第三章

多元线性回归模型

ECONOMETRICS

本章主要讨论:

- **多元线性回归模型及古典假定**
- **多元线性回归模型的估计**
- **多元线性回归模型的检验**
- **多元线性回归模型的预测**

第一节 多元线性回归模型及古典假定

一、多元线性回归模型的意义

一般形式：对于有**K-1**个解释变量的线性回归模型

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \cdots + \beta_k X_{ki} + u_i$$

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

注意：模型中的β_j（j=1,2,---k）是偏回归系数

样本容量为n

偏回归系数：

控制其它解释量不变的条件下，第j个解释变量的单位变动对被解释变量平均值的影响，即对Y平均值“直接”或“净”的影响。

ECONOMETRICS

多元线性回归中的“线性”

指对各个回归系数而言是“线性”的，对变量则可以是线性的，也可以是非线性的

例如：生产函数

$$Y = AL^{\alpha}K^{\beta}u$$

取对数

$$\ln Y = \ln A + \alpha \ln L + \beta \ln K + \ln u$$

这也是多元线性回归模型，只是这时变量为ln$\boldsymbol{Y}$、ln$\boldsymbol{L}$、ln$\boldsymbol{K}$

多元总体回归函数

条件期望表现形式：

将Y的总体条件期望表示为多个解释变量的函数，如：

$$E(Y_i|X_{2i},X_{3i},\cdots X_{ki})=\beta_1+\beta_2X_{2i}+\beta_3X_{3i}+\cdots+\beta_kX_{ki}$$

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

注意：这时**Y**总体条件期望的轨迹是**K**维空间的一条线

个别值表现形式：

引入随机扰动项 $u_i=Y_i-E(Y_i|X_{2i},X_{3i}\cdots X_{ki})$

或表示为 $Y_i=\beta_1+\beta_2X_{2i}+\beta_3X_{3i}+\cdots+\beta_kX_{ki}+u_i$

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

ECONOMETRICS

多元样本回归函数

Y 的样本条件均值可表示为多个解释变量的函数

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \cdots + \hat{\beta}_k X_{ki}$$

或回归剩余（残差）： $e_i = Y_i - \hat{Y}_i$

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \cdots + \hat{\beta}_k X_{ki} + e_i$$

其中 $i = 1, 2, \cdots n$

二、多元线性回归模型的矩阵表示

多个解释变量的多元线性回归模型的n组观测值，可表示为

$$
\begin{cases}
Y_1 = \beta_1 + \beta_2 X_{21} + \beta_3 X_{31} + \cdots + \beta_k X_{k1} + u_1 \\
Y_2 = \beta_1 + \beta_2 X_{22} + \beta_3 X_{32} + \cdots + \beta_k X_{k2} + u_2 \\
\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
Y_n = \beta_1 + \beta_2 X_{2n} + \beta_3 X_{3n} + \cdots + \beta_k X_{kn} + u_n
\end{cases}
$$

用矩阵表示

$$
\underset{n\times 1}{\mathbf{Y}} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \underset{n\times k}{\mathbf{X}} = \begin{bmatrix} 1 & X_{21} & \cdots & X_{k1} \\ 1 & X_{22} & \cdots & X_{k2} \\ \vdots & \cdots & \cdots & \vdots \\ 1 & X_{2n} & \cdots & X_{kn} \end{bmatrix} \underset{k\times 1}{\boldsymbol{\beta}} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix} + \underset{n\times 1}{\mathbf{u}} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}
$$

矩阵表示方式

总体回归函数 $E(Y) = X\beta$ 或 $Y = X\beta + u$

样本回归函数 $\hat{Y} = X\hat{\beta}$ 或 $Y = X\hat{\beta} + e$

其中： $Y, \hat{Y}, u, e$ 都是有n个元素的列向量

$\beta, \hat{\beta}$ 是有k 个 元素的列向量

（ k = 解释变量个数 + 1 ）

X 是第一列为1的n×k阶解释变量数据矩阵，

(截距项可视为解释变量总是取值为1)

ECONOMETRICS

三、多元线性回归中的基本假定

假定1：零均值假定

$E(u_i)=0$ （i=1，2，---n） 或 E（$\boldsymbol{u}$）=0

假定2和假定3：同方差和无自相关假定：

$$Cov(u_i,u_j)=E[(u_i-Eu_i)(u_j-Eu_j)]=E(u_iu_j)=\begin{cases}\sigma^2 & (i=j)\\ 0 & (i\neq j)\end{cases}$$

或用方差-协方差矩阵表示为:

$$Cov(u_i,u_j)=E\{[u_i-E(u_i)][u_j-E(u_j)]\}=E(\boldsymbol{uu}')$$

$(i=1,2,\cdots n \quad j=1,2,\cdots n)$

$$=\begin{bmatrix}E(u_1u_1) & E(u_1u_2) & \cdots & E(u_1u_n)\\ E(u_2u_1) & E(u_2u_2) & \cdots & E(u_2u_n)\\ \vdots & \vdots & \vdots & \vdots\\ E(u_nu_1) & E(u_nu_2) & \cdots & E(u_nu_n)\end{bmatrix}=\sigma^2\begin{bmatrix}1 & 0 & \cdots & 0\\ 0 & 1 & \cdots & 0\\ \vdots & \vdots & \vdots & \vdots\\ 0 & 0 & \cdots & 1\end{bmatrix}=\sigma^2\boldsymbol{I}$$

ECONOMETRICS

假定4：随机扰动项与解释变量不相关

$$Cov(X_{ji}, u_i) = 0 \qquad (j = 2, 3, \cdots, k)$$

假定5：无多重共线性假定（多元中增加的)

假定各解释变量之间不存在线性关系，或各个解释变量观测值之间线性无关。或解释变量观测值矩阵**X**的秩为K(注意**X**为n行K列)。

Rank(***X***)= k ⟶ Rank(***X'X***)=k

即 (***X'X***) 可逆

假定6:正态性假定

$$u_i \sim N(0, \sigma^2) \qquad \mathbf{u} \sim N(\mathbf{0}, \sigma^2\mathbf{I})$$

ECONOMETRICS

第二节 多元线性回归模型的估计

一、普通最小二乘法（OLS）

原则：寻求剩余平方和最小的参数估计式 $\min:\sum e_i^2=\sum(Y_i-\hat{Y}_i)^2$

$$\min:\sum e_i^2=\sum[Y_i-(\hat{\beta}_1+\hat{\beta}_2X_{2i}+\hat{\beta}_3X_{3i}+\cdots+\hat{\beta}_kX_{ki})]^2$$

即 $\min:\sum e_i^2=\min:\mathbf{e}'\mathbf{e}=\min:(\mathbf{Y}-\mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{Y}-\mathbf{X}\hat{\boldsymbol{\beta}})$

求偏导，并令其为0 $\partial(\sum e_i^2)/\partial\hat{\beta}_j=0$ 其中 $(i=1,2,\cdots n)$ $(j=1,2,\cdots n)$

