经济学类各专业核心课程

计量经济学

第三章

多元线性回归模型

第三章 多元线性回归模型

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

本节基本内容:

一、多元线性回归模型

二、多元线性回归模型的矩阵表示

三、多元线性回归中的基本假定

一、多元线性回归模型

指对各个回归系数而言是“线性”的，对变量则可是线性的，也可是非线性的

例如：生产函数

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

取自然对数

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

一、多元线性回归模型

一般形式：对于有 k 个解释变量的线性回归模型

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

模型中参数 $\beta_j (j=1,2,\ldots,k)$ 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应变量平均值的影响。

多元总体回归函数

Y 的总体条件均值表示为多个解释变量的函数

$$E(Y_i \mid X_{2i}, X_{3i}, \ldots, X_{ki}) = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \ldots + \beta_k X_{ki}$$

总体回归函数也可表示为:

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

多元样本回归函数

Y 的样本条件均值表示为多个解释变量的函数

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

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

其中 $i = 1, 2, \cdots, n$

回归剩余（残差）： $e_i = Y_i - \hat{Y}_i$

二、多元线性回归模型的矩阵表示

k 个解释变量的多元线性回归模型的 n 个观测样本，可表示为

$$Y_1 = \beta_1 + \beta_2 X_{21} + \beta_3 X_{31} + \ldots + \beta_k X_{k1} + u_1$$

$$Y_2 = \beta_1 + \beta_2 X_{22} + \beta_3 X_{32} + \ldots + \beta_k X_{k2} + u_2$$

$$\vdots$$

$$Y_n = \beta_1 + \beta_2 X_{2n} + \beta_3 X_{3n} + \ldots + \beta_k X_{kn} + u_n$$

用矩阵表示

$$
\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}
=
\begin{bmatrix} 1 & X_{21} & \cdots & X_{k1} \\ 1 & X_{22} & \cdots & X_{k2} \\ \vdots & \vdots & & \vdots \\ 1 & X_{2n} & \cdots & X_{kn} \end{bmatrix}
\begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix}
+
\begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}
$$

$$
\boldsymbol{Y} \qquad \boldsymbol{X} \qquad \boldsymbol{\beta} \qquad \boldsymbol{u}
$$

$$
n\times 1 \qquad n\times k \qquad k\times 1 \qquad n\times 1
$$

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

三、多元线性回归的基本假定

假定1：零均值假定 $\boxed{E(u_i)=0\ \ (i=1,2,\cdots,n)}$

或

$$E(\boldsymbol{u})=\boldsymbol{0}$$

假定2和假定3：同方差和无自相关假定

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

假定4：随机扰动项与解释变量不相关

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

假定5:无多重共线性假定

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列满秩。

假定6：正态性假定

$$\boxed{u_i \sim N(0, \sigma^2)}$$

第二节
多元线性回归模型的估计

本节基本内容:

- 普通最小二乘法（OLS）
- OLS估计式的性质
- OLS估计的分布性质
- 随机扰动项方差σ^2的估计
- 回归系数的区间估计

一、普通最小二乘法（OLS）

最小二乘原则

剩余平方和最小： $\min\sum e_i^2 = \sum(Y_i - \hat{Y}_i)^2$

$$\min\sum e_i^2 = \sum[Y_i - (\hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \ldots + \hat{\beta}_k X_{ki})]^2$$

求偏导，令其为0： $\dfrac{\partial(\sum e_i^2)}{\partial \hat{\beta}_j} = 0$

用矩阵表示

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

因为样本回归函数为 $\boldsymbol{Y} = \boldsymbol{X}\hat{\boldsymbol{\beta}} + \boldsymbol{e}$

两边乘 $\boldsymbol{X}'$ 有： $\boldsymbol{X}'\boldsymbol{Y} = \boldsymbol{X}'\boldsymbol{X}\hat{\boldsymbol{\beta}} + \boldsymbol{X}'\boldsymbol{e}$

