第二章

经典单方程计量经济学模型

一元线性回归模型

计量模型

- 单方程模型：研究单一经济现象，只包含一个方程
- 联立方程模型：研究经济系统，包含多个方程

单方程模型

- 线性模型：变量间呈线性关系
- 非线性模型：变量间呈非线性关系

线性回归模型：用回归分析方法建立的线性模型，旨在揭示经济变量之间的因果关系。

第二章 一元线性回归模型

2.1 回归分析概述

2.2 一元线性模型的参数估计

2.3 一元线性回归模型的统计检验

2.1 回归分析概述

一、回归分析的基本概念

二、总体回归函数

三、样本回归函数

四、随机干扰项

一、回归分析的基本概念

1.变量间的相互关系

变量间关系的表现形式：
- 确定性函数关系
- 不确定的统计相关关系

相关分析与回归分析

确定性函数关系：确定性现象，非随机变量之间的关系

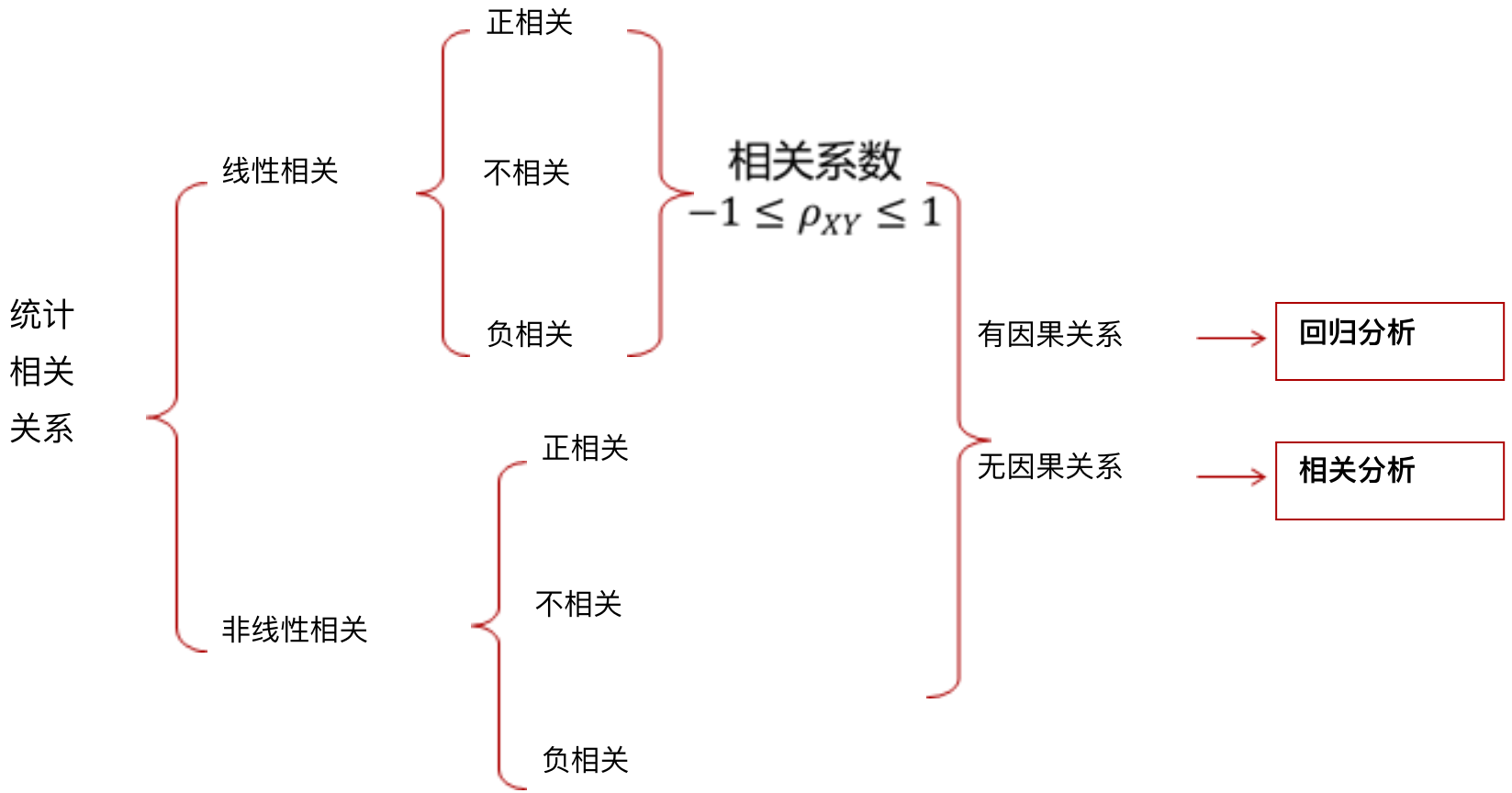
【例】$S = \pi r^2$

不确定的统计相关关系：非确定性现象，随机变量间的关系

【例】农作物产量 $= f$(耕地面积，劳动力，施肥量) $+ \mu$

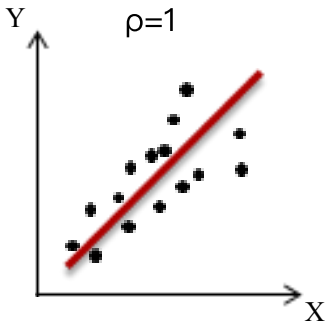
μ=阳光+气温+降雨量+……

【注意】变量间的函数关系和相关关系不是绝对的，在一定条件下两者可以相互转化。

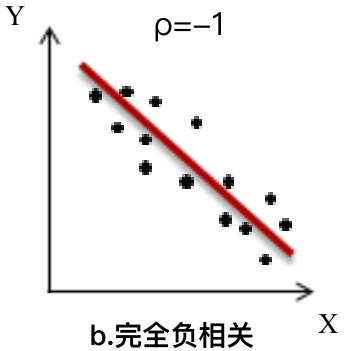

统计
相关
关系
线性相关
非线性相关
正相关
不相关
负相关
正相关
不相关
负相关
相关系数
$-1 \leq \rho_{XY} \leq 1$
有因果关系
无因果关系
回归分析
相关分析

