单方程计量经济学模型理论与方法

Theory and Methodology of Single-Equation Econometric Model

第二章
一元线性回归模型

- 回归分析概述
- 一元线性回归模型的参数估计
- 一元线性回归模型检验
- 一元线性回归模型预测
- 实例

§2.1　回归分析概述

一、变量间的关系及回归分析的基本概念

二、总体回归函数

三、随机扰动项

四、样本回归函数（SRF）

§2.1 回归分析概述

一、变量间的关系及回归分析的基本概念

1、变量间的关系

经济变量之间的关系，大体可分为两类：

（1）**确定性关系**或**函数关系**：研究的是确定现象非随机变量间的关系。

（2）**统计依赖**或**相关关系**：研究的是非确定现象随机变量间的关系。

例如：

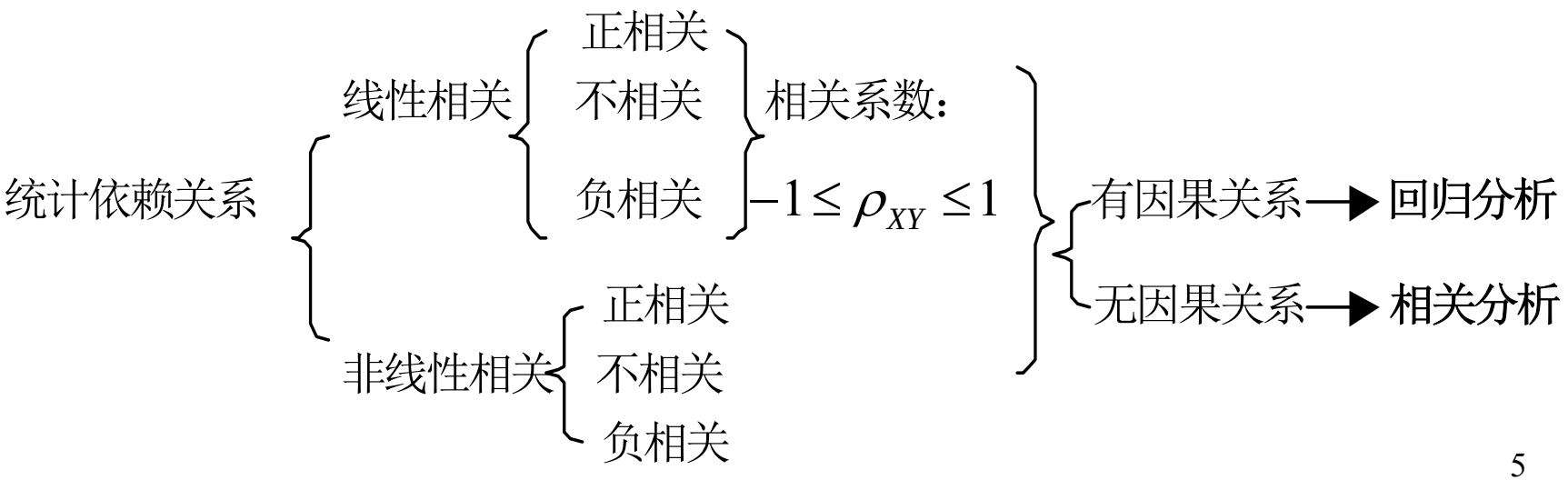
函数关系：

$$\text{圆面积} = f(\pi, \text{半径}) = \pi \cdot \text{半径}^2$$

统计依赖关系/统计相关关系：

$$\text{农作物产量} = f(\text{气温}, \text{降雨量}, \text{阳光}, \text{施肥量})$$

对变量间统计依赖关系的考察主要是通过相关分析(correlation analysis)或回归分析(regression analysis)来完成的：



▲注意：

①不线性相关并不意味着不相关；

②有相关关系并不意味着一定有因果关系；

③回归分析/相关分析研究变量之间的统计依赖关系，并能度量线性依赖程度的大小。

④相关分析对称地对待任何（两个）变量，两个变量都被看作是随机的。回归分析对变量的处理方法存在不对称性，即区分应变量（被解释变量）和自变量（解释变量）：前者是随机变量，后者不是。

2、回归分析的基本概念

回归分析(regression analysis)是研究一个变量关于另一个（些）变量的具体依赖关系的计算方法和理论。

其用意：在于通过后者的已知或设定值，去估计和（或）预测前者的（总体）均值。

这里：前一个变量被称为**被解释变量（Explained Variable）**或**应变量（Dependent Variable）**，后一个（些）变量被称为**解释变量（Explanatory Variable）**或**自变量（Independent Variable）**。

