

1. **总体回归函数：**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，或更一般地称为总体回归曲线。相应的函数： $E(Y | X_i) = f(X_i)$  称为（双变量）总体回归函数（population regression function, PRF）
2. **样本回归函数：**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。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）。
3. **随机的总体回归函数：**函数  $Y = E(Y | X) = \mu$  或者在线性假设下， $Y = \beta_0 + \beta_1 X + \mu$  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。
4. **线性回归模型：**

假设 1、回归模型是正确设定的。

假设 2、解释变量  $X$  是确定性变量，不是随机变量，在重复抽样中取固定值。

假设 3、解释变量  $X$  在所抽取的样本中具有变异性，而且随着样本容量的无限增加，解释变量  $X$  的样本方差趋于一个非零的有限常数，即

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

假设 4、随机误差项  $\mu$  具有零均值、同方差和不序列相关性：

$$E(\mu_i) = 0 \quad i = 1, 2, \dots, n$$

$$\text{Var}(\mu_i) = \sigma_\mu^2 \quad i = 1, 2, \dots, n$$

$$\text{Cov}(\mu_i, \mu_j) = 0 \quad i \neq j \quad i, j = 1, 2, \dots, n$$

假设 5、随机误差项  $\mu$  与解释变量  $X$  之间不相关：

$$\text{Cov}(X_i, \mu_i) = 0 \quad i = 1, 2, \dots, n$$

假设 6、 $\mu$  服从零均值、同方差、零协方差的正态分布

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

以上假设也称为线性回归模型的经典假设，满足该假设的线性回归模型，也称为经典线性回归模型

5. **随机误差项 ( $\mu_i$ ) 和残差项 ( $e_i$ )**：(1)  $\mu_i$  为观察值  $X_i$  围绕它的期望值  $E(Y|X_i)$  的离差，是一个不可观测的随机变量，又称为随机干扰项或随机误差项。

(2)  $e_i$  称为 (样本) 残差或 (剩余) 项代表了其他影响  $Y_i$  的随机因素的集合，可看成是  $\mu_i$  的估计量

6. **条件期望**：即条件均值，指  $X$  取特定值  $X_i$  时  $Y$  的期望值。

定义 3.11 (1) 设  $(X, Y)$  是离散型随机向量, 在  $Y = y_j$  的条件下,  $X$  的条件概率分布为

$$P\{X = x_i | Y = y_j\} = p_{ij}, i = 1, 2, \dots,$$

如果

$$\sum_i |x_i| p_{ij} < \infty,$$

则称  $\sum x_i p_{ij}$  为  $X$  在  $Y = y_j$  的条件下的条件数学期望, 记作  $E[X|Y = y_j]$ .

## 7. 非条件期望:

8. **回归系数或回归参数**：在回归方程中表示自变量  $x$  对因变量  $y$  影响大小的参数。

9. **回归系数的估计量**：指用  $\hat{\beta}_0, \hat{\beta}_1$  等表示的用已知 样本提供的信息所估计出来总体未知参数的结果。

10. **最小平方法**：给定一组样本观测值  $(X_i, Y_i)$  ( $i=1, 2, \dots, n$ ) 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  $Q = \sum_1^n (Y_i - \hat{Y}_i)^2 = \sum_1^n (Y_i - (\beta_0 + \beta_1 X_i))^2$  最小

11. **最大似然法**：最大或然法，也称最大似然法，是指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。

12. **估计量的标准差**：度量一个变量变化大小的测量值。

$$TSS = \sum y_i^2 = \sum (Y_i - \bar{Y})^2$$

13. **总离差平方和**：记为总离差平方和（TSS）：说明实际的 Y 值围绕其均值的总变异。

14. **回归平方和**：记  $ESS = \sum \hat{y}_i^2 = \sum (\hat{Y}_i - \bar{Y})^2$  为回归平方和，反映由模型中解释变量所解释的那部分离差的大小。

15. **残差平方和**：记  $RSS = \sum e_i^2 = \sum (Y_i - \hat{Y}_i)^2$  为残差平方和，反映样本观测值与估计值偏离的大小，也是模型中解释变量未解释的那部分离差的大小。

16. **协方差**：期望值分别为  $E(X) = \mu$  与  $E(Y) = v$  的两个实数，随机变量 X 与 Y 之间的协方差定义为： $COV(X, Y) = E[(X - E(X))(Y - E(Y))]$ ，用来度量 X, Y 两个变量关联程度的统计量