因为 $\boldsymbol{X}'\boldsymbol{e} = \boldsymbol{0}$ ，则正规方程为：

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

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

多元回归中 $\boxed{\hat{\beta} = (X'X)^{-1} X'Y}$

二、OLS估计式的性质

OLS估计式

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：$\mathrm{E}(\hat{\beta}_k) = \beta_k$

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

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**）

三、OLS估计的分布性质

基本思想

- $\hat{\beta}_i$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，决定了Y_i也是服从正态分布的随机变量

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

$\hat{\beta}$ 的期望 $\mathrm{E}(\hat{\beta})=\beta$ （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

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

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

这里是 c_{jj} 矩阵 $(\boldsymbol{X}'\boldsymbol{X})^{-1}$ 中第 j 行第 j 列的元素

故有： $\hat{\beta}_j \sim N(\beta_j, \sigma^2 c_{jj}) \qquad j=1,2,\ldots,k$

四、随机扰动项方差 σ^2 的估计

多元回归中 σ^2 的无偏估计为：

$$\hat{\sigma}^2 = \frac{\sum e_i^2}{n-k} \qquad \hat{\sigma}^2 = \frac{\mathbf{e}'\mathbf{e}}{n-k}$$

或表示为 $$z_k = \frac{\hat{\beta}_k - \beta_k}{\mathrm{SE}(\hat{\beta}_k)} = \frac{\hat{\beta}_k - \beta_k}{\sigma\sqrt{c_{jj}}} \sim N(0,1)$$

将 $\hat{\beta}_k$ 作标准化变换：

因σ^2是未知的，可用$\hat{\sigma}^2$代替σ^2去估计参数$\hat{\beta}$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:

$$\boxed{t = \frac{\hat{\beta}_k - \beta_k}{\hat{\mathrm{SE}}(\hat{\beta}_k)} \sim t(n-k)}$$

五、回归系数的区间估计

由于

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

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

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

$$(j = 1, \ldots, k)$$

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

或:

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

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

第三节
多元线性回归模型的检验

本节基本内容:

- 多元回归的拟合优度检验
- 回归方程的显著性检验（F检验）
- 各回归系数的显著性检验（t检验）

一、多元回归的拟合优度检验

多重可决系数：多元回归中：

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

多重可决系数也可表示为

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

多重可决系数的矩阵表示

$$\text{TSS} = \boldsymbol{Y}'\boldsymbol{Y} - n\overline{Y}^2 \qquad \text{ESS} = \hat{\boldsymbol{\beta}}\boldsymbol{X}'\boldsymbol{Y} - n\overline{Y}^2$$

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

可以证明：$R^2 = \dfrac{\hat{\beta}_2 \sum x_{2i} y_i + \hat{\beta}_3 \sum x_{3i} y_i + \ldots + \hat{\beta}_k \sum x_{ki} y_i}{\sum y_i^2}$

