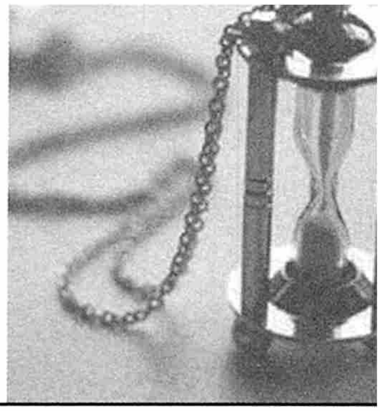

第三章　经典单方程计量经济学模型：多元线性回归模型

在实际经济问题中，一个变量往往受到多个变量的影响。例如，家庭消费支出，除了受家庭可支配收入的影响外，还受诸如家庭所拥有的财富、物价水平、金融机构存款利息，甚至广告、就业状况等多种因素的影响，表现在线性回归模型中的解释变量有多个。这样的模型被称为**多元线性回归模型**。多元线性回归模型参数估计的原理与一元线性回归模型相同，只是计算更为复杂。

§3.1　多元线性回归模型

一、多元线性回归模型的形式

多元线性回归模型的一般形式为

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + \mu \tag{3.1.1}$$

其中 k 为解释变量的数目，β_j ($j = 1, 2, \cdots, k$)称为**回归系数**(regression coefficient)。人们习惯上把常数项看作一个虚变量的参数，在参数估计过程中该虚变量的样本观测值始终取 1，这样，模型中解释变量的数目为 $k+1$ 。

同一元回归分析一样，(3.1.1)式也被称为**总体回归函数**的随机表达形式。它的非随机表达式为

$$\mathrm{E}(Y \mid X_1, X_2, \cdots, X_k) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k \tag{3.1.2}$$

可见，多元回归分析是以多个解释变量的给定值为条件的回归分析，(3.1.2)式表示各解释变量 X 值给定时 Y 的平均响应。β_j 也被称为**偏回归系数**(partial regression coefficient)，表示在其他解释变量保持不变的情况下，X_j 每变化一个单位时，Y 的均值 $\mathrm{E}(Y)$ 的变化，或者说 β_j 给出 X_j 的单位变化对 Y 均值的“直接”或“净”(不含其他变量)影响。

如果给出一组观测值$\{(X_{i1}, X_{i2}, \cdots, X_{ik}, Y_i): i = 1, 2, \cdots, n\}$，则总体回归模型还可写成如下形式

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_k X_{ik} + \mu_i \qquad i = 1, 2, \cdots, n \tag{3.1.3}$$

或

$$Y_i = \boldsymbol{X_i \beta} + \mu_i \qquad i = 1, 2, \cdots, n \tag{3.1.4}$$

其中，$\boldsymbol{X_i} = (1, X_{i1}, X_{i2}, \cdots, X_{ik})$，$\boldsymbol{\beta} = (\beta_0, \beta_1, \cdots, \beta_k)'$ 。

由(3.1.3)式或(3.1.4)式表示的 n 个随机方程的矩阵表达式为

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\mu} \tag{3.1.5}$$

其中

$$\boldsymbol{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}_{n\times 1}, \qquad \boldsymbol{X} = \begin{pmatrix} 1 & X_{11} & X_{12} & \cdots & X_{1k} \\ 1 & X_{21} & X_{22} & \cdots & X_{2k} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & X_{n1} & X_{n2} & \cdots & X_{nk} \end{pmatrix}_{n\times(k+1)}$$

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix}_{(k+1)\times 1}, \qquad \boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{pmatrix}_{n\times 1}$$

与一元回归分析相仿，在给出总体中的一个样本时，估计**样本回归函数**，并让它近似代表未知的总体回归函数。

样本回归函数可表示为

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \cdots + \hat{\beta}_k X_k \tag{3.1.6}$$

其随机表达式为

$$Y = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \cdots + \hat{\beta}_k X_k + e \tag{3.1.7}$$

其中 e 称为**残差**或**剩余项**(residual)，可看成是总体回归函数中随机干扰项 μ 的近似替代。

在一个容量为 n 的样本下，样本回归函数(3.1.6)式与(3.1.7)式也可表示如下

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{i1} + \hat{\beta}_2 X_{i2} + \cdots + \hat{\beta}_k X_{ik} \quad i = 1, 2, \cdots, n \tag{3.1.8}$$

$$Y_i = \hat{\beta}_0 + \hat{\beta}_1 X_{i1} + \hat{\beta}_2 X_{i2} + \cdots + \hat{\beta}_k X_{ik} + e_i \quad i = 1, 2, \cdots, n \tag{3.1.9}$$

同样地，(3.1.8)式与(3.1.9)式中样本回归函数的矩阵表达式分别为

$$\hat{\boldsymbol{Y}} = \boldsymbol{X}\hat{\boldsymbol{\beta}} \tag{3.1.10}$$

$$\boldsymbol{Y} = \boldsymbol{X}\hat{\boldsymbol{\beta}} + \boldsymbol{e} \tag{3.1.11}$$

其中

$$\hat{\boldsymbol{\beta}} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_k \end{pmatrix}, \qquad \boldsymbol{e} = \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix}$$

二、多元线性回归模型的基本假设

为了使参数估计量具有良好的统计性质，对多元线性回归模型可做出类似于一元线性回归分析那样的若干基本假设。

假设 1：回归模型是正确设定的。

假设 2：解释变量 $X_1, X_2, \cdots, X_k$ 在所抽取的样本中具有变异性，且各 X_j 之间不存在严格线性相关性(无完全多重共线性)。

假设 3：随机干扰项具有条件零均值性

$$\mathrm{E}(\mu_i \mid X_1, X_2, \cdots, X_k) = 0 \qquad i = 1, 2, \cdots, n$$

假设 4：随机干扰项具有条件同方差及不序列相关性

$$\mathrm{Var}(\mu_i \mid X_1, X_2, \cdots, X_k) = \sigma^2 \qquad i = 1, 2, \cdots, n$$

$$\mathrm{Cov}(\mu_i, \mu_j \mid X_1, X_2, \cdots, X_k) = 0 \qquad i \neq j \qquad i, j = 1, 2, \cdots, n$$

假设 5：随机干扰项满足正态分布

$$\mu_i \mid X_1, X_2, \cdots, X_k \sim N(0, \sigma^2)$$

与一元线性回归模型的假设相比，假设 2 中对各 X_j 之间不存在严格线性相关性的要求是多元回归模型所特有的。同样地，与一元线性回归模型相类似，由假设 3 可以得到随机干扰项的非条件零均值特性，以及与各解释变量间的不相关特性：

$$\mathrm{E}(\mu_i) = 0$$

$$\mathrm{Cov}(\mu_i, \mathrm{X}_j) = \mathrm{E}(X_j \mu_i) = 0 \qquad j = 1, 2, \cdots, k$$

为了书写方便，上述假设 2 至假设 5 还可用矩阵符号来表示：

假设 2：$n \times (k+1)$ 矩阵 $\boldsymbol{X}$ 的秩 $R(\boldsymbol{X}) = k+1$，即 $\boldsymbol{X}$ 列满秩。

假设 3：$\mathrm{E}(\boldsymbol{\mu} \mid \boldsymbol{X}) = \boldsymbol{0}$ (3.1.12)

假设 4：

$$\mathrm{Var}(\boldsymbol{\mu} \mid \boldsymbol{X}) = \mathrm{E}(\boldsymbol{\mu\mu}' \mid \boldsymbol{X}) = \mathrm{E}\left(\begin{matrix} \mu_1^2 & \cdots & \mu_1\mu_n \\ \vdots & & \vdots \\ \mu_n\mu_1 & \cdots & \mu_n^2 \end{matrix} \middle| \boldsymbol{X}\right)$$

$$= \begin{pmatrix} \sigma^2 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & \sigma^2 \end{pmatrix} = \sigma^2 \boldsymbol{I} \tag{3.1.13}$$

其中，$\boldsymbol{I}$ 为一 n 阶单位矩阵。

假设 5：向量 $\boldsymbol{\mu}$ 服从一多维正态分布，即

$$\boldsymbol{\mu} \mid \boldsymbol{X} \sim N(\boldsymbol{0}, \sigma^2 \boldsymbol{I}_n) \tag{3.1.14}$$

需要说明的是，由假设 3 得到的随机干扰项与各解释变量不相关特性，意味着任何观测点处的 μ 与任何观测点处的各 X 都是不相关的，其中自然包括了第 i 个观测点处的 μ 与各 X 的不相关性，即各 X 是**同期外生**的或与 μ **同期不相关**。记 $\boldsymbol{X}_i = (1, X_{i1}, X_{i2}, \cdots, X_{ik})$，则随机干扰项的非条件零均值特征及与各解释变量的同期不相关特征一并可写成如下矩阵形式：

$$\mathrm{E}(\boldsymbol{X}_i' \mu_i) = \boldsymbol{0} \qquad i = 1, 2, \cdots, n \tag{3.1.15}$$

（3.1.15）式将在多元回归模型参数的估计中扮演重要的角色。

§ 3.2　多元线性回归模型的参数估计

同一元线性回归模型的参数估计一样，多元线性回归模型参数估计的任务仍有两项：一是求得反映变量之间数量关系的结构参数的估计量 $\hat{\beta}_j$ ($j=0,1,\cdots,k$)；二是求得随机干扰项的方差估计 $\hat{\sigma}^2$ 。多元线性回归模型在满足 § 3.1 所列的基本假设的情况下，可以采用普通最小二乘法、最大似然法或者矩估计法估计参数。

一、普通最小二乘估计

1. 普通最小二乘估计及其矩阵表达

随机抽取容量为 n 的样本观测值 $\{(X_{i1},X_{i2},\cdots,X_{ik},Y_i):i=1,2,\cdots,n\}$ ，如果样本函数的参数估计值已经得到，则有

$$\hat{Y}_i=\hat{\beta}_0+\hat{\beta}_1X_{i1}+\hat{\beta}_2X_{i2}+\cdots+\hat{\beta}_kX_{ik}\,,\qquad i=1,2,\cdots,n \tag{3.2.1}$$

根据最小二乘原理，参数估计值应使

$$\begin{aligned}Q&=\sum_{i=1}^{n}e_i^2=\sum_{i=1}^{n}(Y_i-\hat{Y}_i)^2\\&=\sum_{i=1}^{n}[Y_i-(\hat{\beta}_0+\hat{\beta}_1X_{i1}+\hat{\beta}_2X_{i2}+\cdots+\hat{\beta}_kX_{ik})]^2\end{aligned} \tag{3.2.2}$$

达到最小。由微积分知识可知，只需求 Q 关于待估参数 $\hat{\beta}_j(j=0,1,\cdots,k)$ 的偏导数，并令其值为零，就可得到**待估参数估计值的正规方程组**：

$$\begin{cases}\sum(\hat{\beta}_0+\hat{\beta}_1X_{i1}+\hat{\beta}_2X_{i2}+\cdots+\hat{\beta}_kX_{ik})=\sum Y_i\\\sum(\hat{\beta}_0+\hat{\beta}_1X_{i1}+\hat{\beta}_2X_{i2}+\cdots+\hat{\beta}_kX_{ik})X_{i1}=\sum Y_iX_{i1}\\\sum(\hat{\beta}_0+\hat{\beta}_1X_{i1}+\hat{\beta}_2X_{i2}+\cdots+\hat{\beta}_kX_{ik})X_{i2}=\sum Y_iX_{i2}\\\cdots\cdots\cdots\cdots\\\sum(\hat{\beta}_0+\hat{\beta}_1X_{i1}+\hat{\beta}_2X_{i2}+\cdots+\hat{\beta}_kX_{ik})X_{ik}=\sum Y_iX_{ik}\end{cases} \tag{3.2.3}$$

解这 $k+1$ 个方程组成的线性代数方程组，即可得到 $k+1$ 个待估参数的估计值 $\hat{\beta}_j(j=0,1,2,\cdots,k)$ 。

(3.2.3)式的矩阵形式如下：

$$\begin{pmatrix}n&\sum X_{i1}&\cdots&\sum X_{ik}\\\sum X_{i1}&\sum X_{i1}^2&\cdots&\sum X_{i1}X_{ik}\\\vdots&\vdots&&\vdots\\\sum X_{ik}&\sum X_{ik}X_{i1}&\cdots&\sum X_{ik}^2\end{pmatrix}\begin{pmatrix}\hat{\beta}_0\\\hat{\beta}_1\\\vdots\\\hat{\beta}_k\end{pmatrix}=\begin{pmatrix}1&1&\cdots&1\\X_{11}&X_{21}&\cdots&X_{n1}\\\vdots&\vdots&&\vdots\\X_{1k}&X_{2k}&\cdots&X_{nk}\end{pmatrix}\begin{pmatrix}Y_1\\Y_2\\\vdots\\Y_n\end{pmatrix}$$

即
$$(\boldsymbol{X}'\boldsymbol{X})\hat{\boldsymbol{\beta}} = \boldsymbol{X}'\boldsymbol{Y} \tag{3.2.4}$$

由 $\boldsymbol{X}$ 的列满秩性可得 $\boldsymbol{X}'\boldsymbol{X}$ 为满秩对称矩阵，故有
$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y} \tag{3.2.5}$$

将上述过程用矩阵表示如下。

根据最小二乘原理，需寻找一组参数估计值 $\hat{\boldsymbol{\beta}}$，使得残差平方和
$$Q = \sum_{i=1}^{n} e_i^2 = \boldsymbol{e}'\boldsymbol{e} = (\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})'(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})$$
最小，即参数估计值应该是方程组
$$\frac{\partial}{\partial \hat{\boldsymbol{\beta}}}(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})'(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}) = \boldsymbol{0}$$
的解。求解过程如下：
$$\frac{\partial}{\partial \hat{\boldsymbol{\beta}}}(\boldsymbol{Y}'\boldsymbol{Y} - \hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{Y} - \boldsymbol{Y}'\boldsymbol{X}\hat{\boldsymbol{\beta}} + \hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{X}\hat{\boldsymbol{\beta}}) = \boldsymbol{0}$$
$$\frac{\partial}{\partial \hat{\boldsymbol{\beta}}}(\boldsymbol{Y}'\boldsymbol{Y} - 2\boldsymbol{Y}'\boldsymbol{X}\hat{\boldsymbol{\beta}} + \hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{X}\hat{\boldsymbol{\beta}}) = \boldsymbol{0}$$
$$-\boldsymbol{X}'\boldsymbol{Y} + \boldsymbol{X}'\boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{0}$$
即得到
$$\boldsymbol{X}'\boldsymbol{Y} = \boldsymbol{X}'\boldsymbol{X}\hat{\boldsymbol{\beta}}$$
于是，参数的最小二乘估计值为
$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}$$

例 3.2.1

在例 2.1.1 的家庭可支配收入-消费支出例子中，
$$(\boldsymbol{X}'\boldsymbol{X}) = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ X_1 & X_2 & \cdots & X_n \end{pmatrix}\begin{pmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{pmatrix}$$
$$= \begin{pmatrix} n & \sum X_i \\ \sum X_i & \sum X_i^2 \end{pmatrix} = \begin{pmatrix} 10 & 21\,500 \\ 21\,500 & 53\,650\,000 \end{pmatrix}$$
$$\boldsymbol{X}'\boldsymbol{Y} = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ X_1 & X_2 & \cdots & X_n \end{pmatrix}\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}$$
$$= \begin{pmatrix} \sum Y_i \\ \sum X_i Y_i \end{pmatrix} = \begin{pmatrix} 15\,830 \\ 39\,010\,300 \end{pmatrix}$$

可求得

$$(\boldsymbol{X'X})^{-1}=\begin{pmatrix}0.722\ 6 & -0.000\ 3\\ -0.000\ 3 & 1.35\times 10^{-7}\end{pmatrix}$$

于是

$$\hat{\boldsymbol{\beta}}=\begin{pmatrix}\hat{\beta}_1\\ \hat{\beta}_2\end{pmatrix}=\begin{pmatrix}0.722\ 6 & -0.000\ 3\\ -0.000\ 3 & 1.35\times 10^{-7}\end{pmatrix}\begin{pmatrix}15\ 830\\ 39\ 010\ 300\end{pmatrix}=\begin{pmatrix}-142.2\\ 0.670\ 1\end{pmatrix}$$

---

2. 离差形式的普通最小二乘估计

对于正规方程组(3.2.3)式的矩阵形式

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

将 $\boldsymbol{Y}=\boldsymbol{X}\hat{\boldsymbol{\beta}}+\boldsymbol{e}$ 代入得

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

于是

$$\boldsymbol{X'e}=\boldsymbol{0} \tag{3.2.6}$$

或

$$\begin{cases}\sum e_i=0\\ \sum\limits_i X_{ij}e_i=0\ , \qquad j=1,2,\cdots,k\end{cases}$$

(3.2.6)式是多元线性回归模型正规方程组的另一种写法，由此容易得到多元回归分析中的样本回归模型的离差形式：

$$y_i=\hat{\beta}_1x_{i1}+\hat{\beta}_2x_{i2}+\cdots+\hat{\beta}_kx_{ik}+e_i\ , \qquad i=1,2,\cdots,n \tag{3.2.7}$$

其矩阵形式为

$$\boldsymbol{y}=\boldsymbol{x}\hat{\boldsymbol{\beta}}+\boldsymbol{e} \tag{3.2.8}$$

其中

$$\boldsymbol{y}=\begin{pmatrix}y_1\\ y_2\\ \vdots\\ y_n\end{pmatrix},\quad \boldsymbol{x}=\begin{pmatrix}x_{11} & x_{12} & \cdots & x_{1k}\\ x_{21} & x_{22} & \cdots & x_{2k}\\ \vdots & \vdots & & \vdots\\ x_{n1} & x_{n2} & \cdots & x_{nk}\end{pmatrix},\quad \hat{\boldsymbol{\beta}}=\begin{pmatrix}\hat{\beta}_1\\ \hat{\beta}_2\\ \vdots\\ \hat{\beta}_k\end{pmatrix}$$

于是容易推出，离差形式下参数的最小二乘估计结果：

$$\begin{cases}\hat{\boldsymbol{\beta}}=(\boldsymbol{x'x})^{-1}\boldsymbol{x'y}\\ \hat{\beta}_0=\overline{Y}-\hat{\beta}_1\overline{X}_1-\cdots-\hat{\beta}_k\overline{X}_k\end{cases} \tag{3.2.9}$$

3. 随机干扰项μ的方差的普通最小二乘估计

可以证明(参见《计量经济学学习指南与练习(第二版)》，潘文卿，李子奈编著，高等教育出版社，2015)，在普通最小二乘法下，随机干扰项 μ 的方差的无偏估计量为

$$\hat{\sigma}^2 = \frac{\sum e_i^2}{n-k-1} = \frac{\boldsymbol{e}'\boldsymbol{e}}{n-k-1} \tag{3.2.10}$$