2. 相关分析与回归分析

相关分析：研究随机变量间的相关形式及相关程度。变量的地位是对称的，均为随机变量。



a.完全正相关

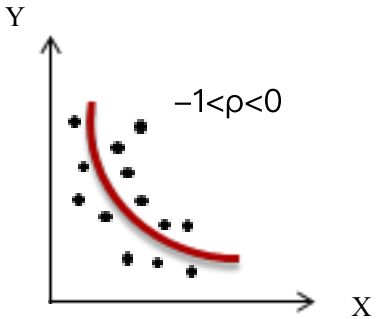


b.完全负相关

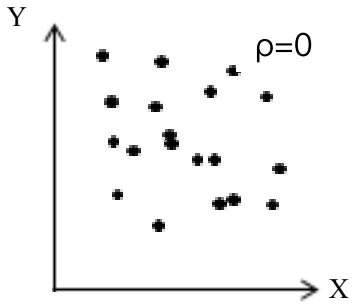
【注意】

①不线性相关不意味着不相关；

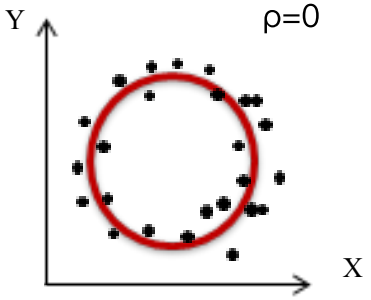
②有相关关系不一定有因果关系。



c.非完全（负）相关



d.完全不相关



e.不线性相关，但高度相关

相关关系与因果关系的再讨论

【例】看到街上人们带雨伞，于是预测天要下雨。

- 这是相关关系，“人们带伞”并不导致“下雨”。

【例】根据与流感相关的海量词条搜索记录，Google公司通过分析大数据，很快地预测流行病的地域传播。

- 这也是相关关系，上网搜索流感信息并不导致流感传播。

相关关系可用于预测，但如果要分析**X导致Y的作用机制**，

即变量之间的**因果关系**，则必须建立计量经济模型。

双向因果关系

【例】FDI促进经济增长，FDI也被吸引到快速增长的地区。

【例】收入增加引起消费增长，消费增长也拉动收入增加。

回归分析：研究一个变量关于另一个(几个)变量的依赖关系。变量的地位不对称，有解释变量与被解释变量之分。

回归分析的目的①：通过设定当自变量取某个确定的值，估计因变量所有可能出现的对应值的平均值，即总体均值。

回归分析构成计量经济学的方法论基础，主要内容包括：

（1）根据样本观察值对计量经济学模型参数进行估计，求得样本回归函数；

（2）对样本回归函数、参数估计值进行统计显著性检验；

（3）利用样本回归函数进行分析、评价和预测。

二、总体回归函数

【例】一个假想的社区由99户家庭组成，要研究该社区每月家庭消费支出Y与每月家庭可支配收入X之间的关系。

即根据家庭的月收入，考察该社区家庭月消费支出的平均水平。

为研究方便，将该99户家庭划分为组内收入接近的10组，以分析每一收入组的家庭消费支出（见下表）。

表2.1.1　某社区家庭每月收入与消费支出统计表

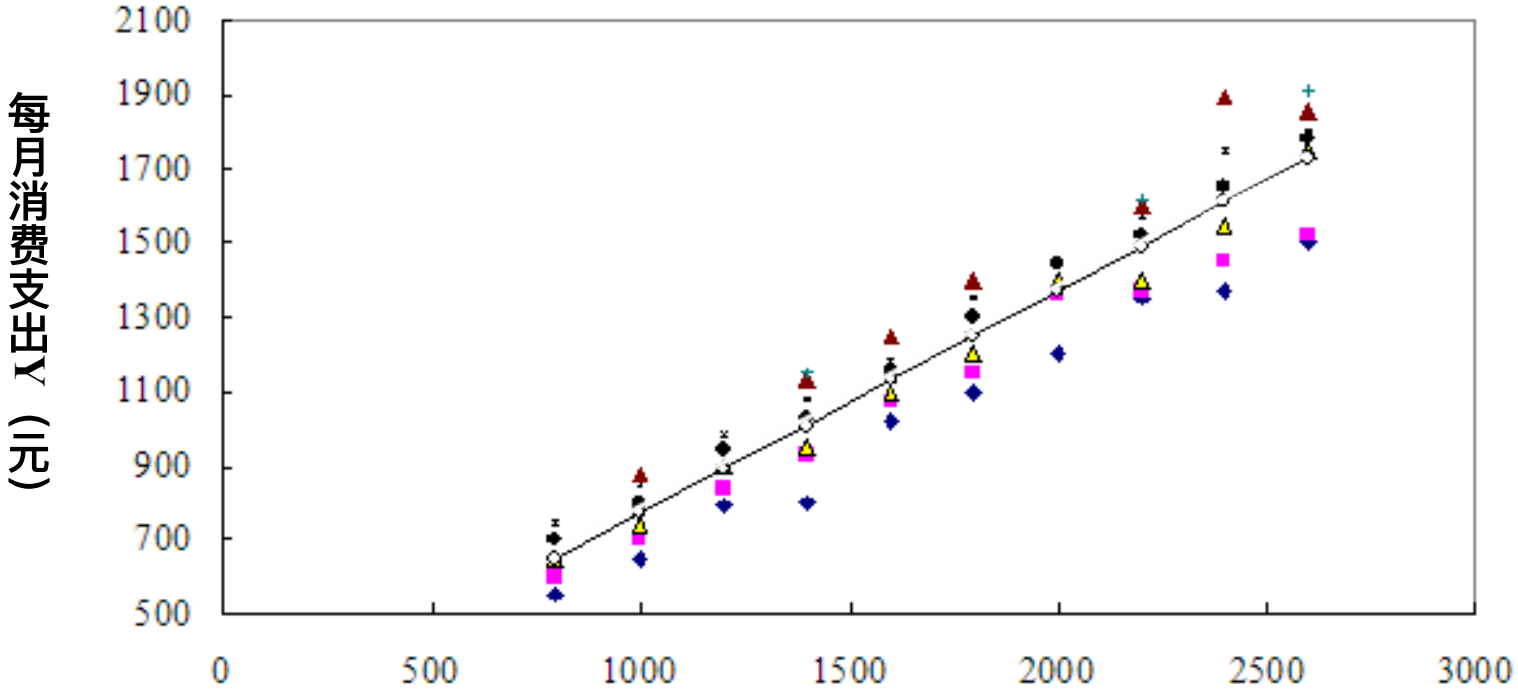
	每月家庭可支配收入X（元）									
	800	1100	1400	1700	2000	2300	2600	2900	3200	3500
每月家庭消费支出Y（元）	561	638	869	1023	1254	1408	1650	1969	2090	2299
	594	748	913	1100	1309	1452	1738	1991	2134	2321
	627	814	924	1144	1364	1551	1749	2046	2178	2530
	638	847	979	1155	1397	1595	1804	2068	2266	2629
		935	1012	1210	1408	1650	1848	2101	2354	2860
		968	1045	1243	1474	1672	1881	2189	2486	2871
			1078	1254	1496	1683	1925	2233	2552	
			1122	1298	1496	1716	1969	2244	2585	
			1155	1331	1562	1749	2013	2299	2640	
			1188	1364	1573	1771	2035	2310		
			1210	1408	1606	1804	2101			
				1430	1650	1870	2112			
				1485	1716	1947	2200			
						2002				
共计	2420	4950	11495	16445	19305	23870	25025	21450	21285	15510

- 由于不确定性因素的影响，对同一收入水平X，不同家庭的消费支出Y并不完全相同。
- 给定收入的值X_i，可得消费支出Y的条件均值或条件期望：

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

表1 各可支配收入水平组相应家庭消费支出的条件概率与条件均值

收入水平	800	1100	1400	1700	2000	2300	2600	2900	3200	3500
条件概率	1/4	1/6	1/11	1/13	1/13	1/14	1/13	1/10	1/9	1/6
条件均值	605	825	1045	1265	1485	1705	1925	2145	2365	2585