即

$$-2\sum\left[Y_i-(\hat{\beta}_1+\hat{\beta}_2X_{2i}+\hat{\beta}_3X_{3i}+\cdots+\hat{\beta}_{ki}X_{ki})\right]=0 \rightarrow \sum e_i=0$$

$$-2\sum X_{2i}\left[Y_i-(\hat{\beta}_1+\hat{\beta}_2X_{2i}+\hat{\beta}_3X_{3i}+\cdots+\hat{\beta}_{ki}X_{ki})\right]=0 \rightarrow \sum X_{2i}e_i=0$$

$$\vdots$$

$$-2\sum X_{ki}\left[Y_i-(\hat{\beta}_1+\hat{\beta}_2X_{2i}+\hat{\beta}_3X_{3i}+\cdots+\hat{\beta}_{ki}X_{ki})\right]=0 \rightarrow \sum X_{ki}e_i=0$$

ECONOMETRICS

用矩阵表示的正规方程

偏导数

$$\begin{bmatrix} \sum e_i \\ \sum X_{2i}e_i \\ \vdots \\ \sum X_{ki}e_i \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & 1 & \cdots & 1 \\ X_{21} & X_{22} & \cdots & X_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ X_{k1} & X_{k2} & \cdots & X_{kn} \end{bmatrix}}_{\boldsymbol{X'}} \underbrace{\begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}}_{\boldsymbol{e}} = \mathbf{X'e} = \underbrace{\begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}}_{\mathbf{0}}$$

因为样本回归函数为

$$\mathbf{Y} = \mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{e}$$

两边左乘 X'

$$X'Y = X'X\hat{\beta} + X'e$$

根据最小二乘原则

$$X'e = 0$$

则正规方程为

$$X'X\hat{\beta} = X'Y$$

ECONOMETRICS

OLS估计式

由正规方程 $X'X\hat{\beta} = X'Y$ $(X'X)_{k\times k}$是满秩矩阵,其逆存在

多元回归的**OLS**估计量为 $\hat{\beta} = (X'X)^{-1} X'Y$

当只有两个解释变量时为:

$$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X}_2 - \hat{\beta}_3 \bar{X}_3$$

$$\hat{\beta}_2 = \frac{(\sum y_i x_{2i})(\sum x_{3i}^2) - (\sum y_i x_{3i})(\sum x_{2i} x_{3i})}{(\sum x_{2i}^2)(\sum x_{3i}^2) - (\sum x_{2i} x_{3i})^2}$$

$$\hat{\beta}_3 = \frac{(\sum y_i x_{3i})(\sum x_{2i}^2) - (\sum y_i x_{2i})(\sum x_{2i} x_{3i})}{(\sum x_{2i}^2)(\sum x_{3i}^2) - (\sum x_{2i} x_{3i})^2}$$

对比

简单线性回归中

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

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

注意: x、y 为*X*、*Y*的离差

ECONOMETRICS

$$Y_i = \hat\beta_1 + \hat\beta_2 X_{2i} + \hat\beta_3 X_{3i} + e_i \qquad \bar Y = \hat\beta_1 + \hat\beta_2 \bar X_2 + \hat\beta_3 \bar X_3$$

$$Y_i - \bar Y = \hat\beta_2 (X_{2i} - \bar X_2) + \hat\beta_3 (X_{3i} - \bar X_3) + e_i$$

$$y_i = \hat\beta_2 x_{2i} + \hat\beta_3 x_{3i} + e_i \qquad \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} x_{21} & x_{31} \\ \vdots & \vdots \\ x_{2n} & x_{3n} \end{pmatrix} \begin{pmatrix} \hat\beta_2 \\ \hat\beta_3 \end{pmatrix} + \begin{pmatrix} e_1 \\ \vdots \\ e_n \end{pmatrix}$$

$$y = x\hat\beta + e \qquad \hat\beta = (x^T x)^{-1} x^T y$$

$$x^T x = \begin{pmatrix} x_{21} & \cdots & x_{2n} \\ x_{31} & \cdots & x_{3n} \end{pmatrix} \begin{pmatrix} x_{21} & x_{31} \\ \vdots & \vdots \\ x_{2n} & x_{3n} \end{pmatrix} = \begin{pmatrix} \sum x_{2i}^2 & \sum x_{2i} x_{3i} \\ \sum x_{2i} x_{3i} & \sum x_{3i}^2 \end{pmatrix}$$

$$(x^T x)^{-1} = \begin{pmatrix} \sum x_{3i}^2 & -\sum x_{2i} x_{3i} \\ -\sum x_{2i} x_{3i} & \sum x_{2i}^2 \end{pmatrix} \Big/ \left(\sum x_{2i}^2 \sum x_{3i}^2 - \left(\sum x_{2i} x_{3i}\right)^2 \right)$$

$$X^T y = \begin{pmatrix} x_{21} & \cdots & x_{2n} \\ x_{31} & \cdots & x_{3n} \end{pmatrix} \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} \sum x_{2i} y_i \\ \sum x_{3i} y_i \end{pmatrix}$$

$$\hat{\beta} = (X^T X)^{-1} X^T y = \frac{1}{\sum x_{2i}^2 \sum x_{3i}^2 - \left(\sum x_{2i} x_{3i}\right)^2} \begin{pmatrix} \sum x_{3i}^2 & -\sum x_{2i} x_{3i} \\ -\sum x_{2i} x_{3i} & \sum x_{2i}^2 \end{pmatrix} \begin{pmatrix} \sum x_{2i} y_i \\ \sum x_{3i} y_i \end{pmatrix}$$

$$= \frac{1}{\sum x_{2i}^2 \sum x_{3i}^2 - \left(\sum x_{2i} x_{3i}\right)^2} \begin{pmatrix} \sum x_{3i}^2 \sum x_{2i} y_i - \sum x_{2i} x_{3i} \sum x_{3i} y_i \\ \sum x_{2i}^2 \sum x_{3i} y_i - \sum x_{2i} x_{3i} \sum x_{2i} y_i \end{pmatrix}$$

$$\hat{\beta}_2 = \frac{\sum x_{3i}^2 \sum x_{2i} y_i - \sum x_{2i} x_{3i} \sum x_{3i} y_i}{\sum x_{2i}^2 \sum x_{3i}^2 - \left(\sum x_{2i} x_{3i}\right)^2}$$

P_{79} 页 公式 3.27

3.28

OLS回归线的数学性质（与简单线性回归相同）

●回归线通过样本均值 $\overline{Y}=\hat{\beta}_1+\hat{\beta}_2\overline{X}_2+\hat{\beta}_3\overline{X}_3+\cdots+\hat{\beta}_k\overline{X}_k$

