第三章 多元线性回归模型

引子：为什么各地区的地方财政性教育经费支出会差异明显?

要实现到2020年基本实现教育现代化，进入人力资源强国的战略目标，需要大幅度增加教育投入，提高国家财政性教育支出占国内生产总值的比例。

在中央和地方一般公共预算教育支出中，地方财政的支出占94.84%，地方财政支出是财政教育经费的主要来源。然而，由于各地区社会经济发展不平衡不充分，各地区的一般公共预算教育支出差异明显：2016年各地区人均地方一般公共预算教育支出中，最高的是北京市，为4083.62元，最低的是河南省，仅为1409.74元，北京市是河南省的2.90倍。

为什么各地区的地方财政性教育经费支出差异如此明显呢?

怎样分析多种因素的影响呢？

影响中国地方财政性教育经费支出的因素很多，人口数量、经济发展、地方财力、政府教育投入意愿、价格因素、居民对高质量教育的需求等，都可能会影响地方财政教育支出。

很明显，简单线性回归不能进行这类多因素问题的分析，还需要进一步寻求有多个解释变量的回归分析方法。

简单线性回归模型主要讨论一个被解释变量和一个解释变量之间的线性关系，但是，由于实际经济问题的复杂性，一个经济变量可能会同多个变量相联系。大部分经济问题均是如此：如商品需求量影响因素、货币需求量影响因素等。

因此，有必要将只有一个解释变量的一元回归模型推广到有多个解释变量的情况。本章将把上一章讨论的结论推广到包含多个解释变量的多元回归模型。

多元线性回归模型

本章讨论:

将简单线性回归的的研究方式推广到多元的情况

- 多元线性回归模型及古典假定
- 多元线性回归参数的估计
- 多元线性回归模型的检验
- 多元线性回归预测

（一）多元线性回归模型及古典假定

1、多元线性回归模型的意义

一般形式：对于有**K-1**个解释变量的线性回归模型

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

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

注意：模型中的β_j（j=2,3,---k）是偏回归系数

样本容量为*n*

偏回归系数：

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

对偏回归系数的理解

例如 $$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i \xleftrightarrow{\text{对比}} Y_i = \alpha_1 + \alpha_2 X_{2i} + u_{1i}$$

并且 $$X_{3i} = b_2 + b_{32} X_{2i} + u_{2i}$$

可证明 $$\alpha_2 = \beta_2 + \beta_3 b_{32} + \underbrace{\frac{\sum x_{2i}(u_i - \bar{u})}{\sum x_{2i}^2}}_{\text{(误差项)}}$$

$$E(\alpha_2) = \beta_2 + \beta_3 b_{32}$$

(证明见古加拉蒂《计量经济学》第三版附录7A.5)

结论:只要 $b_{32} \neq 0$, β_2与α_2 是有区别的。α_2不仅包括 X_{2i} 对Y平均值的“直接”影响，还包括由于 X_{3i} 的变动对Y平均值的“间接”影响。

从残差理解偏回归系数

例如 $Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$ 对比 $Y_i = \alpha_1 + \alpha_2 X_{2i} + u_i$

1、作回归 $Y_i = a_1 + a_{13} X_{3i} + u_{1i}$，则 $\hat{\alpha}_{13} = \sum y_i x_{3i} / \sum x_{3i}^2$

$e_{1i} = Y_i - \hat{a}_1 - \hat{a}_{13} X_{3i} = y_i - \hat{a}_{13} x_{3i}$，$e_{1i}$表示除去$X_{3i}$影响后的 Y_i

2、作回归 $X_{2i} = b_2 + b_{23} X_{3i} + u_{2i}$，则 $\hat{b}_{23} = \sum x_{2i} x_{3i} / \sum x_{3i}^2$

$e_{2i} = X_{2i} - \hat{b}_2 - \hat{b}_{23} X_{3i} = x_{2i} - \hat{b}_{23} x_{3i}$，$e_{2i}$ 表示除去X_{3i} 影响后的 X_{2i}

将 e_{1i} 对 e_{2i} 回归（因 e_{1i}和 e_{2i}的均值为0，为过原点的回归）

$$e_{1i} = \gamma_2 e_{2i} + v_i \quad \text{则} \quad \hat{\gamma}_2 = \sum e_{1i} e_{2i} / \sum e_{2i}^2$$

可以证明，这样估计的 $\hat{\gamma}_2$与估计的 $\hat{\beta}_2$是一致的，而与 $\hat{\alpha}_2$是不一致的。

(证明见古加拉蒂《计量经济学》第四版附录7A.2)