随着收入X的增加，消费Y“平均地说”也在增加，且Y的条件均值均落在一条斜率为正的直线上，这条直线就被称为总体回归线。

总体回归线：在给定解释变量X_i的条件下，被解释变量Y_i的期望轨迹。

总体回归函数：$E(Y|X_i) = f(X_i)$

总体回归函数说明：被解释变量Y的平均状态随解释变量X变化的规律。

在本例中，由散点图不难看出，居民消费支出是其可支配收入的线性函数，因此，总体回归函数为：

$$E(Y|X_i) = \beta_0 + \beta_1 X_i$$

其中，β_0和β_1是未知参数，称为回归系数。

由于总体的信息往往无法掌握，现实的情况只能是在一次观测中得到总体的一组样本。

问题：能从一次抽样中获得总体的近似的信息吗？如果可以，该如何做？

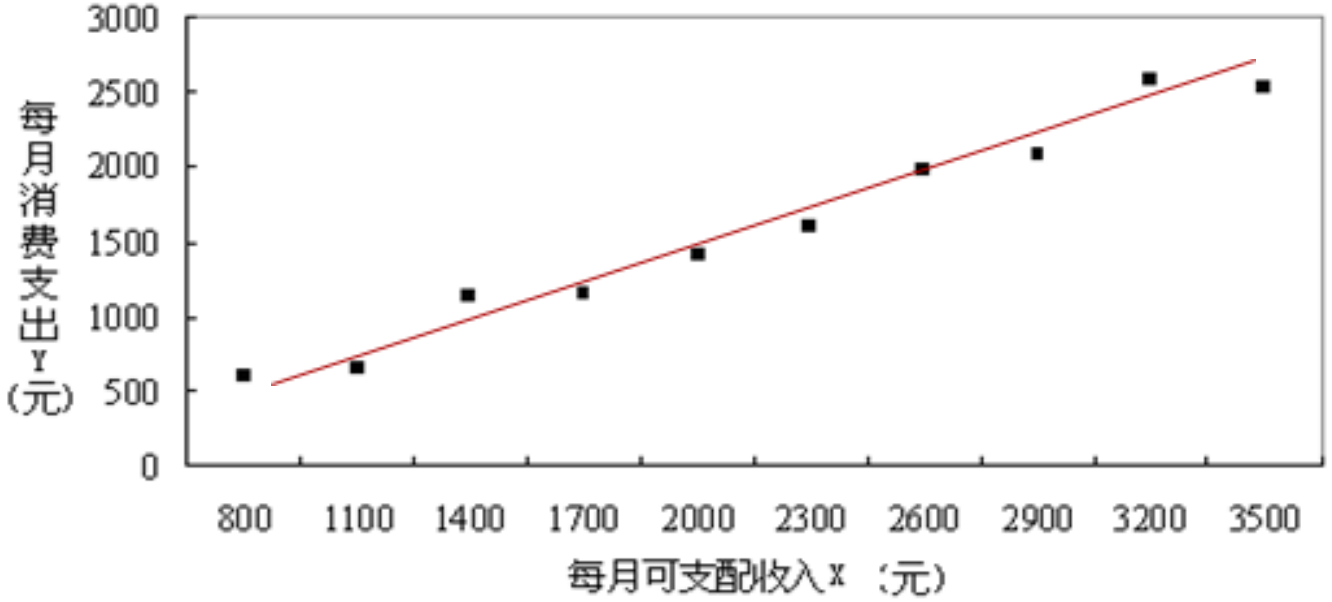
回答：能，利用样本得到样本回归函数，将样本回归函数看作总体回归函数的近似替代。

三、样本回归函数

回归分析的目的2：根据样本回归函数，估计总体回归函数。

在上例的总体中随机抽出一个样本如下：

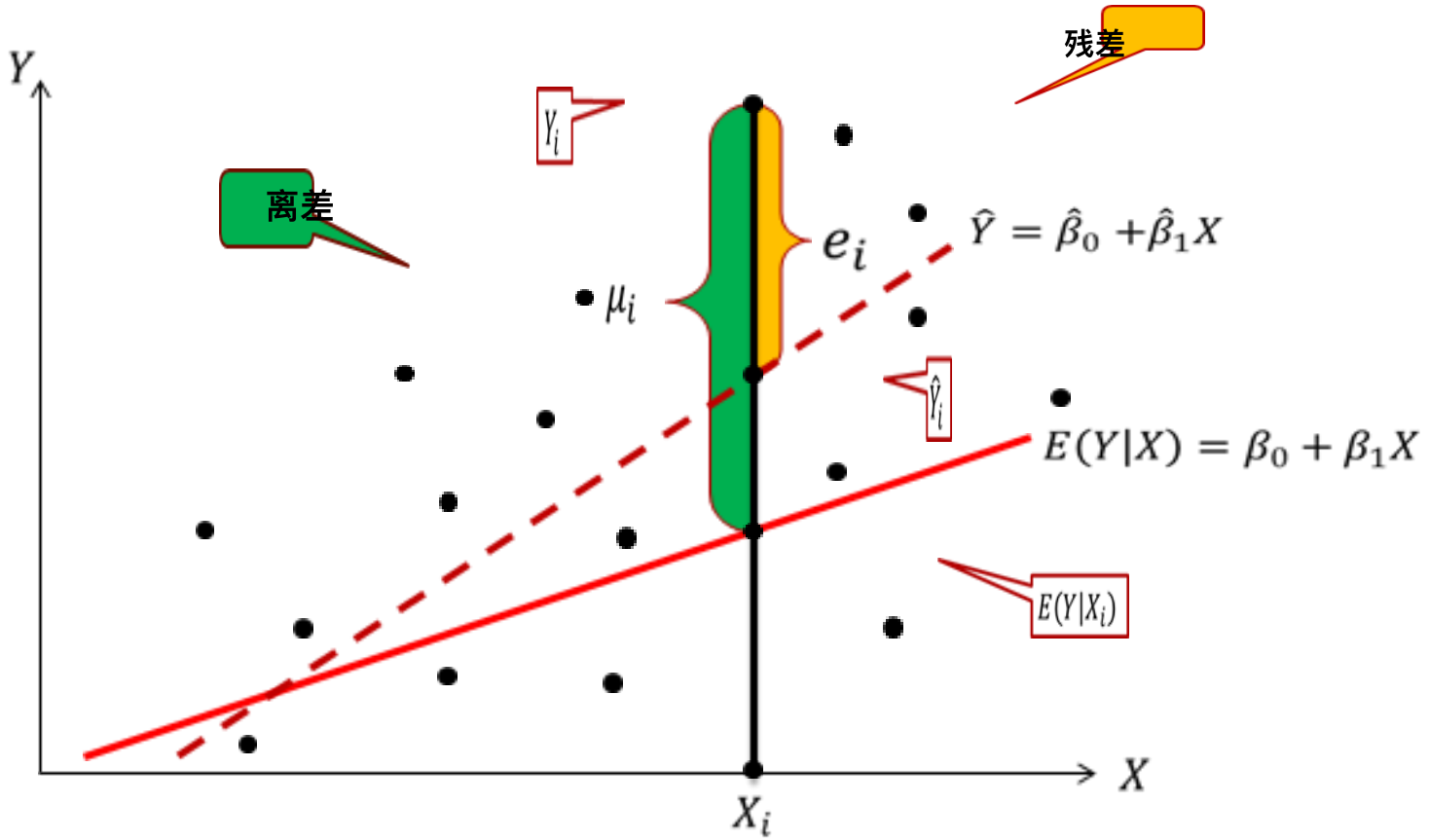
Y	800	1100	1400	1700	2000	2300	2600	2900	3200	3500
X	594	638	1122	1155	1408	1595	1969	2078	2585	2530



- 画一条直线尽可能地拟合该散点图，该直线称为样本回归线。
- 由于样本取自总体，可用样本回归线近似地代表总体回归线。
- 样本回归线的函数形式为：$\hat{Y} = f(X_i) = \hat{\beta}_0 + \hat{\beta}_1 X_i$

称其为样本回归函数。

样本回归线与总体回归线之间的关系如图所示：



1. 总体回归函数的确定性形式：$E(Y|X)=\beta_0+\beta_1X$

2. 总体回归函数的随机形式：$Y=\beta_0+\beta_1X+\mu$

3. 样本回归函数的确定性形式：$\hat{Y}=\hat{\beta}_0+\hat{\beta}_1X$

4. 样本回归函数的随机形式：$Y=\hat{\beta}_0+\hat{\beta}_1X+e$

【52】

四、随机干扰项

- 该案例中，总体回归函数说明在给定的收入水平下，该社区家庭平均的消费支出水平。
- 但对某一个别的家庭，其消费支出可能与该平均水平有偏差。
- 该偏差称为观测值围绕它的期望值的离差，是一个不可观测的随机变量，又称为随机误差项或随机干扰项。

$$\mu_i = Y_i - E(Y|X_i) = Y_i - \beta_0 - \beta_1 X_i$$

$$\boldsymbol{Y_i = E(Y|X_i) + \mu_i = \beta_0 + \beta_1 X_i + \mu_i}$$

引入随机误差项的主要原因：

1. 代表未知的影响因素；
2. 代表残缺数据；
3. 代表细小的影响因素；
4. 代表数据观测误差；
5. 代表模型设定误差；
6. 代表变量内在的随机性。

第二章 一元线性回归模型

2.1 回归分析概述

2.2 一元线性模型的参数估计

2.3 一元线性回归模型的统计检验

一元线性回归模型的一般形式：

$$Y_i = \beta_0 + \beta_1 X_i + \mu_i$$

其中：β_0和β_1为待估参数；

μ_i为随机扰动项；

Y为被解释变量，X为解释变量。

估计任务 $\begin{cases} \text{模型结构参数}\beta_0\text{和}\beta_1 \\ \text{模型分布函数}\sigma^2\text{（}\sigma^2\text{为随机扰动项的方差）} \end{cases}$