回归分析构成计量经济学的方法论基础，其主要内容包括：

（1）根据样本观察值对经济计量模型参数进行估计，求得**回归方程；**

（2）对回归方程、参数估计值进行显著性检验；

（3）利用回归方程进行分析、评价及预测。

二、总体回归函数

由于变量间关系的随机性，回归分析关心的是根据解释变量的已知或给定值，考察被解释变量的总体均值，即当解释变量取某个确定值时，与之统计相关的被解释变量所有可能出现的对应值的平均值。

例2.1：一个假想的社区有100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**的关系。

即如果知道了家庭的月收入，能否预测该社区家庭的平均月消费支出水平。

为达到此目的，将该100户家庭划分为组内收入差不多的10组，以分析每一收入组的家庭消费支出。

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

分析：

（1）由于不确定因素的影响，对同一收入水平X，不同家庭的消费支出不完全相同；

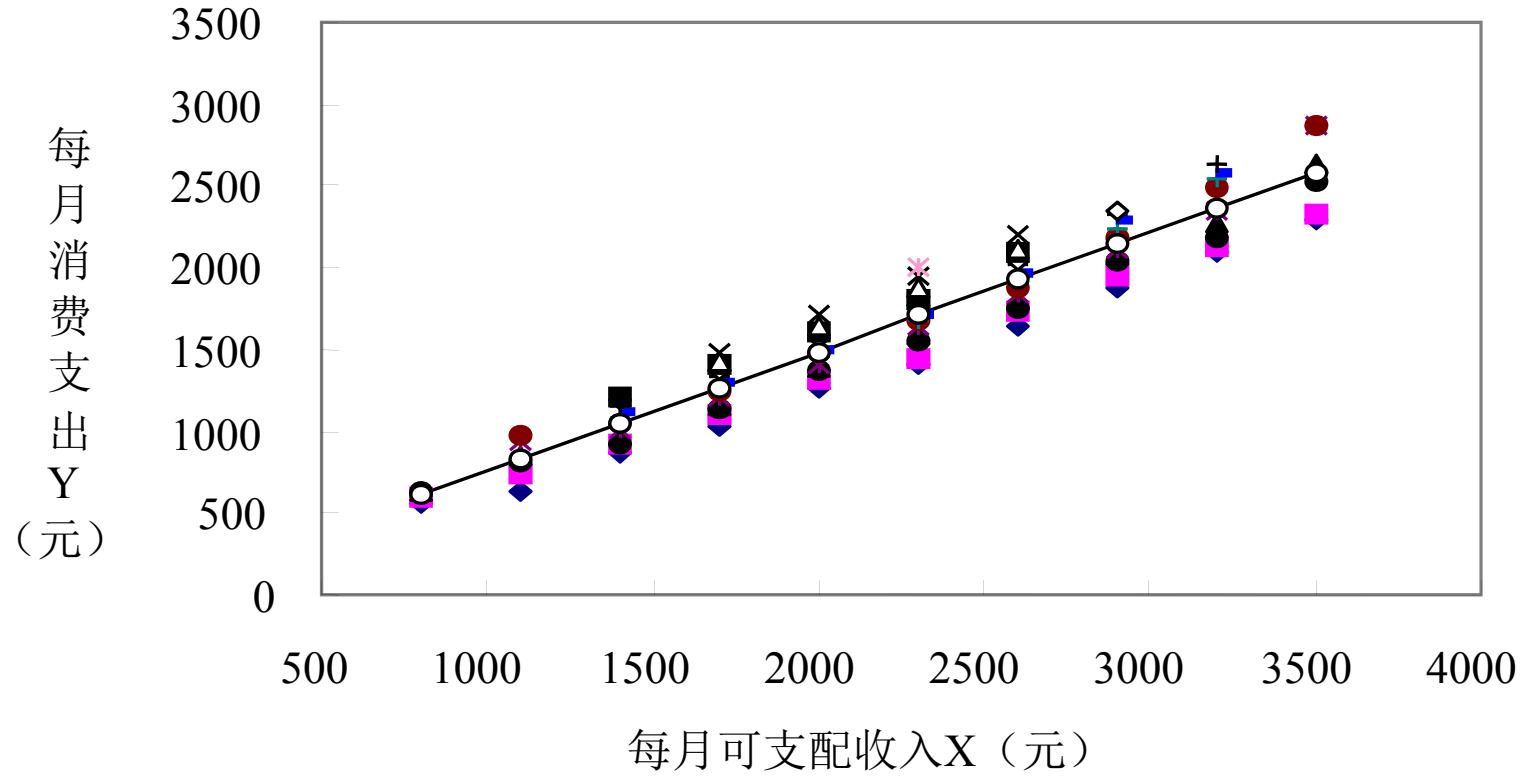
（2）但由于调查的完备性，给定收入水平X的消费支出Y的分布是确定的，即以X的给定值为条件的Y的**条件分布**（**Conditional distribution**）是已知的，如：　　P(Y=561|X=800）=1/4。

