收入每增加1元，平均说来城镇居民每百户计算机拥有量将增加0.002873台，这与预期的经济意义相符。

2007st.com 版权设计网

经济预测

点预测：

如果西部地区某省城镇居民家庭人均总收入能达到25000元/人，利用所估计的模型可预测城镇居民每百户计算机拥有量，点预测值为

$$\hat{Y}_f = 11.9580 + 0.002873 \times 25000 = 83.7846 \quad (台)$$

区间预测： 平均值区间预测上下限：

$$Y_f = \hat{Y}_f \mp t_{\alpha/2}\,\hat{\sigma}\sqrt{\frac{1}{n} + \frac{(X_f - \overline{X})^2}{\sum x_i^2}}$$

已知：

$Y_f = 83.7846$　　$t_{0.025}(29) = 2.045$　　$\hat{\sigma} = 8.027957$　　$n = 31$

2007st.com 版

平均值区间预测区间预测

由X和Y的描述统计结果

	X	Y
Mean	22666.97	77.08161
Median	20094.18	71.66000
Maximum	40532.29	137.7000
Minimum	16267.37	52.65000
Std. Dev.	6112.965	19.25503
Skewness	1.515854	1.185095
Kurtosis	4.384257	4.259649
Jarque-Bera	14.34708	9.305832
Probability	0.000767	0.009534
Sum	702676.0	2389.530
Sum Sq. Dev.	1.12E+09	11122.69
Observations	31	31

$$\bar{X} = 22666.97$$

$$(X_f - \bar{X})^2 = (25000 - 22666.97)^2 = 5443028.981$$

$$\sum x_i^2 = \sum (X_i - \bar{X})^2 = \sigma_X^2 (n-1) = 6112.965^2 \times (31-1) = 1121050233$$

$X_f = 25000$时

$$83.7846 \mp 2.045 \times 8.027957 \times \sqrt{\frac{1}{31} + \frac{5443028.981}{1121050233}} = 83.7846 \mp 3.1627$$

即是说：当地区城镇居民人均总收入达到**25000**元时，城镇居民每百户计算机拥有量 平均值置信度**95%**的预测区间为（**80.6219，86.9473**）台。

2007st.com 版权设计网

个别值区间预测：

$$Y_f = \hat{Y}_f \mp t_{\alpha/2}\,\hat{\sigma}\sqrt{1+\frac{1}{n}+\frac{(X_f-\overline{X})^2}{\sum x_i^2}}$$

$X_F = 25000$时：

$$83.7846 \mp 2.045 \times 8.027957 \times \sqrt{1+\frac{1}{31}+\frac{5443028.981}{1121050233}} = 83.7846 \mp 16.7190$$

即是说：当地区城镇居民人均总收入达到**25000**元时，城镇居民每百户计算机拥有量 个别值置信度**95%**的预测区间为（**67.0656，100.5036**）台。