回归分析的目的之一是根据样本回归函数，估计总体回归函数。

这就要求：设计一种方法构造样本回归函数，使其尽可能接近总体回归函数；或者说，使$\hat{\beta}_0$、$\hat{\beta}_1$和$\hat{\sigma}^2$尽可能接近β_0、β_1和σ^2。

- 估计方法有多种，其中最广泛使用的是**普通最小二乘法**（Ordinary Least Squares, OLS）。
- 为保证参数估计量具有良好的性质，通常对模型提出若干基本假设。满足基本假设，模型可用OLS方法估计，否则OLS方法不适用，应发展新的方法。

【注意】基本假设是针对计量经济模型的估计方法提出来的，而不是针对模型本身。

2.2 一元线性回归模型的参数估计

一、参数估计的普通最小二乘法

二、关于模型的基本假设

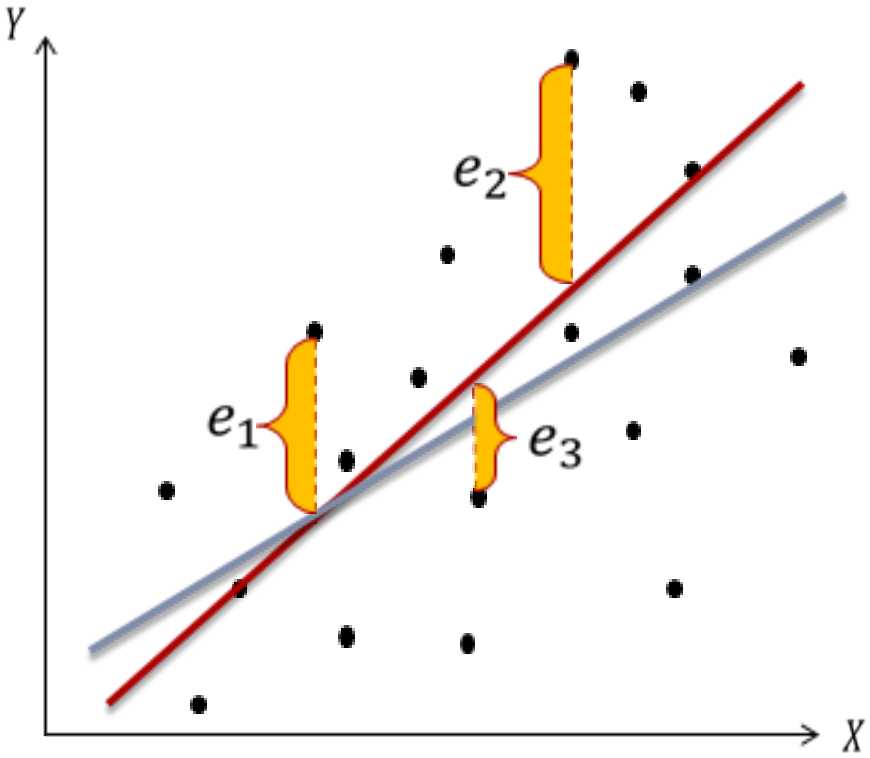
三、随机误差项的方差估计

一、参数估计的普通最小二乘法（OLS）

（一）普通最小二乘估计法

OLS方法基本原理：给定观测值下，样本回归函数尽可能好的拟合该组数值。

OLS方法判定标准：点到直线距离最小。



$$\sum e_i^2 = \sum (Y_i - \hat{Y}_i)^2 = \sum (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)^2 = Q$$

为使Q达到最小（$Q \to min$），应满足

$$\begin{cases} \dfrac{\partial Q}{\partial \hat{\beta}_0} = 0 \\ \dfrac{\partial Q}{\partial \hat{\beta}_1} = 0 \end{cases} \Rightarrow \begin{cases} \sum (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i) = 0 \\ \sum (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i) X_i = 0 \end{cases} \Rightarrow \begin{cases} \sum Y_i = n\hat{\beta}_0 + \hat{\beta}_1 \sum X_i \\ \sum Y_i X_i = \hat{\beta}_0 \sum X_i + \hat{\beta}_1 \sum X_i^2 \end{cases}$$

得到：

$$\begin{cases} \hat{\beta}_0 = \dfrac{\Sigma X_i^2 \Sigma Y_i - \Sigma X_i \Sigma Y_i X_i}{n\Sigma X_i^2 - (\Sigma X_i)^2} \\ \hat{\beta}_1 = \dfrac{n\Sigma Y_i X_i - \Sigma Y_i \Sigma X_i}{n\Sigma X_i^2 - (\Sigma X_i)^2} \end{cases}$$

记：$\begin{cases} \sum x_i^2 = \sum (X_i - \bar{X})^2 \\ \sum x_i y_i = \sum (X_i - \bar{X})(Y_i - \bar{Y}) \end{cases}$

得到：$\begin{cases} \hat{\beta}_1 = \dfrac{\sum x_i y_i}{\sum x_i^2} \\ \hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X} \end{cases}$