证明过程(参考)

由 $$\hat{\gamma}_2=\sum e_{1i}e_{2i}\Big/\sum e_{2i}^2=\frac{\sum(y_i-\hat{a}_{13}x_{3i})(x_{2i}-\hat{b}_{23}x_{3i})}{\sum(x_{2i}-\hat{b}_{23}x_{3i})^2}$$

将$\hat{a}_{13}$、$\hat{b}_{23}$代入得

$$\hat{\gamma}_2=\frac{\sum[y_i-(\sum y_ix_{3i}/\sum x_{3i}^2)x_{3i}][x_{2i}-(\sum x_{2i}x_{3i}/\sum x_{3i}^2)x_{3i}]}{\sum[x_{2i}-(\sum x_{2i}x_{3i}/\sum x_{3i}^2)x_{3i}]^2}$$

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

由后面将讲的OLS估计量可知，这就是多元回归中β_2的估计式$\hat{\beta}_2$

多元线性回归中的“线性”

指针对各个回归系数而言是“线性”的，对变量则可以是线性的，也可以是非线性的

例如：生产函数 $$Y = AL^{\alpha}K^{\beta}u$$

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

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

多元总体回归函数

条件期望表现形式：

将Y的总体条件期望表示为多个解释变量的函数，如：

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

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

注意：这时Y总体条件期望的轨迹是K维空间的一个面

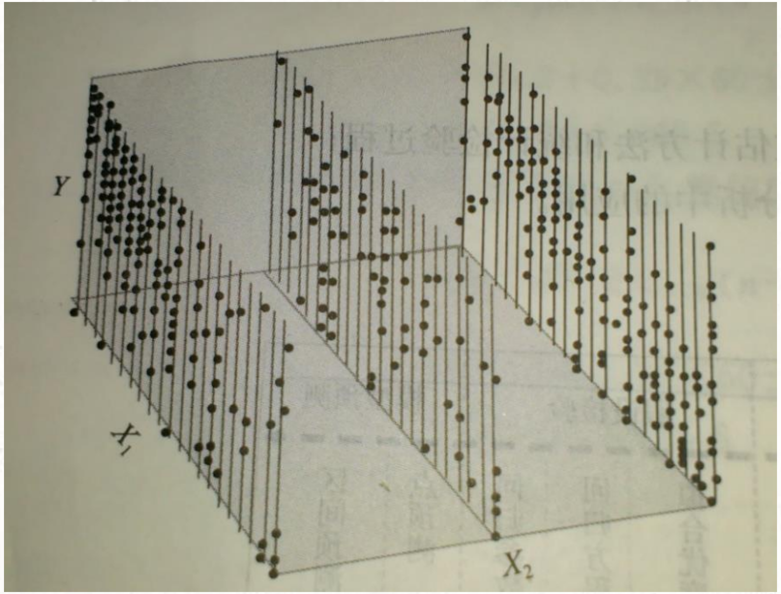
个别值表现形式：

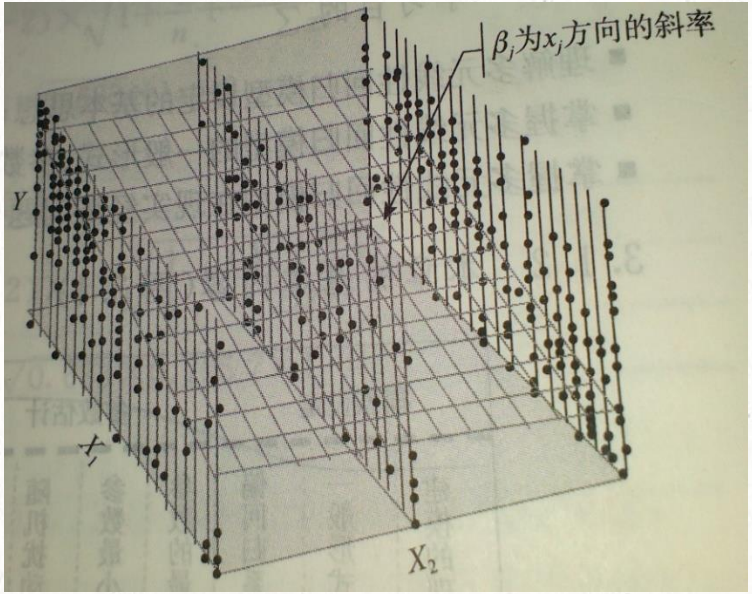
引入随机扰动项 $u_i=Y_i-E(Y_i|X_{2i},X_{3i}\cdots X_{ki})$