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。

可决系数的修正方法

总变差 $\text{TSS}=\sum_{i=1}^{n}(Y_i-\bar{Y})^2=\sum_{i=1}^{n}Y_i^2$ 自由度为 $n-1$

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

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

修正的可决系数为

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

修正的可决系数$\bar{R}^2$与可决系数R^2的关系：

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

特点

可决系数 R^2必定非负，但修正的可决系数$\bar{R}^2$可能为负值，这时规定 $\bar{R}^2 = 0$

二、回归方程显著性检验（F检验）

基本思想

要说明所有解释变量联合起来对应变量影响的总显著性,或整个方程总的联合显著性。对方程总显著性检验需要在方差分析的基础上进行F检验。

方差分析表

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

模型解释了的变差 $RSS=\sum(Y_i-\hat{Y}_i)^2$ 自由度 $k-1$

剩余变差 $ESS=\sum(\hat{Y}_i-\bar{Y})^2$ 自由度 $n-k$

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

F检验

原假设 $\mathrm{H}_0: \beta_2 = \beta_3 = \ldots = \beta_k = 0$

备择假设 $\mathrm{H}_1: \beta_j (j = 1, 2, \ldots, k)$ 不全为0

建立统计量(可以证明)：

$$F = \frac{\mathrm{ESS}/(k-1)}{\mathrm{RSS}/(n-k)} \sim \mathrm{F}(k-1, n-k)$$

给定显著性水平 α，查F分布表得临界值 $F_\alpha(k-1, n-k)$

并通过样本观测值计算 F 值

◆如果 $F > F_{\alpha}(k-1, n-k)$ （小概率事件发生了）则拒绝 $H_0: \beta_2 = \beta_3 = \ldots = \beta_k = 0$，说明回归模型有显著意义，即所有解释变量联合起来对 Y 有显著影响。

◆反之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。

可决系数与F检验

二者都建立在对应变量变差分解的基础上。

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

可看出：当 $R^2=0$ 时，$F=0$

R^2 越大，F 值也越大

当 $R^2=1$ 时，$F \to \infty$

结论：对方程联合显著性检验的F检验，实际上也是对R^2的显著性检验。

三、各回归系数的显著性检验（t 检验）

方法：

原假设 $H_0: \beta_j = 0 \qquad j = 1, 2, \ldots, k$

备择假设 $H_1: \beta_j \neq 0$

统计量

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

t检验的方法

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

就不拒绝 $\mathrm{H}_0: \beta_j = 0$ 而拒绝 $\mathrm{H}_1: \beta_j \ne 0$

即认为 β_j 所对应的解释变量 X_j 对应变量 Y 的影响不显著；

反之则反是。

注意:

1、在多元回归中，可分别对每个回归系数逐个地进行t检验。

2、在一元回归中F检验与t检验等价,

但在多元回归中F检验与t检验作用不同。

$$F = t^2$$

第四节 多元线性回归模型的预测

本节基本内容:

- 应变量平均值预测
- 应变量个别值预测

一、应变量平均值预测

1、Y 平均值的点预测

将解释变量预测值代入估计的方程：

多元回归时：

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

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

2、平均值的区间预测

基本思想：

为对 $\mathrm{E}(Y_F|X_F)$ 作区间预测，必须确定平均值预测值 $\hat{Y}_F$ Y 的抽样分布。必须找出与 $\hat{Y}_F$ 和 $\mathrm{E}(Y_F|X_F)$ 都有关的统计量。

具体作法（回顾一元回归）

一元中已知

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

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

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

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

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

多元回归时可构造t统计量

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

二、应变量个别值预测

基本思想：

- 为了对 Y_F 的个别值 $\hat{Y}_F$ 作区间预测，需要寻找与预测值 Y_F 和个别值 Y 有关的统计量，并要明确其概率分布

具体作法

构造t统计量

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

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

因此，多元回归时 Y 的个别值的置信度 $1-\alpha$ 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

案例：中国税收增长的分析

提出问题

改革开放以来，随着经济体制改革的深化和经济的快速增长，中国的财政收支状况发生很大变化，为了研究影响中国税收收入增长的主要原因，分析中央和地方税收收入的增长规律，预测中国税收未来的增长趋势，需要建立计量经济模型。

参数估计

注意这些指标的统计含义

Dependent Variable: Y
Method: Least Squares
Date: 07/05/05 Time: 16:54
Sample: 1978 2002
Included observations: 25

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-2582.791	940.6128	-2.745860	0.0121
X2	0.022067	0.005577	3.956605	0.0007
X3	0.702104	0.033236	21.12466	0.0000
X4	23.98541	8.738302	2.744859	0.0121

R-squared	0.997430	Mean dependent var	4848.366
Adjusted R-squared	0.997063	S.D. dependent var	4870.971
S.E. of regression	263.9599	Akaike info criterion	14.13512
Sum squared resid	1463172.	Schwarz criterion	14.33014
Log likelihood	-172.6890	F-statistic	2717.238
Durbin-Watson stat	0.948542	Prob(F-statistic)	0.000000