●估计值 $\hat{Y}_i$ 的均值等于实际观测值 Y_i 的均值 $\sum\hat{Y}_i/n=\overline{Y}$

●剩余项 e_i 的均值为零 $\overline{e_i}=\sum e_i/n=0$

●被解释变量估计值 $\hat{Y}_i$ 与剩余项 e_i 不相关

$$Cov(\hat{Y}_i,e_i)=0 \quad 或 \quad \sum(e_i\hat{y}_i)=0$$

●解释变量 X_i 与剩余项 e_i 不相关

$$Cov(X_{ji},e_i)=0$$

（j=1,2,---k）

16

ECONOMETRICS

二、OLS估计式的统计性质

1、 **线性特征** $\hat{\beta} = (X'X)^{-1} X'Y$

$\hat{\beta}$ 是**Y**的线性函数，因 $(X'X)^{-1} X'$ 是非随机或取固定值的矩阵

2、 **无偏特性** $E(\hat{\beta}_K) = \beta_K$ (证明见教材**P101**附录**3.1**)

3、 **最小方差特性**

在 β_K 所有的线性无偏估计中，OLS估计 $\hat{\beta}_K$ 具有最小方差 (证明见教材**P101**或附录**3.2**)

结论：在古典假定下，多元线性回归的 **OLS**估计式是最佳线性无偏估计式（**BLUE**）

$$\hat{\beta} = (X^T X)^{-1} X^T Y \qquad \text{又 } Y = X\beta + u$$

$$= (X^T X)^{-1} X^T (X\beta + u)$$

$$= (X^T X)^{-1} X^T X \beta + (X^T X)^{-1} X^T u$$

$$= \beta + (X^T X)^{-1} X^T u$$

$$\therefore E(\hat{\beta}) = \beta + E\left[(X^T X)^{-1} X^T u\right] = \beta + \underbrace{(X^T X)^{-1} X^T E(u)}_{0} = \beta$$

这便是 OLS 估计量的无偏性.

下面看最小方差性.

设存在另外一个线性无偏估计量. β^*.

由于是线性估计量 ∴ $\beta^* = AY$　A为矩阵.

又由于是无偏估计量. ∴ $E(\beta^*) = \beta$

$$\beta^* = AY = A(X\beta + u) = AX\beta + Au$$

$$E(\beta^*) = E[AX\beta + Au] = AX\beta + AE(u) = AX\beta$$

∴ $AX = I$

下面看 β^* 的方差

$$\mathrm{Var\text{-}Cov}(\beta^*) = E\left[\left(\beta^* - E(\beta^*)\right)\left(\beta^* - E(\beta^*)\right)^T\right]$$

$$Var\text{-}Cov(\beta^*) = E\left[(\beta^*-\beta)(\beta^*-\beta)^T\right]$$

又 $\beta^* = \beta + Au$ $\therefore$ $Var\text{-}Cov(\beta^*) = E\left[Auu^TA^T\right]$

$$= A\,E(uu^T)A^T = \sigma^2 AA^T$$

又OLS估计量 $\hat{\beta}$ 的方差 $Var\text{-}Cov(\hat{\beta}) = \sigma^2(X^TX)^{-1}$

$$Var(\beta^*) - Var(\hat{\beta}) = \sigma^2\left[AA^T - (X^TX)^{-1}\right]$$

又 $\left[A-(X^TX)^{-1}X^T\right]\left[A-(X^TX)^{-1}X^T\right]^T = \left[A-(X^TX)^{-1}X^T\right]\left[A^T - X(X^TX)^{-1}\right]$

$$= AA^T - (X^TX)^{-1}X^TA^T - AX(X^TX)^{-1} + (X^TX)^{-1}X^TX(X^TX)^{-1}$$

$$= AA^T - (X^TX)^{-1}$$

令 $A-(X^TX)^{-1}X^T = C$ $\therefore$ $Var(\beta^*) - Var(\hat{\beta}) = \sigma^2 CC^T$

CC'为半正定矩阵. 又半正定矩阵对角线上元素非负.

$$\therefore E(\beta_j^* - \beta_j)^2 - E(\hat{\beta}_j - \beta_j)^2 \geq 0$$

β^*的方差较OLS估计量$\hat{\beta}$大.

即OLS估计量$\hat{\beta}$具有最小方差性.

三、 OLS估计的分布性质

基本思想：

- $\hat{\beta}$ 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， $Y = X\beta + u$
决定了Y也是服从正态分布的随机变量

- $\hat{\beta}$ 是Y的线性函数，决定了 $\hat{\beta}$ 也是服从正态分布的随机变量

ECONOMETRICS

$\hat{\beta}$ 的期望与方差

- $\hat{\beta}$ 的期望　　$E(\hat{\beta})=\beta$　　（由无偏性）

- $\hat{\beta}$ 的方差和标准误差：

可以证明 $\hat{\beta}$ 的方差—协方差矩阵为（见下页）

$$\mathrm{Var\text{-}Cov}(\hat{\beta})=\sigma^2(X'X)^{-1}$$

$$Var(\hat{\beta}_j)=\sigma^2 c_{jj}$$

这里的 $(X'X)^{-1}=\begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1k} \\ c_{21} & c_{22} & \cdots & c_{2k} \\ \vdots & \vdots & \vdots & \vdots \\ c_{k1} & c_{k2} & \cdots & c_{kk} \end{bmatrix}$

$$SD(\hat{\beta}_j)=\sigma\sqrt{c_{jj}}$$

（其中 c_{jj} 是矩阵 $(X'X)^{-1}$ 中第 j 行第 j 列的元素）

所以　$\hat{\beta}_j \sim N(\beta_j,\sigma^2 c_{jj})$　（j=1,2,---k）

$\hat{\boldsymbol{\beta}}$ 的方差-协方差

$$COV(\hat{\boldsymbol{\beta}}) = E\{[\hat{\boldsymbol{\beta}} - E(\hat{\boldsymbol{\beta}})][\hat{\boldsymbol{\beta}} - E(\hat{\boldsymbol{\beta}})]'\}$$

注意 $\hat{\boldsymbol{\beta}}$是向量 $(i = 1, 2, \cdots n)$ $(j = 1, 2, \cdots n)$

$$= E[(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})']$$

(由无偏性)

$$= E[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{u}\boldsymbol{u}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}]$$

(由OLS估计式)

$$= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'E(\boldsymbol{u}\boldsymbol{u}')\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}$$

$$= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\sigma^2\boldsymbol{I}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}$$

(由同方差性)

$$= \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}$$

其中:

$$\hat{\beta} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u}) = \boldsymbol{\beta} + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{u}$$

$$E(\boldsymbol{u}\boldsymbol{u}') = \sigma^2\boldsymbol{I}$$

ECONOMETRICS

四、 随机扰动项方差σ^2的估计

σ^2 一般未知，可证明多元回归中σ^2 的无偏估计为：(证明见**P103**附录**3.3**)

$$\hat{\sigma}^2=\frac{\sum e_i^2}{n-k}\qquad \text{或表示为}\quad \hat{\sigma}^2=\frac{\boldsymbol{e'e}}{n-k}$$