二、最大似然估计

对于多元线性回归模型(3.1.4)，由于

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

所以

$$Y_i \sim N(\boldsymbol{X}_i \boldsymbol{\beta},\sigma^2)$$

其中

$$\boldsymbol{X}_i = (1 \quad X_{i1} \quad X_{i2} \quad \cdots \quad X_{ik})$$

$\boldsymbol{Y}$ 的随机抽取的 n 组样本观测值的联合概率为

$$\begin{aligned} L(\boldsymbol{\beta},\sigma^2) &= \mathrm{P}(Y_1,Y_2,\cdots,Y_n) \\ &= \frac{1}{(2\pi)^{\frac{n}{2}}\sigma^n}\mathrm{e}^{-\frac{1}{2\sigma^2}\sum[Y_i-(\beta_0+\beta_1X_{i1}+\beta_2X_{i2}+\cdots+\beta_kX_{ik})]^2} \\ &= \frac{1}{(2\pi)^{\frac{n}{2}}\sigma^n}\mathrm{e}^{-\frac{1}{2\sigma^2}(\boldsymbol{Y}-\boldsymbol{X\beta})'(\boldsymbol{Y}-\boldsymbol{X\beta})} \end{aligned} \tag{3.2.11}$$

这就是变量 $\boldsymbol{Y}$ 的似然函数。对数似然函数为

$$L^* = \ln L = -n\ln(\sqrt{2\pi}\sigma) - \frac{1}{2\sigma^2}(\boldsymbol{Y}-\boldsymbol{X\beta})'(\boldsymbol{Y}-\boldsymbol{X\beta}) \tag{3.2.12}$$

对似然函数求极大值，即对对数似然函数求极大值，也就是对

$$(\boldsymbol{Y}-\boldsymbol{X\beta})'(\boldsymbol{Y}-\boldsymbol{X\beta})$$

求极小值，就可以得到一组参数估计量 $\hat{\boldsymbol{\beta}}$，即为参数的最大似然估计

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y} \tag{3.2.13}$$

显然，其结果与参数的普通最小二乘估计是相同的。

与一元回归相仿，容易得出多元线性回归下随机干扰项方差的估计如下：

$$\hat{\sigma}^2 = \frac{(\boldsymbol{Y}-\boldsymbol{X}\hat{\boldsymbol{\beta}})'(\boldsymbol{Y}-\boldsymbol{X}\hat{\boldsymbol{\beta}})}{n} = \frac{\sum e_i^2}{n} \tag{3.2.14}$$

*三、矩估计

普通最小二乘估计是通过得到一个关于参数估计值的**正规方程组**并对它进行求解而完成的。正规方程组 (3.2.3)或(3.2.4)可以从矩估计的思路来导出。

矩估计的基本思想是寻找一组总体矩条件，并通过对应的样本矩条件来推导出未知参数的解。对多元线性总体回归模型(3.1.4)，存在如下一组矩条件：

$$\mathrm{E}(\boldsymbol{X}_i'\mu_i) = \boldsymbol{0} \tag{3.1.15}$$

于是，对应的样本矩条件可写为

$$\frac{1}{n}\sum \boldsymbol{X}_i'(Y_i - \boldsymbol{X}_i\hat{\boldsymbol{\beta}}) = \boldsymbol{0} \tag{3.2.15}$$

可以证明（参见《计量经济学学习指南与练习（第二版）》，潘文卿，李子奈编著，高等教育出版社，2015），（3.2.15）式可等价地写为如下矩阵形式

$$\frac{1}{n}\boldsymbol{X}'(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}) = \boldsymbol{0}$$

由此得到正规方程组(3.2.4)：

$$(\boldsymbol{X}'\boldsymbol{X})\hat{\boldsymbol{\beta}} = \boldsymbol{X}'\boldsymbol{Y}$$

解此正规方程组即得样本估计参数 $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}$ 。

可见矩估计的结果与普通最小二乘法以及最大似然估计法的结果一致。

值得一提的是，**矩估计法是工具变量方法**(Instrumental Variable, IV)和**广义矩估计法**(Generalized Moment Method, GMM)的基础。在矩估计法中关键是利用了由随机干扰项的条件零均值假设所推出的非条件零均值特性，以及随机干扰项与各解释变量间同期不相关特性

$$\mathrm{E}(\boldsymbol{X}_i'\mu_i) = \boldsymbol{0}$$

作为总体矩条件。**如果某个解释变量与随机干扰项相关，只要能找到 1 个工具变量，仍然可以构成一组矩条件，这就是工具变量法。**如果存在大于 k+1 个变量与随机干扰项不**相关，可以构成一组包含大于 k+1 个方程的矩条件，这就是广义矩估计法。**这些将在计量经济学高级课程中介绍。

四、参数估计量的统计性质

当多元线性回归模型满足基本假设时，其参数的普通最小二乘估计、最大似然估计及矩估计仍具有线性性、无偏性和有效性。同时，随着样本容量增加，即当 $n\to+\infty$ 时，参数估计量具有渐近无偏性、一致性及渐近有效性。

1. 线性性

由于

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

其中 $\boldsymbol{C} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$ 仅与固定的 $\boldsymbol{X}$ 有关。可见，参数估计量是被解释变量 $\boldsymbol{Y}$ 的线性组合。

2. 无偏性

在各解释变量样本值给定的条件下，参数估计量 $\hat{\boldsymbol{\beta}}$ 具有无偏性。证明如下：

$$\begin{aligned}
\mathrm{E}(\hat{\boldsymbol{\beta}} \mid \boldsymbol{X}) &= \mathrm{E}((\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y} \mid \boldsymbol{X}) \\
&= \mathrm{E}((\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\mu}) \mid \boldsymbol{X}) \\
&= \boldsymbol{\beta} + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'E(\boldsymbol{\mu} \mid \boldsymbol{X}) \\
&= \boldsymbol{\beta}
\end{aligned} \tag{3.2.16}$$

这里利用了随机干扰项条件零均值的假设 $\mathrm{E}(\boldsymbol{\mu} \mid \boldsymbol{X}) = \mathbf{0}$ 。

3. 有效性

首先给出参数估计量 $\hat{\boldsymbol{\beta}}$ 的方差-协方差矩阵：

$$
\begin{aligned}
\operatorname{Var}(\hat{\boldsymbol{\beta}} \mid \boldsymbol{X}) &= \mathrm{E}[(\hat{\boldsymbol{\beta}} - E(\hat{\boldsymbol{\beta}}))(\hat{\boldsymbol{\beta}} - \mathrm{E}(\hat{\boldsymbol{\beta}}))' \mid \boldsymbol{X}] \\
&= \mathrm{E}[(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})' \mid \boldsymbol{X}] \\
&= \mathrm{E}[(\boldsymbol{X}\boldsymbol{X}')^{-1} \boldsymbol{X}'\boldsymbol{\mu}\boldsymbol{\mu}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1} \mid \boldsymbol{X}] \\
&= (\boldsymbol{X}'\boldsymbol{X})^{-1} \boldsymbol{X}' \mathrm{E}(\boldsymbol{\mu}\boldsymbol{\mu}' \mid \boldsymbol{X}) \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1} \\
&= (\boldsymbol{X}'\boldsymbol{X})^{-1} \boldsymbol{X}' \sigma^2 \boldsymbol{I}_n \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1} \\
&= \sigma^2 (\boldsymbol{X}'\boldsymbol{X})^{-1}
\end{aligned} \tag{3.2.17}
$$

其中利用了

$$
\begin{aligned}
\hat{\boldsymbol{\beta}} &= (\boldsymbol{X}'\boldsymbol{X})^{-1} \boldsymbol{X}'\boldsymbol{Y} \\
&= (\boldsymbol{X}'\boldsymbol{X})^{-1} \boldsymbol{X}'(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\mu}) \\
&= \boldsymbol{\beta} + (\boldsymbol{X}'\boldsymbol{X})^{-1} \boldsymbol{X}'\boldsymbol{\mu}
\end{aligned}
$$

和

$$
\mathrm{E}(\boldsymbol{\mu}\boldsymbol{\mu}' \mid \boldsymbol{X}) = \sigma^2 \boldsymbol{I}_n
$$

$\boldsymbol{I}$ 为 n 阶单位矩阵。

根据高斯-马尔可夫定理，(3.2.17)式表示的方差在所有无偏估计量的方差中是最小的，所以该参数估计量具有有效性(证明见《计量经济学学习指南与练习（第二版）》，潘文卿，李子奈编著，高等教育出版社，2015)。

*4. 一致性

可以证明（参见《计量经济学学习指南与练习（第二版）》，潘文卿，李子奈编著，高等教育出版社，2015），$n \to \infty$时，

$$
\begin{aligned}
P\lim \hat{\boldsymbol{\beta}} &= P\lim (\boldsymbol{X}'\boldsymbol{X})^{-1} \boldsymbol{X}'\boldsymbol{Y} \\
&= P\lim \boldsymbol{\beta} + P\lim (\boldsymbol{X}'\boldsymbol{X})^{-1} \boldsymbol{X}'\boldsymbol{\mu} = \boldsymbol{\beta}
\end{aligned}
$$

五、样本容量问题

模型参数估计是在样本观测值的支持下完成的。计量经济学模型，说到底是从表现已经发生的经济活动的样本数据中寻找经济活动中蕴涵的规律性，所以，它对样本数据具有很强的依赖性。收集与整理样本数据是一件非常困难的工作，于是，怎样选择合适的样本容量，使其既能满足建模的需要，又能减轻收集数据的困难，就成为一个重要的实际问题。

从建模需要来讲，样本容量越大越好，这是显而易见的。这里需要讨论的是最小样本容量和满足基本要求的样本容量。

1. 最小样本容量

所谓"最小样本容量"，即从最小二乘原理和最大似然原理出发，欲得到参数估计量，不管其质量如何，所要求的样本容量的下限。

从参数估计量

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

中可以看到，欲使 $\hat{\beta}$ 存在，必须使得 $(\boldsymbol{X'X})^{-1}$ 存在。为使得 $(\boldsymbol{X'X})^{-1}$ 存在，必须满足

$$|\boldsymbol{X'X}|\neq 0$$

即矩阵 $\boldsymbol{X'X}$ 为 $k+1$ 阶满秩矩阵。矩阵乘积的秩不超过各个因子矩阵的秩，即

$$R(\boldsymbol{AB})\leqslant \min[R(\boldsymbol{A}),R(\boldsymbol{B})]$$

其中符号 R 表示矩阵的秩。所以，只有当

$$R(\boldsymbol{X})\geqslant k+1$$

时，矩阵 $\boldsymbol{X'X}$ 才为 $k+1$ 阶满秩矩阵。而 $\boldsymbol{X}$ 为 $n\times(k+1)$阶矩阵，其秩最大为 $k+1$，此时必须有

$$n\geqslant k+1$$

即样本容量必须不少于模型中解释变量的数目(包括常数项)，这就是最小样本容量。

2. 满足基本要求的样本容量

虽然当 $n\geqslant k+1$ 时可以得到参数估计量，但在 n 较小时，除了参数估计量质量不好外，一些建立模型所必需的后续工作也无法进行。例如，参数的统计检验要求样本容量必须足够大，z 检验在 $n<30$ 时不能应用；t 检验为检验变量显著性的最常用方法，经验表明，当 $n-k\geqslant 8$ 时 t 分布较为稳定，检验才较为有效。所以，**一般经验认为，当 $n\geqslant 30$ 或者至少 $n\geqslant 3(k+1)$ 时，才能说满足模型估计的基本要求。**

如果出现样本容量较小，甚至少于"最小样本容量"，那么只依靠样本信息是无法完成模型估计的。这时需要引入非样本信息，如先验信息和后验信息，并采用其他估计方法，如贝叶斯(Bayes)估计方法，才能完成模型的参数估计。

六、多元线性回归模型的参数估计实例

例 3.2.2

居民的收入水平决定了其消费支出水平，但不同收入来源水平的变动对消费水平的影响是有差异的。从中国的统计资料看，中国城镇居民人均收入来源主要包括工资性收入、营业净收入、财产性收入与转移性收入 4 大项，而从当前的情况看，广大居民的收入主要来源于工资性收入，其他 3 项来源的收入相对来说要小得多。表 3.2.1 给出了 2013 年中国 31 个省区城镇居民人均工资性收入、其他收入以及人均现金消费支出的数据，从 31 个省区简单算术平均数据看，人均工资性收入为 17 872.4 元，而人均其他收入只有 9 798.7 元，前者是后者的 2 倍。为了考察人均工资性收入与其他

收入的变动如何具体影响城镇居民的人均消费性支出，我们考虑建立二元线性模型。

表 3.2.1 2013 年中国内地各地区城镇居民人均收入与人均消费性支出 单位：元

地区	现金消费支出 Y	工资性收入 X_1	其他收入 X_2	地区	现金消费支出 Y	工资性收入 X_1	其他收入 X_2
北　京	26 274.9	30 273.0	15 000.8	湖　北	15 749.5	15 571.8	9 608.7
天　津	21 711.9	23 231.9	12 423.7	湖　南	15 887.1	13 951.4	10 691.6
河　北	13 640.6	14 588.4	9 554.4	广　东	24 133.3	25 286.5	11 217.5
山　西	13 166.2	16 216.4	7 797.2	广　西	15 417.6	15 647.8	9 381.0
内蒙古	19 249.1	18 377.9	8 600.1	海　南	15 593.0	15 773.0	9 146.8
辽　宁	18 029.7	15 882.0	12 022.9	重　庆	17 813.9	16 654.7	10 195.7
吉　林	15 932.3	14 388.3	9 155.9	四　川	16 343.5	14 976.0	8 917.9
黑龙江	14 161.7	12 525.8	8 623.4	贵　州	13 702.9	13 627.6	7 785.5
上　海	28 155.0	33 235.4	15 643.9	云　南	15 156.1	15 140.7	9 557.6
江　苏	20 371.5	21 890.0	13 241.0	西　藏	12 231.9	19 604.0	2 956.7
浙　江	23 257.2	24 453.0	16 788.0	陕　西	16 679.7	16 441.0	7 667.8
安　徽	16 285.2	15 535.3	9 470.8	甘　肃	14 020.7	13 329.7	6 819.3
福　建	20 092.7	21 443.4	11 939.3	青　海	13 539.5	14 015.6	8 115.4
江　西	13 850.5	14 767.5	8 181.9	宁　夏	15 321.1	15 363.9	8 402.8
山　东	17 112.2	21 562.1	9 066.0	新　疆	15 206.2	15 585.3	6 802.6
河　南	14 822.0	14 704.2	8 982.3				

资料来源：根据《中国统计年鉴》(2014）整理。

Eviews 软件估计结果如表 3.2.2 所示。两个解释变量前的参数估计值分别为 0.486 5、0.601 7，都为正数，且都处于 0 与 1 之间，这些参数估计值的经济含义是合理的。随机干扰项的方差的估计值为

$$\hat{\sigma}^2 = \frac{36\,549\,482}{31-3} = 1\,305\,338.6$$

表 3.2.2 中国内地城镇居民人均消费支出二元回归估计

Dependent Variable: Y
Method: Least Squares
Sample: 1 31

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	2599.145	827.3419	3.141561	0.0039
X1	0.486512	0.057588	8.448182	0.0000
X2	0.601749	0.104244	5.772494	0.0000

R-squared	0.922460	Mean dependent var	17190.60
Adjusted R-squared	0.916921	S.D. dependent var	3963.845
S.E. of regression	1142.514	Akaike info criterion	17.01162
Sum squared resid	36549482	Schwarz criterion	17.15039
Log likelihood	-260.6800	Hannan-Quinn criter.	17.05685
F-statistic	166.5516	Durbin-Watson stat	1.914732
Prob(F-statistic)	0.000000		

§3.3 多元线性回归模型的统计检验

多元线性回归模型的参数估计出来，即求出样本回归函数后，还需进一步对该样本回归函数进行统计检验，以判定估计的可靠程度，包括拟合优度检验、方程总体线性的显著性检验、变量的显著性检验，以及参数的置信区间估计等方面。

一、拟合优度检验

1. 可决系数与调整的可决系数

在一元线性回归模型中，使用可决系数 R^2 来衡量样本回归线对样本观测值的拟合程度。在多元线性回归模型中，也可用该统计量来衡量样本回归线对样本观测值的拟合程度。

记 $\text{TSS}=\sum(Y_i-\bar{Y})^2$ 为总离差平方和，$\text{ESS}=\sum(\hat{Y}_i-\bar{Y})^2$ 为回归平方和，$\text{RSS}=\sum(Y_i-\hat{Y}_i)^2$ 为残差平方和，则

$$\begin{aligned}\text{TSS}&=\sum(Y_i-\bar{Y})^2\\&=\sum[(Y_i-\hat{Y}_i)+(\hat{Y}_i-\bar{Y})]^2\\&=\sum(Y_i-\hat{Y}_i)^2+2\sum(Y_i-\hat{Y}_i)(\hat{Y}_i-\bar{Y})+\sum(\hat{Y}_i-\bar{Y})^2\end{aligned}$$

由于

$$\begin{aligned}&\sum(Y_i-\hat{Y}_i)(\hat{Y}_i-\bar{Y})\\&=\sum e_i(\hat{Y}_i-\bar{Y})\\&=\hat{\beta}_0\sum e_i+\hat{\beta}_1\sum e_iX_{1i}+\cdots+\hat{\beta}_k\sum e_iX_{ki}+\bar{Y}\sum e_i\\&=0\end{aligned}$$

所以有

$$\text{TSS}=\sum(Y_i-\hat{Y}_i)^2+\sum(\hat{Y}_i-\bar{Y})^2=\text{RSS}+\text{ESS} \tag{3.3.1}$$

即总离差平方和可分解为回归平方和与残差平方和两部分。回归平方和反映了总离差平方和中可由样本回归线解释的部分，它越大，残差平方和越小，表明样本回归线与样本观测值的拟合程度越高。因此，可用回归平方和占总离差平方和的比重来衡量样本回归线对样本观测值的拟合程度：

$$R^2=\frac{\text{ESS}}{\text{TSS}}=1-\frac{\text{RSS}}{\text{TSS}} \tag{3.3.2}$$

该统计量越接近于 1，模型的拟合优度越高。

在应用过程中发现，如果在模型中增加一个解释变量，R^2 往往增大。这是因为残差平方和往往随着解释变量个数的增加而减少，至少不会增加。这就给人一个错觉：要使

模型拟合得好，只要增加解释变量即可。但是，现实情况往往是，由增加解释变量个数引起的 R^2 的增大与拟合好坏无关，因此在多元回归模型之间比较拟合优度，R^2 就不是一个合适的指标，必须加以调整。

在样本容量一定的情况下，增加解释变量必定使得自由度减少，所以调整的思路是将残差平方和与总离差平方和分别除以各自的自由度，以剔除变量个数对拟合优度的影响。记 $\bar{R}^2$ 为**调整的可决系数**(adjusted coefficient of determination)，则有

$$\bar{R}^2 = 1 - \frac{\mathrm{RSS}/(n-k-1)}{\mathrm{TSS}/(n-1)} \tag{3.3.3}$$

其中 $n-k-1$ 为残差平方和的自由度，$n-1$ 为总离差平方和的自由度。显然，如果增加的解释变量没有解释能力，则对残差平方和 RSS 的减小没有多大帮助，但增加了待估参数的个数，从而使 $\bar{R}^2$ 有较大幅度的下降。

