

## 第二章 单方程计量经济学模型 理论与方法

### Theory and Methodology of Single-Equation Econometric Model

- 线性单方程计量经济学模型的理论基础和参数估计
- 线性单方程计量经济学模型的统计检验和区间估计
- 违背古典假设的计量经济学问题

## 教学基本要求

本章是课程的重点和主要内容，应占总课内学时的1/3以上。通过教学，要求学生达到：

- **了解（最低要求）：**线性单方程计量经济学模型的基本理论与方法；运用矩阵描述、推导和证明与普通最小二乘法有关的参数估计过程和结论；应用计量经济学软件进行线性单方程模型的普通最小二乘估计；独立完成建立线性单方程计量经济学模型的全过程工作。

- **掌握（较高要求）：**关于线性单方程计量经济模型的假设，以及各类违背基本假设的模型的经济背景；最大或然法的基本原理，以及最小二乘法的区别；各类违背基本假设的模型的主要检验方法和主要克服的方法及它们的基本原理；软件包中有关加权最小二乘法、广义差分法和工具变量法的应用。
- **应用（对应用能力的要求）：**应用所学知识，在本章结束前独立完成一个综合练习，自己选择研究对象，自己建立理论模型，自己收集样本数据，进行模型的估计和检验，最后提交一篇报告。

## § 2.1 回归分析概述

### Introduction to Regression Analysis

- 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

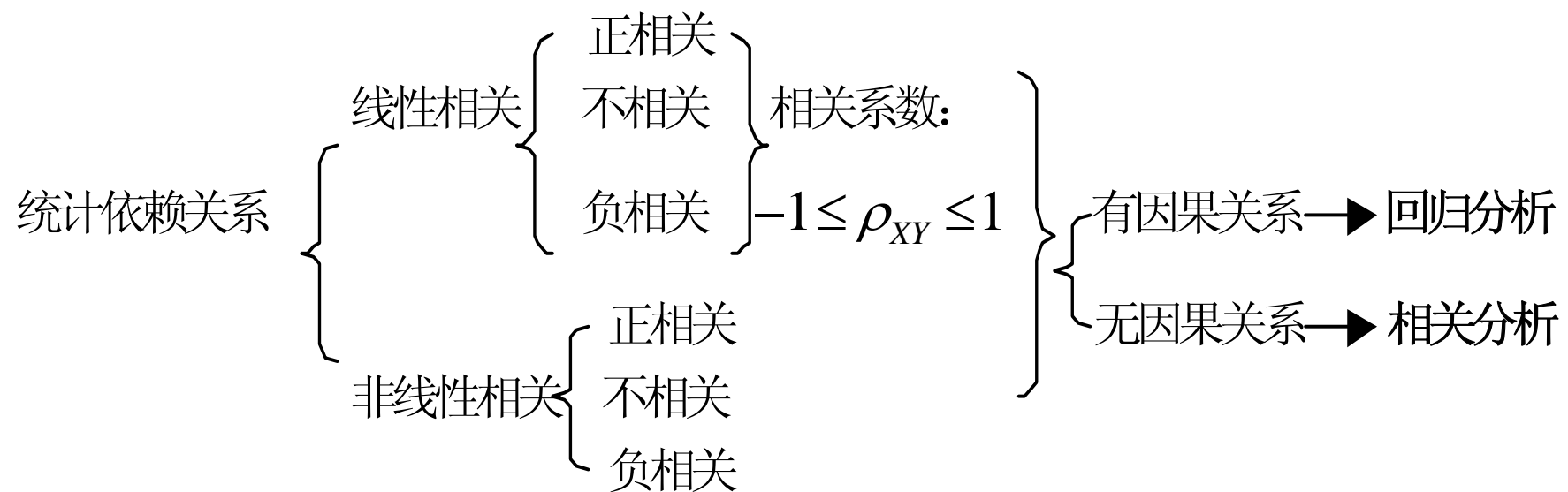
# 一、变量间的关系及回归分析的基本概念

## 1、变量间的关系

△ 经济变量之间的关系，大体可分为两类：

- **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。
- **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。

△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：



## △几点注意

- 非线性相关并不意味着不相关；
- 有相关关系并不意味着一定有因果关系；
- **回归分析/相关分析**研究一个变量对另一个（些）变量的统计依赖关系，但它们并不意味着一定有因果关系；
- **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。
- 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）。