对比： 一元回归中 $\hat{\sigma}^2=\sum e_i^2/(n-2)$

将 $\hat{\boldsymbol{\beta}}$ 作标准化变换：

$$z_k=\frac{\hat{\beta}_k-\beta_k}{SD(\hat{\beta}_k)}=\frac{\hat{\beta}_k-\beta_k}{\sigma\sqrt{c_{jj}}}\sim N(0,1)$$

25

σ^2 未知时$\hat{\beta}$ 的标准化变换

因 σ^2是未知的， 可用 $\hat{\sigma}^2$ 代替 σ^2去估计参数的标准误差:

● 当为大样本时，用估计的参数标准误差对$\hat{\beta}$ 作标准化变换，所得 **Z** 统计量仍可视为服从正态分布

●当为小样本时，用估计的参数标准误差对 $\hat{\beta}$ 作标准化变换，所得的 **t** 统计量服从 **t** 分布：

$$t^* = \frac{\hat{\beta}_j - \beta_j}{SE(\hat{\beta}_j)} = \frac{\hat{\beta}_j - \beta_j}{\hat{\sigma}\sqrt{c_{jj}}} \sim t(n-k)$$

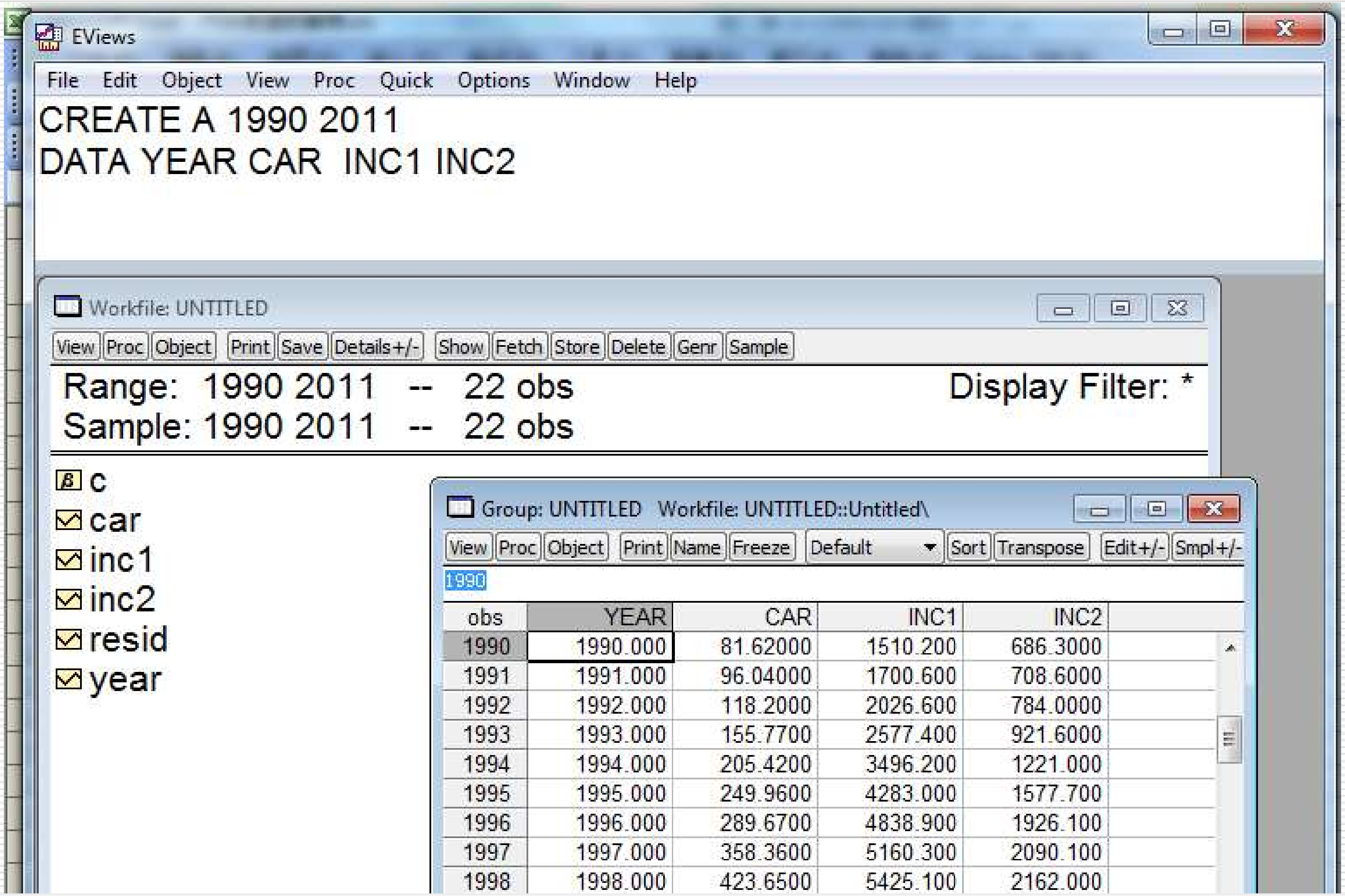
26

举例：中国1990年至2011年私人拥有汽车的数量（car）与城市居民人均收入（inc1）、农村人均收入（inc2）的回归模型如下：

CAR = -1373 + 0.15*INC1 + 0.63*INC2

用Eviews软件进行多元线性回归的操作过程：

1、建立新文件；2、定义序列；3、导入数据；



ECONOMETRICS

用**Eviews**软件进行多元线性回归的操作过程：

4、采用**OLS**方法回归；

LS CAR C INC1 INC2

Equation: EQ01 Workfile: UNTITLED::Untitled\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: CAR
Method: Least Squares
Date: 10/08/13 Time: 23:44
Sample: 1990 2011
Included observations: 22

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-1373.904	298.3567	-4.604905	0.0002
INC1	0.151215	0.210519	0.718298	0.4813
INC2	0.628150	0.718736	0.873965	0.3931

R-squared	0.936089	Mean dependent var	1635.362
Adjusted R-squared	0.929362	S.D. dependent var	2033.957
S.E. of regression	540.5811	Akaike info criterion	15.54929
Sum squared resid	5552331.	Schwarz criterion	15.69807
Log likelihood	-168.0422	Hannan-Quinn criter.	15.58434
F-statistic	139.1453	Durbin-Watson stat	0.202771
Prob(F-statistic)	0.000000		

Path = c:\users\w\documents | DB = none | WF = ur

Sum squared resid 5552331

$$\sum e^2 = e' e = 5552331$$

S.E. of regression 540.5811

$$\hat{\sigma} = \sqrt{\frac{\sum e_i^2}{n-k}} = \sqrt{\frac{5552331}{22-3}} = 540.5811$$