17. **拟合优度检验**：拟合优度检验是指对样本回归线与样本观测值之间拟合程度的检验。度量拟合程度的指标是可决系数  $R^2$ 。

18. **t 检验**：t 检验是针对每个解释变量进行的显著性检验，即构造一个 t 统计量，如果该统计量的值落在置信区间外，就拒绝原假设。

19. **F 检验**：两个独立的卡方变量之商的分布（注意自由度）

$$\text{若 } X_1 \sim \chi^2(n), X_2 \sim \chi^2(m), X_1, X_2 \text{ 相互独立, 则 } F = \frac{X_1/n}{X_2/m} \sim F(n, m)$$

20. **异方差性**：对于不同的样本点，随机干扰项的方差不再是常数，而是互不相同，则认为出现了异方差性。

21. **序列相关性**：多元线形回归模型的基本假设之一是模型的随机干扰项相互独立或不相关。如果模型的随机干扰项违背了相互独立的基本假设，称为存在序列相关性。

22. **多重共线性**：如果某两个或多个解释变量之间出现了相关性，则称为多重共线性

23. **偏回归系数**：在多元回归模型中，每一个解释变量前的参数即为偏回归系数，它测度了当其他解释变量保持不变时，该变量增加 1 个单位对解释变量带来的平均影响程度。

24. **完全多重共线性**：如果存在

$$c_1X_{1i}+c_2X_{2i}+\cdots+c_kX_{ki}=0$$

$$i=1, 2, \cdots, n$$

其中：  $c_i$  不全为 0，则称为解释变量间存在完全共线性。

25. **不完全多重共线性**：如果存在

$$c_1X_{1i}+c_2X_{2i}+\cdots+c_kX_{ki}+v_i=0 \quad i=1, 2, \cdots, n$$

其中  $c_i$  不全为 0， $v_i$  为随机误差项，则称为 近似共线性或交互相关。

26. **随机解释变量**：如果存在一个或多个随机变量作为解释变量，则称原模型出现随机解释变量问题。

27. **差分法**：差分法是将原模型变换为满足 OLS 法的差分模型，再进行 OLS 估计。

28. **广义最小二乘法 (GLS)**：是一种常见的消除异方差的方法. 它的主要思想是为解释变量加上一个权重, 从而使得加上权重后的回归方程方差是相同的. 因此在 GLS 方法下我们可以得到估计量的无偏和一致估计, 并可以对其进行 OLS 下的 t 检验和 F 检验.

29. **D. W. 检验**：简述 D. W. 检验的步骤：

(1) 计算 DW 值；

(2) 给定  $\alpha$ ，由  $n$  和  $k$  的大小查 DW 分布表，得临界值  $dL$  和  $dU$

(3) 比较、判断。若  $0 < D.W. < dL$ ，存在正自相关； $dL < D.W. < dU$ ，不能确定； $dU < D.W. < 4 - dU$ ，无自相关； $4 - dU < D.W. < 4 - dL$ ，不能确定； $4 - dL < D.W. < 4$ ，存在负自相关。

当 D. W. 值在 2 左右时，模型不存在一阶自相关。

30. **多元线性回归**：在现实经济活动中往往存在一个变量受到其他多个变量的影响的现象，表现为在线性回归模型中有多个解释变量，这样的模型成为多元线性回归模型，多元指多个变量。

31. **正规方程组**：指采用 OLS 法估计线性回归模型时，对残差平方和关于各参数求偏导，并令偏导数为 0 后得到的一组方程：

$$\begin{cases} \sum (\hat{\beta}_0 + \hat{\beta}_1 X_i - Y_i) = 0 \\ \sum (\hat{\beta}_0 + \hat{\beta}_1 X_i - Y_i) X_i = 0 \end{cases} \quad \text{或} \quad \begin{cases} \sum Y_i = n\hat{\beta}_0 + \hat{\beta}_1 \sum X_i \\ \sum Y_i X_i = \hat{\beta}_0 \sum X_i + \hat{\beta}_1 \sum X_i^2 \end{cases}$$

32. **无偏性**：即它的均值或期望值是否等于总体的真实值；

33. **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

34. **参数估计量的置信区间**：置信区间是指由样本统计量所构造的总体参数的估计区间。参数的置信区间用来考察：在一次抽样中所估计的参数值离参数的真实值有多“近”。

35. **被解释变量预测值的置信区间**：

36. **受约束回归**：在实际经济活动中，常常需要根据经济理论对模型中变量的参数施加一定的约束条件，对模型参数施加约束条件后进行回归。

37. **无约束回归**：无需对模型中变量的参数施加约束条件进行的回归。

38. **参数稳定性检验**：(书上 98 页)

$(RSS_1 + RSS_2) / (n_1 + n_2)$

因此，对参数稳定性的原假设(3.6.21)式的检验步骤为：首先，分别以两个连续的时间序列作为两个样本运用(3.6.1)式进行回归，得到相应的残差平方和  $RSS_1$  与  $RSS_2$ ；然后，将两序列并为一个大样本后运用(3.6.1)式进行回归，得到大样本下的残差平方和  $RSS_R$ ；最后，通过(3.6.24)式的  $F$  统计量，在事先给定的显著性水平下进行假设检验。如果  $F$  大于相应的临界值，则拒绝原假设，认为发生了结构变化，参数是非稳定的。该检验方法也称为邹氏参数稳定性检验(Chow)