因此，给定收入X的值X_i，可得消费支出Y的**条件均值**（conditional mean）或**条件期望**（conditional expectation）：

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

该例中：E(Y | X=800)=605

描出散点图发现：随着收入的增加，消费“**平均地说**”也在增加，且Y的条件均值均落在一根正斜率的直线上。这条直线称为**总体回归线**。



- **概念：**

在给定解释变量X_i条件下被解释变量Y_i的期望轨迹称为**总体回归线**（population regression line），或更一般地称为**总体回归曲线**（population regression curve）。

相应的函数：

$$E(Y \mid X_i) = f(X_i)$$

称为（双变量）**总体回归函数**（population regression function, **PRF**）。

- **含义：**

回归函数（PRF）说明被解释变量Y的平均状态（总体条件期望）随解释变量X变化的规律。

- **函数形式：**

可以是线性或非线性的。

例2.1中，将居民消费支出看成是其可支配收入的线性函数时：

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

为一**线性函数**。其中，β_0，β_1是未知参数，称为**回归系数**（regression coefficients）。 。

三、随机扰动项

总体回归函数说明在给定的收入水平X_i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。

记 $$\mu_i = Y_i - E(Y \mid X_i)$$

称μ_i为观察值Y_i围绕它的期望值$E(Y|X_i)$的**离差**（**deviation**），是一个不可观测的随机变量，又称为**随机干扰项**（**stochastic disturbance**）或**随机误差项**（**stochastic error**）。

例2.1中，个别家庭的消费支出为：

$$Y_i = E(Y \mid X_i) + \mu_i = \beta_0 + \beta_1 X_i + \mu_i \qquad (*)$$

即，给定收入水平X_i,个别家庭的支出可表示为两部分之和:

（1）该收入水平下所有家庭的平均消费支出$E(Y|X_i)$，称为系统性（systematic）或确定性（deterministic）部分。

（2）其他随机或非确定性（nonsystematic)部分μ_i。

（*）式称为**总体回归函数（方程）PRF的随机设定形式。表明被解释变量除了受解释变量的系统性影响外，还受其他因素的随机性影响。**

由于方程中引入了随机项，成为计量经济学模型，因此也称为**总体回归模型**。

随机误差项主要包括下列因素的影响：

1）在解释变量中被忽略的因素的影响；

2）变量观测值的观测误差的影响；

3）模型关系的设定误差的影响；

4）其它随机因素的影响。

产生并设计随机误差项的主要原因：

1）理论的含糊性；

2）数据的欠缺；

3）节省原则。

四、样本回归函数（SRF）

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个样本。

问题：能从一次抽样中获得总体的近似的信息吗？如果可以，如何从抽样中获得总体的近似信息？

例2.2：在例2.1的总体中有如下一个样本，

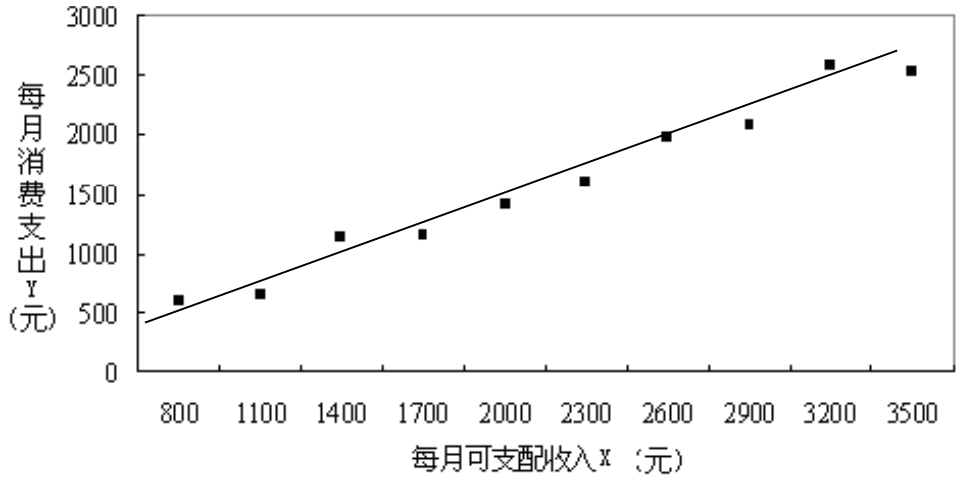
问：能否从该样本估计总体回归函数PRF？

表 2.1.3　家庭消费支出与可支配收入的一个随机样本

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

回答：能

核样本的散点图（scatter diagram)：



样本散点图近似于一条直线，画一条直线以尽好地拟合该散点图，由于样本取自总体，可以该线近似地代表总体回归线。该线称为**样本回归线（sample regression lines）**。

