第 5 章　多元线性回归

5.1 二元线性回归

一元回归很可能遗漏了其他因素。

比如，在第 4 章关于教育投资回报率的研究中，将工资对数对教育年限回归。

但工资还依赖于个人能力，而个人能力未包括在回归方程中，故被纳入扰动项。

能力强的人通常上学更久(二者正相关)，故一元回归所估计的教育回报率事实上也包括了对能力的回报，导致估计出现偏差。

其他遗漏变量还包括年龄、工龄、性别、种族、相貌等。

本章考虑多元回归。首先考察比较简单的二元回归。

$$y_i = \alpha + \beta x_{i1} + \gamma x_{i2} + \varepsilon_i \quad (i = 1, \cdots, n) \qquad (5.1)$$

x_{i1}与x_{i2}为解释变量；α为截距项；

β为在给定x_2条件下，x_1对y的边际效应(忽略扰动项ε_i)；

γ为在给定x_1条件下，x_2对y的边际效应。

OLS 估计量的最优化问题仍为残差平方和最小化：

$$\min_{\hat{\alpha},\hat{\beta},\hat{\gamma}} \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \hat{\alpha} - \hat{\beta} x_{i1} - \hat{\gamma} x_{i2})^2 \qquad (5.2)$$

几何上，这意味着要寻找一个回归平面 $\hat{y}_i \equiv \hat{\alpha} + \hat{\beta} x_{1i} + \hat{\gamma} x_{2i}$，即估计参数 $\hat{\alpha}, \hat{\beta}, \hat{\gamma}$，使得所有样本点 $\{(x_{1i}, x_{2i}, y_i)\}_{i=1}^{n}$ 离此回归平面最近，参见图 5.1。

将表达式(5.2)分别对 $\hat{\alpha}, \hat{\beta}, \hat{\gamma}$ 求偏导数，可得此最小化问题的一阶条件，求解可获得 $\hat{\alpha}, \hat{\beta}, \hat{\gamma}$ 的 OLS 估计量。