调整的可决系数与未经调整的可决系数之间存在如下关系：

$$\bar{R}^2 = 1 - (1-R^2)\frac{n-1}{n-k-1} \tag{3.3.4}$$

在实际应用中，$\bar{R}^2$ 达到多大才算模型通过了检验？没有绝对的标准，要看具体情况而定。模型的拟合优度并不是判断模型质量的唯一标准，有时甚至为了追求模型的经济意义，可以牺牲一点拟合优度。在下一部分中，将推导出 $\bar{R}^2$ 与另一个统计量 F 的关系，那时会对 $\bar{R}^2$ 有新的认识。

在例 3.2.2 中，$\bar{R}^2$ =0.916 9，这对截面数据来说应是很好的拟合结果了。如果去掉其他来源的收入项，只保留工资性收入项，则回归结果显示 $\bar{R}^2$ =0.824 3。因此，模型中引入其他来源的收入项提高了模型的解释能力，换言之，其他收入项应该作为重要的解释变量引入到模型中来。

*2. 赤池信息准则和施瓦茨准则

为了比较所含解释变量个数不同的多元回归模型的拟合优度，常用的标准还有**赤池信息准则**(Akaike Information Criterion, AIC)和**施瓦茨准则**(Schwarz Criterion，SC)，其定义分别为

$$\mathrm{AIC} = \ln\frac{\boldsymbol{e}'\boldsymbol{e}}{n} + \frac{2(k+1)}{n} + 1 + \ln(2\pi) \tag{3.3.5}$$

$$\mathrm{SC} = \ln\frac{\boldsymbol{e}'\boldsymbol{e}}{n} + \frac{k+1}{n}\ln n + 1 + \ln(2\pi) \tag{3.3.6}$$

这两个准则均要求仅当所增加的解释变量能够减少 AIC 值或 SC 值时才在原模型中增加该解释变量。显然，与调整的可决系数相仿，如果增加的解释变量没有解释能力，则对残差平方和 $\boldsymbol{e}'\boldsymbol{e}$ 的减小没有多大帮助，但增加了待估参数的个数，这时可能导致 AIC 或 SC 的值增加。

在例 3.2.2 中，EViews 软件的估计结果显示 AIC 值与 SC 值分别为 17.01 与 17.15，

如果去掉其他来源的收入项，只保留工资性收入项，则回归结果显示 AIC 值与 SC 值分别变化成 17.73 与 17.82。从这点看，可以说其他来源的收入项应该作为解释变量包括在模型中。

二、方程总体线性的显著性检验(F 检验)

方程总体线性的显著性检验，旨在对模型中被解释变量与解释变量之间的线性关系在总体上是否显著成立作出推断。

从上面的拟合优度检验中可以看出，拟合优度高，解释变量对被解释变量的解释程度就高，可以推测模型总体线性关系成立；反之，就不成立。但这只是一个模糊的推测，不能给出一个在统计上严格的结论。这就要求进行方程的显著性检验。方程的显著性检验所应用的方法仍是数理统计学中的假设检验。

1. 方程显著性的 F 检验

方程显著性的 F 检验是要检验模型

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_k X_{ik} + \mu_i \ , \qquad i = 1, 2, \cdots, n$$

中参数 $\beta_1, \cdots, \beta_k$ 是否显著不为零。按照假设检验的原理与程序，原假设与备择假设分别为

$$H_0: \beta_1 = 0, \beta_2 = 0, \cdots, \beta_k = 0$$
$$H_1: \beta_j\ (j = 1, 2, \cdots, k)\text{不全为零}$$

F 检验的思想来自于总离差平方和的分解式

$$\text{TSS=ESS+RSS}$$

由于回归平方和 $\text{ESS} = \sum \hat{y}_i^2$ 是解释变量 X 的联合体对被解释变量 Y 的线性作用的结果，考虑比值

$$\frac{\text{ESS}}{\text{RSS}} = \frac{\sum \hat{y}_i^2}{\sum e_i^2}$$

如果这个比值较大，则 X 的联合体对 Y 的解释程度高，可认为总体存在线性关系；反之总体上可能不存在线性关系。因此可通过该比值的大小对总体线性关系进行推断。

根据数理统计学中的知识，在原假设 H_0 成立的条件下，统计量

$$F = \frac{\text{ESS}/k}{\text{RSS}/(n-k-1)} \tag{3.3.7}$$

服从自由度为 $(k, n-k-1)$ 的 F 分布。因此，给定显著性水平 α，查表得到临界值 $F_\alpha(k, n-k-1)$，根据样本求出 F 统计量的数值后，可通过

$$F > F_\alpha(k, n-k-1) \quad (\text{或 } F \leqslant F_\alpha(k, n-k-1))$$

来拒绝(或不拒绝)原假设 H_0，以判定原方程总体上的线性关系是否显著成立。

对于例 3.2.2，计算得到 F =166.55，给定显著性水平 α =0.05，查 F 分布表，得到

临界值 $F_{0.05}(2,28)=3.34$(例子中解释变量数目为 2，样本容量为 31)，显然有

$$F > F_{\alpha}(k,n-k-1)$$

表明模型的线性关系在 5%的显著性水平下显著成立。

2. 关于拟合优度检验与方程总体线性的显著性检验关系的讨论

拟合优度检验和方程总体线性的显著性检验是从不同原理出发的两类检验，前者是从已经得到估计的模型出发，检验它对样本观测值的拟合程度，后者是从样本观测值出发检验模型总体线性关系的显著性。但是二者又是关联的，模型对样本观测值的拟合程度高，模型总体线性关系的显著性就强。那么，找出两个用作检验标准的统计量之间的数量关系，在实际应用中互为验证，是有实际意义的。

用(3.3.3)式和(3.3.7)式分别表示的两个统计量之间存在下列关系：

$$\bar{R}^2 = 1 - \frac{n-1}{n-k-1+kF} \tag{3.3.8}$$

$$F = \frac{R^2/k}{(1-R^2)/(n-k-1)} \tag{3.3.9}$$

由(3.3.9)式可知 F 与 R^2 同向变化：当 $R^2=0$ 时，$F=1$； R^2 越大，F 值也越大；当 $R^2=1$ 时，F 为无穷大。因此，F 检验是所估计回归的总显著性的一个度量，也是 R^2 的一个显著性检验，亦即，检验原假设 $H_0:\beta_1=0,\beta_2=0,\cdots,\beta_k=0$，等价于检验 $R^2=0$ 这一虚拟假设。

对于例 3.2.2，给定显著性水平 $\alpha=0.05$ 时，查 F 分布表，得到临界值 $F_{0.05}(2,28)=3.34$，即是说，只要 F 统计量的值大于 3.34，模型的线性关系在 95%的置信度下是显著成立的。将该数值代入(3.3.8)式，计算得到对应的 $\bar{R}^2$ 为 0.134 9。如果我们首先得到 $\bar{R}^2$ 为 0.134 9，肯定认为该模型质量不高，殊不知它的总体线性关系的置信度达到 95%。这样，**在应用中不必对 $\bar{R}^2$ 过分苛求，重要的是需考察模型的经济关系是否合理。**

三、变量的显著性检验(t 检验)

对于多元线性回归模型，方程的总体线性关系是显著的，并不能说明每个解释变量对被解释变量的影响都是显著的，因此，必须对每个解释变量进行显著性检验，以决定是否作为解释变量被保留在模型中。如果某个变量对被解释变量的影响并不显著，应该将它剔除，以建立更为简单的模型。变量显著性检验中应用最为普遍的是 t 检验，在目前使用的计量经济学软件包中，都有关于 t 统计量的计算结果。

1. t 统计量

在上一节中，已经导出了参数估计量的方差为

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

以 c_{jj} 表示矩阵 $(\boldsymbol{X'X})^{-1}$ 主对角线上的第 j 个元素，于是参数估计量 $\hat{\beta}_j$ 的方差为

$$\mathrm{Var}(\hat{\beta}_j)=\sigma^2 c_{jj}\,,\qquad j=1,2,\cdots,k \tag{3.3.10}$$

其中σ^2为随机干扰项的方差，在实际计算时，用它的估计量$\hat{\sigma}^2$代替。这样，当模型参数估计完成后，就可以计算每个参数估计量的方差值。

因为$\hat{\beta}_j$服从如下正态分布：

$$\hat{\beta}_j \sim N(\beta_j,\sigma^2 c_{jj})$$

因此，可构造如下t统计量：

$$t=\frac{\hat{\beta}_j-\beta_j}{S_{\hat{\beta}_j}}=\frac{\hat{\beta}_j-\beta_j}{\sqrt{c_{jj}\dfrac{\boldsymbol{e'e}}{n-k-1}}}\sim t(n-k-1) \tag{3.3.11}$$

该统计量即为用于变量显著性检验的t统计量。

2. t 检验

在变量显著性检验中，针对某变量X_j ($j=1,2,\cdots,k$)设计的原假设与备择假设为

$$H_0:\beta_j=0$$
$$H_1:\beta_j\neq 0$$

给定一个显著性水平α，得到临界值$t_{\frac{\alpha}{2}}(n-k-1)$，于是可根据

$$|t|>t_{\frac{\alpha}{2}}(n-k-1)\qquad (\text{或}\,|t|\leqslant t_{\frac{\alpha}{2}}(n-k-1))$$

来决定拒绝(或不拒绝)原假设H_0，从而判定对应的解释变量是否应包含在模型中。

需注意的是，**在一元线性回归中，t 检验与 F 检验是一致的。**

一方面，t检验与F检验都是对相同的原假设$H_0:\beta_1=0$进行检验；另一方面，两个统计量之间有如下关系：

$$\begin{aligned}F&=\frac{\sum\hat{y}_i^2}{\sum e_i^2/(n-2)}=\frac{\hat{\beta}_1^2\sum x_i^2}{\sum e_i^2/(n-2)}\\&=\frac{\hat{\beta}_1^2}{\sum e_i^2/(n-2)\sum x_i^2}=\left[\frac{\hat{\beta}_1}{\sqrt{\sum e_i^2/(n-2)\sum x_i^2}}\right]^2\\&=\left(\hat{\beta}_1\bigg/\sqrt{\frac{\sum e_i^2}{n-2}\cdot\frac{1}{\sum x_i^2}}\right)^2=t^2\end{aligned}$$

在例 3.2.2 中，已经由应用软件计算出两个变量X_1，X_2的t值，分别为

$$|t_1|=8.448,\qquad |t_2|=5.772$$

给定显著性水平α=0.05，查t分布表中自由度为 28(在这个例子中$n-k-1$=28)的相应临界值，得到$t_{\frac{\alpha}{2}}(28)$=2.048。可见，两变量的t值都大于该临界值，所以拒绝原假设，即是说，模型中引入的两个解释变量都在 5%的显著性水平下通过了显著性检验。该检验结

果意味着，对中国城镇居民来说，工资性收入及其他收入的变化都会影响现金消费支出的变动。

经常遇到一些实际问题，各个变量的 t 值相差较大，有的在很高的显著性水平下影响显著，有的则在不太高的显著性水平下影响显著，是否都认为通过显著性检验？**没有绝对的显著性水平。关键仍然是考察变量在经济关系上是否对解释变量有影响，显著性检验起到验证的作用；同时还要看显著性水平不太高的变量在模型中及模型应用中的作用，不要简单地剔除变量。**

四、参数的置信区间估计

参数的假设检验用来判别所考察的解释变量是否对被解释变量有显著的线性影响，但并未回答在一次抽样中，所估计的参数值离参数的真实值有多“近”。这需要进一步通过对参数的置信区间的估计来考察。

在变量显著性检验中已经知道

$$t=\frac{\hat{\beta}_j-\beta_j}{S_{\hat{\beta}_j}}\sim t(n-k-1)$$

容易推出：在 $1-\alpha$ 的置信度下 β_j 的置信区间是

$$(\hat{\beta}_j-t_{\frac{\alpha}{2}}\times S_{\hat{\beta}_j},\hat{\beta}_j+t_{\frac{\alpha}{2}}\times S_{\hat{\beta}_j}) \tag{3.3.12}$$

其中，$t_{\frac{\alpha}{2}}$ 为 t 分布表中显著性水平为 α，自由度为 $n-k-1$ 的临界值。

在例 3.2.2 中，如果给定 α=0.05，查表得

$$t_{\frac{\alpha}{2}}(n-k-1)=t_{0.025}\ (28)=2.048$$

从回归计算中得到

$$\hat{\beta}_1=0.486\,5 \qquad S_{\hat{\beta}_1}=0.057\,6$$

$$\hat{\beta}_2=0.601\,7 \qquad S_{\hat{\beta}_2}=0.104\,2$$

根据(3.3.12)式计算得到 β_1 和 β_2 的置信区间分别为(0.368 5, 0.604 5)和(0.388 3, 0.815 1)。显然，参数 β_1 的置信区间比 β_2 要小，这意味着在同样的置信度下，β_1 的估计结果精度更高一些。

同样地，在实际应用中，我们希望置信度越高越好，置信区间越小越好。**如何才能缩小置信区间呢？**从(3.3.12)式中可以看出：(1)**增大样本容量 n**，在同样的置信度下，n 越大，临界值 $t_{\frac{\alpha}{2}}$ 越小，同时，增大样本容量，在一般情况下可使 $S_{\hat{\beta}_j}=\sqrt{c_{jj}\dfrac{\boldsymbol{e'e}}{n-k-1}}$ 减小，因为式中分母的增大是肯定的，分子并不一定增大；(2)更主要的是**提高模型的拟合优度**，以减小残差平方和 $\boldsymbol{e'e}$，设想一种极端情况，如果模型完全拟合样本观测值，残差平方和

为 0，则置信区间的长度也为 0；(3)**提高样本观测值的分散度**，在一般情况下，样本观测值越分散，c_{jj} 越小。

值得注意的是，置信度的高低与置信区间的大小存在此消彼长的关系。置信度越高，在其他情况不变时，临界值 $t_{\frac{\alpha}{2}}$ 越大，置信区间越大。如果要求缩小置信区间，在其他情况不变时，就必须降低对置信度的要求。

§3.4 多元线性回归模型的预测

计量经济学模型的一个重要应用是经济预测。对于模型

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

如果给定样本以外的解释变量的观测值 $\boldsymbol{X}_0 = (1, X_{01}, X_{02}, \cdots, X_{0k})$，可以得到被解释变量的预测值

$$\hat{Y}_0 = \boldsymbol{X}_0\hat{\boldsymbol{\beta}}$$

同样地，**严格地说，这只是被解释变量预测值的估计值，而不是预测值。**原因在于模型中参数估计量的不确定性及随机干扰项的影响两个方面。因此，我们得到的仅是预测值的一个估计值。为了进行科学预测，还需求出预测值的置信区间，包括均值 $\mathrm{E}(Y_0)$ 和点预测值 Y_0 的置信区间。

一、$\mathrm{E}(Y_0)$的置信区间

从参数估计量性质的讨论中易知，在 $\boldsymbol{X} = \boldsymbol{X}_0$ 的条件下

$$\mathrm{E}(\hat{Y}_0) = \mathrm{E}(\boldsymbol{X}_0\hat{\boldsymbol{\beta}}) = \boldsymbol{X}_0\mathrm{E}(\hat{\boldsymbol{\beta}}) = \boldsymbol{X}_0\boldsymbol{\beta} = \mathrm{E}(Y_0)$$

$$\mathrm{Var}(\hat{Y}_0) = \mathrm{E}[(\boldsymbol{X}_0\hat{\boldsymbol{\beta}} - \boldsymbol{X}_0\boldsymbol{\beta})^2] = \mathrm{E}[\boldsymbol{X}_0(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})\boldsymbol{X}_0(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})]$$

由于 $\boldsymbol{X}_0(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ 为标量，因此

$$\begin{aligned}\mathrm{Var}(\hat{Y}_0) &= \mathrm{E}[\boldsymbol{X}_0(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'\boldsymbol{X}_0'] \\ &= \boldsymbol{X}_0\mathrm{E}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'\boldsymbol{X}_0' \\ &= \sigma^2\boldsymbol{X}_0(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}_0'\end{aligned}$$

容易证明

$$\hat{Y}_0 \sim N[\boldsymbol{X}_0\boldsymbol{\beta}, \sigma^2\boldsymbol{X}_0(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}_0']$$

取随机干扰项的样本估计量 $\hat{\sigma}^2$，可构造如下 t 统计量：

$$\frac{\hat{Y}_0 - \mathrm{E}(Y_0)}{\hat{\sigma}\sqrt{\boldsymbol{X}_0(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}_0'}} \sim t(n-k-1)$$

于是，得到$1-\alpha$ 的置信度下 $\mathrm{E}(Y_0)$ 的置信区间：

$$\hat{Y}_0 - t_{\frac{\alpha}{2}} \times \hat{\sigma}\sqrt{\boldsymbol{X}_0(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}_0'} < E(Y_0) < \hat{Y}_0 + t_{\frac{\alpha}{2}} \times \hat{\sigma}\sqrt{\boldsymbol{X}_0(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}_0'} \tag{3.4.1}$$

二、Y_0的置信区间

如果已经知道实际的预测值Y_0，那么预测误差为

$$e_0 = Y_0 - \hat{Y}_0$$

容易证明，在$\boldsymbol{X} = \boldsymbol{X}_0$的条件下

$$\begin{aligned}\mathrm{E}(e_0) &= \mathrm{E}(\boldsymbol{X}_0 \boldsymbol{\beta} + \mu_0 - \boldsymbol{X}_0 \hat{\boldsymbol{\beta}}) \\ &= \mathrm{E}[\mu_0 - \boldsymbol{X}_0(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})] \\ &= \mathrm{E}[\mu_0 - \boldsymbol{X}_0(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\mu}] \\ &= 0\end{aligned}$$

$$\begin{aligned}\mathrm{Var}(e_0) &= \mathrm{E}(e_0^2) \\ &= \mathrm{E}[\mu_0 - \boldsymbol{X}_0(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\mu}]^2 \\ &= \sigma^2[1 + \boldsymbol{X}_0(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}_0']\end{aligned}$$

e_0服从正态分布，即

$$e_0 \sim N\{0, \sigma^2[1 + \boldsymbol{X}_0(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}_0']\}$$

取随机干扰项的样本估计量$\hat{\sigma}^2$，可得e_0的方差的估计量

$$\hat{\sigma}_{e_0}^2 = \hat{\sigma}^2[1 + \boldsymbol{X}_0(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}_0')]$$

构造t统计量

$$t = \frac{\hat{Y}_0 - Y_0}{\hat{\sigma}_{e_0}} \sim t(n-k-1)$$

可得给定$1-\alpha$的置信水平下Y_0的置信区间：

$$\hat{Y}_0 - t_{\frac{\alpha}{2}} \times \hat{\sigma}\sqrt{1 + \boldsymbol{X}_0(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}_0'} < Y_0 < \hat{Y}_0 + t_{\frac{\alpha}{2}} \times \hat{\sigma}\sqrt{1 + \boldsymbol{X}_0(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}_0'} \tag{3.4.2}$$