上述方程组称为OLS估计量的离差形式。

二、关于模型的基本假设

关于模型设定的假设

假设1：回归模型是正确设定的。

正确设定模型包括：

① 模型选择了正确的变量

② 模型选择了正确的函数形式

误设定包括下列类型：

① 选择错误的函数形式

② 遗漏重要的解释变量

③ 选择无关的解释变量

关于解释变量的假设

假设2：随着样本容量的增加，解释变量X的样本方差趋于一个非零的有限常数。

$$\sum_{i=1}^{n}(X_i-\bar{X})^2/n \to Q \quad , \quad n \to \infty$$

假设目的：排除时间序列出现持续上升或下降的变量作为解释变量，这类数据往往出现伪回归问题。

【注意】该假设是唯一与多元线性回归模型基本假设不同的假设。

关于随机干扰项

假设3：随机干扰项具有给定X条件下的零均值。

$$E(\mu_i|X) = 0$$

也就是说，μ的期望不依赖X的变化而变化，总为常数零。即μ与X不存在任何形式的相关性。

$$Cov(X, \mu_i) = E(X\mu_i) - E(X)E(\mu_i) = E(X\mu_i) = 0$$

重要推论：$Cov(X_i, \mu_i) = 0$，X与μ同期不相关。

假设4：随机干扰项具有给定X条件下的同方差和不序列相关性。

$$Var(\mu_i|X) = \sigma^2 \qquad i = 1, 2, \dots, n$$

即μ的方差不随X的变化而变化，总为常数σ^2。

$$Cov(\mu_i, \mu_j|X) = 0 \qquad i \neq j$$

即给定X条件下，任意两个不同观测点的随机干扰项不相关。

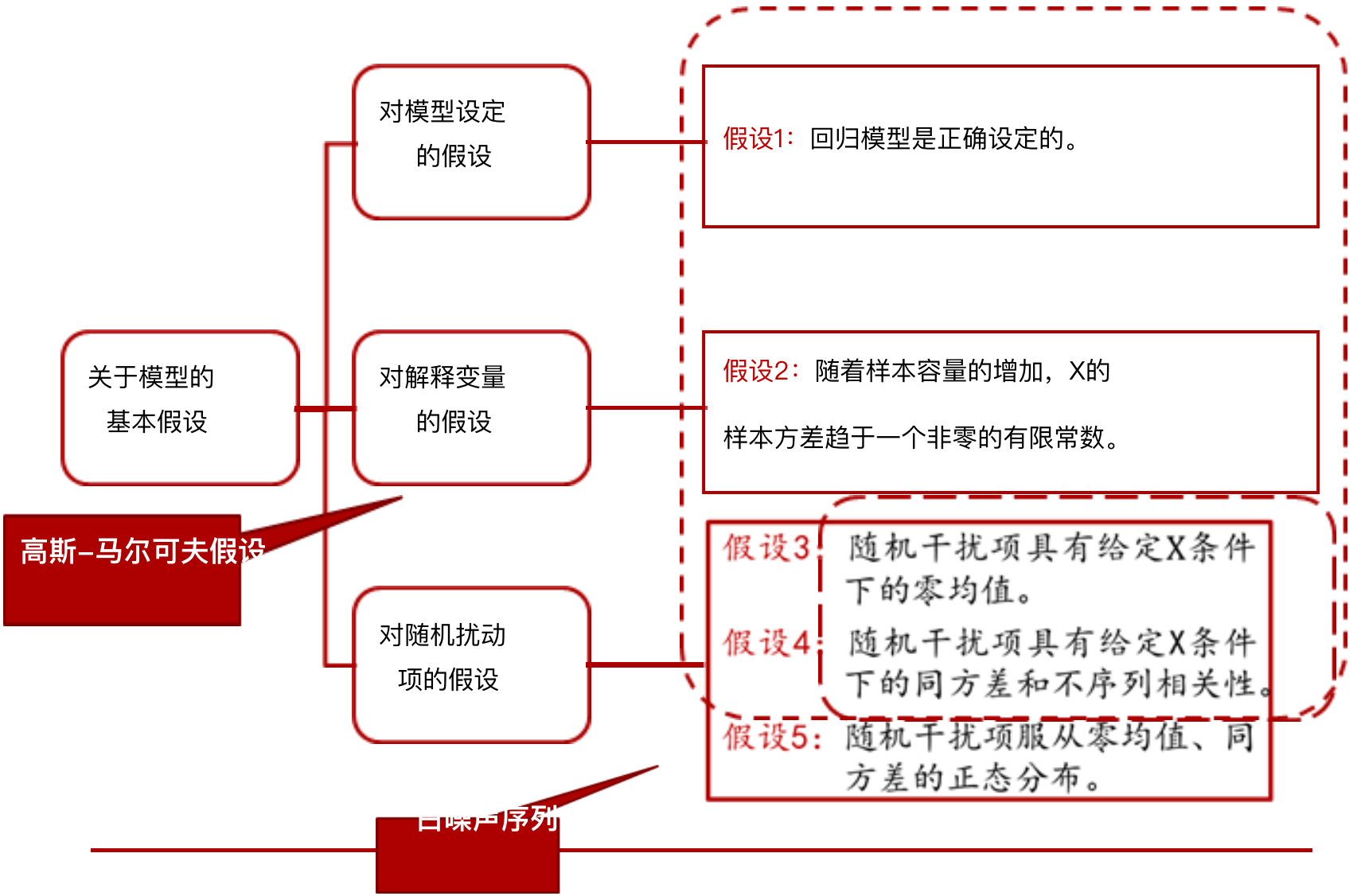
假设5：随机干扰项服从零均值、同方差的正态分布。

$$\mu_i|X \sim N(0, \sigma^2)$$

根据中心极限定理，在大样本下，可以放宽该假设。

【注意】以上所有假设的目的：保证估计方法的良好效果。

总结：回归模型的经典假设



（二）最小二乘估计量的统计性质

估计得到模型参数后，需考虑参数估计值的精度，即是否能代表总体参数的真值。

小样本性质

一旦具有，不以样本大小而改变

线性性：是否是另一个随机变量的线性函数

无偏性：均值或者期望是否等于总体的真实值

有效性：是否在所有线性无偏估计中有最小方差

评价估计量的标准

渐进无偏性：样本容量无穷大时，均值序列是否趋于总体真值

一致性：样本容量无穷大时，是否依概率收敛于总体真值

渐进有效性：样本无穷大时，所有的一致估计量是否有最小渐进方差

最佳线性无偏估计量（BLUE）

- **线性性**

由

$$\begin{cases}\hat{\beta}_1 = \dfrac{\sum x_i y_i}{\sum x_i^2} \\ \hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}\end{cases}$$

得到

$$\hat{\beta}_1 = \frac{\sum x_i y_i}{\sum x_i^2} = \frac{\sum x_i (Y_i - \bar{Y})}{\sum x_i^2} = \frac{\sum x_i Y_i}{\sum x_i^2} - \frac{\bar{Y} \sum x_i}{\sum x_i^2}$$

因为$\sum x_i = \sum (X_i - \bar{X}) = 0$，因此

$$\hat{\beta}_1 = \sum k_i Y_i$$

其中，$k_i = \frac{x_i}{\sum x_i^2}$。即估计量$\hat{\beta}_1$是$Y_i$的线性组合。

同样，可以得到

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$$

$$= \frac{1}{n}\sum Y_i - \sum k_i Y_i \bar{X}$$

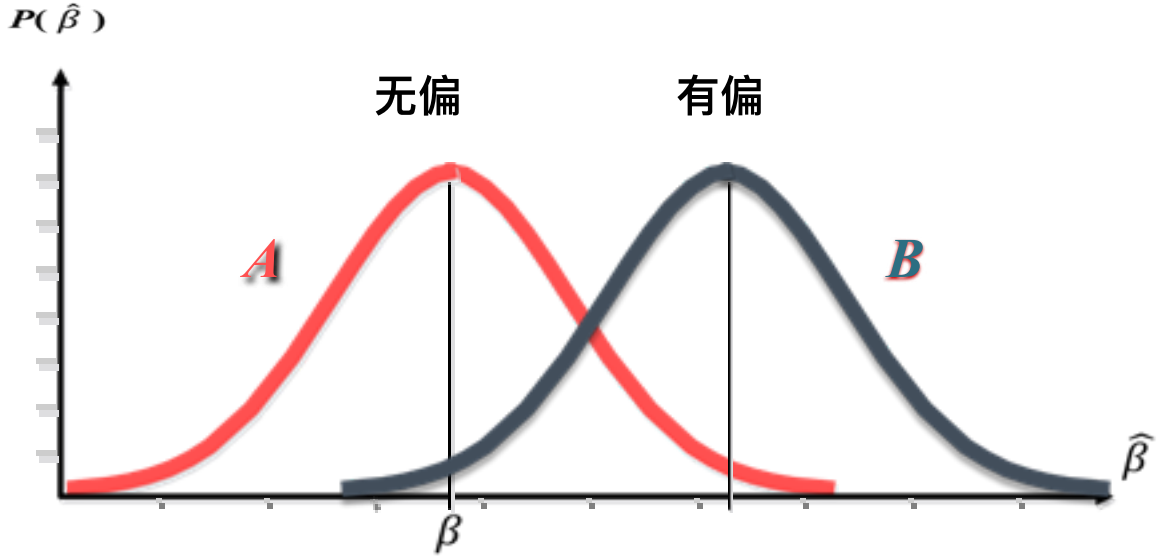
$$= \sum \left(\frac{1}{n} - \bar{X} k_i\right) Y_i$$

$$= \sum \omega_i Y_i$$

其中，$k_i = \frac{x_i}{\sum x_i^2}$ ，$\omega_i = \frac{1}{n} - \bar{X} k_i$。即估计量$\hat{\beta}_0$是$Y_i$的线性组合。

无偏性

具体来说，无偏性是指以X的所有样本值为条件，估计量$\hat{\beta}_0$、$\hat{\beta}_1$的均值（期望）等于总体回归参数真值β_0、β_1。即$E(\hat{\beta})=\beta$。