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

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

直观理解（供参考）





多元样本回归函数

Y 的样本条件均值可表示为多个解释变量的函数

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

或回归剩余（残差）： $e_i = Y_i - \hat{Y}_i$

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

其中 $i = 1, 2, \cdots n$

多元线性回归分析要解决的主要问题，仍然是如何根据变量的样本观测值去估计回归模型中的各个参数，即要用样本回归函数去估计总体回归函数，并且对估计的参数及回归方程进行统计检验，最后利用回归模型进行预测和经济分析。

2、多元线性回归模型的矩阵表示

多个解释变量的多元线性回归模型的n组样本观测值，可表示为

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

用矩阵表示

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

矩阵表示方式

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

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

其中：$\boldsymbol{Y},\hat{\boldsymbol{Y}},\boldsymbol{u},\boldsymbol{e}$ 都是有n个元素的列向量

$\boldsymbol{\beta},\hat{\boldsymbol{\beta}}$ 是有k 个 元素的列向量

（ k = 解释变量个数 + 1 ）

$\boldsymbol{X}$ 是第一列为1的n×k阶解释变量数据矩阵，

(截距项可视为解释变量总是取值为1)

3、多元线性回归中的基本假定

假定1：零均值假定

$$E(u_i)=0$$ （i=1，2，---n） 或 E（**U**）=0

假定2和假定3：同方差和无自相关假定：

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

或 $$\mathrm{Var}(\boldsymbol{U})=E(\boldsymbol{U}\boldsymbol{U}')=\sigma^2\boldsymbol{I}$$

假定4：随机扰动项与解释变量不相关

$$Cov\left(X_{ji},u_i\right)=0$$ j=2,3,----k

假定5: **无多重共线性假定** (多元中增加的)

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

Ran(***X***)= k → Ran(***X'X***)=k

即 (***X'X***) 可逆

假定6: **正态性假定**

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

（二）多元线性回归模型的估计

1.普通最小二乘法（OLS）

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

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

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

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

即

$$-2\sum\left[Y_i - (\hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \cdots + \hat{\beta}_{ki} X_{ki})\right] = 0 \rightarrow \sum e_i = 0$$

$$-2\sum X_{2i}\left[Y_i - (\hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \cdots + \hat{\beta}_{ki} X_{ki})\right] = 0 \rightarrow \sum X_{2i} e_i = 0$$

$$\vdots$$

$$-2\sum X_{ki}\left[Y_i - (\hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \cdots + \hat{\beta}_{ki} X_{ki})\right] = 0 \rightarrow \sum X_{ki} e_i = 0$$

用矩阵表示的正规方程

偏导数

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

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

两边左乘 $\boldsymbol{X}'$ $\boldsymbol{X}'\boldsymbol{Y} = \boldsymbol{X}'\boldsymbol{X}\hat{\boldsymbol{\beta}} + \boldsymbol{X}'\boldsymbol{e}$

根据最小二乘原则 $\boldsymbol{X}'\boldsymbol{e} = \boldsymbol{0}$

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

$(\mathbf{X'X})_{k\times k}$ 是满秩矩阵，其逆存在

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

只有两个解释变量时：

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

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

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

注意：x、y 为X、Y的离差

对比

简单线性回归中

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

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

OLS回归线的数学性质 （与简单线性回归相同）

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

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

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

- 被解释变量估计值 $\hat{Y}_i$ 与剩余项 e_i 不相关

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

- 解释变量 X_i 与剩余项 e_i 不相关

$$Cov(X_{ji},e_i)=0 \qquad (j=2,3,\text{---}k)$$

2、OLS估计式的统计性质

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

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

2、 **无偏特性** $E(\hat{\beta}_K) = \beta_K$ **(证明见教材P70)**

3、 **最小方差特性**

在β_K所有的线性无偏估计中，OLS估计 $\hat{\beta}_K$

具有最小方差 **(证明见教材P91附录3.1)**

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）

3、OLS估计的分布性质

基本思想：

- $\hat{\beta}$是随机变量，必须确定其分布性质才可能进行区间估计和假设检验
- u_i是服从正态分布的随机变量，决定了Y也是服从正态分布的随机变量
- $\hat{\beta}$是Y的线性函数，决定了$\hat{\beta}$也是服从正态分布的随机变量