记样本回归线的函数形式为：

$$\hat{Y}_i = f(X_i) = \hat{\beta}_0 + \hat{\beta}_1 X_i$$

称为**样本回归函数（sample regression function，SRF）**。

注意：

这里将**样本回归线**看成**总体回归线**的近似替代

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$$

$$\Downarrow$$

$$\begin{aligned} Y_i &= E(Y \mid X_i) + \mu_i \\ &= \beta_0 + \beta_1 X_i + \mu_i \end{aligned}$$

则 $\hat{Y}_i$ 为 $E(Y \mid X_i)$ 的估计量；

$\hat{\beta}_i$ 为 β_i 的估计量，$i = (0,1)$

样本回归函数的随机形式/样本回归模型：

同样地，样本回归函数也有如下的随机形式：

$$Y_i = \hat{Y}_i + \hat{\mu}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i + e_i$$

式中，e_i 称为**（样本）残差**（或**剩余**）**项**（residual），代表了其他影响 Y_i 的随机因素的集合，可看成是 μ_i 的估计量 $\hat{\mu}_i$ 。

由于方程中引入了随机项，成为计量经济模型，因此也称为**样本回归模型**（**sample regression model**）。

▼**回归分析的主要目的**：根据样本回归函数SRF，估计总体回归函数PRF。

即，根据　　$Y_i = \hat{Y}_i + e_i = \hat{\beta}_0 + \hat{\beta}_1 X_i + e_i$

估计　　$Y_i = E(Y \mid X_i) + \mu_i = \beta_0 + \beta_1 X_i + \mu_i$

这就要求：

设计一“方法”构造SRF，以使SRF尽可能“接近”PRF，或者说使$\hat{\beta}_i (i = 0,1)$尽可能接近$\beta_i (i = 0,1)$。

注意：这里PRF可能永远无法知道。

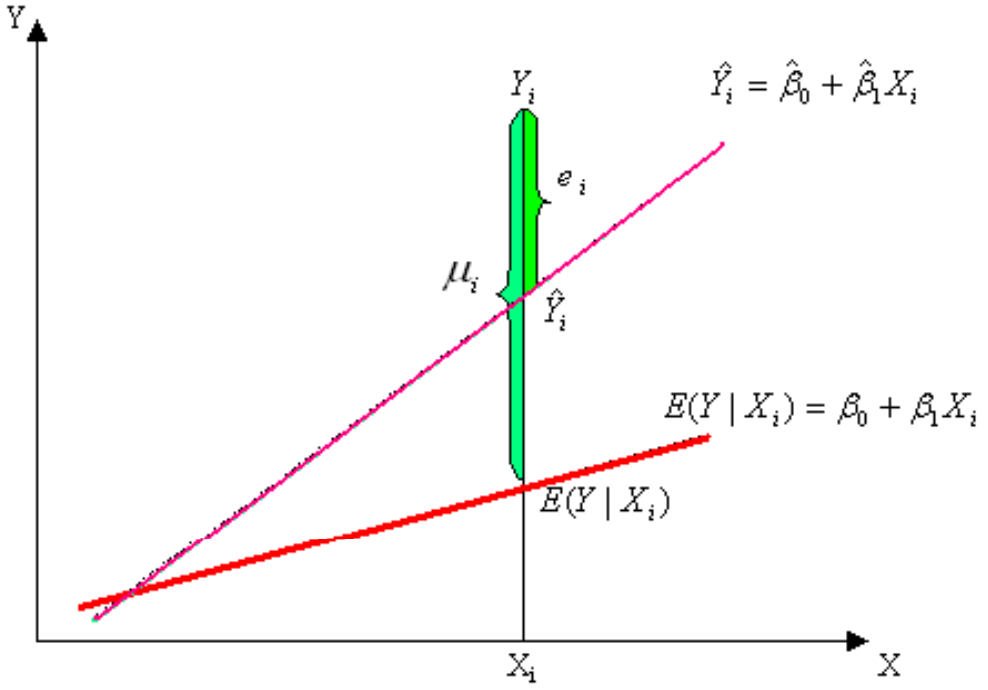


图 2.1.3　总体回归线与样本回归线的基本关系

§2.2 一元线性回归模型的参数估计

一、一元线性回归模型的基本假设

二、参数的普通最小二乘估计（OLS）

三、参数估计的最大或然法(ML)