例 3.2.2 中，假设某城镇居民 2013 年工资性收入为 20 000 元，其他收入为 10 000 元，则该居民 2013 年现金消费支出的预测值为

$\hat{Y}$ =2 599.1+0.486 5×20 000+0.601 7×10 000=18 346.1（元）

而就全国平均情况看，2013 年具有人均工资性收入 20 000 元、其他收入 10 000 元的城镇居民，当年平均的现金消费支出预测值的置信区间可如下求出：

在 95%的置信度下，临界值 $t_{0.025}(28)$=2.048，随机干扰项方差的估计值为 $\hat{\sigma}^2$=1 305 338.6，由于

$$\boldsymbol{X}_0 = (1,\quad 20\,000,\quad 10\,000)$$

$$(\boldsymbol{X}'\boldsymbol{X})^{-1} = \begin{pmatrix} 0.524\,380\,10 & -0.000\,014\,11 & -0.000\,024\,49 \\ -0.000\,014\,11 & 0.000\,000\,00 & 0.000\,000\,00 \\ -0.000\,024\,49 & 0.000\,000\,00 & 0.000\,000\,01 \end{pmatrix}$$

$$\boldsymbol{X}_0(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}_0' = 0.041\,4$$

于是 $E(\hat{Y})$ 的 95%的置信区间为

$$18\,346.1 \pm 2.048 \times \sqrt{1\,305\,338.6} \times \sqrt{0.041\,4}$$

或
$$(17\,870.0,\ 18\,822.2)$$

同样地，就工资性收入 20 000 元、其他收入 10 000 元的某城镇居民来说，也易得其 2013 年消费支出 $\hat{Y}$ 的 95%的置信区间：

$$18\,346.1 \pm 2.048 \times \sqrt{1\,305\,338.6} \times \sqrt{1.041\,4}$$

或
$$(15\,958.3,\ 20\,733.9)$$

需要指出的是，经常听到这样的说法："如果给定解释变量值，根据模型就可以得到被解释变量的预测值为……"，这种说法是不科学的，也是计量经济学模型无法达到的。如果一定要给出一个具体的预测值，那么它的置信度则为 0；如果一定要回答以 100%的置信度处在什么区间中，那么这个区间是$(-\infty, +\infty)$。

§3.5 可化为线性的多元非线性回归模型

迄今为止，我们都假设未知的总体回归线是线性的，拟合优度检验及变量显著性检验也都是对函数形式的线性检验。然而，在实际经济活动中，经济变量的关系是复杂的，直接表现为线性关系的情况并不多见。例如，著名的恩格尔曲线(Engle Curve)表现为幂函数曲线形式，宏观经济学中的菲利普斯曲线(Pillips Cuve)表现为双曲线形式等。但是，它们中的大部分又可以通过一些简单的数学处理，使之化为数学上的线性关系，从而可以运用线性回归的方法建立线性计量经济学模型。下面通过一些常见的例子说明常用的数学处理方法。

一、模型的类型与变换

1. 倒数模型、多项式模型与变量的直接置换法

例如，商品的需求曲线是一种双曲线形式，商品需求量 Q 与商品价格 P 之间的关系表现为非线性关系：

$$\frac{1}{Q} = a + b\frac{1}{P} + \mu \tag{3.5.1}$$

显然，可以用 $Y=\dfrac{1}{Q}$ 和 $X=\dfrac{1}{P}$ 的置换，将方程变成

$$Y = a + bX + \mu \tag{3.5.2}$$

再如，著名的拉弗曲线(Laffer Curve)描述的税收 s 和税率 r 的关系是一种抛物线形式：

$$s = a + br + cr^2 + \mu\,, \qquad c<0 \tag{3.5.3}$$

可以用 $X_1 = r, X_2 = r^2$ 进行置换，将方程变成

$$s = a + bX_1 + cX_2 + \mu\,, \qquad c < 0 \tag{3.5.4}$$

一般地，关于解释变量的非线性问题都可以通过变量置换变成线性问题。

2. 幂函数模型、指数函数模型与函数变换法

如果是关于参数的非线性问题，变量置换方法就无能为力了，函数变换是常用的方法。

例如，著名的 Cobb-Dauglas 生产函数将产出量 Q 与投入要素(K,L)之间的关系描述为幂函数的形式：

$$Q = AK^{\alpha} L^{\beta} \mathrm{e}^{\mu} \tag{3.5.5}$$

方程两边取对数后，即成为一个线性形式：

$$\ln Q = \ln A + \alpha \ln K + \beta \ln L + \mu \tag{3.5.6}$$

再如，生产中成本 C 与产量 Q 的关系呈现指数关系：

$$C = ab^{Q} \mathrm{e}^{\mu} \tag{3.5.7}$$

方程两边取对数后，即成为一个线性形式：

$$\ln C = \ln a + Q \ln b + \mu \tag{3.5.8}$$

3. 复杂函数模型与级数展开法

例如,著名的 CES 生产函数将产出量 Q 与投入要素(K,L)之间的关系描述为如下的复杂函数形式：

$$Q = A(\delta_1 K^{-\rho} + \delta_2 L^{-\rho})^{-\frac{1}{\rho}} \mathrm{e}^{\mu}\,, \qquad (\delta_1 + \delta_2 = 1) \tag{3.5.9}$$

方程两边取对数后，得到

$$\ln Q = \ln A - \frac{1}{\rho} \ln (\delta_1 K^{-\rho} + \delta_2 L^{-\rho}) + \mu \tag{3.5.10}$$

将式中 $\ln(\delta_1 K^{-\rho} + \delta_2 L^{-\rho})$ 在 $\rho = 0$ 处展开泰勒(Taylor)级数，取关于 ρ 的线性项，即得到一个线性近似式。如取 0 阶、1 阶、2 阶项，可得

$$\ln Y = \ln A + \delta_1 \ln K + \delta_2 \ln L - \frac{1}{2} \rho\, \delta_1 \delta_2 \left[\ln\left(\frac{K}{L}\right) \right]^2$$

当然，并非所有的非线性函数形式都可以线性化。无法线性化模型的一般形式为

$$Y = f(X_1, X_2, \cdots, X_k) + \mu \tag{3.5.11}$$

其中 $f(X_1, X_2, \cdots, X_k)$ 为非线性函数。形如

$$Q = AK^{\alpha} L^{\beta} + \mu \tag{3.5.12}$$

的生产函数模型就无法线性化，需要采用非线性方法估计其参数。

二、可化为线性的非线性回归实例

例 3.5.1

建立中国工业生产函数(production function)模型。生产函数是指，在既定的工程技术知识水平下，给定投入之后的最大产出。因此，在仅考虑资本与劳动这两类主要的要素投入时，生产函数就是产出关于资本与劳动投入的函数：

$$Y = f(K, L) \tag{3.5.13}$$

其中，Y 为总产出，K, L 分别为资本、劳动投入要素。生产函数的一个最大特征是能够刻画要素的边际收益递减规律(law of diminishing returns)，即当其他要素投入不变时，随着某一要素投入量的增加，获得的产出增量越来越少。当然，通过生产函数，还可以考察当所有要素投入等比例变化时，产出量是否也会等比例变化，即考察所谓的规模收益(returns to scale)问题。

首先确定生产函数的具体函数形式。根据要素投入的边际收益递减规律，可将生产函数设定为幂函数的形式：

$$Y = AK^{\beta_1}L^{\beta_2} \tag{3.5.14}$$

其中，A 代表了既定的工程技术水平，β_1，β_2 分别为资本与劳动投入的产出弹性。(3.5.5)式就是著名的 Cobb-Dauglas 生产函数，简称 C-D 生产函数。当 $\beta_1 + \beta_2 = 1$ 时，表明规模收益不变；当大于 1 或小于 1 时，表明规模收益递增或递减。显然，如果规模收益不变，则(3.5.14)式可等价地变换为

$$Y / L = A(K / L)^{\beta_1} \tag{3.5.15}$$

为了进行比较，我们将分别估计(3.5.14)式与(3.5.15)式。经对数变换，(3.5.14)式可用如下双对数线性回归模型进行估计：

$$\ln(Y) = \beta_0 + \beta_1 \ln K + \beta_2 \ln L + \mu \tag{3.5.16}$$

式中，$\beta_0 = \ln A$。同样地，（3.5.15）式可用如下线性回归模型进行估计：

$$\ln(Y/L) = \beta_0 + \beta_1 \ln(K / L) + \mu \tag{3.5.17}$$

采用双对数线性回归模型，能够方便地考察生产函数中规模收益的特征。显然，对(3.5.16)式施加 $\beta_1 + \beta_2 = 1$ 的约束，即可化为(3.5.17)式。因此，对(3.5.17)式进行回归，就意味着原生产函数具有规模收益不变的特征。

表 3.5.1 列出了 2010 年中国 39 个制造行业的工业总产值(Y)与固定资产净值(K_1)、流动资产(K_2)以及年均的从业人员(L)。将固定资产净值与流动资产合计成总的资本投入(K)后建立 2010 年中国制造业的生产函数。

表 3.5.1 2010 年中国制造业各行业的总产出及要素投入

编号	行业	*Y* (亿元)	*K* (亿元)	*L* (万人)	*K*1 (亿元)	*K*2 (亿元)
1	煤炭开采和洗选业	22 109.3	21 785.1	527.2	9 186.86	12 598.27
2	石油和天然气开采业	9 917.8	12 904.0	106.1	9 381.72	3 522.31
3	黑色金属矿采选业	5 999.3	4 182.5	67.0	1 630.93	2 551.53
4	有色金属矿采选业	3 799.4	2 317.5	55.4	1 109.92	1 207.54
5	非金属矿采选业	3 093.5	1 424.4	56.5	675.55	748.83
6	其他采矿业	31.3	14.2	0.5	7.64	6.58
7	农副食品加工业	34 928.1	14 373.1	369.0	5 493.82	8 879.32
8	食品制造业	11 350.6	6 113.6	175.9	2 515.71	3 597.87
9	饮料制造业	9 152.6	6 527.0	130.0	2 540.24	3 986.79
10	烟草制品业	5 842.5	4 569.6	21.1	859.08	3 710.47
11	纺织业	28 507.9	16 253.0	647.3	6 276.68	9 976.28
12	纺织服装、鞋、帽制造业	12 331.2	6 044.7	447.0	1 791.52	4 253.18
13	皮革、毛皮、羽毛(绒)及其制品业	7 897.5	3 410.6	276.4	963.81	2 446.80
14	木材加工及木、竹、藤、棕、草制品业	7 393.2	3 037.7	142.3	1 404.12	1 633.60
15	家具制造业	4 414.8	2 261.3	111.7	741.82	1 519.47
16	造纸及纸制品业	10 434.1	7 949.1	157.9	3 797.64	4 151.47
17	印刷业和记录媒介的复制	3 562.9	2 801.6	85.1	1 146.82	1 654.82
18	文教体育用品制造业	3 135.4	1 602.1	128.1	517.56	1 084.54
19	石油加工、炼焦及核燃料加工业	29 238.8	13 360.6	92.2	6 561.08	6 799.50
20	化学原料及化学制品制造业	47 920.0	31 948.6	474.1	14 679.02	17 269.53
21	医药制造业	11 741.3	9 017.0	173.2	3 023.11	5 993.89
22	化学纤维制造业	4 954.0	3 526.1	43.9	1 361.12	2 164.94
23	橡胶制品业	5 906.7	3 595.5	102.9	1 503.38	2 092.10
24	塑料制品业	13 872.2	8 033.2	283.3	2 808.75	5 224.49
25	非金属矿物制品业	32 057.3	21 490.5	544.6	10 382.38	11 108.09
26	黑色金属冶炼及压延加工业	51 833.6	37 101.9	345.6	17 309.25	19 792.66
27	有色金属冶炼及压延加工业	28 119.0	16 992.7	191.6	6 768.77	10 223.92
28	金属制品业	20 134.6	11 477.4	344.6	3 701.16	7 776.22
29	通用设备制造业	35 132.7	24 005.6	539.4	7 200.64	16 804.98
30	专用设备制造业	21 561.8	16 879.4	334.2	4 426.12	12 453.31
31	交通运输设备制造业	55 452.6	40 224.8	573.7	10 364.94	29 859.82
32	电气机械及器材制造业	43 344.4	27 454.8	604.3	6 467.85	20 986.90
33	通信设备、计算机及其他电子设备制造业	54 970.7	34 005.4	772.8	10 437.66	23 567.72
34	仪器仪表及文化、办公用机械制造业	6 399.1	4 565.8	124.9	1 140.44	3 425.38
35	工艺品及其他制造业	5 662.7	2 904.5	140.4	819.12	2 085.35
36	废弃资源和废旧材料回收加工业	2 306.1	829.8	13.9	206.13	623.67
37	电力、热力的生产和供应业	40 550.8	58 989.3	275.6	47 901.41	11 087.90
38	燃气生产和供应业	2 393.4	2 263.8	19.0	1 255.33	1 008.42
39	水的生产和供应业	1 137.1	4 207.7	45.9	2 858.79	1 348.86

资料来源：根据《中国统计年鉴》（2011 年）整理。

按(3.5.16)式回归，Eviews 软件的输出结果如表 3.5.2 所示。(3.5.18)式给出了通常的报告式。

表 3.5.2 中国工业生产函数估计（1）

Dependent Variable: LOG(Y)

Sample: 1 39

Included observations: 39

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.818240	0.403628	4.504739	0.0001
LOG(K)	0.676657	0.081642	8.288068	0.0000
LOG(L)	0.289731	0.085470	3.389857	0.0017

R-squared	0.940802	Mean dependent var	9.198437
Adjusted R-squared	0.937513	S.D. dependent var	1.399184
S.E. of regression	0.349759	Akaike info criterion	0.810660
Sum squared resid	4.403933	Schwarz criterion	0.938626
Log likelihood	-12.80786	Hannan-Quinn criter.	0.856573
F-statistic	286.0643	Durbin-Watson stat	1.251692
Prob(F-statistic)	0.000000		

$$\ln \hat{Y} = 1.818 + 0.677 \ln(K) + 0.290 \ln(L) \tag{3.5.18}$$

$$(4.50) \quad (8.29) \quad (3.39)$$

$$R^2 = 0.940\,8 \quad \bar{R}^2 = 0.937\,5 \quad F = 286.06$$

回归结果表明，在 2010 年，$\ln Y$ 变化的 94.1%可由资本与劳动投入的变化来解释。在 5%的显著性水平下，F 统计量的临界值为 $F_{0.05}(2,36)=5.25$，表明模型的线性关系显著成立。自由度 $n-k-1=36$ 的 t 统计量的临界值为 $t_{0.025}(36)=2.03$，因此 $\ln K$ 与 $\ln L$ 的参数显著地异于零。

从 $\ln K$ 前的参数估计看，2010 年，中国工业总产出关于资本投入的产出弹性为 0.677，表明当其他因素保持不变时，工业的资本投入增加 1%，总产出将增加 0.677%；同样地，$\ln L$ 前的参数估计为 0.290，表明在其他因素保持不变时，劳动力投入每增长 1%，工业总产出将增加 0.290%。可见资本投入的增加对工业总产出的增长起到了更大的作用。

估计的资本投入 K 与劳动投入 L 的产出弹性之和 $\hat{\beta}_1+\hat{\beta}_2=0.967$，很接近于 1，但不为 1。在§3.7 节将进一步从统计学的意义上考察，看它是否为 1，即估计的生产函数是否具有规模收益不变特征。

按(3.5.17)式回归，Eviews 软件的输出结果如表 3.5.3 所示。

表 3.5.3 中国工业生产函数估计（2）

Dependent Variable: LOG(Y/L)
Sample: 1 39
Included observations: 39

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.611681	0.312957	5.149844	0.0000
LOG(K/L)	0.687127	0.080263	8.560961	0.0000
R-squared	0.664521	Mean dependent var		4.248038
Adjusted R-squared	0.655454	S.D. dependent var		0.593165
S.E. of regression	0.348176	Akaike info criterion		0.777704
Sum squared resid	4.485384	Schwarz criterion		0.863014
Log likelihood	-13.16522	Hannan-Quinn criter.		0.808312
F-statistic	73.29005	Durbin-Watson stat		1.315995
Prob(F-statistic)	0.000000			

$$\ln\frac{\hat{Y}}{L}=1.612+0.687\ln\left(\frac{K}{L}\right) \tag{3.5.19}$$

（5.15） (8.56)

R^2=0.664 5 $\bar{R}^2$=0.655 5 F=73.29

从回归结果看，$\ln(K/L)$前的参数在1%的显著性水平下显著地异于零，表明劳均资本增加，会促使劳均工业总产值的增加，劳均产出关于劳均资本投入的弹性值为0.687。

为了与(3.5.18)式作比较，将(3.5.19)式改写为

$$\ln\left(\frac{\hat{Y}}{L}\right)=1.612+0.687(\ln k-\ln L)$$

或
$$\ln\hat{Y}=1.612+0.687\ln k+0.313\ln L \tag{3.5.20}$$

可看出(3.5.20)式与(3.5.18)式比较接近，这意味着(3.5.18)式资本与劳动投入的弹性和可能为1，即所建立的2010年的中国工业生产函数具有规模收益不变的特性。

*三、非线性普通最小二乘法

可化为线性的多元回归模型也可以直接采用非线性普通最小二乘法或者非线性最大似然法估计。下面简单介绍非线性普通最小二乘法的原理。

1. 普通最小二乘原理

将可化为线性的多元回归模型的一般形式表示为

$$Y=f(X_1,X_2,\cdots,X_k,\beta_0,\beta_1,\cdots,\beta_k)+\mu \tag{3.5.21}$$

模型(3.5.21)中，如果随机误差项服从零均值、同方差的正态分布，且不序列相关，

则可以从普通最小二乘原理出发，构造模型的估计方法。为了简单，下面只对有一个参数的非线性模型进行讨论。

对于只有一个参数的非线性模型，在有 n 组观测值的情况下，(3.5.21)式可写成：

$$Y_i = f(X_i, \beta) + \mu_i \qquad i = 1, 2, \cdots, n \tag{3.5.22}$$

如果参数估计值已经得到，则应使得残差平方和最小，即

$$Q(\hat{\beta}) = \sum_{i=1}^{n}(Y_i - f(X_i, \hat{\beta}))^2 \tag{3.5.23}$$

最小。(3.5.23)式取极小值的一阶条件为

$$\frac{\mathrm{d}Q}{\mathrm{d}\hat{\beta}} = -2\sum_{i=1}^{n}(Y_i - f(X_i, \hat{\beta})\left(\frac{-\mathrm{d}f(X_i, \hat{\beta})}{\mathrm{d}\hat{\beta}}\right) = 0$$

即

$$\sum_{i=1}^{n}(Y_i - f(X_i, \hat{\beta})(\frac{\mathrm{d}f(X_i, \hat{\beta})}{\mathrm{d}\hat{\beta}}) = 0 \tag{3.5.24}$$