根据公式，$SE(\hat{\beta}_j) = \hat{\sigma}\sqrt{c_{jj}}$

又知（$\mathrm{X}^T\mathrm{X}$）$^{-1}$为

0.30461406	0.00014398	-0.00053370
0.00014398	0.00000015	-0.00000052
-0.00053370	-0.00000052	0.00000177

五、回归系数的区间估计

由于

$$t^* = \frac{\hat{\beta}_j - \beta_j}{SE(\hat{\beta}_j)} = \frac{\hat{\beta}_j - \beta_j}{\hat{\sigma}\sqrt{c_{jj}}} \sim t(n-k)$$

给定 α，查t分布表的自由度为 n-k 的临界值 $t_{\alpha/2}(n-k)$

$$P[-t_{\alpha/2}(n-k) \le t^* = \frac{\hat{\beta}_j - \beta_j}{\mathrm{SE}(\hat{\beta}_j)} \le t_{\alpha/2}(n-k)] = 1-\alpha \qquad (j = 1 \cdots k)$$

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

或

$$P[\hat{\beta}_j - t_{\alpha/2}\hat{\sigma}\sqrt{c_{jj}} \le \beta_j \le \hat{\beta}_j + t_{\alpha/2}\hat{\sigma}\sqrt{c_{jj}}] = 1-\alpha$$

或表示为

$$\beta_j = (\hat{\beta}_j - t_{\alpha/2(n-k)}\hat{\sigma}\sqrt{c_{jj}}, \hat{\beta}_j + t_{\alpha/2(n-k)}\hat{\sigma}\sqrt{c_{jj}})$$

33

第三节 多元线性回归模型的检验

一、多元回归的拟合优度检验

多重可决系数：在多元回归模型中，由各个解释变量联合起来解释了的Y的变差，在Y的总变差中占的比重，用 R^2 表示

与简单线性回归中可决系数 r^2 的区别只是 $\hat{Y}_i$ 不同

多元回归中 $$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \cdots + \hat{\beta}_k X_{ki}$$

多重可决系数可表示为

$$R^2 = \frac{ESS}{TSS} = \frac{\sum (\hat{Y}_i - \bar{Y})^2}{\sum (Y_i - \bar{Y})^2} = \frac{TSS - RSS}{TSS} = 1 - \frac{\sum e_i^2}{\sum y_i^2}$$

(注意:红色字体是与一元回归不同的部分)

34

多重可决系数的矩阵表示

$$TSS = \sum (Y_i - \bar{Y})^2 = \mathbf{Y'Y} - n\bar{Y}^2 \qquad ESS = \sum (\hat{Y}_i - \bar{Y})^2 = \hat{\boldsymbol{\beta}}'\mathbf{X'Y} - n\bar{Y}^2$$

$$R^2 = \frac{ESS}{TSS} = \frac{\hat{\boldsymbol{\beta}}'\mathbf{X'Y} - n\bar{Y}^2}{\mathbf{Y'Y} - n\bar{Y}^2}$$

可用代数式表达为

$$R^2 = \frac{\hat{\beta}_2 \sum x_{2i} y_i + \hat{\beta}_3 \sum x_{3i} y_i + \cdots + \hat{\beta}_k \sum x_{ki} y_i}{\sum y_i^2}$$

特点: 多重可决系数是模型中解释变量个数的不减函数，这给对比不同模型的多重可决系数带来缺陷，所以需要修正。

ECONOMETRICS

修正的可决系数

思想： 可决系数只涉及变差，没有考虑自由度。如果用自由度去校正所计算的变差，可纠正解释变量个数不同引起的对比困难。

回顾：

自由度： 统计量的自由度指可自由变化的样本观测值个数，它等于所用样本观测值的个数减去对观测值的约束个数。

可决系数的修正方法

总变差 **TSS** $=\sum(Y_i-\bar{Y})^2=\sum y_i^2$ 自由度为 **n-1**

解释了的变差 **ESS** $=\sum(\hat{Y}_i-\bar{Y})^2$ 自由度为 **k-1**

剩余平方和 **RSS** $=\sum(Y_i-\hat{Y}_i)^2=\sum e_i^2$ 自由度为 **n-k**

修正的可决系数为

$$\bar{R}^2=1-\frac{\sum e_i^2/(n-k)}{\sum y_i^2/(n-1)}=1-\frac{n-1}{n-k}\frac{\sum e_i^2}{\sum y_i^2}=1-\frac{n-1}{n-k}(1-R^2)$$

ECONOMETRICS 修正的可决系数 $\bar{R}^2$ 与可决系数 R^2 的关系

已经导出：

$$\bar{R}^2 = 1-(1-R^2)\frac{n-1}{n-k}$$

注意：

可决系数 R^2 必定非负，但所计算的修正可决系数 $\bar{R}^2$ 有可能为负值

解决办法： 若计算的 $\bar{R}^2 < 0$，规定 $\bar{R}^2$ 取值为0

ECONOMETRICS

二、回归方程的显著性检验（F检验）

基本思想：

在多元回归中包含多个解释变量，它们与被解释变量是否有显著关系呢?

当然可以分别检验各个解释变量对被解释变量影响的显著性。

但是我们首先关注的是所有解释变量联合起来对被解释变量影响的显著性，或整个方程总的联合显著性，需要对方程的总显著性在方差分析的基础上进行F检验。

ECONOMETRICS

1.方差分析

在讨论可决系数时已经分析了被解释变量总变差

TSS的分解及自由度：

TSS=ESS+RSS

注意：　　　**Y**的样本方差**=** 总变差**/**自由度

即

$$\hat{\sigma}_{Y_i}^2 = \frac{TSS}{n-1} = \frac{\sum (Y_i - \bar{Y})^2}{n-1}$$

显然，**Y**的样本方差也可分解为两部分，可用方差分析表分解

ECONOMETRICS

方差分析表

总变差 TSS= $\sum (Y_i - \bar{Y})^2$ 自由度 N−1

模型解释了的变差 ESS= $\sum (\hat{Y}_i - \bar{Y})^2$ 自由度 K−1

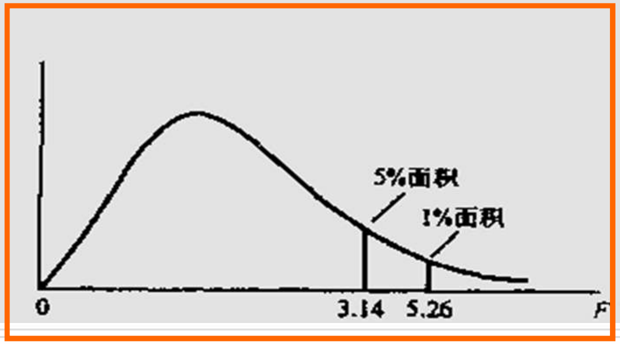
剩余变差 RSS= $\sum (Y_i - \hat{Y}_i)^2$ 自由度 N−K