四、最小二乘估计量的性质

五、参数估计量的概率分布及随机干扰项方差的估计

一元线性回归模型：

$$Y_i = \beta_0 + \beta_1 X_i + \mu_i \qquad i=1,2,\ldots,n$$

Y为被解释变量，X为解释变量，β_0与β_1为**待估参数**，μ为**随机干扰项**

回归分析的主要目的是要通过样本回归函数（模型）SRF尽可能准确地估计总体回归函数（模型）PRF。

估计方法有多种，其种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。

注：实际这些假设与所采用的估计方法紧密相关。

一、线性回归模型的基本假设

假设1、解释变量X是确定性变量，不是随机变量；

假设2、随机误差项μ具有零均值、同方差和不序列相关性：

$$E(\mu_i)=0 \qquad i=1,2,\ldots,n$$

$$Var(\mu_i)=\sigma_\mu^2 \qquad i=1,2,\ldots,n$$

$$Cov(\mu_i,\mu_j)=0 \qquad i\neq j \quad i,j=1,2,\ldots,n$$

假设3、随机误差项μ与解释变量X之间不相关：

$$Cov(X_i,\mu_i)=0 \qquad i=1,2,\ldots,n$$

假设4、μ服从零均值、同方差、零协方差的正态分布

$$\mu_i \sim N(0,\sigma_\mu^2) \qquad i=1,2,\ldots,n$$

注意：

1、如果假设1、2满足，则假设3也满足;

2、如果假设4满足，则假设2也满足。

以上假设也称为线性回归模型的**经典假设**或**高斯（Gauss）假设**，满足该假设的线性回归模型，也称为**经典线性回归模型**（Classical Linear Regression Model, CLRM）。

另外，在进行模型回归时，还有两个暗含的假设：

假设5：随着样本容量的无限增加，解释变量X的样本方差趋于一有限常数。即

$$\sum (X_i - \bar{X})^2 / n \to Q, \qquad n \to \infty$$

假设6：回归模型是正确设定的

假设5旨在排除时间序列数据出现持续上升或下降的变量作为解释变量，因为这类数据不仅使大样本统计推断变得无效，而且往往产生所谓的伪回归问题（spurious regression problem）。

假设6也被称为模型没有设定偏误（specification error）

二、参数的普通最小二乘估计（OLS）

给定一组样本观测值（X_i, Y_i）（i=1,2,…n）要求样本回归函数尽可能好地拟合这组值.

普通最小二乘法（Ordinary least squares, OLS）给出的判断标准是：二者之差的平方和

$$Q = \sum_1^n (Y_i - \hat{Y}_i)^2 = \sum_1^n (Y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_i))^2$$

最小。

即在给定样本观测值之下，选择出 $\hat{\beta}_0$、$\hat{\beta}_1$ 能使 Y_i 与 $\hat{Y}_i$ 之差的平方和最小。

根据微分运算，可推得用于估计$\hat{\beta}_0$ 、$\hat{\beta}_1$ 的下列方程组：

$$\begin{cases} \sum(\hat{\beta}_0 + \hat{\beta}_1 X_i - Y_i) = 0 \\ \sum(\hat{\beta}_0 + \hat{\beta}_1 X_i - Y_i)X_i = 0 \end{cases} \qquad (*)$$

或

$$\begin{cases} \Sigma Y_i = n\hat{\beta}_0 + \hat{\beta}_1 \Sigma X_i \\ \Sigma Y_i X_i = \hat{\beta}_0 \Sigma X_i + \hat{\beta}_1 \Sigma X_i^2 \end{cases}$$

解得：

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

方程组（*）称为**正规方程组**（**normal equations**）。

记 $$\sum x_i^2 = \sum (X_i - \bar{X})^2 = \sum X_i^2 - \frac{1}{n}\left(\sum X_i\right)^2$$

$$\sum x_i y_i = \sum (X_i - \bar{X})(Y_i - \bar{Y}) = \sum X_i Y_i - \frac{1}{n}\sum X_i \sum Y_i$$

上述参数估计量可以写成：
$$\begin{cases} \hat{\beta}_1 = \dfrac{\Sigma x_i y_i}{\Sigma x_i^2} \\ \hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X} \end{cases}$$

称为OLS估计量的离差形式（**deviation form**）。

由于参数的估计结果是通过最小二乘法得到的，故称为**普通最小二乘估计量（ordinary least squares estimators）**。

三、参数估计的最大或然法(ML)

最大或然法(Maximum Likelihood,简称ML)，也称最大似然法，是不同于最小二乘法的另一种参数估计方法，是从最大或然原理出发发展起来的其它估计方法的基础。