$\hat{\beta}$ 的期望与方差

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

- $\hat{\beta}$的方差和标准误差：

可以证明$\hat{\beta}$ 的方差—协方差矩阵为（见下页）

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

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

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

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

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

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

$\hat{\beta}$ 的方差—协方差矩阵

$$
\begin{aligned}
COV(\hat{\beta}) &= E\{[\hat{\beta}-E(\hat{\beta})][\hat{\beta}-E(\hat{\beta})]'\} \\
&= E[(\hat{\beta}-\beta)(\hat{\beta}-\beta)'] \qquad \text{(由无偏性)} \\
&= E[(\boldsymbol{X'X})^{-1}\boldsymbol{X'uu'X}(\boldsymbol{X'X})^{-1}] \qquad \text{(由OLS估计式)} \\
&= (\boldsymbol{X'X})^{-1}\boldsymbol{X'}E(\boldsymbol{uu'})\boldsymbol{X}(\boldsymbol{X'X})^{-1} \\
&= (\boldsymbol{X'X})^{-1}\boldsymbol{X'}\sigma^2\boldsymbol{IX}(\boldsymbol{X'X})^{-1} \qquad \text{(由同方差和无自相关)} \\
&= \sigma^2(\boldsymbol{X'X})^{-1}
\end{aligned}
$$

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

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

4、 随机扰动项方差 σ^2 的估计

σ^2一般未知，可证明多元回归中 σ^2 的无偏估计为：（P92附录3.2证明有误，请参考专门文档）

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

将 $\hat{\boldsymbol{\beta}}$ 作标准化变换：

$$z_j = \frac{\hat{\beta}_j - \beta_j}{SE(\hat{\beta}_j)} = \frac{\hat{\beta}_j - \beta_j}{\sigma\sqrt{c_{jj}}} \sim N(0,1)$$

σ^2 未知时$\hat{\beta}$ 的标准化变换

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

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

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

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

5. 回归系数的区间估计

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

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

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

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

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

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

（三）多元线性回归模型的检验

1、多元回归的拟合优度检验

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

与简单线性回归中可决系数 r^2 的区别只是 $\hat{Y}_i$ 不同

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

多重可决系数可表示为

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

(注意:红色字体是与一元回归不同的部分)

多重可决系数的矩阵表示

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

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

可用代数式表达为

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

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

可决系数的修正方法

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

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

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

修正的可决系数为

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

修正的可决系数

思想：可决系数只涉及变差，没有考虑自由度。

如果用自由度去校正所计算的变差，可纠正解释变量个数不同引起的对比困难。

回顾：

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

修正的可决系数$\bar{R}^2$与可决系数R^2的关系

已经导出：

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

注意：

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

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

修正可决系数的特点

- 修正后$R^2 \leq R^2$，且随着解释变量个数增加两者差距变大。
- 修正后R^2与R^2同增同减（在其他条件不变的前提下），具有同样的两层含义。
- 修正后R^2不再是解释变量个数的不减函数，而要视正面影响（对拟合优度贡献）和负面影响（自由度损失）的相对大小。
- 修正后R^2也只能做描述性判断。
- 修正后R^2使用原则与R^2相同。

2、回归方程的显著性检验（F检验）

基本思想：

在多元回归中包含多个解释变量，它们与被解释变量是否有显著关系呢?

当然可以分别检验各个解释变量对被解释变量影响的显著性。

但为了说明所有解释变量联合起来对被解释变量影响的显著性，或整个方程总的联合显著性，需要对方程的总显著性在方差分析的基础上进行F检验。

(1)方差分析

在讨论可决系数时已经分析了被解释变量总变差TSS的分解及自由度：

TSS=ESS+RSS

注意： Y的样本方差= 总变差/自由度

即

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

显然，Y的样本方差也可分解为两部分，可用方差分析表分解

方差分析表

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

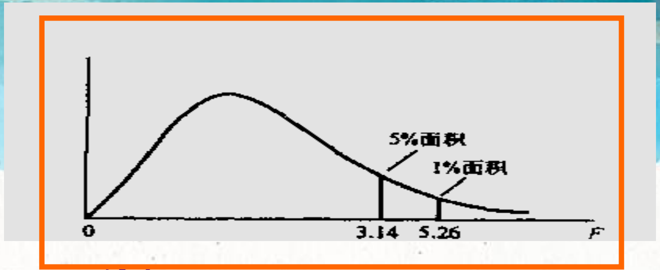
模型解释了的变差 ESS= $\sum (\hat{Y}_i - \bar{Y})^2$ 自由度 k－1