通过线性性得到

$$\hat{\beta}_1 = \sum k_i Y_i = \sum k_i (\beta_0 + \beta_1 X_i + \mu_i)$$

$$= \beta_0 \sum k_i + \beta_1 \sum k_i X_i + \sum k_i \mu_i$$

其中

$$\sum k_i = \frac{\sum x_i}{\sum x_i^2} = 0 \quad , \quad \sum k_i X_i = 1$$

因此

$$\hat{\beta}_1 = \beta_1 + \sum k_i \mu_i$$

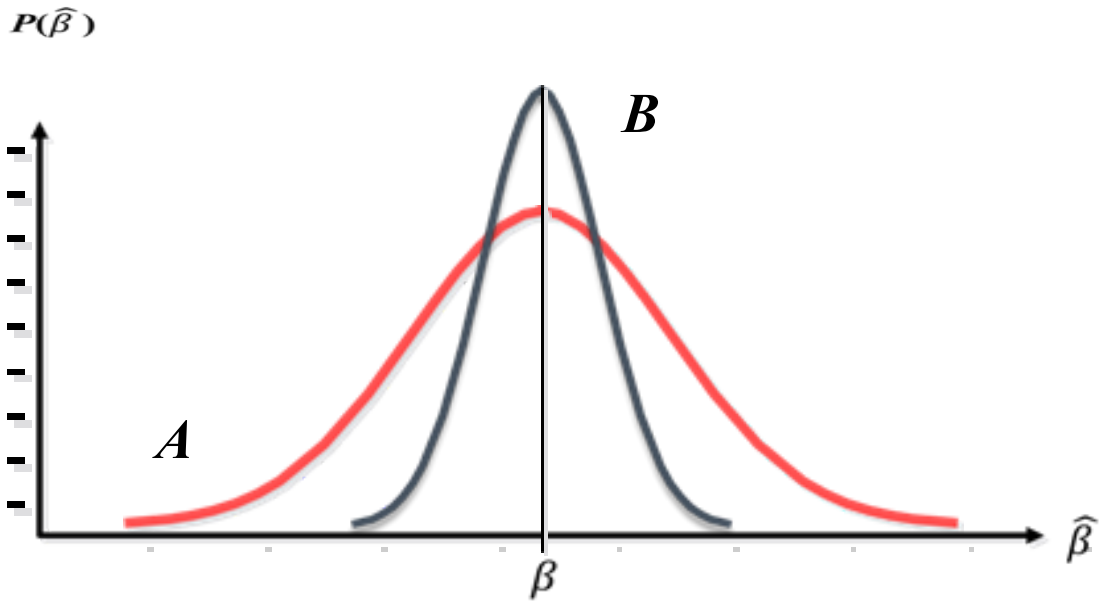
$$E(\hat{\beta}_1 | X) = E\left[\left(\beta_1 + \sum k_i \mu_i\right) \middle| X\right] = \beta_1 + \sum k_i E(\mu_i) = \beta_1$$

同理可以证明

$$E(\hat{\beta}_0 | X) = E\left[\left(\beta_0 + \sum \omega_i \mu_i\right) \middle| X\right] = \beta_0 + \sum \omega_i E(\mu_i | X) = \beta_0$$

• 有效性（最小方差性）

在所有线性无偏估计量中，普通最小二乘估计量是具有最小方差的估计量。$Var(\hat{\beta}) \rightarrow min$



估计量$\hat{\beta}_0$、$\hat{\beta}_1$是关于Y_i的线性函数，条件方差为

$$Var(\hat{\beta}_1|X) = Var\left(\sum k_i Y_i|X\right) = \sum {k_i}^2 \mathrm{Var}[(\beta_0 + \beta_1 X_i + \mu_i)|\mathrm{X}]$$

$$= \sum {k_i}^2 Var(\mu_i|X) = \sum (\frac{x_i}{\sum {x_i}^2})^2 \sigma^2 = \frac{\sigma^2}{\sum {x_i}^2}$$

同理可得

$$Var(\hat{\beta}_0|X) = \frac{\sum {X_i}^2}{n\sum {x_i}^2}\sigma^2$$

能够证明，在总体回归参数β_0、β_1的所有无偏估计中，$\frac{\sum {X_i}^2}{n\sum {x_i}^2}\sigma^2$和$\frac{\sigma^2}{\sum {x_i}^2}$是最小方差。

- 普通最小二乘估计量具有线性性、无偏性和有效性等优良性质，是**最佳线性无偏估计量**，这就是著名的**高斯–马尔可夫定理**。

- 也就是说，在满足基本假设的条件下，最小二乘估计量是具有最小方差的线性无偏估计量，即BLUE 。

参数估计的最大似然法*

- **最大似然法**(Maximum Likelihood, ML)，是不同于最小二乘法的另一种参数估计方法，是从极大似然原理出发发展起来的估计方法。
- 普通最小二乘法是从模型总体抽取容量为n的样本观测值后，求得能最好拟合样本数据的参数估计量。而对于最大似然法，当从模型总体随机抽取容量为n的样本观测值后，最合理的参数估计量应该使得从模型中抽取该样本观测值的概率最大。
- **理论基础：**大概率事件发生的可能性比小概率事件大，因此实际发生的更有可能是大概率事件。
- **最大似然法的前提：**总体分布已知。

在满足基本假设条件下，对一元线性回归模型：

$$Y_i = \beta_0 + \beta_1 X_i + \mu_i$$

随机抽取n组样本观测值$(X_i\ ,\ Y_i)$，$(i=1,2,\cdots n)$。

假如模型的参数估计量已经求得，为$\hat{\beta}_0$、$\hat{\beta}_1$和$\hat{\sigma}^2$，那么Y_i服从正态分布：$Y_i \sim N(\hat{\beta}_0 + \hat{\beta}_1 X_i, \hat{\sigma}^2)$

于是，Y的概率密度函数为：

$$P(Y_i) = \frac{1}{\sqrt{2\pi}\hat{\sigma}} e^{-\frac{1}{2\hat{\sigma}^2}(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)^2} \qquad (i = 1,2,\cdots,n)$$

因为Y_i是相互独立的，因此所有样本观测值的联合概率，也即似然函数为：

$$L(\hat{\beta}_0,\hat{\beta}_1,\hat{\sigma}) = P(Y_1,Y_2,\cdots,Y_n)$$

$$= \frac{1}{(2\pi)^{\frac{n}{2}}\hat{\sigma}^n} e^{-\frac{1}{2\hat{\sigma}^2}\Sigma(Y_i-\hat{\beta}_0-\hat{\beta}_1X_i)^2}$$

将似然函数最大化，即可求得模型参数的最大似然估计量。

将似然函数两边取对数，得到对数似然函数如下：

$$L^* = \ln(L)$$

$$= -n\ln(\sqrt{2\pi}\hat{\sigma}) - \frac{1}{2\hat{\sigma}^2}\sum(Y_i-\hat{\beta}_0-\hat{\beta}_1X_i)^2$$

对L^*求最大值，等价于对$\sum(Y_i-\hat{\beta}_0-\hat{\beta}_1X_i)^2$求最小值，得

$$\begin{cases} \dfrac{\partial}{\partial\hat{\beta}_0}\Sigma(Y_i-\hat{\beta}_0-\hat{\beta}_1X_i)^2 = 0 \\ \dfrac{\partial}{\partial\hat{\beta}_1}\Sigma(Y_i-\hat{\beta}_0-\hat{\beta}_1X_i)^2 = 0 \end{cases}$$

解得模型的参数估计量为：

$$\begin{cases} \hat{\beta}_1 = \dfrac{\Sigma x_i y_i}{\Sigma x_i^2} \\ \hat{\beta}_0 = \overline{Y} - \hat{\beta}_1 \overline{X} \end{cases}$$

可见，在满足基本假设的情况下，模型结构参数的极大似然估计量与普通最小二乘估计量是相同的。

三、随机误差项的方差估计

- 为了测定估计值精度，需要对其概率分布进行确定。
- 由于$\hat{\beta}_0$、$\hat{\beta}_1$分别是Y_i的线性组合，因此$\hat{\beta}_0$、$\hat{\beta}_1$的概率分布取决于Y_i。
- 在μ_i是正态分布的前提下，Y_i是正态分布，则$\hat{\beta}_0$、$\hat{\beta}_1$也服从正态分布，其分布由其均值和方差唯一决定。
- 即$\hat{\beta}_1 \sim N(\beta_1, \frac{\sigma^2}{\sum x_i^2})$，$\hat{\beta}_0 \sim N(\beta_0, \frac{\sum X_i^2}{n\sum x_i^2}\sigma^2)$

- 上式中：σ^2 ——随机干扰项的方差、总体方差。
- 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 σ^2进行估计。
- 得到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，即： $E(\hat{\sigma}^2) = \sigma^2$