39. **虚拟变量**：许多经济变量是可以定量度量的，为了在模型中反映对模型的影响因素，并提高模型的精度，需要将它们“量化”，这种“量化”是通过

引入“虚拟变量”来完成的。根据这些因素的属性类型，构造只取“0”或“1”的人工变量，通常称为虚拟变量。

40. **虚拟因变量模型**：同时含有一般解释变量与虚拟变量的模型称为虚拟变量模型或者方差分析模型。

41. **滞后变量**：把过去时期的，具有滞后作用的变量叫做滞后变量。

42. **滞后效应**：

一般说来，被解释变量与解释变量的因果关系不一定就在瞬时发生，可能存在时间上的滞后，或者说解释变量的变化可能需要经过一段时间才能完全对被解释变量产生影响。同样地，被解释变量当前的变化也可能受其自身过去水平的影响。这种被解释变量受到自身或另一解释变量的前几期值影响的现象称为**滞后效应**，表示前几期值的变量称为**滞后变量**。

43. **分布滞后模型**：如果滞后变量模型中没有滞后被解释变量，仅有解释变量X的当期值及其若干期的滞后值，则成为分布滞后模型。

44. **自回归模型**：解释变量仅包含X的当期值与被解释变量Y的一个或多个滞后值的模型。

45. **h 检验**：

■ 为了解决自回归分布滞后模型的自相关检验问题，Durbin在1970年提出了一个新的检验方法，即**h检验法**，也称之为**德宾h检验**，其h统计量为：

$$h = \hat{\rho} \sqrt{\frac{n}{1 - n \text{Var}(\hat{\beta}_1)}} \quad (12.2.15)$$

46. **有限最小二乘法**：

47. **联立问题**：同时研究多个经济关系，反映在模型的建立上，就需要一个由多个方程组成的联立方程组来解决问题，这就是模型的联立性问题。

48. **行为方程**：是描述经济系统中变量之间行为关系的结构式方程。

49. **间接最小二乘法**：先对关于内生解释变量的简化式方程采用普通最小二乘法估计简化式参数，得到简化式参数估计量，然后通过参数关系体系，计算得到结构式参数的估计量。间接最小二乘法只适用于恰好识别的结构方程的参数估计，因为只有恰好识别的结构方程，才能从参数关系体系得

到唯一一组结构参数的估计量。

50. **识别问题：**如果联立方程计量经济学模型中某个结构方程不具有确定的统计形式，则称该方程为不可识别

51. **二阶段最小二乘法：**是一种既适用于恰好识别的结构方程，又适用于过度识别的结构方程的单方程估计方法。

52. **三阶段最小二乘法：**

53. **简化式模型：**将联立方程计量经济学模型的每个内生变量表示成所有先决变量和随机干扰项的函数，即用所有先决变量作为每个内生变量的解释变量，所形成的模型称为简化式模型。

54. **不可识别：**若联立方程计量经济学模型中某个结构不具有确定性的统计形式，则称该方程为不可识别

55. **恰度识别：**是指对联立方程模型，我们能够唯一地估计出模型的参数。

56. **过度识别：**是指模型方程中有一个或几个参数有若干个估计值。

57. **结构式模型：**根据经济理论和行为规律建立的描述经济变量之间直接关系结构的计量经济学方程系统统称为结构式模型。

58. **递归系统模型：**一种在形式是属于联立方程模型但仍然可以采用单方程模型的估计方法估计每个方程的特殊情况，即递归系统模型。

59. **先决变量：**外生变量与滞后内生变量统称为先决变量。

60. **伪回归：**如果一组非平稳时间序列之间不存在协整关系，则这一组变量构造的回归模型就是伪回归。

61. **单整：**如果一个时间序列经过一次差分变成平稳的，就称原序列是1阶单整序列，记为  $I(1)$ 。如果一个时间序列经过  $d$  次差分变成平稳序列，则称原序列是  $d$  阶单整序列记为  $I(d)$ 。

62. 如果  $X_t$  满足下列条件：

(1) 均值  $E(X_t) = \mu$ ，与时间  $t$  无关的常数

(2) 方差 $\text{Var}(X_t) = \sigma^2$ ，与时间  $t$  无关的常数

(3) 协方差 $\text{Cov}(X_t X_{t+k}) = \gamma_k$ ，只与时期间隔  $k$  有关，与时间  $t$  无关的常数。

则称该时间序列是**平稳的**，而该随机过程是一个**平稳随机过程**。