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

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

基本思想：如果多个解释变量联合起来对被解释变量的影响不显著，“归于回归的方差”应该比“归于剩余的方差”显著地小（即这应是大概率事件）。

(2)F检验



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

（所有解释变量联合起来对被解释变量的影响不显著）

备择假设: $H_1: \beta_j (j = 2, 3, \cdots, k)$ 不全为0

建立统计量:

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

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

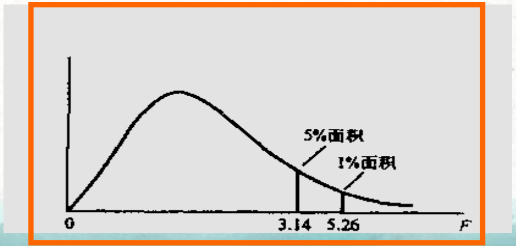
F检验方式

▼如果计算的F值≥临界值 $F_{\alpha}(k-1,n-k)$ (小概率事件发生)

则拒绝 $H_0:\beta_2=\beta_3=\cdots=\beta_k=0$ ，说明回归模型有显著意义，即所有解释变量联合起来对Y确有显著影响。

▼如果计算的F值<临界值 $F_{\alpha}(k-1,n-k)$ (大概率事件发生)

则不拒绝 $H_0:\beta_2=\beta_3=\cdots=\beta_k=0$ ，说明回归模型没有显著意义，即所有解释变量联合起来对Y没有显著影响。



(3) 可决系数的显著性检验

拟合优度检验与对线性回归的总体显著性的 F 检验是从不同原理出发的两类检验，但二者有内在联系：

拟合优度检验——从已估计的模型出发，检验对样本观测值的拟合程度。

总体显著性的F检验——从样本观测值出发，检验模型总体线性关系的显著性。

F检验与多重可决系数有密切关系：二者都建立在对被解释变量变差分解的基础上，实际上 F 统计量也可通过可决系数去计算：

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

可以看出：当 $R^2=0$ 时，F=0 ；当 $R^2=1$时，F→∞；

当 R^2越大时，F值也越大

修正的可决系数与F检验的关系

由方差分析可以看出，F统计量与修正的多重可决系数都建立在对被解释变量变差分解的基础上，而且都与自由度有关。

二者关系：

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

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

F与 $\bar{R}^2$ 同方向变化，F检验等价于对 $\bar{R}^2 = 0$ 的显著性检验。

3、各回归系数的假设检验（t 检验）

注意: 在一元回归中F检验与 t 检验等价，且 $F=t^2$ **（见教材P74证明）**

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

方法：

原假设 $H_0: \beta_j = 0$ （j=1,2,……k）

备择假设 $H_1: \beta_j \neq 0$

统计量t为：

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

对各回归系数假设检验的作法

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

如果 $-t_{\alpha/2}(n-k) < t^* < t_{\alpha/2}(n-k)$ (大概率事件发生)

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

即认为 β_j 所对应的解释变量 X_j 对被解释变量Y的影响不显著。

如果 $t^* \leq -t_{\alpha/2}(n-k)$ 或 $t^* \geq t_{\alpha/2}(n-k)$ (小概率事件发生)

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

即认为 β_j 所对应的解释变量 X_j 对被解释变量Y的影响是显著的。

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

F检验与t检验的关系

- 在一元回归中F检验与t检验等价，且 $F = t^2$（证明见本科教材P74）。
- 在多元回归中，F检验与t检验的关系是：
  - 整体的F检验显著并不见得个别系数的t检验显著。
  - 个别系数的t检验显著则整体F检验通常也显著。
- 在多元回归中，既要作F检验，又要进一步分别对每个回归系数逐个地进行t检验。

（四）多元线性回归模型的预测

1、被解释变量平均值预测

(1)Y平均值的点预测

方法：将解释变量预测值代入估计的方程：

多元回归时：

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

或

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

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

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

(2)Y平均值的区间预测

基本思想：　（与简单线性回归时相同）

- 由于存在抽样波动，预测的平均值 $\hat{Y}_F$ 不一定等于真实平均值 $E(Y_F|X_F)$，还需要对 $E(Y_F|X_F)$ 作区间估计。
- 为了对Y作区间预测，必须确定平均值预测值 $\hat{Y}_F$ 的抽样分布。
- 必须找出与 $\hat{Y}_F$ 和 $E(Y_F|X_F)$ 都有关的统计量，并要明确其概率分布性质。