模型估计的结果可表示为

$$\hat{Y}_i = -2582.791 + 0.022067X_2 + 0.702104X_3 + 23.98541X_4$$

$$(940.6128) \quad (0.0056) \quad (0.0332) \quad (8.7363)$$

$$t = (-2.7459) \quad (3.9566) \quad (21.1247) \quad (2.7449)$$

$R^2 = 0.9974$ $\quad \bar{R}^2 = 0.9971$ $\quad F = 2717.238$ $\quad df = 21$

拟合优度：可决系数 $R^2 = 0.9974$ 较高，

修正的可决系数 $\bar{R}^2 = 0.9971$ 也较高，

表明模型拟合较好。

显著性检验

F检验： 针对 $H_0: \beta_2=\beta_3=\beta_4=0$，取 $\alpha=0.05$

查自由度为 $k-1=3$ 和 $n-k=21$ 的临界值 $F_\alpha(3,21)$ 。

由于 $F=2717.238>F_\alpha(3,21)=3.075$ ，应拒绝 H_0 ，说明回归方程显著，即“国内生产总值”、“财政支出”、“商品零售物价指数”等变量联合起来确实对“税收收入”有显著影响。

t检验：给定 $\alpha=0.05$，查**t**分布表，在自由度为 $n\text{-}3=25-4=21$ 时临界值为 $t_{0.025}(21)=2.080$ ，因为 X_2, X_3, X_4 的参数对应的**t**统计量均大于**2.080**，这说明在**5%**的显著性水平下，斜率系数均显著不为零，表明国内生产总值、财政支出、商品零售价格指数对财政收入分别都有显著影响。

经济意义检验

本模型中 $\hat{\beta}_2 = 0.022067, \hat{\beta}_3 = 0.702104, \hat{\beta}_4 = 23.98541$ 所估计的参数的符号与经济理论分析一致，说明在其他因素不变的情况下，国内生产总值每增加1亿元，平均说来财政收入将增加220.67万元；财政支出每增加1亿元，平均说来财政收入将增加7021.04万元;商品零售物价指数每增加**1%,**平均说来财政收入将增加**23.98541**亿元。

第三章　小结

1. 多元线性回归模型是将总体回归函数描述为一个被解释变量与多个解释变量之间线性关系的模型。

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_p X_{pi} + u_i$$

通常多元线性回归模型可以用矩阵形式表示：

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{u}$$

2. 多元线性回归模型中对随机扰动项u的假定:零均值假定、同方差假定、无自相关假定、随机扰动与解释变量不相关假定、正态性假定、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式及期望、方差和标准误差：

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

$$\hat{\mathrm{Var}}(\hat{\beta}_j) = \hat{\sigma}^2 C_{jj} = \left(\frac{\sum e_i^2}{n-k}\right) C_{jj} \qquad \mathrm{SE}(\hat{\beta}_j) = \sigma \sqrt{C_{jj}}$$

4. 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式。

5. 多元线性回归模型中参数区间估计的方法。

6. 多重可决系数的意义和计算方法：

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

修正可决系数的作用和方法：

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

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

7. F检验是对多元线性回归模型中所有解释变量联合显著性的检验，F检验是在方差分析基础上进行的。

$$F = \frac{\mathrm{ESS}/(k-1)}{\mathrm{RSS}/(n-k)} \sim F(k-1, n-k)$$

8. 多元回归分析中，为了分别检验当其它解释变量不变时，各个解释变量是否对被解释变量有显著影响，需要分别对所估计的各个回归系数作t检验。

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

9. 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。

点预测：$\hat{Y}_f = \boldsymbol{X}_F \hat{\boldsymbol{\beta}}$

平均值：

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

个别值：

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

本章内容结束！