变差来源	平方和	自由度	方差
归于回归模型	ESS= $\sum (\hat{Y}_i - \bar{Y})^2$	k-1	$\sum (\hat{Y}_i - \bar{Y})^2 /(k-1)$
归于剩余	RSS= $\sum (Y_i - \hat{Y}_i)^2$	n-k	$\sum (Y_i - \hat{Y}_i)^2 /(n-k)$
总变差	TSS= $\sum (Y_i - \bar{Y})^2$	n-1	$\sum (Y_i - \bar{Y})^2 /(n-1)$

基本思想:如果多个解释变量联合起来对被解释变量的影响不显著,"归于回归的方差"比"归于剩余的方差"显著地小应是大概率事件。

ECONOMETRICS

2. F检验



原假设：$H_0: \beta_2 = \beta_3 = \cdots = \beta_k = 0$

（所有解释变量联合起来对被解释变量的影响不显著）

备择假设：$H_1: \beta_j (j = 2, \cdots k)$ 不全为**0**

建立统计量(可以证明)：

$$F = \frac{ESS/(k-1)}{RSS/(n-k)} = \frac{\sum(\hat{Y}_i - \bar{Y})^2/(k-1)}{\sum(Y_i - \hat{Y}_i)^2/(n-k)} \sim F(k-1, n-k)$$

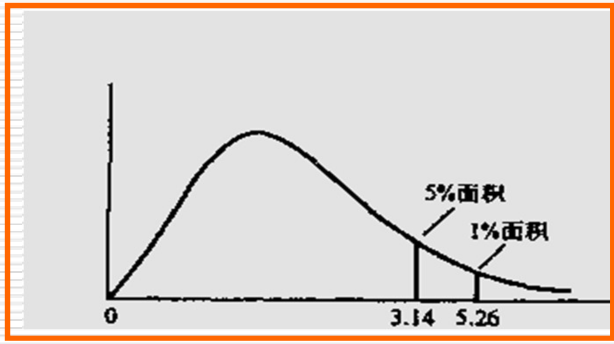
给定显著性水平 α，查**F**分布表中自由度为 **k-1** 和 **n-k** 的临界值 $F_\alpha(k-1, n-k)$，并通过样本观测值计算**F**值

42

ECONOMETRICS

F检验方式

- ▼如果计算的**F**值大于临界值 $F_\alpha(k-1,n-k)$， 则拒绝 $H_0:\beta_2=\beta_3=\cdots=\beta_k=0$， 说明回归模型有显著意义，即所有解释变量联合起来对**Y**确有显著影响。
- ▼如果计算的**F**值小于临界值 $F_\alpha(k-1,n-k)$ ，则不拒绝 $H_0:\beta_2=\beta_3=\cdots=\beta_k=0$ ，说明回归模型没有显著意义，即所有解释变量联合起来对**Y**没有显著影响。



三、各回归系数的假设检验

注意： 在一元回归中F检验与t检验等价，且 $F = t^2$

(见教材**P87**证明)

但在多元回归中，**F**检验显著，不一定每个解释变量都对**Y**有显著影响。还需要分别检验当其他解释变量保持不变时，各个解释变量**X**对被解释变量**Y**是否有显著影响。

方法：

原假设 $H_0: \beta_j = 0$

备择假设 $H_1: \beta_j \neq 0$

（j=1,2,……k）

统计量**t**为：
$$t^* = \frac{\hat{\beta}_j - \beta_j}{SE(\hat{\beta}_j)} = \frac{\hat{\beta}_j}{\hat{\sigma}\sqrt{c_{jj}}} \sim t(n-k)$$

对各回归系数假设检验的作法

给定显著性水平α，查t分布表的临界值为 $t_{\alpha/2}(n-k)$

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

就不拒绝 $H_0: \beta_j = 0$ ，而拒绝 $H_1: \beta_j \neq 0$

即认为 β_j 所对应的解释变量 X_j 对被解释变量Y的影响不显著。

如果 $t^* < -t_{\alpha/2}(n-k)$ 或 $t^* > t_{\alpha/2}(n-k)$

就拒绝 $H_0: \beta_j = 0$ 而不拒绝 $H_1: \beta_j \neq 0$

即认为 β_j 所对应的解释变量 X_j 对被解释变量Y的影响是显著的。

讨论：在多元回归中，可以作F检验，也可以分别对每个回归系数逐个地进行 t 检验。F 检验与t检验的关系是什么？

第四节 多元线性回归模型的预测

一、被解释变量平均值预测

1. Y平均值的点预测

方法：将解释变量预测值代入估计的方程：

多元回归时：

$$\hat{Y}_F = \hat{\beta}_1 + \hat{\beta}_2 X_{F2} + \hat{\beta}_3 X_{F3} + \cdots + \hat{\beta}_K X_{Fk}$$

或 $$\hat{Y}_F = \boldsymbol{X}_F \hat{\boldsymbol{\beta}}$$

注意：预测期的 $\boldsymbol{X}_F$ 是第一个元素为1的行向量,不是矩阵,也不是列向量

$$\boldsymbol{X}_F = (1 \quad X_{F2} \quad X_{F3} \quad \cdots X_{Fk} \quad)$$

ECONOMETRICS

2. Y平均值的区间预测

基本思想： **(与简单线性回归时相同)**

- 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。

- 为了对**Y**作区间预测，必须确定平均值预测值 $\hat{Y}_F$ 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，并要明确其概率分布性质。

47

ECONOMETRICS

区间预测的具体作法 （回顾简单线性回归）

简单线性回归中

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

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

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

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

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

48

区间预测的具体作法（多元时）

多元回归时，与预测的平均值 $\hat{Y}_F$ 和真实平均值 $E(Y_F|\boldsymbol{X_F})$ 都有关的是二者的偏差 w_F：

$$w_F = \hat{Y}_F - E(Y_F|\boldsymbol{X_F})$$

w_F 服从正态分布，可证明

$$E(w_F) = 0 \qquad Var(w_F) = \sigma^2 \boldsymbol{X_F}(\boldsymbol{X'X})^{-1}\boldsymbol{X'_F}$$

用 $\hat{\sigma}^2 = \sum e_i^2 / (n-k)$ 代替 σ^2，可构造t统计量

$$t^* = \frac{w_F - E(w_F)}{SE(w_F)} = \frac{\hat{Y}_F - E(Y_F|\boldsymbol{X_F})}{\hat{\sigma}\sqrt{\boldsymbol{X_F}(\boldsymbol{X'X})^{-1}\boldsymbol{X'_F}}} \sim t(n-k)$$

ECONOMETRICS

或者

$\hat{Y}_F$ 服从正态分布，可证明

$$E(\hat{Y}_F)=E(Y_F\,|\,\boldsymbol{X_F})$$

$$Var(\hat{Y}_F)=\sigma^2\boldsymbol{X_F}(\boldsymbol{X'X})^{-1}\boldsymbol{X'_F}$$