区间预测的具体作法（回顾简单线性回归）

简单线性回归中 $$E(\hat{Y}_F) = E(Y_F | X_F) = \beta_1 + \beta_2 X_F$$

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

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

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

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

区间预测的具体作法（多元时）

$\hat{Y}_F$ 服从正态分布，可证明

$$E(\hat{Y}_F) = E(Y_F | \mathbf{X}_F)$$

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

即

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

标准化

$$t^* = \frac{\hat{Y}_F - E(\hat{Y}_F)}{SE(\hat{Y}_F)} = \frac{\hat{Y}_F - E(Y_F | \mathbf{X}_F)}{\sigma\sqrt{\mathbf{X}_F (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}_F'}} \sim N(0,1)$$

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

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

区间预测的具体作法

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

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

$$=1-\alpha$$

或

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

$$=1-\alpha$$

2、被解释变量个别值预测

基本思想： （与简单线性回归时相同）

- 由于存在随机扰动u_i的影响，*Y*的平均值并不等于*Y*的个别值。
- 为了对*Y*的个别值Y_F作区间预测，需要寻找与预测值$\hat{Y}_F$和个别值Y_F有关的统计量，并要明确其概率分布性质。

个别值区间预测具体作法

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

并且已知 e_F 服从正态分布，且多元回归时可证明

$$E(e_F) = 0$$

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

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

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

个别值预测具体作法（续）

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

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

因此，多元回归时Y的个别值的置信度1-α的预测区间的上下限为

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

（六）案例分析

接引子：为什么各地区的地方财政性教育经费支出会差异明显？

提出问题： 为了研究影响中国各地区地方一般公共预算教育支出变动的主要原因，分析地方财政教育支出增长的数量规律，预测地方一般公共预算教育支出的增长趋势，可以建立计量经济模型。

理论分析： 影响地方财政教育支出的因素很多，据分析主要的因素可能有：（1）各地区居民对教育数量、质量、品质的需求有差异，地区的人口数量不同决定了各地区教育规模不同；（2）由地区经济规模决定的地方整体财力是地方财政教育支出的基本源泉，关系到地方政府对教育投入的能力；（3）地方政府对教育投入的意愿和重视程度的不同，可能构成地方财政教育支出差异的原因；（4）物价水平，特别是教育消费的价格变动，会影响地方财政对教育的支出；（5）现在社会及民办教育发展很快，可以适应部分居民对教育消费的需求，可能对政府公共预算教育投入形成一定的补充。