基本原理：

对于最大或然法，当从模型总体随机抽取n组样本观测值后，最合理的参数估计量应该使得从模型中抽取该n组样本观测值的概率最大。

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,…n）。

假如模型的参数估计量已经求得，为$\hat{\beta}_0$ 、$\hat{\beta}_1$

那么Y_i服从如下的正态分布：

$$Y_i \sim N(\hat{\beta}_0 + \hat{\beta}_1 X_i, \sigma^2)$$

于是，Y的概率函数为

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

因为Y_i是相互独立的，所以的所有样本观测值的联合概率，也即**或然函数**(likelihood function)为：

$$L(\hat{\beta}_0,\hat{\beta}_1,\sigma^2)=P(Y_1,Y_2,\cdots,Y_n)$$

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

将该或然函数极大化，即可求得到模型参数的极大或然估计量。

由于或然函数的极大化与或然函数的对数的极大化是等价的，所以，取对数或然函数如下：

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

$$= -n\ln(\sqrt{2\pi}\sigma) - \frac{1}{2\sigma^2}\Sigma(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)^2$$

对L^*求极大值，等价于对$\Sigma(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)^2$ 求极小值:

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

解得模型的参数估计量为：

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

可见，在满足一系列基本假设的情况下，模型结构参数的最大或然估计量与普通最小二乘估计量是相同的。

例2.2.1：在上述家庭**可支配收入-消费支出**例中，对于所抽出的一组样本数，参数估计的计算可通过下面的表2.2.1进行。

表 2.2.1　　参数估计的计算表

	X_i	Y_i	x_i	y_i	$x_i y_i$	x_i^2	y_i^2	X_i^2	Y_i^2
1	800	594	-1350	-973	1314090	1822500	947508	640000	352836
2	1100	638	-1050	-929	975870	1102500	863784	1210000	407044
3	1400	1122	750	445	334050	562500	198381	1960000	1258884
4	1700	1155	-450	-412	185580	202500	170074	2890000	1334025
5	2000	1408	-150	-159	23910	22500	25408	4000000	1982464
6	2300	1595	150	28	4140	22500	762	5290000	2544025
7	2600	1969	450	402	180720	202500	161283	6760000	3876961
8	2900	2078	750	511	382950	562500	260712	8410000	4318084
9	3200	2585	1050	1018	1068480	1102500	1035510	10240000	6682225
10	3500	2530	1350	963	1299510	1822500	926599	12250000	6400900
求和	21500	15674			5769300	7425000	4590020	53650000	29157448
平均	2150	1567							

$$\hat{\beta}_1 = \frac{\sum x_i y_i}{\sum x_i^2} = \frac{5769300}{7425000} = 0.777$$

$$\hat{\beta}_0 = \overline{Y} - \hat{\beta}_0 \overline{X} = 1567 - 0.777 \times 2150 = -103.172$$

因此，由该样本估计的回归方程为：

$$\hat{Y}_i = -103.172 + 0.777 X_i$$

四、最小二乘估计量的性质

当模型参数估计出后，需考虑参数估计值的精度，即是否能代表总体参数的真值，或者说需考察参数估计量的统计性质。

一个用于考察总体的估计量，可从如下几个方面考察其优劣性：

（1）线性性，即它是否是另一随机变量的线性函数；

（2）无偏性，即它的均值或期望值是否等于总体的真实值；

（3）有效性，即它是否在所有线性无偏估计量中具有最小方差。

这三个准则也称作估计量的小样本性质。

拥有这类性质的估计量称为最佳线性无偏估计量（**best liner unbiased estimator, BLUE**）。

当不满足小样本性质时，需进一步考察估计量的大样本或渐近性质：

（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）一致性，即样本容量趋于无穷大时，它是否依概率收敛于总体的真值；

（6）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。

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。

1、线性性，即估计量$\hat{\beta}_0$ 、 $\hat{\beta}_1$ 是Y_i 的线性组合。

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

令 $k_i = \frac{x_i}{\sum x_i^2}$，因 $\sum x_i = \sum (X_i - \bar{X}) = 0$，故有

$$\hat{\beta}_1 = \sum \frac{x_i}{\sum x_i^2} Y_i = \sum k_i Y_i$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X} = \frac{1}{n}\sum Y_i - \sum k_i Y_i \bar{X} = \sum (\frac{1}{n} - \bar{X}k_i) Y_i = \sum w_i Y_i$$

2、无偏性，即估计量$\hat{\beta}_0$、$\hat{\beta}_1$的均值（期望）等于总体回归参数真值β_0与β_1

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

易知 $$\sum k_i = \frac{\sum x_i}{\sum x_i^2} = 0 \qquad \sum k_i X_i = 1$$

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

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

同样地，容易得出

$$E(\hat{\beta}_0) = E(\beta_0 + \sum w_i \mu_i) = E(\beta_0) + \sum w_i E(\mu_i) = \beta_0$$