σ^2估计后，得到：

- $\hat{\beta}_1$的样本方差 $S^2_{\hat{\beta}_1} = \dfrac{\hat{\sigma}^2}{\sum x_i^2}$
- $\hat{\beta}_1$的样本标准差 $S_{\hat{\beta}_1} = \dfrac{\hat{\sigma}}{\sqrt{\sum x_i^2}}$
- $\hat{\beta}_0$的样本方差 $S^2_{\hat{\beta}_0} = \dfrac{\hat{\sigma}^2 \sum X_i^2}{n \sum x_i^2}$
- $\hat{\beta}_0$的样本标准差 $S_{\hat{\beta}_0} = \hat{\sigma}\sqrt{\dfrac{\sum X_i^2}{n \sum x_i^2}}$

第二章 一元线性回归模型

2.1 回归分析概述

2.2 一元线性模型的参数估计

2.3 一元线性回归模型的统计检验

2.3 一元线性回归模型的统计检验

- 回归分析是要通过样本所估计的参数来代替总体的真实参数，或者说是用样本回归线代替总体回归线。
- 尽管从统计性质上已知，如果有足够多的重复抽样，参数的估计值的期望（均值）就等于其总体的参数真值，但在一次抽样中，估计值不一定就等于该真值。那么，在一次抽样中，参数的估计值与真值的差异有多大，是否显著，这就需要进一步进行统计检验。

主要包括： 拟合优度检验、方程的显著性检验、变量的显著性检验。

检验目的： 考察在一次抽样中，参数的估计值与真值的差异有多大，是否显著。

2.3 一元线性回归模型的统计检验

一、拟合优度检验

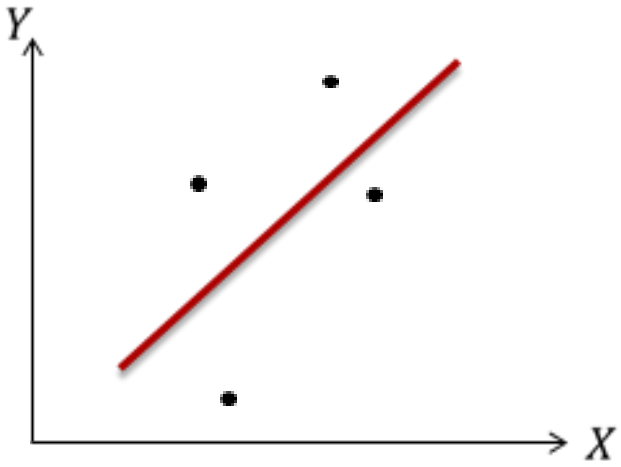
二、方程的显著性检验*

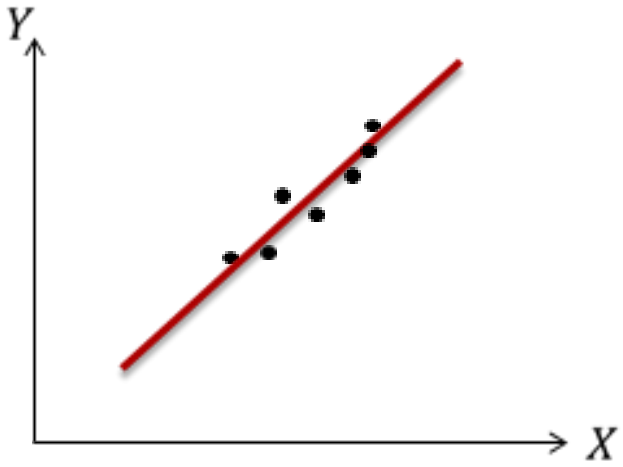
三、变量的显著性检验

一、拟合优度检验

定义：对样本回归线与样本观测值之间拟合程度的检验，度量样本回归线与实际观测值点的远近。

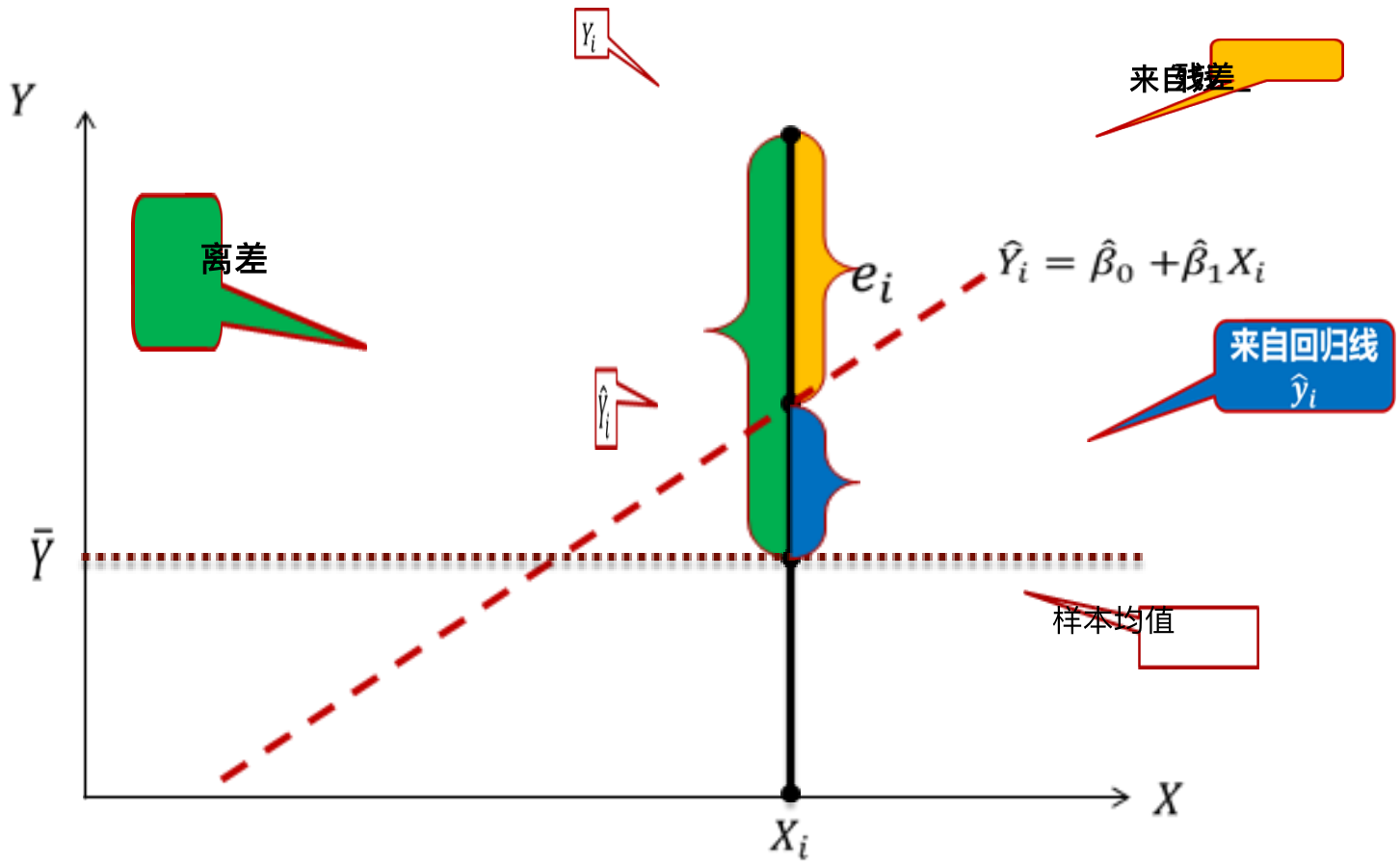
【注意】普通最小二乘法是对同一个问题的内部比较，拟合优度检验结果所表示的是不同问题间的比较。





度量拟合优度的指标：判定系数——R^2，又称可决系数。

（一）总离差平方和的分解


Y_i
来自残差
离差
e_i
$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$
来自回归线
$\hat{y}_i$
$\hat{Y}_i$
Y
$\bar{Y}$
样本均值
X
X_i

由一组样本观测值(X_i, Y_i),得到样本回归线：$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$