- 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。
- 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) 根据样本观察值对计量经济模型参数进行估计，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

## 1. 例子

**例2. 1：**一个假想的社区有60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60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）。

表 2.1 某社区每月家庭收入与消费支出查统计表

	每月家庭收入X（元）									
	800	1000	1200	1400	1600	1800	2000	2200	2400	2600
每月 家庭 消费 支出 Y（元）	550	650	790	800	1020	1100	1200	1350	1370	1500
	600	700	840	930	1070	1150	1360	1370	1450	1520
	650	740	900	950	1100	1200	1400	1400	1550	1750
	700	800	940	1030	1160	1300	1440	1520	1650	1780
	750	850	980	1080	1180	1350	1450	1570	1750	1800
	0	880	0	1130	1250	1400	0	1600	1890	1850
	0	0	0	1150	0	0	0	1620	0	1910
共计	3250	4620	4450	7070	6780	7500	6850	10430	9660	12110
条件概率	1/5	1/6	1/5	1/7	1/6	1/6	1/5	1/7	1/6	1/7
条件均值	650	770	890	1010	1130	1250	1370	1490	1610	1730

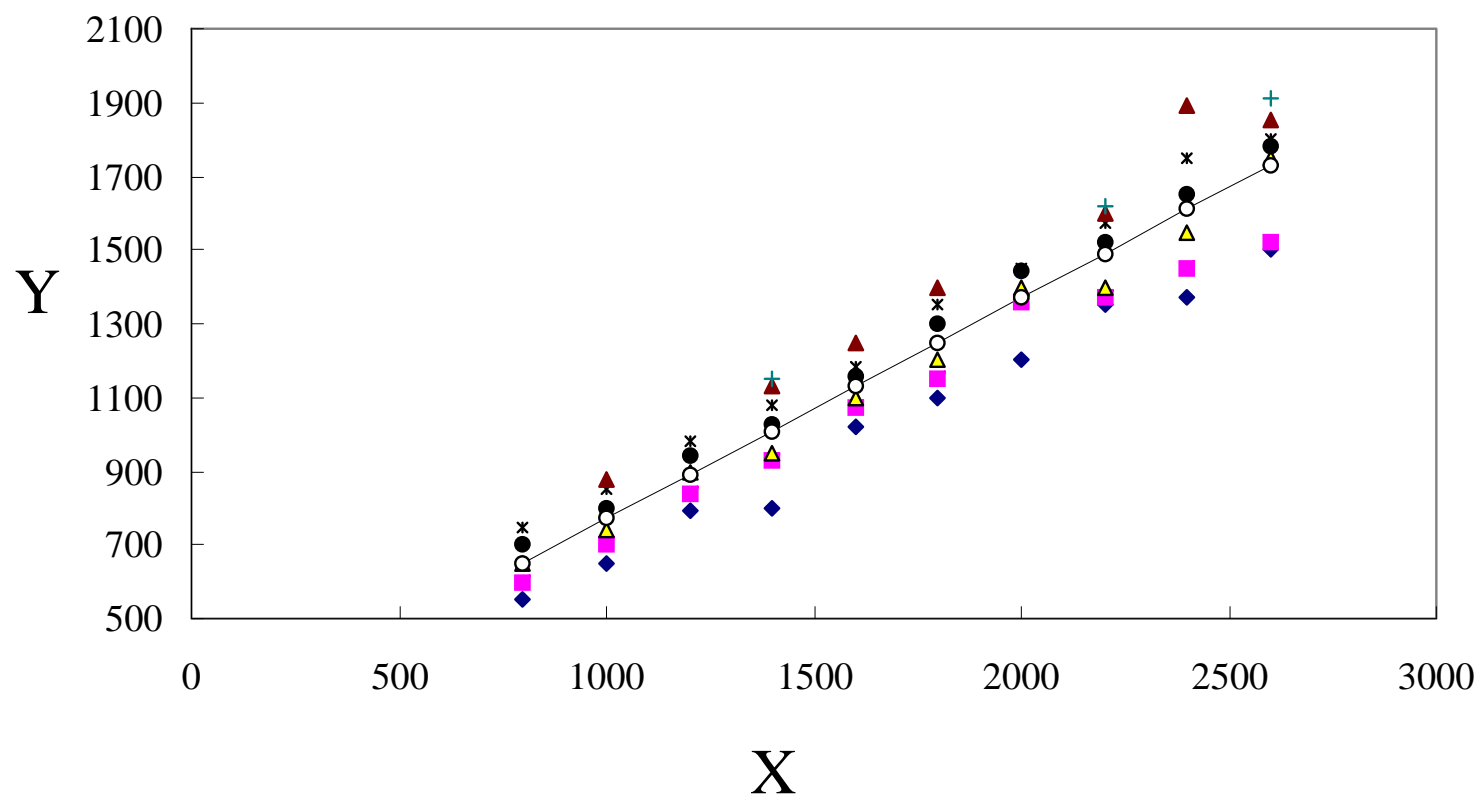
## 2. 分析

- 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；
- 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：  
 $P(Y=550|X=800) = 1/5$ 。
- 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： $E(Y | X=800) = 650$

- 从散点图发现：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**。