130
2007st.com 版面设计网

第二章 小 结

1、变量间的关系： 函数关系——相关关系

相关系数——对变量间线性相关程度的度量

2、现代意义的回归：一个被解释变量对若干个解释变量依存关系的研究

实质：由固定的解释变量去估计被解释变量的平均值

2007st.com 版面设计网

3、总体回归函数（*PRF*）：将总体被解释变量*Y*的条件均值表现为解释变量*X* 的某种函数

样本回归函数（*SRF*）：将被解释变量*Y* 的样本条件均值表示为解释变量*X* 的某种函数。

总体回归函数与样本回归函数的区别与联系

4、随机扰动项：被解释变量实际值与条件均值的偏差，代表排除在模型以外的所有因素对*Y*的影响。

2007st.com 版图设计网

5、简单线性回归的基本假定：

对模型和变量的假定　　$E(Y_i) = \beta_1 + \beta_2 X_i$

对随机扰动项u的假定

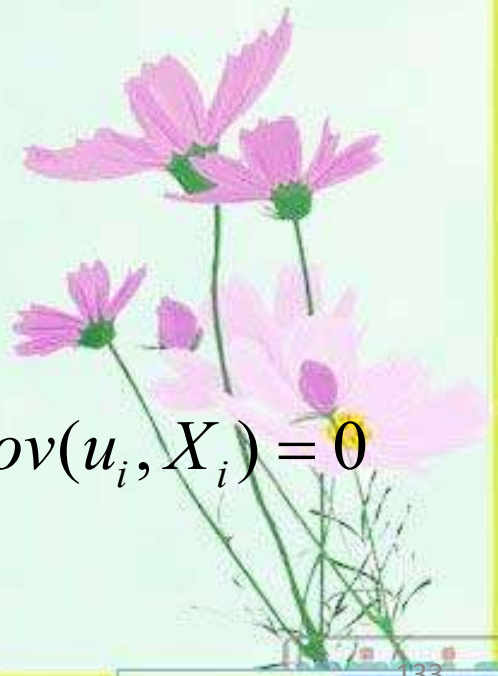
零均值假定：　$E(u_i) = 0$

同方差假定：　$Var(u_i) = Var(Y_i) = \sigma^2$

无自相关假定：　$Cov(u_i, u_j) = E(u_i u_j) = 0$

随机扰动与解释变量不相关假定：　$Cov(u_i, X_i) = 0$

正态性假定：　$u_i \sim N(0, \sigma^2)$



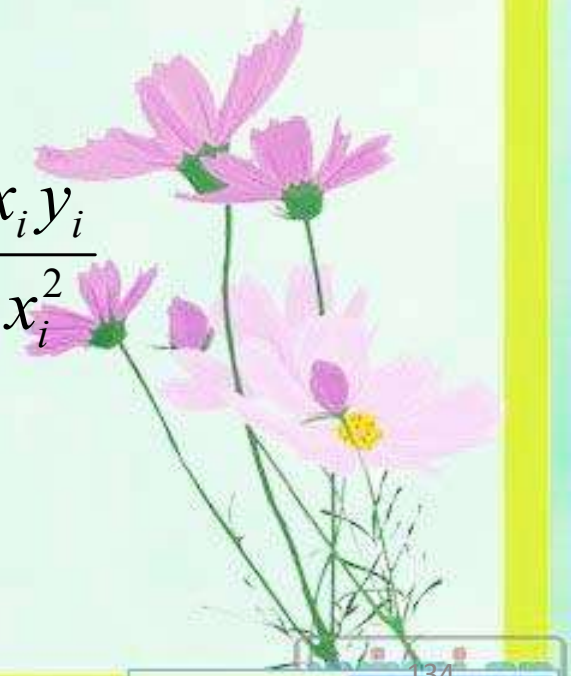
2007st.com 版图设计网

6、普通最小二乘法（OLS）估计参数的基本思想及估计式；

$$\hat{\beta}_1 = \frac{\sum X_i^2 \sum Y_i - \sum X_i \sum X_i Y_i}{N\sum X_i^2 - (\sum X_i)^2}$$

$$\hat{\beta}_2 = \frac{N\sum X_i Y_i - \sum X_i \sum Y_i}{N\sum X_i^2 - (\sum X_i)^2} = \frac{\sum x_i y_i}{\sum x_i^2}$$

$$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X}$$



2007st.com 版面设计网

OLS 估计式的分布性质

期望：$$E(\hat{\beta}_k) = \beta_k$$

方差：$$Var(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_i^2}$$

$$Var(\hat{\beta}_1) = \sigma^2 \frac{\sum X_i^2}{N\sum x_i^2}$$

标准差：$$SD(\hat{\beta}_2) = \frac{\sigma}{\sqrt{\sum x_i^2}}$$

$$SD(\hat{\beta}_1) = \sigma\sqrt{\frac{\sum X_i^2}{N\sum x_i^2}}$$

OLS估计式是最佳线性无偏估计式。

2007st.com 版图设计网

7、σ^2的无偏估计

$$\hat{\sigma}^2 = \frac{\sum e_i^2}{n-2}$$

8、对回归系数区间估计的思想和方法

$$P[\hat{\beta}_2 - t_{\alpha/2} SE(\hat{\beta}_2) \le \beta_2 \le \hat{\beta}_2 + t_{\alpha/2} SE(\hat{\beta}_2)] = 1-\alpha$$

2007st.com 版面设计网

9、拟合优度：样本回归线对样本观测数据拟合的优劣程度，

可决系数：在总变差分解基础上确定的，模型解释了的变差在总变差中的比重

可决系数的计算方法、特点与作用。

$$1=\frac{\sum \hat{y}^2}{\sum y_i^2}+\frac{\sum e_i^2}{\sum y_i^2} \qquad r^2=\frac{\sum \hat{y}^2}{\sum y_i^2} \qquad r^2=1-\frac{\sum e_i^2}{\sum y_i^2}$$

2007st.com 版图设计网

10、对回归系数的假设检验

假设检验的基本思想

对回归系数 t 检验的思想与方法

$$t^* = \frac{\hat{\beta}_2 - \beta_2}{SE(\hat{\beta}_2)} = \frac{\hat{\beta}_2}{SE(\hat{\beta}_2)} \sim t(n-2)$$

用 ***P*** 值判断参数的显著性

2007st.com 版图设计网

11、对被解释变量的预测

被解释变量平均值预测与个别值预测的关系

被解释变量平均值的点预测和区间预测的方法

$$[\hat{Y}_F - t_{\alpha/2}\hat{\sigma}\sqrt{\frac{1}{n}+\frac{(X_F-\overline{X})^2}{\sum x_i^2}}, \hat{Y}_F + t_{\alpha/2}\hat{\sigma}\sqrt{\frac{1}{n}+\frac{(X_F-\overline{X})^2}{\sum x_i^2}}]$$

2007st.com 版图设计网

被解释变量个别值区间预测的方法

$$Y_F = \hat{Y}_F \mp t_{\alpha/2}\hat{\sigma}\sqrt{1+\frac{1}{n}+\frac{(X_F-\overline{X})^2}{\sum x_i^2}}$$

12、运用**EViews**软件对简单的线性回归模型进行估计和检验

2007st.com 版面设计网