即 $\hat{Y}_F \sim N\{E(Y_F\,|\,\boldsymbol{X_F}),\sigma^2\boldsymbol{X_F}(\boldsymbol{X'X})^{-1}\boldsymbol{X'_F}\}$

标准化 $t^*=\dfrac{\hat{Y}_F-E(\hat{Y}_F)}{SE(\hat{Y}_F)}=\dfrac{\hat{Y}_F-E(Y_F\,|\,\boldsymbol{X_F})}{\sigma\sqrt{\boldsymbol{X_F}(\boldsymbol{X'X})^{-1}\boldsymbol{X'_F}}}\sim N(0,1)$

当用 $\hat{\sigma}^2=\sum e_i^2/(n-k)$ 代替 σ^2 时，可构造 **t** 统计量

$$t=\frac{\hat{Y}_F-E(\hat{Y}_F)}{SE(\hat{Y}_F)}=\frac{\hat{Y}_F-E(Y_F\,|\,\boldsymbol{X_F})}{\hat{\sigma}\sqrt{\boldsymbol{X_F}(\boldsymbol{X'X})^{-1}\boldsymbol{X'_F}}}\sim t(n-k)$$

50

区间预测的具体作法

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

$$P\{[(\hat{Y}_F - t_{\alpha/2}\,\hat{SE}(w_F)] \le E(Y_F|\boldsymbol{X_F}) \le [(\hat{Y}_F + t_{\alpha/2}\,\hat{SE}(w_F)]\}$$

$$= 1-\alpha$$

或

$$P\{[\hat{Y}_F - t_{\alpha/2}\hat{\sigma}\sqrt{\boldsymbol{X_F(X`X)^{-1}X'_F}}] \le E(Y_F|\boldsymbol{X_F}) \le [\hat{Y}_F + t_{\alpha/2}\hat{\sigma}\sqrt{\boldsymbol{X_F(X`X)^{-1}X'_F}}]\}$$

$$= 1-\alpha$$

ECONOMETRICS

二、被解释变量个别值预测

基本思想： （与简单线性回归时相同）

●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。

●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性质。

ECONOMETRICS

个别值区间预测具体作法

已知剩余项 e_F 是与预测值 $\hat{Y}_F$ 和个别值 Y_F 都有关的变量 $e_F = Y_F - \hat{Y}_F$

并且已知 e_F服从正态分布，且多元回归时可证明

$$E(e_F) = 0$$

$$Var(e_F) = \sigma^2[1 + \mathbf{X}_F(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_F']$$

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

$$t = \frac{e_F - E(e_F)}{SE(e_F)} = \frac{Y_F - \hat{Y}_F}{\hat{\sigma}\sqrt{1 + \mathbf{X}_F(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_F'}} \sim t(n-k)$$

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

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

因此，多元回归时**Y**的个别值的置信度**1-α**的预测区间的上下限为

$$Y_F = \hat{Y}_F \mp t_{\alpha/2}\hat{\sigma}\sqrt{1+\boldsymbol{X}_F(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}_F'}$$

第五节 案例分析

研究的目的要求

为了研究影响中国地方财政教育支出差异的主要原因，分析地方财政教育支出增长的数量规律，预测中国地方财政教育支出的增长趋势，需要建立计量经济模型。

研究范围： 2011年31个省市区的数据为样本

理论分析： 影响中国地方财政教育支出的主要的因素有：

（1）由地区经济规模决定的地方整体财力；

（2）地区人口数量不同决定各地教育规模不同；

（3）人民对教育质量的需求对以政府教育投入为代表的公共财政的需求会有相当的影响。

（4）物价水平，影响地方财政对教育的支出。

（5）地方政府对教育投入的能力与意愿

模型设定

选择地方财政教育支出为被解释变量。

选择“地区生产总值（**GDP**）”作为地区经济规模的代表；

选择各地区的“年末人口数量”作为各地区居民对教育规模的需求的代表；

选择“居民平均每人教育现金消费”作为代表居民对教育质量的需求；

选择居民教育消费价格指数作为价格变动影响的因素；

由于地方政府教育投入的能力与意愿难以直接量化，选择“教育支出在地方财政支出中的比重”作为其代表。

探索将模型设定为线性回归模型形式：

$$Y_{1i} = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \beta_5 X_{5i} + \beta_6 X_{6i} + u_i$$

样本数据：2011年各地区地方财政教育支出及主要影响因素

地区	地方财政教育支出（亿元）	地区生产总值（亿元）	年末人口数（万人）	居民平均每人教育现金消费（元/人）	居民教育消费价格指数	教育支出在地方财政支出中的比重%
	Y	X2	X3	X4	X5	X6
北京	520.08	16251.93	2018.6	1171.28	101.154	0.1603
天津	302.32	11307.28	1355	780.85	100.2248	0.1683
河北	652.11	24515.76	7240.51	523.54	100.8815	0.1843
山西	421.79	11237.55	3593	722.92	102.8131	0.1784
内蒙古	390.69	14359.88	2481.71	740.41	101.3471	0.1307
辽宁	544.09	22226.7	4383	757.94	101.2031	0.1393
吉林	319.82	10568.83	2749.41	739.14	100.4868	0.1453
黑龙江	373.83	12582	3834	598.7	102.0862	0.1338
上海	549.24	19195.69	2347.46	1285.61	101.7647	0.1403
江苏	1093.22	49110.27	7898.8	1005.65	102.3058	0.1757
浙江	751.42	32318.85	5463	1332.63	99.82673	0.1956
安徽	564.71	15300.65	5968	776.9	100.6223	0.1710
福建	406.73	17560.18	3720	629.11	99.34025	0.1850
江西	474.43	11702.82	4488.437	610.69	101.3277	0.1872
山东	1047.9	45361.85	9637	656.83	100.7661	0.2095
河南	857.14	26931.03	9388	544.6	101.7941	0.2017

地区	地方财政教育支出（亿元）	地区生产总值（亿元）	年末人口数（万人）	居民平均每人教育现金消费（元/人）	居民教育消费价格指数	教育支出在地方财政支出中的比重%
	Y	X2	X3	X4	X5	X6
湖北	488.16	19632.26	5757.5	690.86	101.0219	0.1519
湖南	540.83	19669.56	6595.6	626.86	102.5655	0.1536
广东	1227.87	53210.28	10504.85	929.03	101.2925	0.1829
广西	456.89	11720.87	4645	561.81	102.8054	0.1795
海南	127.27	2522.66	877.34	565.08	100.8938	0.1634
重庆	318.7	10011.37	2919	460.09	101.8485	0.1240
四川	684.66	21026.68	8050	534.18	101.5956	0.1465

三、估计参数

Dependent Variable: Y
Method: Least Squares
Date: 10/02/13 Time: 20:41
Sample: 1 31
Included observations: 31