对于一个点来说，离差可以分解为：$y_i = Y_i - \bar{Y} = e_i + \hat{y}_i$

对所有点：$\sum y_i^2 = \sum \hat{y}_i^2 + \sum e_i^2 + 2\sum \hat{y}_i e_i = \sum \hat{y}_i^2 + \sum e_i^2$

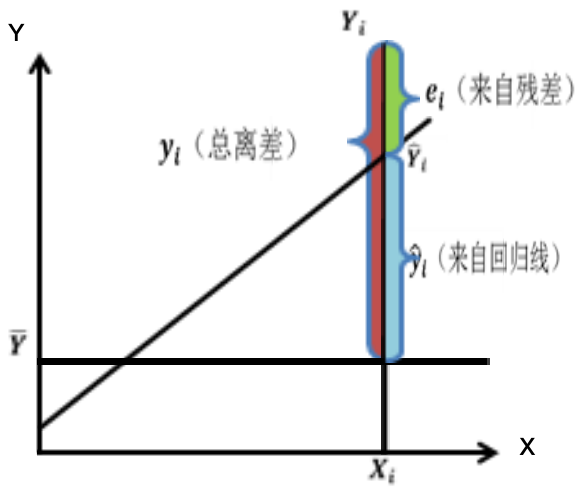
记：

$TSS = \sum y_i^2 = \sum (Y_i - \bar{Y})^2$ ——总离差平方和

$ESS = \sum \hat{y}_i^2 = \sum (\hat{Y}_i - \bar{Y})^2$ ——回归平方和

$RSS = \sum e_i^2 = \sum (Y_i - \hat{Y}_i)^2$ ——残差平方和

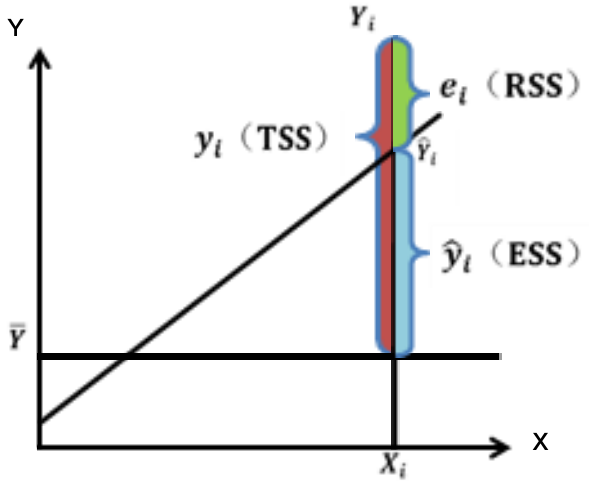
$TSS = ESS + RSS$



- Y的观测值围绕其均值的总离差(TSS)可分解为两部分：一部分来自回归线(ESS)，另一部分则来自随机势力(RSS)。
- 在给定样本中，TSS不变。**ESS在TSS中所占的比重越大，则模型的拟合效果越好；**

（二）R^2统计量

$$R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS}$$



R^2的取值范围：[0，1]

R^2越接近1，说明实际观测值离样本回归线越近，拟合效果越好。

R^2的意义：解释变量X的变动可以解释被解释变量Y多大程度上的变动。

二、方程的显著性检验

一元线性回归模型，只有一个解释变量，方程的显著性检验等价于变量的显著性检验。

方程的显著性检验（略）

三、变量的显著性检验

检验目标：在一元线性回归模型中，变量的显著性检验就是判断X是否对Y具有显著的线性影响。

检验方法：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。变量的显著性检验所用的方法是数理统计学中的假设检验。

（一）假设检验

检验原理：事先对总体参数（总体分布形式）做出一个假设，然后利用样本信息来判断原假设是否合理，即判断样本信息与原假设是否有显著差异，从而决定接受或拒绝原假设。

检验方法：反证法。先假定原假设正确，然后根据样本信息，观察由此假设而导致的结果是否合理，从而判断是否接受原假设。

检验依据：判断结果合理与否，是基于“小概率事件不易发生”这一原理的。

（二）变量的显著性检验

$\hat{\beta}_1$的分布函数

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

由于其中真实的σ^2是未知的，用它的无偏估计量$\frac{\sum {e_i}^2}{n-2}$替代时，可构造统计量

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

用该统计量作为β_1显著性检验的**t统计量**

变量显著性检验的步骤：

<1> 对总体参数提出原假设和备选假设。

$$H_0 : \beta_1 = 0 \quad , \quad H_1 : \beta_1 \neq 0 .$$

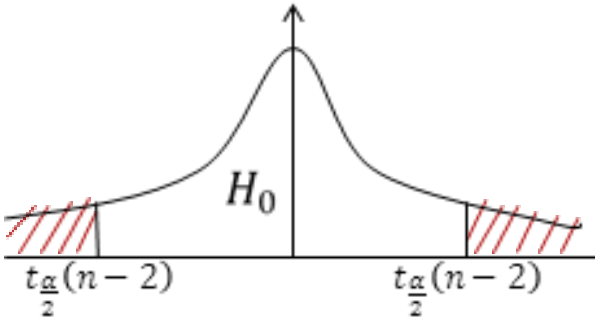
<2> 以原假设H_0构造t统计量，并计算t值. $t = \dfrac{\hat{\beta}_1}{S_{\hat{\beta}_1}}$

<3> 给定显著性水平α，查t分布表得临界值$t_{\frac{\alpha}{2}}(n-2)$

<4> 比较，判断：

若|t| $> t_{\frac{\alpha}{2}}(n-2)$ ，接受H_1。

若|t| $\leq t_{\frac{\alpha}{2}}(n-2)$ ，接受H_0。



第二章·小结

1、建立回归模型

- 研究某一经济现象，先根据经济理论，选择具有因果关系的两个变量（Y，X），建立一元线性回归模型。
- 如果不明确两个变量是否为线性关系，也可以根据散点图来分析。
- 建立模型时，不仅要有理论依据，同时也要考虑数据的可得性。

2、收集样本数据

收集样本数据并经过适当的加工整理，得到适于回归分析的样本数据。

3、估计模型参数

利用样本数据，以OLS得到模型参数的估计值。

4、表述回归结果

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X$$

$$[se(\hat{\beta}_0)] \quad [se(\hat{\beta}_1)] \qquad R^2 = ?$$

或者

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X$$

$$[t_1] \qquad [t_2] \qquad R^2 = ?$$

5、对回归模型进行检验

经济意义检验：参数的符号、大小及相互关系是否与经济理论相符。若不符，寻找原因。（数据？模型设定？）

统计检验：拟合优度检验；变量的显著性检验。

例题

为考察中国居民2013年人均可支配收入与人均消费支出的关系，统计了中国内地31个省、市、自治区以当年价格测算的居民家庭年人均可支配收入与人均消费支出，如下表所示。

请据此建立模型并对模型进行预测。

表2-1 中国各地区居民家庭人均全年可支配收入与人均全年消费性支出（元）

地区	可支配收入	消费支出	地区	可支配收入	消费支出
	X	Y		X	Y
北 京	40830	29175.6	湖 北	16472.5	11760.8
天 津	26359.2	20418.7	湖 南	16004.9	11945.9
河 北	15189.6	10872.2	广 东	23420.7	17421
山 西	15119.7	10118.3	广 西	14082.3	9596.5
内蒙古	18692.9	14877.7	海 南	15733.3	11192.9
辽 宁	20817.8	14950.2	重 庆	16568.7	12600.2
吉 林	15998.1	12054.3	四 川	14231	11054.7
黑龙江	15903.4	12037.2	贵 州	11083.1	8288
上 海	42173.6	30399.9	云 南	12577.9	8823.8
江 苏	24775.5	17925.8	西 藏	9746.8	6310.6
浙 江	29775	20610.1	陕 西	14371.5	11217.3
安 徽	15154.3	10544.1	甘 肃	10954.4	8943.4
福 建	21217.9	16176.6	青 海	12947.8	11576.5
江 西	15099.7	10052.8	宁 夏	14565.8	11292
山 东	19008.3	11896.8	新 疆	13669.6	11391.8
河 南	14203.7	10002.5			

资料来源：《中国统计年鉴》（2014）。