数据： 以国家统计局已经公布的2016年31个省市区的数据为样本，从《中国统计年鉴2017》中收集到以下数据（见表3.3，P80-81）。

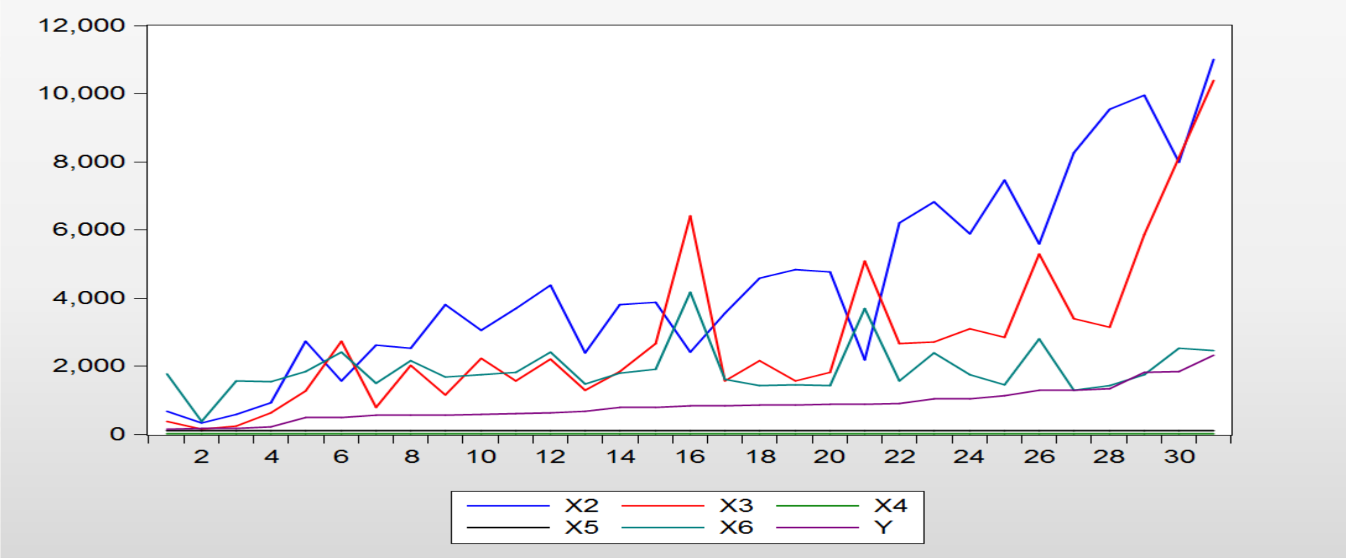
模型设定：

为了全面反映中国地方一般公共预算教育支出的差异，选择“地方一般公共预算教育支出”为被解释变量。根据对影响中国地方一般公共预算教育支出主要因素的分析，选择各地区的“年末人口数量”作为居民对教育需求的代表性变量；选择“地方一般公共预算收入”度量地区财政教育投入的能力；地方政府教育投入意愿难以直接量化，选择“教育支出在地方财政支出中的比重”作为其代表；选择“教育消费价格指数”作为价格变动影响的因素；选择“居民人均教育文化娱乐消费”代表居民对高质量、高品质教育的需求和社会及民办教育的补充因素。

模型设定：

为利用EViews软件分析和估计模型的参数，用第二章案例中介绍的方法，建立工作文件。在EViews命令框直接键入“data Y X2 X3 X4 X5 X6 ”，在对应的“Y、X2、X3、X4、 X5、 X6”下输入或粘贴相应的数据。为初步观察数据的关系，在命令栏输入“Sort Y”，其中Y为按其递增排序的变量名称，实现数据按Y递增排序。在“workfile”中按住“ctrl”键，点击“Y、X2、X3、X4 、X5、 X6”，在双击的菜单中点“open group”，出现“Y、X2、X3、X4 、X5 、X6”数据表“Group”。在数据表“Group” 中点“View/graph/line/ok”，出现序列Y、X2、X3、X4、X5、X6的线性图：

模型设定：



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

估计参数：

利用EViews估计模型参数，可点击“Quick”下拉菜单中的“Estimate Equation”，在出现的对话框的“Equation Specification”栏中键入“Y C X2 X3 X4 X5 X6”，在“Estimation Settings”栏中选择“Least Sqares”(最小二乘法)，点“ok”，或者在命令栏中直接输入“LS Y C X2 X3 X4 X5 X6”，回车即出现回归结果（见下页图）。估计结果按规范格式写为：

$$\hat{Y}_i = 1807.610 + 0.078621X_2 + 0.125787X_3 + 25.78046X_4 - 19.21102X_5 - 0.054383X_6$$

(967.2413)　(0.007994)　(0.010103)　(5.426062)　（9.461518）　(0.025146)

t= (1.868831)　(9.835376)　(12.45009)　(4.751229)　(-2.030438)　(-2.162720)

$R^2 = 0.988633$　$\bar{R}^2 = 0.98636$　F=434.8814　n=31

Eviews输出结果

Equation: UNTITLED Workfile: UNTITLED::Untitled\

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

Dependent Variable: Y
Method: Least Squares
Date: 09/08/19 Time: 16:01
Sample: 1 31
Included observations: 31

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1807.610	967.2413	1.868831	0.0734
X2	0.078621	0.007994	9.835376	0.0000
X3	0.125787	0.010103	12.45009	0.0000
X4	25.78046	5.426062	4.751229	0.0001
X5	-19.21102	9.461518	-2.030438	0.0531
X6	-0.054383	0.025146	-2.162720	0.0403

R-squared	0.988633	Mean dependent var	858.8729
Adjusted R-squared	0.986360	S.D. dependent var	498.2165
S.E. of regression	58.18699	Akaike info criterion	11.13719
Sum squared resid	84643.16	Schwarz criterion	11.41473
Log likelihood	-166.6264	Hannan-Quinn criter.	11.22766
F-statistic	434.8814	Durbin-Watson stat	2.140347
Prob(F-statistic)	0.000000		

注：估计方程还有一种方法，按着“Ctrl”键，先选被解释变量，依次选择需要的解释变量，右键“Open as equation”同样可以估计参数（默认带截距项c），尤其适用于变量较多的情况下。