### 3. 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 (方程):

$$E(Y | X_i) = f(X_i) \quad (2.1.1)$$

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



- 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：  $E(Y | X_i) = \beta_0 + \beta_1 X_i$  为一线性函数。

其中，  $\beta_1$  与  $\beta_2$  为未知然而固定的参数，称为**回归系数**（**regression coefficients**）。

### 三、随机扰动项

# 1. 随机扰动项的引入

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

$$\begin{aligned}\mu_i &= Y_i - E(Y | X_i) \\ &= Y_i - (\beta_0 + \beta_1 X_i)\end{aligned}\tag{2.1.2}$$

称 $\mu_i$ 为观察值 $Y_i$ 围绕它的期望值 $E(Y | X_i)$ 的**离差 (deviation)**，它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)**。

由（2.1.2）式，个别家庭的消费支出为：

$$\begin{aligned} Y_i &= E(Y | X_i) + \mu_i \\ &= \beta_0 + \beta_1 X_i + \mu_i \end{aligned} \quad (2.1.3)$$

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**  $\mu_i$ 。

- (2.1.3) 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**。

## 2. 随机误差项的影响因素

- 在解释变量中被忽略的因素的影响；
- 变量观测值的观测误差的影响；
- 模型关系的设定误差的影响；
- 其它随机因素的影响。

### 3. 产生并设计随机误差项的主要原因

- 理论的含糊性;
- 数据的欠缺;
- 节省原则。

## 四、样本回归函数 (SRF)

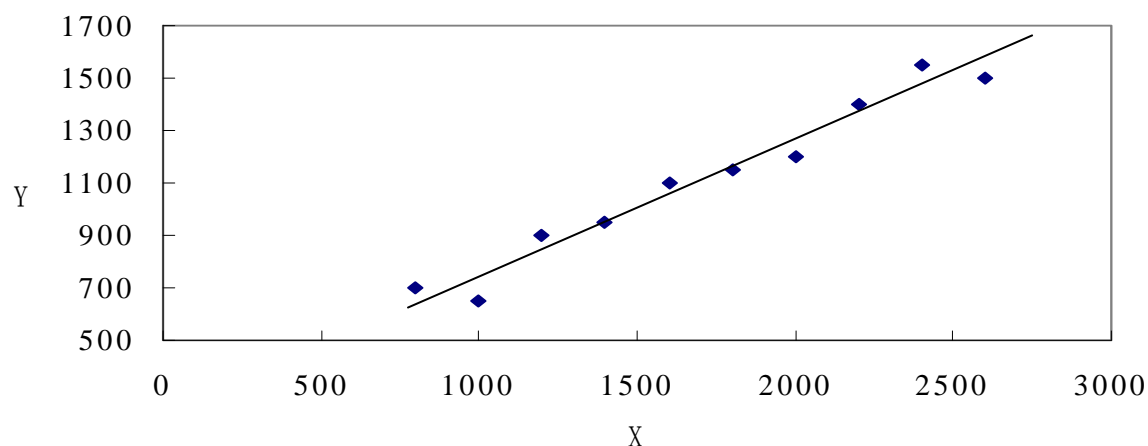


## 1. 问题的提出

- 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。
- 问题是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？