现在的问题在于如何求解非线性方程(3.5.24)。

对于多参数非线性模型，用矩阵形式表示(3.5.21)式为

$$\boldsymbol{Y} = f\left(\boldsymbol{X}, \boldsymbol{\beta}\right) + \boldsymbol{\mu} \tag{3.5.25}$$

其中各个符号的意义与线性模型相同。向量 $\boldsymbol{\beta}$ 的普通最小平方估计值 $\hat{\boldsymbol{\beta}}$ 应该使得残差平方和

$$Q(\hat{\boldsymbol{\beta}}) = (\boldsymbol{Y} - f(\boldsymbol{X}, \hat{\boldsymbol{\beta}}))'(\boldsymbol{Y} - f(\boldsymbol{X}, \hat{\boldsymbol{\beta}}))$$

达到最小值，即 $\hat{\boldsymbol{\beta}}$ 应该满足下列条件：

$$\frac{\partial}{\partial \hat{\boldsymbol{\beta}}}(Q(\hat{\boldsymbol{\beta}})) = -2\frac{\partial}{\partial \hat{\boldsymbol{\beta}}}(f(\boldsymbol{X}, \hat{\boldsymbol{\beta}})')(\boldsymbol{Y} - f(\boldsymbol{X}, \hat{\boldsymbol{\beta}})) = \boldsymbol{0}$$

即

$$\frac{\partial}{\partial \hat{\boldsymbol{\beta}}}(f(\boldsymbol{X}, \hat{\boldsymbol{\beta}})')(\boldsymbol{Y} - f(\boldsymbol{X}, \hat{\boldsymbol{\beta}})) = \boldsymbol{0} \tag{3.5.26}$$

其中，$\frac{\partial}{\partial \hat{\boldsymbol{\beta}}}(f(\boldsymbol{X}, \hat{\boldsymbol{\beta}})')$ 是一个 $k \times n$ 偏微分矩阵，其第 (j,i) 个元素为 $\frac{\partial}{\partial \hat{\beta}_j}(f(\boldsymbol{X}_i, \hat{\boldsymbol{\beta}})')$。求解(3.5.26)式的原理和方法与求解(3.5.24)式相同，只是数学描述更为复杂。在下面关于求解方法的讨论中，我们只以(3.5.24)式为例，即以单参数非线性模型为例。

2．高斯-牛顿(Gauss-Newton)迭代法

对于非线性方程组(3.5.24)，直接解法已不适用，只能采用迭代解法，高斯-牛顿迭代法就是较为实用的一种。

（1）高斯-牛顿迭代法的原理。迭代是从(3.5.23)式出发的。

根据经验给出参数估计值 $\hat{\beta}$ 的初值 $\hat{\beta}_{(0)}$，将(3.5.23)式中的 $f(X_i,\hat{\beta})$ 在 $\hat{\beta}_{(0)}$ 处展开泰勒级数，取一阶近似值，即有

$$f(X_i,\hat{\beta}) \approx f(X_i,\hat{\beta}_{(0)}) + \frac{\mathrm{d}f(X_i,\hat{\beta})}{\mathrm{d}\hat{\beta}}\bigg|_{\hat{\beta}_{(0)}} (\hat{\beta}-\hat{\beta}_{(0)}) \tag{3.5.27}$$

令

$$Z_i(\hat{\beta}) = \frac{\mathrm{d}f(X_i,\hat{\beta})}{\mathrm{d}\hat{\beta}}$$

于是

$$Z_i(\hat{\beta}_{(0)}) = \frac{\mathrm{d}f(X_i,\hat{\beta})}{\mathrm{d}\hat{\beta}}\bigg|_{\hat{\beta}_{(0)}}$$

代入(3.5.23)式，得

$$\begin{aligned} Q(\hat{\beta}) &= \sum_{i=1}^{n}\left[Y_i - f(X_i,\hat{\beta}_{(0)}) - Z_i(\hat{\beta}_{(0)})(\hat{\beta}-\hat{\beta}_{(0)})\right]^2 \\ &= \sum_{i=1}^{n}\left[Y_i - f(X_i,\hat{\beta}_{(0)}) + Z_i(\hat{\beta}_{(0)})\hat{\beta}_{(0)} - Z_i(\hat{\beta}_{(0)})\hat{\beta}\right]^2 \\ &= \sum_{i=1}^{n}\left[\tilde{Y}_i(\hat{\beta}_{(0)}) - Z_i(\hat{\beta}_{(0)})\hat{\beta}\right]^2 \end{aligned} \tag{3.5.28}$$

其中，$\tilde{Y}_i(\hat{\beta}_{(0)}) = Y_i - f(X_i,\hat{\beta}_{(0)}) + Z_i(\hat{\beta}_{(0)})\hat{\beta}_{(0)}$，可见，一旦给出参数估计值 $\hat{\beta}$ 的初值 $\hat{\beta}_{(0)}$，可以计算出(3.5.28)式中的 $\tilde{Y}_i(\hat{\beta}_{(0)})$ 和 $Z_i(\hat{\beta}_{(0)})$ 的确定的观测值。于是，将(3.5.23)式取极小值就变成了对(3.5.28)式取极小值。

如果有一个线性模型：

$$\tilde{Y}_i(\hat{\beta}_{(0)}) = Z_i(\hat{\beta}_{(0)})\beta + \varepsilon_i \tag{3.5.29}$$

很容易求得其参数 β 的普通最小二乘估计值 $\hat{\beta}_{(1)}$，该估计值使得残差平方和

$$Q(\hat{\beta}_{(1)}) = \sum_{i=1}^{n}(\tilde{Y}_i(\hat{\beta}_{(0)}) - Z_i(\hat{\beta}_{(0)})\hat{\beta}_{(1)})^2 \tag{3.5.30}$$

最小。比较(3.5.28)式与(3.5.30)式后发现，满足使(3.5.30)式达到最小的估计值 $\hat{\beta}_{(1)}$ 同时也是使(3.5.28)式达到最小的 $\hat{\beta}$。换句话说，线性模型(3.5.29)的普通最小二乘估计值就是模型(3.5.22)的一个近似估计值。因为它是在给定参数估计值 $\hat{\beta}$ 的初值 $\hat{\beta}_{(0)}$ 的情况下得到的，将它记为参数估计值 $\hat{\beta}$ 的第一次迭代值 $\hat{\beta}_{(1)}$。它是通过对线性模型(3.5.29)进行普通最小二乘估计而得到的，而线性模型(3.5.29)实际上并不存在，故称之为**线性伪模型**。

将 $\hat{\beta}_{(1)}$ 作为 $\hat{\beta}$ 的新的给定值，将(3.5.23)式中的 $f(X_i,\hat{\beta})$ 在 $\hat{\beta}_{(1)}$ 处展开泰勒级数，取一阶近似值，又可以构造一个新的线性伪模型，对其进行普通最小二乘估计，得到 $\hat{\beta}$ 的第二次迭代值 $\hat{\beta}_{(2)}$ ……如此迭代下去，直到收敛(连续两次得到的参数估计值之差满足确定的标准)。至此完成了非线性模型(3.5.22)的普通最小二乘估计。

（2）高斯-牛顿迭代法的步骤。在对上述采用高斯-牛顿迭代法实现非线性模型参数

最小二乘估计的原理了解之后，可以将高斯-牛顿迭代法的步骤简洁地归纳如下。

第一步：给出参数估计值 $\hat{\beta}$ 的初值 $\hat{\beta}_{(0)}$，将 $f(X_i,\hat{\beta})$ 在 $\hat{\beta}_{(0)}$ 处展开泰勒级数，取一阶近似值；

第二步：计算 $z_i = \left.\dfrac{\mathrm{d}f(X_i,\hat{\beta})}{\mathrm{d}\hat{\beta}}\right|_{\hat{\beta}_{(0)}}$ 和 $\tilde{Y}_i = y_i - f(X_i,\hat{\beta}_{(0)}) + Z_i \cdot \hat{\beta}_{(0)}$ 的样本观测值；

第三步：采用普通最小二乘法估计模型 $\tilde{Y}_i = Z_i\beta + \varepsilon_i$，得到 β 的估计值 $\hat{\beta}_{(1)}$；

第四步：用 $\hat{\beta}_{(1)}$ 代替第一步中的 $\hat{\beta}_{(0)}$，重复这一过程，直至收敛。

3. 牛顿-拉弗森(Newton-Raphson)迭代法

牛顿-拉弗森迭代法作为高斯-牛顿迭代法的改进，当给出参数估计值 $\hat{\beta}$ 的初值 $\hat{\beta}_{(0)}$ 时，将(3.5.23)式在 $\hat{\beta}_{(0)}$ 处展开泰勒级数，取二阶近似值，即

$$Q(\hat{\beta}) \approx Q(\hat{\beta}_{(0)}) + \left.\frac{\mathrm{d}Q(\hat{\beta})}{\mathrm{d}\hat{\beta}}\right|_{\hat{\beta}_{(0)}}(\hat{\beta}-\hat{\beta}_{(0)}) + \frac{1}{2}\left.\frac{\mathrm{d}^2Q(\hat{\beta})}{\mathrm{d}\hat{\beta}^2}\right|_{\hat{\beta}_{(0)}}(\hat{\beta}-\hat{\beta}_{(0)})^2 \tag{3.5.31}$$

这里与高斯-牛顿迭代法有两点不同：一是直接对 $Q(\hat{\beta})$ 展开泰勒级数，而不是对其中的 $f(X_i,\hat{\beta})$ 展开；二是取二阶近似值，而不是取一阶近似值。

使(3.5.31)式达到极小的条件是

$$\frac{\mathrm{d}Q(\hat{\beta})}{\mathrm{d}\hat{\beta}} = 0$$

注意，这里的 $Q(\hat{\beta})$ 已经用(3.5.31)式的近似式代入，而不是(3.5.23)式。再对 $\dfrac{\mathrm{d}Q(\hat{\beta})}{\mathrm{d}\hat{\beta}}$ 取一阶近似，则有

$$\frac{\mathrm{d}Q(\hat{\beta})}{\mathrm{d}\hat{\beta}} \approx \left.\frac{\mathrm{d}Q(\hat{\beta})}{\mathrm{d}\hat{\beta}}\right|_{\hat{\beta}_{(0)}} + \left.\frac{\mathrm{d}^2Q(\hat{\beta})}{\mathrm{d}\hat{\beta}^2}\right|_{\hat{\beta}_{(0)}} \cdot (\hat{\beta}-\hat{\beta}_{(0)}) = 0$$

于是得到

$$\hat{\beta} = \hat{\beta}_{(0)} - \left(\left.\frac{\mathrm{d}^2Q(\hat{\beta})}{\mathrm{d}\hat{\beta}^2}\right|_{\hat{\beta}_{(0)}}\right)^{-1} \cdot \left.\frac{\mathrm{d}Q(\hat{\beta})}{\mathrm{d}\hat{\beta}}\right|_{\hat{\beta}_{(0)}} \tag{3.5.32}$$

由(3.5.32)式得到的 $\hat{\beta}$ 并不是最后的参数估计值，将它作为第一次迭代值 $\hat{\beta}_{(1)}$，再进行上述过程，直至收敛。

无论是高斯-牛顿迭代法还是牛顿-拉弗森迭代法，都存在一个问题，即如何保证迭代所逼近的是总体极小值(即最小值)而不是局部极小值？这就需要选择不同的初值，进行多次迭代求解。

非线性普通最小二乘法早已出现在计量经济学应用软件中，即使是目前使用最为普遍、最为简单的 Eviews 中也有非线性普通最小二乘法估计方法。在选择了该估计方法、给定参数初始值后，只要将非线性方程的形式输入，就可以得到参数的估计量。

例 3.5.2

下面直接用非线性普通最小二乘法进行例 3.5.1 的估计。

首先，估计(3.5.16)式对应的非线性模型

$$Y = AK^{\beta_1}L^{\beta_2}$$

这里需要将等式右边的 A 改写为 exp(β_0)。取各参数 β_0 、β_1、β_2 的初值均为 1，Eviews 软件的估计结果如表 3.5.4 所示。

表 3.5.4 2010 年中国工业生产函数的非线性估计（1）

Dependent Variable: Y

Sample: 1 39

Included observations: 39

Convergence achieved after 10 iterations

Y=EXP(C(1))*K^C(2)*L^C(3)

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	2.197159	0.622341	3.530473	0.0012
C(2)	0.618177	0.062043	9.963646	0.0000
C(3)	0.332525	0.074122	4.486195	0.0001

R-squared	0.912511	Mean dependent var	17912.58
Adjusted R-squared	0.907651	S.D. dependent var	16771.61
S.E. of regression	5096.732	Akaike info criterion	19.98439
Sum squared resid	9.35E+08	Schwarz criterion	20.11236
Log likelihood	–386.6956	Hannan-Quinn criter.	20.03030
Durbin-Watson stat	1.451013		

与原双对数线性模型的估计结果相比，无论是常数项，还是资本投入 K 或劳动投入 L 项，相应参数的估计结果都较为接近，且都通过了 1%显著性水平的检验。

其次，估计（3.5.17）式对应的非线性模型

$$Y/L = A\left(K/L\right)^{\beta_1} \tag{3.5.33}$$

这里仍需要将等式右边的 A 改写为 exp(β_0)。取各参数 β_0 、 β_1 的初值均为 1，Eviews 软件的估计结果如表 3.5.5 所示。可以看出，这里两个参数的估计结果与对应的双对数线性模型的估计结果也较为相似。

表 3.5.5 2010 年中国工业生产函数的非线性估计（2）

Dependent Variable: Y/L
Sample: 1 39
Included observations: 39
Convergence achieved after 4 iterations
Y/L=EXP(C(1))*(K/L)^C(2)

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	1.426539	0.426337	3.346036	0.0019
C(2)	0.747946	0.091945	8.134746	0.0000
R-squared	0.633599	Mean dependent var		84.40605
Adjusted R-squared	0.623696	S.D. dependent var		61.77091
S.E. of regression	37.89252	Akaike info criterion		10.15730
Sum squared resid	53126.18	Schwarz criterion		10.24262
Log likelihood	-196.0674	Hannan-Quinn criter.		10.18791
Durbin-Watson stat	1.799491			

§3.6 含有虚拟变量的多元线性回归模型

到目前为止，我们遇到的多元回归模型，无论是被解释变量还是解释变量，都具有定量的含义，如商品需求量、价格、收入、产量等，变量所赋值的大小都传递了明确的“定量”信息。然而，这些定量的变量还不能刻画经济生活中遇到的全部现象。例如，当考察某一突发事件对经济行为带来的影响时，如何“量化”这一突发事件呢？该突发事件又是如何引入模型并参与到模型的估计的呢？本节将对这一问题进行探讨。

一、含有虚拟变量的模型

在对经济现象的描述中，通常会有一些影响经济变量的因素无法定量度量，如职业、性别对收入的影响，战争、自然灾害对 GDP 的影响，季节对某些产品(如冷饮)销售的影响等。为了能够在模型中反映这些因素的影响，并提高模型的精度，需要将它们“量化”，这种“量化”通常是通过引入“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(dummy variable)，记为 D。例如，反映文化程度的虚拟变量可取为

$$D=\begin{cases}1, & \text{本科学历}\\ 0, & \text{非本科学历}\end{cases}$$

一般地，在虚拟变量的设置中，基础类型和肯定类型取值为 1；比较类型和否定类型取值为 0。同时含有一般解释变量与虚拟变量的模型称为**含有虚拟变量的模型**。一个以

性别为虚拟变量来考察员工薪金的模型如下：

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 D_i + \mu_i \tag{3.6.1}$$

其中，Y_i 为员工的薪金；X_i 为工龄；D_i =1 代表男性，D_i =0 代表女性。

二、虚拟变量的引入

虚拟变量作为解释变量引入模型有两种基本方式：加法方式和乘法方式。

1．加法方式

上述员工薪金模型中性别虚拟变量的引入采取了加法方式，即模型中将虚拟变量以相加的形式引入模型。在该模型中，如果仍假定 $E(\mu_i)=0$，则女职工的平均薪金为

$$E(Y_i \mid X, D = 0) = \beta_0 + \beta_1 X_i \tag{3.6.2}$$

男职工的平均薪金为

$$E(Y_i \mid X, D = 1) = (\beta_0 + \beta_2) + \beta_1 X_i \tag{3.6.3}$$

从几何意义上看(图 3.6.1)，假定 $\beta_2 > 0$，则两个函数有相同的斜率，但有不同的截距。这意味着，男女员工平均薪金对工龄的变化率是一样的，但两者的平均薪金水平相差 β_2。可以通过传统的回归检验，对 β_2 的统计显著性进行检验，以判断男女员工的平均薪金水平是否有显著差异。

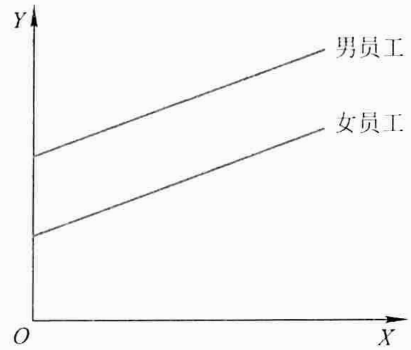

图 3.6.1 男女员工平均薪金示意图

又例如，在截面数据基础上，考虑个人保健支出对个人收入和教育水平的回归。教育水平考虑三个层次：高中以下、高中、大学及其以上。这时需要引入两个虚拟变量：

$$D_1 = \begin{cases} 1, & \text{高中} \\ 0, & \text{其他} \end{cases}$$

$$D_2 = \begin{cases} 1, & \text{大学及其以上} \\ 0, & \text{其他} \end{cases} \tag{3.6.4}$$

模型可设定如下：

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 D_1 + \beta_3 D_2 + \mu_i$$