另外：对回归结果中所有统计量的含义，以及不同统计量之间的关系要全面掌握！可参照练习题3.4进行熟悉。

模型检验：

1. 经济意义检验

模型估计结果表明，在假定其它变量不变的情况下，地区年末人口每增长1万人，平均说来地方一般公共预算教育支出会增长0.078621亿元；地区地方一般公共预算收入每增长1亿元，平均说来地方一般公共预算教育支出将增长0.125787亿元；当教育支出在地方财政支出中的比重增加1个百分点，平均说来地方一般公共预算教育支出会增长25.78046亿元；当教育消费价格指数增加1个百分点，平均说来地方一般公共预算教育支出会减少19.21102亿元；当居民人均教育文化娱乐消费每增加1元，平均说来地方一般公共预算教育支出可减少0.054383亿元；这与理论分析和经验判断基本相一致。

模型检验：

2. 统计检验

（1）拟合优度：由图 3.2 中数据可以得到：$R^2 = 0.988633$，修正的可决系数为 $\bar{R}^2 = 0.98636$，这说明模型对样本的拟合很好。

（2）F 检验：针对 $H_0 : \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = 0$，给定显著性水平 $\alpha = 0.05$，在 F 分布表中查出自由度为 k-1=5 和 n-k=25 的临界值 $F_\alpha(5,25) = 2.61$。由图 3.2 中得到 F=434.8814，由于 F=434.8814> $F_\alpha(5,25) = 2.61$，应拒绝原假设 $H_0 : \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = 0$，说明回归方程显著，即 “年末人口数”、“地方一般公共预算收入”、“教育支出在地方一般公共预算支出中的比重”、“教育消费价格指数”、“居民人均教育文化娱乐消费” 等变量联合起来，确实对 “地方一般公共预算教育支出” 有显著影响。

模型检验：

（3）t 检验：分别针对H_0：$\beta_j=0$（$j=1,2,3,4,5,6$），若给定显著性水平$\alpha=0.05$，查 t 分布表得自由度为 n-k=25 临界值$t_{0.05/2}(n-k)=2.060$；若给定显著性水平$\alpha=0.10$，查 t 分布表得自由度为 n-k=25 临界值$t_{0.10/2}(n-k)=1.708$。由图 3.2 中数据可得，$\hat{\beta}_2$、$\hat{\beta}_3$、$\hat{\beta}_4$、$\hat{\beta}_6$所对应的 t 统计量分别为 9.835376、12.45009、4.751229、-2.162720，其绝对值均大于$t_{0.05/2}(n-k)=2.060$，这说明在显著性水平$\alpha=0.05$下，分别都应当拒绝H_0：$\beta_j=0$（$j=2,3,4,6$），也就是说，在其它解释变量不变的情况下，解释变量 “年末人口数”（X_2）、“地区一般公共预算收入”（X_3）、“教育支出在一般公共预算支出中的比重”（X_4）、“居民人均教育文化娱乐消费”（X_6）分别对被解释变量“地方一般员工工预算教育支出” Y 都有显著的影响。

t 检验（续）

除此以外，与 $\hat{\beta}_1$、 $\hat{\beta}_5$ 对应的 t 统计量分别为 1.868831 和-2.030438，其绝对值均小于 $t_{0.05/2}(n-k)=2.060$，也就是说在给定显著性水平 $\alpha=0.05$ 下，还不能拒绝 $H:\beta_6=0$ 。可是，与 $\hat{\beta}_1$、 $\hat{\beta}_5$ 对应的 t 统计量的绝对值都大于 $t_{0.10/2}(n-k)=1.708$，也就是说在 $\alpha=0.10$ 的显著性水平下，可以拒绝 H_0： $\beta_j=0$ $(j=1,5)$ ，认为“教育消费价格指数”对“地方财政一般公共预算教育支出”Y 有显著的影响。这样的结论从表 3.4 中的 P 值也可以判断，与 $\hat{\beta}_2$、$\hat{\beta}_3$、$\hat{\beta}_4$、$\hat{\beta}_6$ 估计值对应的 P 值均小于 $\alpha=0.05$,表明在 $\alpha=0.05$ 的显著性水平下，对应解释变量对被解释变量影响显著。与 $\hat{\beta}_1$ 和 $\hat{\beta}_5$ 估计值对应的 P 值为 0.0734 和 0.0531，虽然大于 $\alpha=0.05$，但是小于 $\alpha=0.10$，表明在 $\alpha=0.10$ 的显著性水平下，“教育消费价格指数”对“地方财政一般公共预算教育支出”Y 影响是显著的。

本章作业

本科教材练习题3.4和3.5

（P89，注意提交纸质版，要求附计算公式与步骤）