Y	700	650	900	950	1100	1150	1200	1400	1550	1500
X	800	1000	1200	1400	1600	1800	2000	2200	2400	2600

该样本的散点图（scatter diagram）:



样本散点图近似于一条直线，画一条直线以尽可能好地拟合该散点图，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**），其函数形式记为：

$$\hat{Y}_i = f(X_i) = \hat{\beta}_0 + \hat{\beta}_1 X_i \quad (2.1.4)$$

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

注意： 这里

$\hat{Y}_i$  就为  $E(Y|X_i)$  的估计量;

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

将 (2.1.4) 看成 (2.1.1) 的近似替代。

## 2. 样本回归函数的随机形式/样本回归模型

同样地，对某一个体 $Y_i$ ，有

$$Y_i = \hat{Y}_i + \hat{\mu}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i + e_i \quad (2.1.5)$$

式中， $e_i$ 称为**（样本）残差（或剩余）项**（residual），代表了其他影响 $Y_i$ 的随机因素的集合体，可看成为 $\mu_i$ 的估计量。

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**。

### 3. 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

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

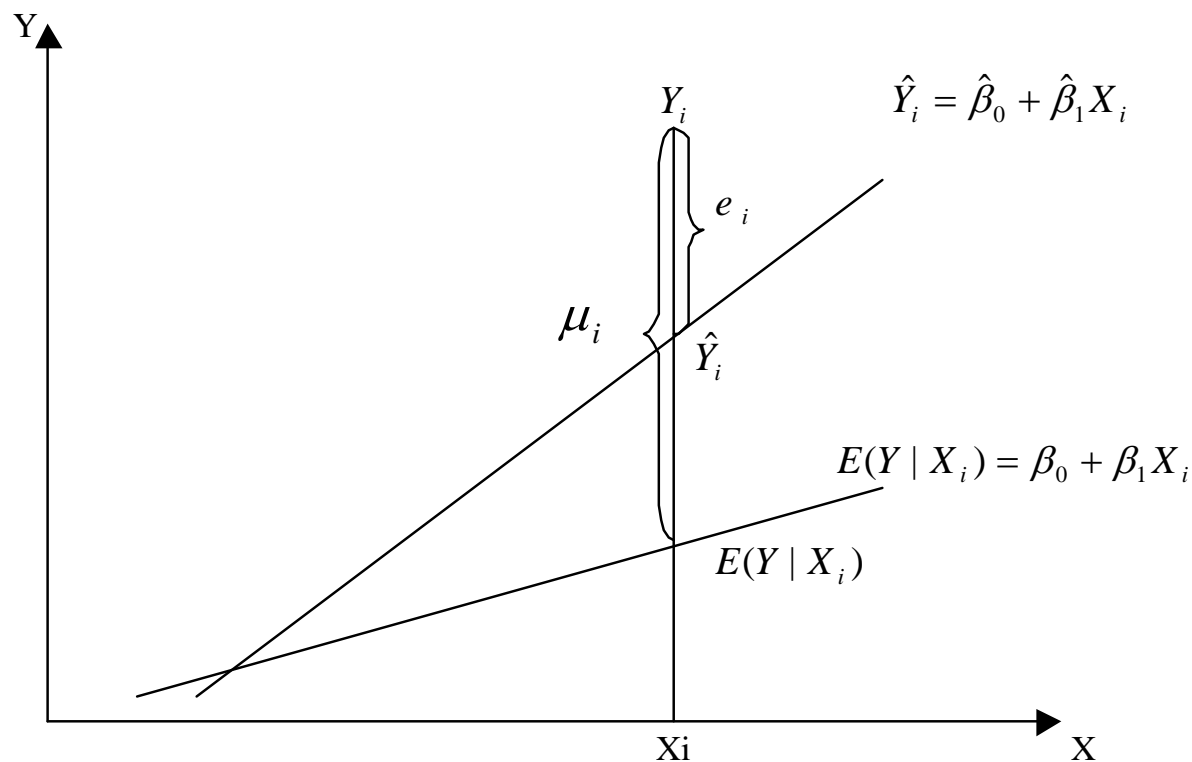
估计

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

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:** 这里真实的 PRF  
可能永远无从知道



样本与总体回归线

30