在 $E(\mu_i | X, D_1, D_2)=0$ 的初始假定下，容易得到高中以下、高中、大学及其以上教育水平个人保健支出的函数：

高中以下：　$\mathrm{E}(Y_i \mid X, D_1=0, D_2=0)=\beta_0+\beta_1 X_i$

高中：　$\mathrm{E}(Y_i \mid X, D_1=1, D_2=0)=(\beta_0+\beta_2)+\beta_1 X_i$

大学及其以上：　$\mathrm{E}(Y_i \mid X, D_1=0, D_2=1)=(\beta_0+\beta_3)+\beta_1 X_i$

假定 $\beta_3>\beta_2$，则其几何意义如图 3.6.2 所示。

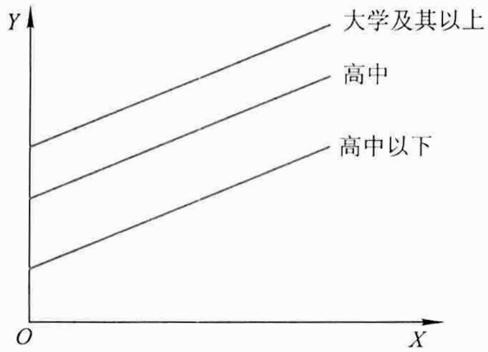

图 3.6.2　不同教育程度人员保健支出示意图

还可将多个虚拟变量引入模型中以考察多种“定性”因素的影响。例如，在上述员工薪金的例子中，再引入学历的虚拟变量 D_2：

$$D_2=\begin{cases}1, & \text{本科及以上学历}\\ 0, & \text{本科以下学历}\end{cases}$$

则员工薪金的回归模型可设计如下：

$$Y_i=\beta_0+\beta_1 X_i+\beta_2 D_1+\beta_3 D_2+\mu_i \tag{3.6.5}$$

于是，不同性别、不同学历员工的平均薪金分别由下面各式给出：

女员工本科以下学历的平均薪金：

$$\mathrm{E}(Y_i \mid X, D_1=0, D_2=0)=\beta_0+\beta_1 X_i$$

男员工本科以下学历的平均薪金：

$$\mathrm{E}(Y_i \mid X, D_1=1, D_2=0)=(\beta_0+\beta_2)+\beta_1 X_i$$

女员工本科以上学历的平均薪金：

$$\mathrm{E}(Y_i \mid X, D_1=0, D_2=1)=(\beta_0+\beta_3)+\beta_1 X_i$$

男员工本科以上学历的平均薪金：

$$\mathrm{E}(Y_i \mid X, D_1=1, D_2=1)=(\beta_0+\beta_2+\beta_3)+\beta_1 X_i$$

2. 乘法方式

加法方式引入虚拟变量，可以考察截距的不同，而在许多情况下，往往是斜率有变化，或斜率、截距同时发生变化。斜率的变化可通过乘法的方式引入虚拟变量来测度。

例如，中国农村居民的边际消费倾向会与城镇居民的边际消费倾向不同吗？这种消费倾向的差异可通过在收入的系数中引入虚拟变量来考察。如，设

$$D_i = \begin{cases} 1, & \text{农村居民} \\ 0, & \text{城镇居民} \end{cases}$$

则全体居民的消费模型可建立如下：

$$C_i = \beta_0 + \beta_1 X_i + \beta_2 D_i X_i + \mu_i \tag{3.6.6}$$

这里，C，X 分别表示居民家庭人均年消费支出与年可支配收入，虚拟变量 D 以与 X 相乘的方式引入模型，从而可用来考察边际消费倾向的差异。在 $E(\mu_i | X,D)=0$ 的假定下，上述模型所表示的函数可化为

农村居民：$\mathrm{E}(C_i \mid X, D=1) = \beta_0 + (\beta_1 + \beta_2) X_i$

城镇居民：$\mathrm{E}(C_i \mid X, D=0) = \beta_0 + \beta_1 X_i$

显然，如果 β_2 显著地异于 0，则可判定农村居民与城镇居民的边际消费倾向有差异。

例 3.6.1

表 3.6.1 中给出了 2013 年中国农村居民家庭与城镇居民家庭人均工资收入、其他收入及生活消费支出的相关数据。可由这组数据来判断 2013 年中国农村居民与城镇居民不同来源的收入对生活消费支出的影响是否有差异。

表 3.6.1 2013 年中国居民人均收入与人均生活消费支出数据 单位：元

	农村居民			城镇居民		
	生活消费	工资收入	其他收入	生活消费	工资收入	其他收入
北 京	13 553.2	12 034.9	6 302.6	26 274.9	30 273.0	15 000.8
天 津	10 155.0	9 091.5	6 749.5	21 711.9	23 231.9	12 423.7
河 北	6 134.1	5 236.7	3 865.2	13 640.6	14 588.4	9 554.4
山 西	5 812.7	4 041.1	3 112.4	13 166.2	16 216.4	7 797.2
内蒙古	7 268.3	1 694.6	6 901.1	19 249.1	18 377.9	8 600.1
辽 宁	7 159.0	4 209.4	6 313.3	18 029.7	15 882.0	12 022.9
吉 林	7 379.7	1 813.2	7 808.0	15 932.3	14 388.3	9 155.9
黑龙江	6 813.6	1 991.4	7 642.8	14 161.7	12 525.8	8 623.4
上 海	14 234.7	12 239.4	7 355.6	28 155.0	33 235.4	15 643.9
江 苏	9 909.8	7 608.5	5 989.2	20 371.5	21 890.0	13 241.0
浙 江	11 760.2	9 204.3	6 901.7	23 257.2	24 453.0	16 788.0
安 徽	5 724.5	3 733.5	4 364.3	16 285.2	15 535.3	9 470.8
福 建	8 151.2	5 193.9	5 990.2	20 092.7	21 443.4	11 939.3
江 西	5 653.6	4 422.1	4 359.4	13 850.5	14 767.5	8 181.9
山 东	7 392.7	5 127.2	5 492.8	17 112.2	21 562.1	9 066.0
河 南	5 627.7	3 581.6	4 893.8	14 822.0	14 704.2	8 982.3
湖 北	6 279.5	3 868.2	4 998.7	15 749.5	15 571.8	9 608.7
湖 南	6 609.5	4 595.6	3 776.6	15 887.1	13 951.4	10 691.6

续表

	农村居民			城镇居民		
	生活消费	工资收入	其他收入	生活消费	工资收入	其他收入
广东	8 343.5	7 072.4	4 596.9	24 133.3	25 286.5	11 217.5
广西	5 205.6	2 712.3	4 078.6	15 417.6	15 647.8	9 381.0
海南	5 465.6	3 001.5	5 341.0	15 593.0	15 773.0	9 146.8
重庆	5 796.4	4 089.2	4 242.8	17 813.9	16 654.7	10 195.7
四川	6 308.5	3 542.8	4 352.6	16 343.5	14 976.0	8 917.9
贵州	4 740.2	2 572.6	2 861.4	13 702.9	13 627.6	7 785.5
云南	4 743.6	1 729.2	4 412.1	15 156.1	15 140.7	9 557.6
西藏	3 574.0	1 475.3	5 102.9	12 231.9	19 604.0	2 956.7
陕西	5 724.2	3 151.2	3 351.4	16 679.7	16 441.0	7 667.8
甘肃	4 849.6	2 203.4	2 904.4	14 020.7	13 329.7	6 819.3
青海	6 060.2	2 347.5	3 848.9	13 539.5	14 015.6	8 115.4
宁夏	6 489.7	2 878.4	4 052.6	15 321.1	15 363.9	8 402.8
新疆	6 119.1	1 311.8	5 984.6	15 206.2	15 585.3	6 802.6

资料来源：《中国统计年鉴》(2014)。

以 Y 为人均生活消费支出，X_1 为人均工资收入，X_2 为人均其他来源的收入。农村与城镇居民人均消费函数可写成：

农村居民： $Y_i = \alpha_0 + \alpha_1 X_{i1} + \alpha_2 X_{i2} + \mu_{1i} \qquad i = 1, 2, \cdots, n_1$

城镇居民： $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \mu_{2i} \qquad i = 1, 2, \cdots, n_2$

我们可能会关心是否截距项不同、斜率项不同或者截距项与斜率项都不同。而当我们关心是否截距项与斜率项都不同时，可以通过引入加法以及乘法形式的虚拟变量来进行考察。将本例中的 n_1 与 n_2 次观察值合并，并用以估计以下回归模型：

$$Y_i = \beta_0 + \delta_0 D_i + \beta_1 X_{i1} + \delta_1 (D_i X_{i1}) + \beta_2 X_{i2} + \delta_2 (D_i X_{i2}) + \mu_i \tag{3.6.7}$$

其中 D_i 为引入的虚拟变量，农村居民取值 1，城镇居民取值 0，则有

$$E(Y_i \mid D = 0, X_1, X_2) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2}$$

$$E(Y_i \mid D = 1, X_1, X_2) = (\beta_0 + \delta_0) + (\beta_1 + \delta_1) X_{i1} + (\beta_2 + \delta_2) X_{i2}$$

分别表示城镇居民消费函数与农村居民消费函数。在显著性检验中，如果 $\delta_0 = 0$ 或者 $\delta_1 = 0$ 或者 $\delta_2 = 0$ 的假设被拒绝，则说明农村居民与城镇居民的消费行为是不同的。

由表 3.6.1 中数据得到的具体回归结果为：

$$\hat{Y}_i = 2\,599.1 - 1\,573.9 D_i + 0.487 X_{i1} + 0.190 D_i X_{i1} + 0.602 X_{i2} - 0.006 D_i X_{i2}$$

$$(3.82) \quad (-1.69) \quad (10.27) \quad (2.40) \quad (7.02) \quad (-0.04)$$

$$R^2 = 0.977\,985 \qquad \bar{R}^2 = 0.976\,0 \qquad F = 497.56$$

在 5%与 10%的显著性水平下，自由度为 56 的 t 统计量的临界值分别为 $t_{0.025}(56) = 2.00$ 与 $t_{0.05}(56) = 1.67$。可见，δ_0 在 10%的显著性水平下异于 0，δ_1 在 5%的显著性水平下

异于 0，但即使在 10%的显著性水平下也无法拒绝 $\delta_2 = 0$ 的假设。从这里可以看出，2013 年中国农村居民的平均消费支出要比城镇居民少 1 573.9 元，同时，在其他条件不变的情况下，农村居民与城镇居民的工资收入都增加 100 元时，农村居民要比城镇居民多支出 19 元用于生活消费；但农村居民与城镇居民在其他来源的收入方面有相同的增加量时，两者增加的消费支出没有显著差异。

三、虚拟变量的设置原则

虚拟变量的个数需按以下原则确定：每一定性变量所需的虚拟变量个数要比该定性变量的类别数少 1，即如果有 m 个定性变量，只在模型中引入 m–1 个虚拟变量。

例如，已知冷饮的销售量 Y 除受 k 种定量变量 X_k 的影响外，还受春、夏、秋、冬四季变化的影响。要考察该四季的影响，只需引入 3 个虚拟变量即可：

$$D_{i1} = \begin{cases} 1 & \text{春季} \\ 0 & \text{其他} \end{cases}$$

$$D_{i2} = \begin{cases} 1 & \text{夏季} \\ 0 & \text{其他} \end{cases}$$

$$D_{i3} = \begin{cases} 1 & \text{秋季} \\ 0 & \text{其他} \end{cases}$$

则冷饮销售量的模型为

$$Y_i = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_k X_{ik} + \alpha_1 D_{i1} + \alpha_2 D_{i2} + \alpha_3 D_{i3} + \mu_i$$

在上述模型中，若再引入第 4 个虚拟变量

$$D_{i4} = \begin{cases} 1 & \text{冬季} \\ 0 & \text{其他} \end{cases}$$

则冷饮销售模型变量为

$$Y_i = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_k X_{ik} + \alpha_1 D_{i1} + \alpha_2 D_{i2} + \alpha_3 D_{i3} + \alpha_4 D_{i4} + \mu_i$$

其矩阵形式为

$$\boldsymbol{Y} = (\boldsymbol{X} \quad \boldsymbol{D}) \begin{pmatrix} \boldsymbol{\beta} \\ \boldsymbol{\alpha} \end{pmatrix} + \boldsymbol{\mu}$$

如果只取 6 个观测值，其中春季与夏季取了两次，秋、冬各取到一次观测值，则其中

$$(\boldsymbol{X} \quad \boldsymbol{D}) = \begin{pmatrix} 1 & X_{11} & \cdots & X_{1k} & 1 & 0 & 0 & 0 \\ 1 & X_{21} & \cdots & X_{2k} & 0 & 1 & 0 & 0 \\ 1 & X_{31} & \cdots & X_{3k} & 0 & 0 & 1 & 0 \\ 1 & X_{41} & \cdots & X_{4k} & 0 & 0 & 0 & 1 \\ 1 & X_{51} & \cdots & X_{5k} & 0 & 1 & 0 & 0 \\ 1 & X_{61} & \cdots & X_{6k} & 1 & 0 & 0 & 0 \end{pmatrix}$$

$$\boldsymbol{\beta}=\begin{pmatrix}\beta_0\\ \beta_1\\ \vdots\\ \beta_k\end{pmatrix},\qquad \boldsymbol{\alpha}=\begin{pmatrix}\alpha_1\\ \alpha_2\\ \alpha_3\\ \alpha_4\end{pmatrix}$$

显然，$(\boldsymbol{X}\quad \boldsymbol{D})$中的第 1 列可表示成后 4 列的线性组合，从而$(\boldsymbol{X}\quad \boldsymbol{D})$不满秩，参数无法唯一求出。这就是所谓的“虚拟变量陷阱”，应该避免这种情况发生。

§ 3.7 受约束回归

在建立回归模型时，有时根据经济理论需要对模型中变量的参数施加一定的约束条件。例如，在估计以幂函数的形式表示的生产函数模型时，有时也施加产出关于资本与劳动的弹性和为 1 的约束。模型施加约束条件后进行回归，称为**受约束回归**（restricted regression），与此对应，不加任何约束的回归称为**无约束回归**（unrestricted regression）。

一、模型参数的线性约束

一般地，估计线性模型时可对模型参数施加若干个线性约束条件。例如，对模型

$$Y=\beta_0+\beta_1X_1+\beta_2X_2+\cdots+\beta_kX_k+\mu \tag{3.7.1}$$

可施加

$$\beta_1+\beta_2=1,\qquad \beta_{k-1}=\beta_k \tag{3.7.2}$$

于是，对(3.7.1)式的回归可转化为对施加上述条件后如下模型的回归：

$$Y=\beta_0+\beta_1X_1+(1-\beta_1)X_2+\cdots+\beta_{k-1}X_{k-1}+\beta_{k-1}X_k+\mu^* \tag{3.7.3}$$

或

$$Y^*=\beta_0+\beta_1X_1^*+\beta_3X_3+\cdots+\beta_{k-2}X_{k-2}+\beta_{k-1}X_{k-1}^*+\mu^* \tag{3.7.4}$$

其中

$$Y^*=Y-X_2,\qquad X_1^*=X_1-X_2,\qquad X_{k-1}^*=X_{k-1}+X_k$$

如果运用普通最小二乘法得到参数的估计结果$\hat\beta_0,\hat\beta_1,\hat\beta_3,\cdots,\hat\beta_{k-1}$，可由上述约束条件得到

$$\hat\beta_2=1-\hat\beta_1,\qquad \hat\beta_k=\hat\beta_{k-1}$$

然而，对所考察的具体问题能否施加约束条件，或者说能否直接对施加约束后的模型进行回归，还需进一步进行相应的检验。常用的检验有 F 检验、χ^2 检验与 t 检验，这里主要介绍 F 检验。

在同一数据样本下，记无约束样本回归模型的矩阵式为

$$\boldsymbol{Y}=\boldsymbol{X}\hat{\boldsymbol{\beta}}+\boldsymbol{e} \tag{3.7.5}$$

记受约束样本回归模型的矩阵式为

$$Y = X\hat{\beta}_* + e_* \tag{3.7.6}$$

于是，受约束样本回归模型的残差项可写为

$$e_* = Y - X\hat{\beta}_* = X\hat{\beta} + e - X\hat{\beta}_* = e - X(\hat{\beta}_* - \hat{\beta})$$

得到受约束样本回归模型的残差平方和 $\mathrm{RSS_R}$ 为

$$e_*'e_* = e'e + (\hat{\beta}_* - \hat{\beta})'X'X(\hat{\beta}_* - \hat{\beta}) \tag{3.7.7}$$

式中第二项为一个非负标量，于是

$$e_*'e_* \geqslant e'e \tag{3.7.8}$$

其中，$e'e$ 为无约束样本回归模型的残差平方和 $\mathrm{RSS_U}$。

在(3.7.5)式与(3.7.6)式两个回归模型中，有着相同的被解释变量 Y 与相同的数据样本，于是 Y 的总离差平方和 TSS 也相同。(3.7.8)式表明受约束样本回归模型的残差平方和不小于无约束样本回归模型的残差平方和，于是，受约束样本回归模型的回归平方和 $\mathrm{ESS_R}$ 不大于无约束样本回归模型的回归平方和 $\mathrm{ESS_U}$。这意味着，通常情况下，对模型施加约束条件会降低模型的解释能力。

但是，如果约束条件为真，则受约束回归模型与无约束回归模型具有相同的解释能力，从而使得 $\mathrm{RSS_U}$ 与 $\mathrm{RSS_R}$ 的差异变小。于是，可用 $\mathrm{RSS_R} - \mathrm{RSS_U}$ 的大小来检验约束条件的真实性。

根据数理统计学的知识，$\dfrac{\mathrm{RSS}}{\sigma^2} \sim \chi^2(n-k-1)$，其中 k 为回归模型中解释变量的个数，σ^2 为回归模型随机干扰项的方差。于是，

$$\frac{\mathrm{RSS_U}}{\sigma^2} \sim \chi^2(n - k_\mathrm{U} - 1) \tag{3.7.9}$$

$$\frac{\mathrm{RSS_R}}{\sigma^2} \sim \chi^2(n - k_\mathrm{R} - 1) \tag{3.7.10}$$

$$\frac{\mathrm{RSS_R} - \mathrm{RSS_U}}{\sigma^2} \sim \chi^2(k_\mathrm{U} - k_\mathrm{R}) \tag{3.7.11}$$

其中，k_U，k_R 分别为无约束与受约束回归模型的解释变量的个数(不包括常数项)。于是可通过计算(3.7.11)式的 χ^2 统计量来进行相应的 χ^2 检验。当然，由于随机干扰项的方差 σ^2 往往未知，检验时需用它的估计量 $\hat{\sigma}^2$ 替代。

当约束条件为真时，由(3.7.9)式与(3.7.11)式可进一步得到如下的 F 统计量：

$$F = \frac{(\mathrm{RSS_R} - \mathrm{RSS_U})/(k_\mathrm{U} - k_\mathrm{R})}{\mathrm{RSS_U}/(n - k_\mathrm{U} - 1)} \sim F(k_\mathrm{U} - k_\mathrm{R}, n - k_\mathrm{U} - 1) \tag{3.7.12}$$

F 统计量无需估计随机干扰项的方差 σ^2。根据该统计量，如果约束条件无效，则 $\mathrm{RSS_R}$ 与 $\mathrm{RSS_U}$ 的差异较大，计算的 F 值也较大。于是，可用计算的 F 统计量的值与所给定的显著性水平下的临界值作比较，来对约束条件的真实性进行检验。需注意的是，$k_\mathrm{U} - k_\mathrm{R}$ 恰为约束条件的个数。

例 3.7.1

在§3.5 的例 3.5.1 中，讨论了 2010 年中国工业生产函数。如果规模收益不变，则（3.5.16）式可等价地变换为（3.5.17）式，前者为无约束模型，后者为受约束模型。从回归结果看，无约束回归模型的残差平方和 $\mathrm{RSS_U}$ =4.403 9，受约束回归模型的残差平方和 $\mathrm{RSS_R}$ =4.485 4，样本容量 n=39，无约束回归模型变量个数 k_{U} =2，约束条件个数 $k_{\mathrm{U}}-k_{\mathrm{R}}$ =2–1=1。于是

$$F=\frac{(4.485\,4-4.403\,9)/1}{4.403\,9/36}=0.666\,2$$

在 5%的显著性水平下，自由度为（1，36）的 F 统计量的临界值为 $F_{0.05}$=4.11。计算的 F 值小于临界值，不能拒绝 2010 年中国工业生产函数具有规模收益不变这一假设。