3、有效性（最小方差性），即在所有线性无偏估计量中，最小二乘估计量$\hat{\beta}_0$、$\hat{\beta}_1$具有最小方差。

(1)先求$\hat{\beta}_0$与$\hat{\beta}_1$的方差

$$\mathrm{var}(\hat{\beta}_1)=\mathrm{var}(\sum k_iY_i)=\sum k_i^2\ \mathrm{var}(\beta_0+\beta_1X_i+\mu_i)=\sum k_i^2\ \mathrm{var}(\mu_i)$$

$$=\sum\left(\frac{x_i}{\sum x_i^2}\right)^2\sigma^2=\frac{\sigma^2}{\sum x_i^2}$$

$$\mathrm{var}(\hat{\beta}_0)=\mathrm{var}(\sum w_iY_i)=\sum w_i^2\ \mathrm{var}(\beta_0+\beta_1X_i+\mu_i)=\sum(1/n-\bar{X}k_i)^2\sigma^2$$

$$=\sum\left[\left(\frac{1}{n}\right)^2-2\frac{1}{n}\bar{X}k_i+\bar{X}^2k_i^2\right]\sigma^2=\left(\frac{1}{n}-\frac{2}{n}\bar{X}\sum k_i+\bar{X}^2\sum\left(\frac{x_i}{\sum x_i^2}\right)^2\right)\sigma^2$$

$$=\left(\frac{1}{n}+\frac{\bar{X}^2}{\sum x_i^2}\right)\sigma^2=\frac{\sum x_i^2+n\bar{X}^2}{n\sum x_i^2}\sigma^2=\frac{\sum X_i^2}{n\sum x_i^2}\sigma^2$$

（2）证明最小方差性

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，$c_i=k_i+d_i$，d_i为不全为零的常数

则容易证明

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

同理，可证明 β_0 的最小二乘估计量 $\hat{\beta}_0$ 具有最的小方差

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）**

由于最小二乘估计量拥有一个“好”的估计量所应具备的小样本特性，它自然也拥有大样本特性。

如考察$\hat{\beta}_1$的一致性

$$P\lim(\hat{\beta}_1) = P\lim(\beta_1 + \sum k_i \mu_i) = P\lim(\beta_1) + P\lim(\frac{\sum x_i \mu_i}{\sum x_i^2})$$

$$= \beta_1 + \frac{P\lim(\sum x_i \mu_i / n)}{P\lim(\sum x_i^2 / n)}$$

$$= \beta_1 + \frac{Cov(X, \mu)}{Q} = \beta_1 + \frac{0}{Q} = \beta_1$$

五、参数估计量的概率分布及随机干扰项方差的估计

1、参数估计量 $\hat{\beta}_0$ 和 $\hat{\beta}_1$ 的概率分布

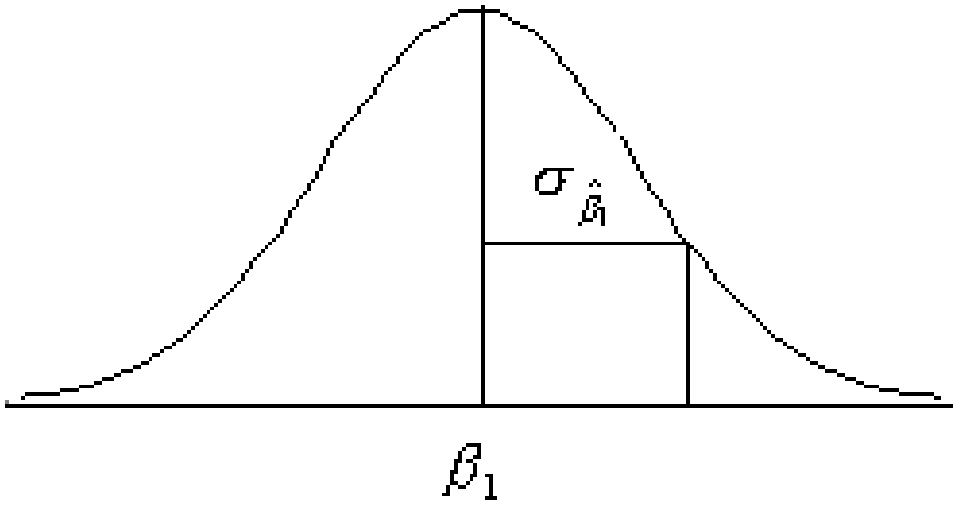
普通最小二乘估计量 $\hat{\beta}_0$、$\hat{\beta}_1$ 分别是 Y_i 的线性组合，因此，$\hat{\beta}_0$ 和 $\hat{\beta}_1$ 的概率分布取决于 Y 的分布特征

在 μ 是正态分布的假设下，Y 是正态分布，则 $\hat{\beta}_0$、$\hat{\beta}_1$ 也服从正态分布，因此

$$\hat{\beta}_1 \sim N(\beta_1, \frac{\sigma^2}{\sum x_i^2}) \qquad \hat{\beta}_0 \sim N(\beta_0, \frac{\sum X_i^2}{n\sum x_i^2}\sigma^2)$$