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-2416.491	935.8816	-2.582048	0.0161
X2	0.011171	0.001768	6.316726	0.0000
X3	0.039473	0.007951	4.964338	0.0000
X4	0.146028	0.051660	2.826690	0.0091
X5	22.81615	9.086687	2.510942	0.0189
X6	866.4100	470.3214	1.842166	0.0773

R-squared	0.973227	Mean dependent var	499.9448
Adjusted R-squared	0.967872	S.D. dependent var	275.3621
S.E. of regression	49.35657	Akaike info criterion	10.80800
Sum squared resid	60901.79	Schwarz criterion	11.08555
Log likelihood	-161.5241	Hannan-Quinn criter.	10.89848
F-statistic	181.7539	Durbin-Watson stat	2.378747
Prob(F-statistic)	0.000000		

模型估计的结果为：

$$\hat{Y}_i = -2416.49 + 0.0112X_2 + 0.0395X_3 + 0.1460X_4 + 22.8162X_5 + 866.4100X_6$$

(935.8816) (0.0018) (0.0080) (0.0517) （9.0867） (470.3214)

t= (-2.5820) (6.3167) (4.9643) (2.8267) (2.5109) (1.8422)

$R^2 = 0.9732$ F=181.7539 n=31

ECONOMETRICS

模型检验：

1、经济意义检验：在假定其它变量不变的情况下，地区生产总值(GDP)每增长1亿元，平均说来地方财政教育支出将增长0.0112亿元；地区年末人口每增长1万人，平均说来地方财政教育支出会增长0.0395亿元；当居民平均每人教育现金消费增加1元，平均说来地方财政教育支出会增长0.1460亿元；当居民教育消费价格指数增加1个百分点，平均说来地方财政教育支出会增长22.8162亿元。当教育支出在地方财政支出中的比重增加1%，平均说来地方财政教育支出会增长866.41亿元。

2、统计检验：

拟合优度：$R^2 = 0.9732$ ，修正的可决系数为 $\bar{R}^2 = 0.9679$ ，说明模型对样本的拟合很好。

F检验：针对 $H_0: \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = 0$，给定显著性水平 $\alpha = 0.05$ ，查F分布表自由度为k-1=5和n-k=25的临界值为 $F_\alpha(5,25) = 2.61$ 。由于F=181.7539>2.61 ，应拒绝原假设 ，说明回归方程显著，

ECONOMETRICS

t检验

分别针对：$H_O:\beta_j=0 \quad (j=1,2,3,4,5,6)$ ，
取 $\alpha=0.05$，查t分布表得自由度为n-k=25临界值 $F_\alpha(5,25)=2.06$。
取 $\alpha=0.10$，查t分布表得自由度为n-k=25临界值$F_\alpha(5,25)=1.708$ 。
与 $\hat{\beta}_1$、$\hat{\beta}_2$、$\hat{\beta}_3$、$\hat{\beta}_4$、$\hat{\beta}_5$对应的t统计量分别为-2.5820、6.3167、4.9643、2.8267、2.5109，其绝对值均大于$F_\alpha(5,25)=2.06$，说明在显著性水平$\alpha=0.05$ 下，分别都应当拒绝：$H_O:\beta_j=0 \quad (j=1,2,3,4,5)$
而与 $\hat{\beta}_6$对应的t统计量 $F_{0.05}(5,25)=2.06>t_{(\hat{\beta}_6)}=1.8422>F_{0.10}(5,25)=1.708$
表明“教育支出在地方财政支出中的比重”对“地方财政教育支出”Y在$\alpha=0.05$的显著性水平下，没有显著的影响。但是在 $\alpha=0.10$ 显著性水平下，“教育支出在地方财政支出中的比重”对“地方财政教育支出”Y有显著的影响。这样的结论从表3.4中的P值也可能判断，与$\hat{\beta}_1$ $\hat{\beta}_2$ $\hat{\beta}_3$ $\hat{\beta}_4$、$\hat{\beta}_5$估计值对应的P值均小于$\alpha=0.05$,表明在显著性水平 $\alpha=0.05$ 下，对应解释变量对被解释变量影响显著。与 $\hat{\beta}_6$ 估计值对应的P值为0.0773，小于 $\alpha=0.10$ ，表明在 $\alpha=0.10$ 的显著性水平下，“教育支出在地方财政支出中的比重”对“地方财政教育支出”Y影响是显著的。

本章小结

1. 多元线性回归模型及其矩阵形式。

2. 多元线性回归模型中对随机扰动项**u**的假定，除了其他基本假定以外，还要求满足无多重共线性假定。

3. 多元线性回归模型参数的最小二乘估计量；在基本假定满足的条件下，多元线性回归模型最小二乘估计式是最佳线性无偏估计量。

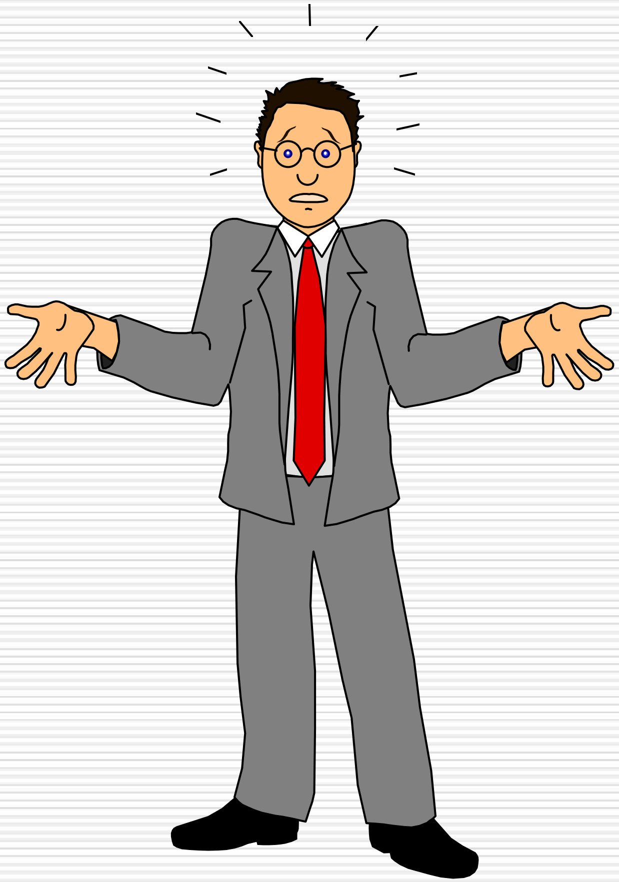
4. 多元线性回归模型中参数区间估计的方法。

5. 多重可决系数的意义和计算方法，修正可决系数的作用和方法。

6. 对多元线性回归模型中所有解释变量联合显著性的**F**检验。

7. 多元回归分析中，对各个解释变量是否对被解释变量有显著影响的**t**检验。

8. 利用多元线性回归模型作被解释变量平均值预测与个别值预测的方法。



第三章结束了！

THANKS