需要指出的是，这里介绍的 F 检验适合所有关于参数线性约束的检验，§3.3 中对回归模型总体的线性检验，可以归结到这里的 F 检验上来。例如，对线性模型

$$Y=\beta_0+\beta_1X_1+\beta_2X_2+\cdots+\beta_kX_k+\mu$$

的总体线性检验，就是要检验联合假设：

$$H_0:\beta_j=0\,,\qquad j=1,2,\cdots,k$$

因此，受约束回归模型为

$$Y=\beta_0+\mu_*$$

由(3.7.12)式，检验的 F 统计量为

$$\begin{aligned}F&=\frac{(\mathrm{RSS_R}-\mathrm{RSS_U})/(k_{\mathrm{U}}-k_{\mathrm{R}})}{\mathrm{RSS_U}/(n-k_{\mathrm{U}}-1)}\\&=\frac{(\mathrm{TSS}-\mathrm{ESS_R}-\mathrm{RSS_U})/k}{\mathrm{RSS_U}/(n-k-1)}\\&=\frac{(\mathrm{TSS}-\mathrm{RSS_U})/k}{\mathrm{RSS_U}/(n-k-1)}\\&=\frac{\mathrm{ESS_U}/k}{\mathrm{RSS_U}/(n-k-1)}\end{aligned}$$

这里，运用了受约束回归模型的回归平方和 $\mathrm{ESS_R}=0$。该 F 统计量的另一个等价式为

$$F=\frac{R_{\mathrm{U}}^2/k}{(1-R_{\mathrm{U}}^2)/(n-k-1)}\tag{3.7.13}$$

其中，R_{U}^2 为无约束模型的可决系数。

二、对回归模型增加或减少解释变量

在建立回归模型时，一个重要的问题是如何判断增加重要的解释变量或去掉不必要

的解释变量。t 检验可对单个变量的取舍进行判断，而上面介绍的 F 检验除能对单个变量的取舍进行判断外，还可对多个变量的同时取舍进行判断。

考虑如下两个回归模型：

$$Y=\beta_0+\beta_1X_1+\cdots+\beta_kX_k+\mu \tag{3.7.14}$$

$$Y=\beta_0+\beta_1X_1+\cdots+\beta_kX_k+\beta_{k+1}X_{k+1}+\cdots+\beta_{k+q}X_{k+q}+\mu \tag{3.7.15}$$

(3.7.14)式可以看成是(3.7.15)式施加如下约束条件的受约束回归：

$$H_0：\ \beta_{k+1}=\beta_{k+2}=\cdots=\beta_{k+q}=0$$

相应的 F 统计量为

$$\begin{aligned}F&=\frac{(\mathrm{RSS_R}-\mathrm{RSS_U})/q}{\mathrm{RSS_U}/[n-(k+q+1)]}\\&=\frac{(\mathrm{ESS_U}-\mathrm{ESS_R})/q}{\mathrm{RSS_U}/[n-(k+q+1)]}\sim F[q,n-(k+q+1)]\end{aligned} \tag{3.7.16}$$

如果约束条件为真，即额外的变量 $X_{k+1},\cdots,X_{k+q}$ 对 Y 没有解释能力，则 F 统计量较小；否则，约束条件为假，意味着额外的变量 $X_{k+1},\cdots,X_{k+q}$ 对 Y 有较强的解释能力，则 F 统计量较大。因此，可通过给定某一显著性水平下 F 分布的临界值与 F 统计量的计算值的比较，来判断额外变量 $X_{k+1},\cdots,X_{k+q}$ 是否应包括在模型中。

由(3.7.16)式可得到 F 统计量的另一个等价的式子（留作练习）：

$$F=\frac{(R_\mathrm{U}^2-R_\mathrm{R}^2)/q}{(1-R_\mathrm{U}^2)/[n-(k+q+1)]} \tag{3.7.17}$$

其中 R_U^2，R_R^2 分别为无约束回归与受约束回归方程的可决系数，表明通过变量增减前后回归方程的可决系数 R^2 是否有“足够大”的变化来判断变量的增减与否。

例 3.7.2

居民的食品消费中，大致有粮食类、油脂类、蔬菜类以及蛋白类的食物。在蛋白类的食品中主要有肉禽类、蛋类及水产类食品。表 3.7.1 中列出了 2012 年中国 31 个省、市、自治区农村居民人均消费蛋类食物的数量以及各类主要食品的价格指数，同时也列出了农村居民人均生活消费支出及农村居民的居民消费价格指数（北京、天津、上海、重庆四个直辖市没有农村居民消费价格指数，用城市居民消费价格指数替代）。为了考察农村居民的生活消费支出（它与纯收入有关）以及各类食物价格变动与蛋类食品的人均消费关系，可建立如下回归模型：

$$\ln Q=\beta_0+\beta_1\ln(X/P_0)+\beta_2(\mathrm{P}_1/P)+\beta_3(P_2/P)+\beta_4P_{01}+\beta_5P_{02}+\beta_6P_{03}+\mu$$

其中，Q 为蛋类食品的人均消费量，X 为人均生活消费支出，P_0 为居民消费价格指数，P 为蛋类食品的居民消费价格指数，P_1、P_2 分别为肉禽类及水产类食品的居民

消费价格指数，P_{01}、P_{02}、P_{03} 分别为粮食类、油脂类及蔬菜类食品的居民消费价格指数。由于三类蛋白类食品有一定的替代性，因此我们需要考察蛋类食品价格与肉禽类、水产类价格的相对变化对蛋类食品消费需求的影响；而粮食类、油脂类、蔬菜类食物与蛋白类食物的替代性不是太强，我们仅考察它们价格的变化对蛋类食物消费可能产生的影响。

表 3.7.1　2012 年中国农村居民对蛋类食物的消费及相关食物的价格指数

	蛋类消费量 Q（千克）	各类食品的消费价格指数（上年=100）						居民消费价格指数 P_0（上年=100）	人均消费支出 X（元）
		蛋类 P	肉禽类 P_1	水产类 P_2	粮食类 P_{01}	油脂类 P_{02}	蔬菜类 P_{03}		
北　京	11.05	96.9	106.7	104.8	102.6	104.5	112.0	103.3	11 878.92
天　津	12.84	101.7	105.7	106.7	102.4	103.7	119.6	102.7	8 336.55
河　北	10.42	96.4	101.1	104.8	102.9	106.3	114.9	102.5	5 364.14
山　西	7.82	96.2	101.4	107.4	103.0	105.2	114.2	102.6	5 566.19
内蒙古	6.45	98.1	105.3	107.7	105.7	105.3	112.3	102.5	6 381.97
辽　宁	8.48	96.2	102.6	107.3	103.6	105.0	117.5	102.5	5 998.39
吉　林	7.90	94.6	103.7	108.5	104.2	105.7	110.5	102.4	6 186.17
黑龙江	6.33	98.3	105.4	104.8	104.6	102.6	115.3	102.9	5 718.05
上　海	8.92	98.2	105.1	105.8	102.9	103.8	111.1	102.8	11 971.50
江　苏	6.96	97.0	102.5	108.4	102.3	104.2	109.0	102.6	9 138.18
浙　江	5.56	97.6	100.9	108.8	103.7	103.7	115.2	102.3	10 652.73
安　徽	7.23	94.3	98.7	110.8	104.2	105.8	113.3	102.4	5 555.99
福　建	5.32	96.8	102.0	107.8	103.0	105.4	116.5	102.4	7 401.92
江　西	4.22	96.9	98.9	112.6	103.8	104.2	118.2	103.0	5 129.47
山　东	12.32	95.9	101.6	108.8	102.5	107.5	111.2	102.0	6 775.95
河　南	9.06	94.4	99.4	108.9	104.1	105.0	113.2	102.4	5 032.14
湖　北	5.02	98.6	101.7	111.1	105.3	105.2	113.2	103.0	5 726.73
湖　南	4.92	100.1	98.5	110.9	105.3	102.5	110.8	101.6	5 870.12
广　东	3.39	98.2	104.4	107.3	105.0	106.0	114.9	102.9	7 458.56
广　西	2.22	97.3	103.0	104.9	103.8	108.2	116.7	103.3	4 933.58
海　南	2.43	102.7	103.8	102.2	104.1	106.2	115.6	103.2	4 776.30
重　庆	5.18	100.6	99.1	106.7	107.7	106.0	112.3	102.6	5 018.64
四　川	4.87	97.7	99.9	111.5	104.9	105.2	118.1	102.0	5 366.71
贵　州	2.35	95.7	101.3	107.6	104.5	104.4	109.0	102.8	3 901.71
云　南	2.82	100.1	103.1	104.9	103.5	102.9	117.8	102.3	4 561.33
西　藏	0.56	102.4	108.9	102.8	103.0	105.5	114.6	103.4	2 967.56
陕　西	3.91	97.6	101.5	110.4	103.3	105.9	111.7	103.1	5 114.68
甘　肃	3.93	97.4	104.2	105.2	102.3	104.5	108.5	103.1	4 146.24
青　海	1.58	99.2	107.6	109.6	102.8	105.6	112.8	103.1	5 338.91
宁　夏	3.40	97.7	104.8	107.2	101.0	103.0	108.7	101.7	5 351.36
新　疆	3.62	102.1	105.9	105.2	107.3	105.3	117.6	104.7	5 301.25

资料来源：《中国统计年鉴》(2013)。

上述无约束模型的 OLS 估计结果如下：

$$\ln\hat{Q} = -11.15 + 1.328\ln(X/P_0) - 1.453(\mathrm{P}_1/P) + 5.127(P_2/P)$$
$$(-0.88) \quad (4.08) \qquad\qquad (-0.35) \quad (2.25)$$
$$+0.015P_{01} + 0.005P_{02} + 0.010P_{03}$$
$$(0.20) \qquad (0.07) \qquad (0.31)$$
$$R^2 = 0.527\,3 \qquad \bar{R}^2 = 0.409\,1 \qquad F = 4.46 \qquad \mathrm{RSS}_{\mathrm{U}}^2 = 6.318\,7$$

运用 t 检验，在 5%的显著性水平下，容易得知除真实的居民生活消费支出项以及水产品与蛋类产品的相对价格项显著异于零外，其他各项均没有通过显著性检验。表明在其他条件不变的情况下，农村居民真实的生活消费支出的增加会带动对蛋类食品的消费增长，而当水产品价格比蛋类产品价格上升更快时，会刺激农村居民消费更多的蛋类产品，意味着在中国农村，水产品与蛋类产品有着一定的替代性，但肉禽类产品与蛋类产品没有明显的替代效应。同样地，粮食类、油脂类、蔬菜类食物价格的上升也不会对蛋类产品的消费有显著的影响。

可以通过 F 检验来考察是否需要将粮食类、油脂类、蔬菜类食物的价格引入到农村居民对蛋类食品的消费需求函数中来，即针对上述无约束模型，检验联合假设

$$H_0:\ \beta_4 = 0,\ \beta_5 = 0,\ \beta_6 = 0$$

该联合假设成立时的受约束模型估计如下

$$\ln\hat{Q} = -7.01 + 1.323\ln(X/P_0) - 2.120(\mathrm{P}_1/P) + 4.973(P_2/P)$$
$$(-1.79) \quad (4.45) \qquad\qquad (-0.64) \quad (2.38)$$
$$R^2 = 0.524\,1 \qquad \bar{R}^2 = 0.471\,2 \qquad F = 9.91 \qquad \mathrm{RSS}_{\mathrm{R}}^2 = 6.361\,7$$

相应的 F 检验为

$$F = \frac{(6.361\,7 - 6.318\,7)/3}{6.318\,7/(31-6)} = 0.056\,7$$

该值小于 5%显著性水平下相应的临界值 $F_{0.05}(3,25)=2.99$，因此，不拒绝上述联合假设，可将 P_{01}、P_{02}、P_{03} 从原模型中去掉。当然，也可以将 $\beta_2 = 0$ 加入到上面的联合假设 H_0 中，进行受约束、无约束的 F 检验，以考察对应的变量项是否应从原模型中去掉。

三、检验不同组之间回归函数的差异

运用受约束、无约束模型的 F 检验，可以考察不同组之间的回归函数是否有差异。假设需要建立的模型为

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + \mu$$

对两个不同的组各自对应的样本（$1,2,\cdots,n_1$）与（$1,2,\cdots,n_2$），相应的模型分别为

$$Y=\beta_0+\beta_1X_1+\cdots+\beta_kX_k+\mu_1 \tag{3.7.18}$$

$$Y=\alpha_0+\alpha_1X_1+\cdots+\alpha_kX_k+\mu_2 \tag{3.7.19}$$

合并两个样本的大样本为$(1,2,\cdots,n_1,n_1+1,\cdots,n_1+n_2)$，则可写出如下无约束回归模型：

$$\begin{pmatrix}\boldsymbol{Y}_1\\\boldsymbol{Y}_2\end{pmatrix}=\begin{pmatrix}\boldsymbol{X}_1&\boldsymbol{0}\\\boldsymbol{0}&\boldsymbol{X}_2\end{pmatrix}\begin{pmatrix}\boldsymbol{\beta}\\\boldsymbol{\alpha}\end{pmatrix}+\begin{pmatrix}\boldsymbol{\mu}_1\\\boldsymbol{\mu}_2\end{pmatrix} \tag{3.7.20}$$

其中，$\boldsymbol{\beta}$，$\boldsymbol{\alpha}$ 分别是两样本组对应模型中的参数列向量，$\boldsymbol{Y}_i(i=1,2)$ 是对应模型的被解释变量以其样本为元素的列向量，$\boldsymbol{X}_i(i=1,2)$ 是对应模型的解释变量矩阵。

如果 $\boldsymbol{\beta}=\boldsymbol{\alpha}$，表示两个组别在回归函数上无差别，因此可针对如下假设进行检验：

$$\boldsymbol{H}_0：\ \boldsymbol{\beta}=\boldsymbol{\alpha} \tag{3.7.21}$$

(3.7.20)式施加该约束条件后变换为受约束回归模型

$$\begin{pmatrix}\boldsymbol{Y}_1\\\boldsymbol{Y}_2\end{pmatrix}=\begin{pmatrix}\boldsymbol{X}_1\\\boldsymbol{X}_2\end{pmatrix}\boldsymbol{\beta}+\begin{pmatrix}\boldsymbol{\mu}_1\\\boldsymbol{\mu}_2\end{pmatrix} \tag{3.7.22}$$

因此，仍可用如下 F 统计量进行检验：

$$F=\frac{(\mathrm{RSS_R}-\mathrm{RSS_U})/(k+1)}{\mathrm{RSS_U}/[n_1+n_2-2(k+1)]}\sim F[k+1,n_1+n_2-2(k+1)] \tag{3.7.23}$$

其中，$\mathrm{RSS_U}$ 与 $\mathrm{RSS_R}$ 分别为对应于无约束模型(3.7.20)式与受约束模型(3.7.22)式的残差平方和。

记 RSS_1 与 RSS_2 为两个组别对应的回归模型(3.7.18)式与(3.7.19)式在各自的样本下分别回归后所得的残差平方和，容易验证（留作练习）

$$\mathrm{RSS_U}=\mathrm{RSS}_1+\mathrm{RSS}_2$$

于是，F 统计量可写为

$$F=\frac{[\mathrm{RSS_R}-(\mathrm{RSS}_1+\mathrm{RSS}_2)]/(k+1)}{(\mathrm{RSS}_1+\mathrm{RSS}_2)/[n_1+n_2-2(k+1)]}\sim F[k+1,n_1+n_2-2(k+1)] \tag{3.7.24}$$

因此，检验两个组别在回归函数上有无差异的步骤为：首先，分别对两个组各自的样本运用（3.7.1）式进行回归，得到相应的残差平方和 RSS_1 与 RSS_2；然后，将组并为一个大样本后运用（3.7.1）式进行回归，得到大样本下的残差平方和 RSS_R；最后，通过（3.7.24）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，认为两组的回归函数有差异，即它们的参数不完全相同。该检验方法也被称为**邹氏参数稳定性检验**（Chow test for parameter stability）。

例 3.7.3

在§3.6 节中国农村居民与城镇居民人均消费支出的实例中，建立了可以考察两者消费行为差异的包含虚拟变量的模型：

$$Y_i = \beta_0 + \delta_0 D_i + \beta_1 X_{i1} + \delta_1(D_i X_{i1}) + \beta_2 X_{i2} + \delta_2(D_i X_{i2}) + \mu_i \qquad (3.7.25)$$

其中，Y 为人均生活消费支出，X_1 为人均工资收入，X_2 为人均其他来源的收入，D_i 为引入的虚拟变量，农村居民取值 1，城镇居民取值 0。可以通过包含虚拟变量项参数的显著性来考察农村居民与城镇居民人均消费行为是否有差异。这也可以通过邹氏参数稳定性检验来完成。

记农村或城镇居民的人均生活消费支出函数为

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \mu_i$$

对农村居民，有如下回归结果

$$\hat{Y}_i = 1\,025.3 + 0.676X_{i1} + 0.554X_{i2}$$

（2.22）(14.79)　　(6.38)

R^2=0.930 5　　$\bar{R}^2$=0.925 6　　F=187.50　　RSS_1=12 891 881

对城镇居民，有如下回归结果

$$\hat{Y}_i = 2\,599.1 + 0.487X_{i1} + 0.602X_{i2}$$

（3.14）　(8.45)　　(5.77)

R^2=0.922 5　　$\bar{R}^2$=0.916 9　　F=166.55　　RSS_2=36 549 482

将农村居民与城镇居民对应的样本合成一个大样本后，有如下回归结果

$$\hat{Y}_i = 1\,765.4 + 0.555X_{i1} + 0.560X_{i2}$$

（5.23）　(18.79)　　(7.74)

R^2=0.975 1　　$\bar{R}^2$=0.974 3　　F=1 154.99　　RSS_R=55 933 406

于是，可进行如下 F 检验

$$F = \frac{[55\,933\,406 - (12\,891\,881 + 36\,549\,482)]/3}{(12\,891\,881 + 36\,549\,482)/(62-6)} = 2.451$$

5%与 10%显著性水平下相应的临界值分别为 $F_{0.05}(3, 56)$=2.77、$F_{0.10}(3, 56)$=2.18。因此，在 5%的显著性水平下不拒绝中国农村居民与城镇居民在生活消费支出上无差异的假设，但在 10%的显著性水平下拒绝该假设。事实上，从例 3.6.1 包含虚拟变量的模型估计中，我们已经得知农村居民组与城镇居民组在截距项上是有差异的，在工资收入项上对应的参数也是有差异的，但其他来源的收入项所对应的参数无显著差异。因此，这里两组间回归函数差异的检验可同例 3.6.1 那样通过引入虚拟变量的方式来完成，但显然，虚拟变量的方法更直观、简单，而且虚拟变量方法还可能更细致地检验出差异是在截距项上，还是某个斜率项上。