$\hat{\beta}_0$ 和 $\hat{\beta}_1$ 的标准差

$$\sigma_{\hat{\beta}_1} = \sqrt{\sigma^2 / \sum x_i^2} \qquad \sigma_{\hat{\beta}_0} = \sqrt{\frac{\sigma^2 \sum X_i^2}{n \sum x_i^2}}$$



2、随机误差项μ的方差σ^2的估计

在估计的参数 $\hat{\beta}_0$ 和 $\hat{\beta}_1$ 的方差表达式中，都含有随机扰动项 μ 的方差 σ^2。σ^2又称为总体方差。

由于 σ^2 实际上是未知的，因此 $\hat{\beta}_0$ 和 $\hat{\beta}_1$ 的方差实际上无法计算，这就需要对其进行估计。

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。

在最大或然估计法中，

解或然方程

$$\frac{\partial}{\partial \sigma^2} L^* = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \Sigma (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)^2 = 0$$

即可得到 σ^2 的最大或然估计量为：

$$\hat{\sigma}^2 = \frac{1}{n} \Sigma (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)^2 = \frac{\sum e_i^2}{n}$$

因此， σ^2的最大或然估计量不具无偏性，但却具有一致性。

在随机误差项μ的方差σ^2 估计出后，参数$\hat{\beta}_0$和$\hat{\beta}_1$的方差和标准差的估计量分别是：

$\hat{\beta}_1$的样本方差：　　$S^2_{\hat{\beta}_1} = \hat{\sigma}^2 \Big/ \sum x_i^2$

$\hat{\beta}_1$的样本标准差：　　$S_{\hat{\beta}_1} = \hat{\sigma} \Big/ \sqrt{\sum x_i^2}$

$\hat{\beta}_0$的样本方差：　　$S^2_{\hat{\beta}_0} = \hat{\sigma}^2 \sum X_i^2 \Big/ n\sum x_i^2$

$\hat{\beta}_0$的样本标准差：　　$S_{\hat{\beta}_0} = \hat{\sigma} \sqrt{\sum X_i^2 \Big/ n\sum x_i^2}$

例1　线性回归模型

$$y_t = \alpha + \beta x_t + \mu_t \qquad t = 1,2,\cdots,n$$

的 0 均值假设是否可以表示为 $\frac{1}{n}\sum_{t=1}^{n}\mu_t = 0$？为什么？

例2　已知两个量X和Y的一组观察值（x_i，y_i），i=1，2，…，n。

证明：Y 的真实值和拟合值有共同的均值。

例3　对没有截距项的一元回归模型

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

称之为过原点回归（regrission through the origin）。试证明

（1）如果通过相应的样本回归模型可得到通常的的正规方程组

$$\sum e_i = 0$$
$$\sum e_i X_i = 0$$

则可以得到 β_1 的两个不同的估计值：$\tilde{\beta}_1 = \overline{Y}/\overline{X}$，$\hat{\beta}_1 = \left(\sum X_i Y_i\right)/\left(\sum X_i^2\right)$。

（2）在基本假设 $E(\mu_i) = 0$ 下，$\tilde{\beta}_1$ 与 $\hat{\beta}_1$ 均为无偏估计量。

（3）拟合线 $\hat{Y} = \hat{\beta}_1 X$ 通常不会经过均值点 $(\overline{X}, \overline{Y})$，但拟合线 $\tilde{Y} = \tilde{\beta}_1 X$ 则相反。

（4）只有 $\hat{\beta}_1$ 是 β_1 的 OLS 估计量。

例4　假设模型为 $Y_t = \alpha + \beta X_t + \mu_t$。给定 n 个观察值 (X_1, Y_1)，(X_2, Y_2)，…，(X_n, Y_n)，按如下步骤建立 β 的一个估计量：在散点图上把第 1 个点和第 2 个点连接起来并计算该直线的斜率；同理继续，最终将第 1 个点和最后一个点连接起来并计算该条线的斜率；最后对这些斜率取平均值，称之为 $\hat{\beta}$，即 β 的估计值。

（1）画出散点图，给出 $\hat{\beta}$ 的几何表示并推出代数表达式。

（2）计算 $\hat{\beta}$ 的期望值并对所做假设进行陈述。这个估计值是有偏的还是无偏的？解释理由。

（3）证明为什么该估计值不如我们以前用 OLS 方法所获得的估计值，并做具体解释。

例5　试证：

(1)模型 $y_i=\beta_0+u_i$ $(i=1,2,\cdots,n)$ 中 β_0 的最小二乘估计量为 $\hat{\beta}_0=\bar{y}$；

(2)如(1)中的随机项满足经典回归的基本假定，则有

$$E(\hat{\beta}_0)=\beta_0,V(\hat{\beta}_0)=\frac{1}{n}\sigma_u^2$$