*四、非线性约束

估计线性模型时也可对模型参数施加非线性约束。例如，对模型

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + \mu$$

施加非线性约束 $\beta_1\beta_2 = 1$。然而，这时再无法找到一个类似于(3.7.3)式或(3.7.4)式那样简单的受约束回归模型，只能得到如下受约束回归模型：

$$Y = \beta_0 + \beta_1 X_1 + \frac{1}{\beta_1} X_2 + \cdots + \beta_k X_k + \mu^* \tag{3.7.26}$$

该模型无法运用通常的普通最小二乘法进行估计，必须采用**非线性最小二乘法**(nonlinear least squares)进行估计。

非线性最小二乘法的主要问题在于其参数估计量并不必然具有期望的小样本性质。然而，从最小二乘估计原理与最大似然估计原理来看，受约束的残差平方和最小化问题，等价于受约束的最大似然函数最大化的问题，且最小二乘估计量就是最大似然估计量，因此，估计量具有一致性与渐近有效性等大样本性质。

有鉴于此，非线性约束检验是建立在最大似然原理基础上的。有三个著名的检验：**最大似然比检验**(Likelihood Ratio Test, LR)、**沃尔德检验**(Wald Test, WD)与**拉格朗日乘数检验**(Lagrange Multiplier Test, LM)。它们的共同特点是：在大样本下，以共同的 χ^2 检验为基础，而自由度就是约束条件的个数。

1. 最大似然比检验(LR)

最大似然比检验仍需估计无约束回归模型与受约束回归模型，但运用的估计方法是最大似然法，检验的是两个似然函数的值的差异是否“足够”大。

记 $L(\boldsymbol{\beta},\sigma^2)$ 为似然函数，对给定的样本数据，无约束回归就是求一组参数 $\hat{\boldsymbol{\beta}}$ 与 $\hat{\sigma}^2$，以使似然函数 $L(\hat{\boldsymbol{\beta}},\hat{\sigma}^2)$ 的值最大。如果给定约束条件 $g(\boldsymbol{\beta}) = \mathbf{0}$，有约束的回归则是在该约束条件下，求另一组参数 $\tilde{\boldsymbol{\beta}}$ 与 $\tilde{\sigma}^2$，以使似然函数 $L(\tilde{\boldsymbol{\beta}},\tilde{\sigma}^2)$ 的值最大。根据拉格朗日乘数法，就是求如下函数的极值问题：

$$\Phi = L(\boldsymbol{\beta},\sigma^2) - \boldsymbol{\lambda}' g(\boldsymbol{\beta}) \tag{3.7.27}$$

其中，$g(\boldsymbol{\beta})$ 是以各约束条件为元素的列向量，$\boldsymbol{\lambda}'$ 则是以相应拉格朗日乘数为元素的行向量。

同样地，受约束的函数值不会超过无约束的函数值，但如果给出的约束条件为真，则两个函数值就非常“接近”。由此，定义**似然比**(likelihood ratio)为

$$\frac{L(\tilde{\boldsymbol{\beta}},\ \tilde{\sigma}^2)}{L(\hat{\boldsymbol{\beta}},\ \hat{\sigma}^2)}$$

如果该比值很小，说明受约束似然函数值与无约束似然函数值差距较大，则应拒绝约束条件为真的假设；如果该比值接近于 1，则受约束似然函数值与无约束似然函数值很接

近，应接受约束条件为真的假设。

在进行检验时，由于在大样本下

$$\mathrm{LR} = -2[\ln\ L(\tilde{\boldsymbol{\beta}},\tilde{\sigma}^2) - \ln\ L(\hat{\boldsymbol{\beta}},\hat{\sigma}^2)] \sim \chi^2(h) \tag{3.7.28}$$

其中 h 是约束条件的个数，则可以在给定的显著性水平下，通过最大似然比检验的计算值与相应的 χ^2 分布的临界值的比较，来判断是否拒绝给定的约束条件的假设。

在例 3.5.1 中，受约束回归模型的最大似然值的对数值为 $\ln\ L(\tilde{\boldsymbol{\beta}},\tilde{\sigma}^2) = -13.165$，无约束回归模型的最大似然值的对数值为 $\ln\ L(\hat{\boldsymbol{\beta}},\hat{\sigma}^2) = -12.808$，于是

$$\mathrm{LR} = -2(-13.165 - (-12.808)) = -0.714$$

该值小于 5%显著性水平下自由度为 1 的 χ^2 分布的临界值 $\chi^2_{0.05}(1)$ =3.84，不拒绝原约束的假设，表明 2010 年的中国工业生产函数具有规模收益不变的特征。

2. 沃尔德检验(WD)

最大似然比检验不仅要估计无约束模型，还要估计受约束模型。而在沃尔德检验中，只需估计无约束模型。例如，对模型

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + \mu$$

要检验 $\beta_1 + \beta_2 = 1$ 的约束，只需对该模型进行回归，并判断 $\hat{\beta}_1 + \hat{\beta}_2$ 与 1 的差距是否足够大。在所有基本假设都成立的条件下，容易证明

$$\hat{\beta}_1 + \hat{\beta}_2 \sim N(\beta_1 + \beta_2, \sigma^2_{\hat{\beta}_1 + \hat{\beta}_2})$$

因此，在 $\beta_1 + \beta_2 = 1$ 的约束条件下

$$z = \frac{\hat{\beta}_1 + \hat{\beta}_2 - 1}{\sigma_{\hat{\beta}_1 + \hat{\beta}_2}} \sim N(0,1) \tag{3.7.29}$$

由于 $\sigma^2_{\hat{\beta}_1 + \hat{\beta}_2}$ 是 $\hat{\beta}_1 + \hat{\beta}_2$ 的方差，其值与随机干扰项 μ 的方差有关：

$$\sigma^2_{\hat{\beta}_1 + \hat{\beta}_2} = \mathrm{Var}(\hat{\beta}_1) + \mathrm{Var}(\hat{\beta}_2) + 2\mathrm{Cov}(\hat{\beta}_1, \hat{\beta}_2) = \sigma^2 f(\boldsymbol{X})$$

以 σ^2 的极大似然估计量 $\tilde{\sigma}^2 = \dfrac{\boldsymbol{e}'\boldsymbol{e}}{n}$ 代入，并记 $\tilde{\sigma}^2_{\hat{\beta}_1 + \hat{\beta}_2} = \tilde{\sigma}^2 f(\boldsymbol{X})$，于是，可建立服从自由度为 1 的 χ^2 分布的沃尔德统计量

$$W = \frac{(\hat{\beta}_1 + \hat{\beta}_2 - 1)^2}{\tilde{\sigma}^2_{\hat{\beta}_1 + \hat{\beta}_2}} \sim \chi^2(1) \tag{3.7.30}$$

如果有 h 个约束条件，则可得到 h 个类似于(3.7.29)式的统计量 $z_1, z_2, \cdots, z_h$，当它们相互独立时，其平方和服从自由度为 h 的 χ^2 分布。然而，一般情况下，由 h 个约束条件得到的 h 个 z 统计量不相互独立，因此，无法得到精确的 χ^2 分布。但是在各约束条件为真的情况下，可建立大样本下的服从自由度为 h 的渐近 χ^2 分布统计量

$$W = \boldsymbol{Z}'\boldsymbol{C}^{-1}\boldsymbol{Z} \sim \chi^2(h) \tag{3.7.31}$$

其中，$\boldsymbol{Z}$ 为以 z_i 为元素的列向量，$\boldsymbol{C}$ 是 z_i 的方差-协方差矩阵。因此，统计量 W 从总体

上测量了无约束回归不满足各约束条件的程度。

对非线性约束，也可类似地建立沃尔德统计量 W，但算法描述要复杂得多。例如，对(3.7.1)式，如果给出非线性约束 $\beta_1\beta_2=1$，也可以证明，在大样本下

$$\hat{\beta}_1\hat{\beta}_2 \sim N(\beta_1\beta_2,\sigma^2_{\hat{\beta}_1\hat{\beta}_2})$$

因此，如果 $\beta_1\beta_2=1$ 的约束为真，则在大样本下

$$W=\frac{(\hat{\beta}_1\hat{\beta}_2-1)^2}{\tilde{\sigma}^2_{\hat{\beta}_1\hat{\beta}_2}} \sim \chi^2(1)$$

这里 $\tilde{\sigma}^2_{\hat{\beta}_1\hat{\beta}_2}$ 的算法比较复杂，可用泰勒公式进行展开计算。

3. 拉格朗日乘数检验(LM)

沃尔德检验只需估计无约束模型，而拉格朗日乘数检验则只需估计受约束模型。

在受约束回归最大似然法的极值问题

$$\Phi=L(\boldsymbol{\beta},\sigma^2)-\boldsymbol{\lambda}'g(\boldsymbol{\beta})$$

中，$\boldsymbol{\lambda}'$ 是以各约束条件相应拉格朗日乘数为元素的行向量，各拉格朗日乘数 λ_i 的大小则衡量了各约束条件对最大似然函数值 $L(\tilde{\boldsymbol{\beta}},\tilde{\sigma}^2)$ 的影响程度。如果某一约束为真，则该约束条件对最大似然函数值 $L(\tilde{\boldsymbol{\beta}},\tilde{\sigma}^2)$ 的影响很小，于是，相应的拉格朗日乘数的值应接近于零。因此，拉格朗日乘数检验就是检验某些拉格朗日乘数的值是否“足够大”，如果“足够大”，则拒绝约束条件为真的假设。

拉格朗日统计量拉格朗日乘数检验本身是一个关于拉格朗日乘数的复杂的函数，在各约束条件为真的情况下，服从自由度恰为约束条件个数的渐近 χ^2 分布。同样地，如果为线性约束，拉格朗日乘数检验则服从于精确的 χ^2 分布：

$$\mathrm{LM}=nR^2 \tag{3.7.32}$$

其中 n 为样本容量，R^2 为如下被称为**辅助回归**(auxiliary regression)的可决系数

$$e_R=\delta_0+\delta_1X_1+\delta_2X_2+\cdots+\delta_kX_k$$

这里，e_R 为受约束回归模型的残差项。

如果约束是非线性的，辅助回归方程的估计比较复杂，但仍可按(3.7.32)式计算拉格朗日乘数检验统计量的值。

最后，一般有 $\mathrm{LM}\leqslant\mathrm{LR}\leqslant W$，因此，在有限样本中，它们的数值结果会有所不同。

本章练习题

1. 多元线性回归模型的基本假设是什么？在证明最小二乘估计量的无偏性和有效性的过程中，哪些基本假设起了作用？

2. 在多元线性回归分析中，t 检验与 F 检验有何不同？在一元线性回归分析中二者是否有等价的作用？

3. 为什么说对模型参数施加约束条件后，其回归的残差平方和一定不比未施加约束的残差平方和小？在什么样的条件下，受约束回归与无约束回归的结果相同？

4. 在一项调查大学生一学期平均成绩(Y)与每周在学习(X_1)、睡觉(X_2)、娱乐(X_3)与其他各种活动(X_4)所用时间的关系的研究中，建立如下回归模型：

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \mu$$

如果这些活动所用时间的总和为一周的总小时数 168。问：保持其他变量不变，而改变其中一个变量的说法是否有意义？该模型是否有违背基本假设的情况？如何修改此模型以使其更加合理？

5. 考虑下列两个模型：

(a) $$Y_i = \alpha_0 + \alpha_1 X_{i1} + \alpha_2 X_{i2} + u_i$$

(b) $$Y_i - X_{i1} = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + v_i$$

（1）证明：

$$\hat{\beta}_1 = \hat{\alpha}_1 - 1\,, \qquad \hat{\beta}_0 = \hat{\alpha}_0\,, \qquad \hat{\beta}_2 = \hat{\alpha}_2$$

（2）证明：两个模型的最小二乘残差相等，即对任何 i，有 $\hat{u}_i = \hat{v}_i$。

（3）在什么条件下，模型(b)的 R^2 小于模型(a)的 R^2？

6. 考虑下列三个试验步骤：

（1）对 $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + u_i$ 进行回归；

（2）对 $X_{i1} = \alpha_0 + \alpha_1 X_{i2} + v_i$ 进行回归，计算残差 $\hat{v}_i$；

（3）对 $Y_i = \gamma_0 + \gamma_1 \hat{v}_i + \gamma_2 X_{i2} + w_i$ 进行回归。

试证明 $\hat{\beta}_1 = \hat{\gamma}_1$，并直观地解释该结果。

7. 考虑以下过原点回归：

$$Y_i = \hat{\beta}_1 X_{i1} + \hat{\beta}_2 X_{i2} + e_i$$

（1）求参数的 OLS 估计量；

（2）对该模型，是否仍有结论

$$\sum e_i = 0\,, \qquad \sum e_i X_{i1} = 0\,, \qquad \sum e_i X_{i2} = 0$$

8. 对下列模型：

(a) $$Y_i = \alpha + \beta X_i + 2Z_i + u_i$$

(b) $$Y_i = \alpha + \beta X_i - \beta Z_i + u_i$$

求出 β 的最小二乘估计值，并将结果与下面的三变量回归方程的最小二乘估计值作比较：

(c) $$Y_i = \alpha + \beta X_i + \gamma Z_i + u_i$$

你认为哪一个估计值更好？

9. 回归模型中引入虚拟变量的作用是什么？有哪两种基本的引入方式，它们各适用于什么情况？

10. 在一项对北京某大学学生月消费支出的研究中，认为学生的消费支出除受其家庭每月收入水平的影响外，还受在学校中是否得到奖学金、来自农村还是城市、是经济发达地区还是欠发达地区，以及性别等因素的影响。试设定适当的模型，并导出如下情形下学生消费支出的平均水平：

（1）来自欠发达农村地区的女生，未得到奖学金；

（2）来自欠发达城市地区的男生，得到奖学金；

（3）来自发达地区的农村女生，得到奖学金；

（4）来自发达地区的城市男生，未得到奖学金。

11. 下表给出三变量模型 $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + u_i$ 的回归结果：

方差来源	平方和(SS)	自由度(d.f.)	平方和的均值(MSS)
来自回归	65 965	—	—
来自残差	—	—	—
来自总离差	66 042	14	

（1）求样本容量 n，残差平方和 RSS，回归平方和 ESS 及残差平方和 RSS 的自由度。

（2）求拟合优度 R^2 及调整的拟合优度 $\bar{R}^2$。

（3）检验假设：X_1 和 X_2 对 Y 无影响。应采用什么假设检验？为什么？

（4）根据以上信息，你能否确定 X_1 和 X_2 各自对 Y 的影响？

12. 试由教材中的(3.7.16)式推导出(3.7.17)式。

13. 在考察两组回归函数是否有差异的 F 检验中，试证明

$$\mathrm{RSS_U} = \mathrm{RSS_1} + \mathrm{RSS_2}$$

其中，$\mathrm{RSS_1}$，$\mathrm{RSS_2}$ 分别为两组对应各自样本下回归的残差平方和，$\mathrm{RSS_U}$ 为两组对应的样本合成的大样本下的回归的残差平方和。

14. 在一项对某社区家庭对某种消费品的消费需要调查中，得到下表所示的资料。

单位：元

序号	对某商品的消费支出 Y	商品单价 X_1	家庭月收入 X_2	序号	对某商品的消费支出 Y	商品单价 X_1	家庭月收入 X_2
1	591.9	23.56	7 620	6	644.4	34.14	12 920
2	654.5	24.44	9 120	7	680.0	35.30	14 340
3	623.6	32.07	10 670	8	724.0	38.70	15 960
4	647.0	32.46	11 160	9	757.1	39.63	18 000
5	674.0	31.15	11 900	10	706.8	46.68	19 300

请用手工与软件两种方式对该社区家庭对该商品的消费支出作二元线性回归分析，其中手工方式要求以矩阵表达式进行运算。

（1）估计回归方程的参数及随机干扰项的方差 $\hat{\sigma}^2$，计算 R^2 及 $\bar{R}^2$。

（2）对方程进行 F 检验，对参数进行 t 检验，并构造参数 95%的置信区间。

（3）如果商品单价变为 35 元，则某一月收入为 20 000 元的家庭的消费支出估计是多少？构造该估计值的 95%的置信区间。

15. 针对教材例 3.6.1，包含虚拟变量的模型（3.6.7）为

$$Y_i = \beta_0 + \delta_0 D_i + \beta_1 X_{i1} + \delta_1 (D_i X_{i1}) + \beta_2 X_{i2} + \delta_2 (D_i X_i) + \mu_i$$

请检验如下两个联合假设，并回答问题：

（1）$\mathrm{H_0}$: $\delta_0 = 0$，$\delta_1 = 0$。并与单参数的 t 检验结果进行对比。

（2）$\mathrm{H_0}$: $\delta_0 = 0$，$\delta_1 = 0$，$\delta_2 = 0$。并与例 3.7.3 中的邹氏参数稳定性检验的结果进行对比。

16. 在教材例 3.7.2 中，你如何检验联合假设：蛋类消费与肉禽类、水产类消费没有显著的替代关系？

17. 下表列出了某年中国按行业分的全部制造业国有企业及规模以上制造业非国有企业的工业总

产值 Y，资产合计 K 及职工人数 L。

序号	工业总产值 Y（亿元）	资产合计 K（亿元）	职工人数 L（万人）	序号	工业总产值 Y（亿元）	资产合计 K（亿元）	职工人数 L（万人）
1	3 722.70	3 078.22	113	17	812.70	1 118.81	43
2	1 442.52	1 684.43	67	18	1 899.70	2 052.16	61
3	1 752.37	2 742.77	84	19	3 692.85	6 113.11	240
4	1 451.29	1 973.82	27	20	4 732.90	9 228.25	222
5	5 149.30	5 917.01	327	21	2 180.23	2 866.65	80
6	2 291.16	1 758.77	120	22	2 539.76	2 545.63	96
7	1 345.17	939.10	58	23	3 046.95	4 787.90	222
8	656.77	694.94	31	24	2 192.63	3 255.29	163
9	370.18	363.48	16	25	5 364.83	8 129.68	244
10	1 590.36	2 511.99	66	26	4 834.68	5 260.20	145
11	616.71	973.73	58	27	7 549.58	7 518.79	138
12	617.94	516.01	28	28	867.91	984.52	46
13	4 429.19	3 785.91	61	29	4 611.39	18 626.94	218
14	5 749.02	8 688.03	254	30	170.30	610.91	19
15	1 781.37	2 798.90	83	31	325.53	1 523.19	45
16	1 243.07	1 808.44	33				

设定模型为

$$Y = AK^{\alpha}L^{\beta}e^{\mu}$$

（1）利用上述资料，进行回归分析。

（2） 回答：中国该年的制造业总体呈现规模报酬不变状态吗？

18. 继续 17 题，如果将 Cobb-Dauglas 生产函数设定为 $Y = AK^{\alpha}L^{\beta} + \mu$，则模型是非线性的，而且无法线性化。试给出这一设定的非线性普通最小二乘估计结果，并与第 17 题的估计结果进行比较。