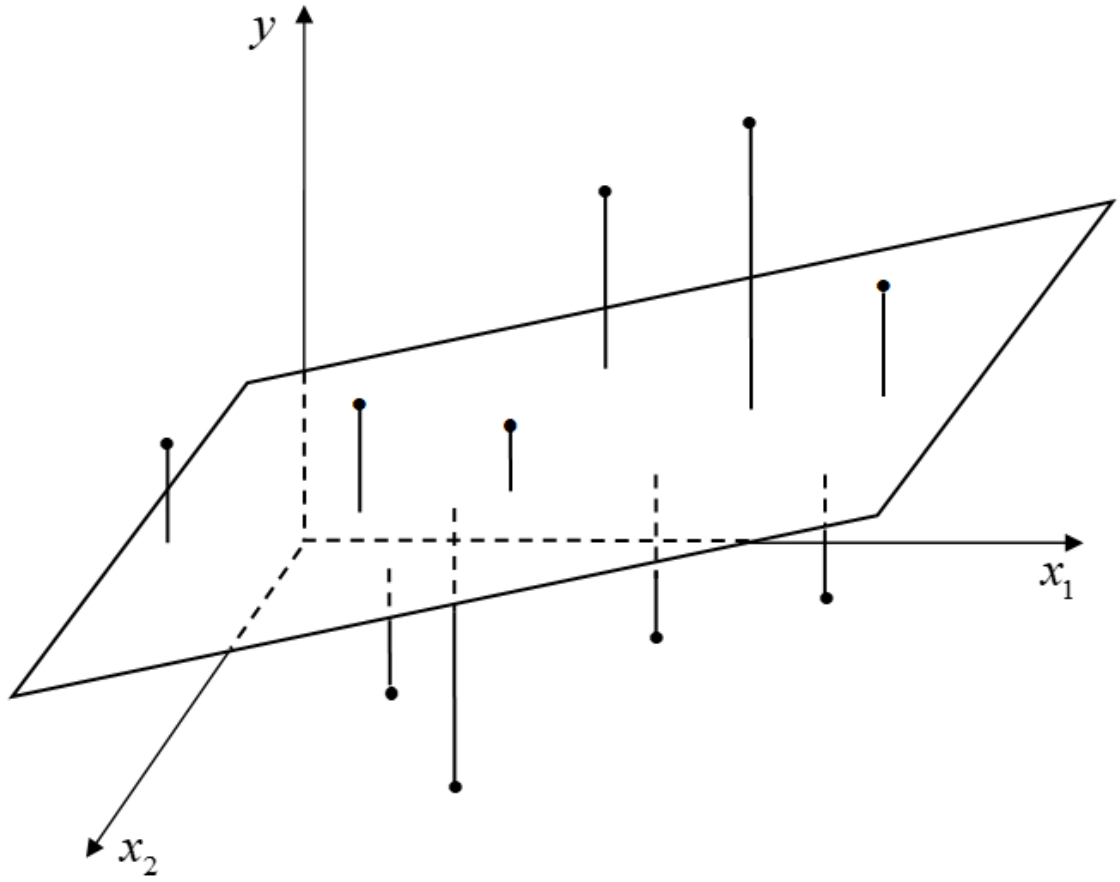

图 5.1 二元线性回归的示意图

例 (Cobb-Douglas 生产函数) Cobb and Douglas (1928)使用美国1899-1922 年制造业产出(y)、资本(k)与劳动力(l)的数据，估计如下生产函数：

$$y_t = \alpha k_t^{\beta} l_t^{\gamma} e^{\varepsilon_t} \qquad (5.3)$$

其中，e^{ε_t} 为乘积形式的扰动项，而下标 t 表示时间(年)。在方程两边同时取对数，可转换为线性模型：

$$\ln y_t = \ln \alpha + \beta \ln k_t + \gamma \ln l_t + \varepsilon_t \qquad (5.4)$$

数据集 cobb_douglas.dta 提供了 Cobb and Douglas (1928)的原始数据。首先看一下数据集中的观测值。

```
. use cobb_douglas.dta, clear
. list
```

	year	k	l	y	lnk	lnl	lny
1.	1899	100	100	100	4.60517	4.60517	4.60517
2.	1900	107	105	101	4.672829	4.65396	4.61512
3.	1901	114	110	112	4.736198	4.70048	4.718499
4.	1902	122	118	122	4.804021	4.770685	4.804021
5.	1903	131	123	124	4.875197	4.812184	4.820282
6.	1904	138	116	122	4.927254	4.75359	4.804021
7.	1905	149	125	143	5.003946	4.828314	4.962845
8.	1906	163	133	152	5.09375	4.890349	5.02388
9.	1907	176	138	151	5.170484	4.927254	5.01728
10.	1908	185	121	126	5.220356	4.795791	4.836282
11.	1909	198	140	155	5.288267	4.941642	5.043425
12.	1910	208	144	159	5.337538	4.969813	5.068904
13.	1911	216	145	153	5.375278	4.976734	5.030438
14.	1912	226	152	177	5.420535	5.02388	5.17615
15.	1913	236	154	184	5.463832	5.036952	5.214936
16.	1914	244	149	169	5.497168	5.003946	5.129899
17.	1915	266	154	189	5.583496	5.036952	5.241747
18.	1916	298	182	225	5.697093	5.204007	5.416101
19.	1917	335	196	227	5.81413	5.278115	5.42495
20.	1918	366	200	223	5.902633	5.298317	5.407172
21.	1919	387	193	218	5.958425	5.26269	5.384495
22.	1920	407	193	231	6.008813	5.26269	5.442418
23.	1921	417	147	179	6.033086	4.990433	5.187386
24.	1922	431	161	240	6.066108	5.081404	5.480639

变量 *k*，*l* 与 *y* 均将 1899 年的取值标准化为 100(以 1899 年为指数的基期)，而 *lnk*，*lnl* 与 *lny* 分别为其对数值。

在 Stata 中进行二元回归的命令为

```
. regress y x1 x2
```

其中，“y”为被解释变量，而“x1 x2”为解释变量。

对方程(5.4)进行二元回归估计，可输入如下命令

```
.reg lny lnk lnl
```

Source	SS	df	MS	Number of obs	=	24
				F(2, 21)	=	236.12
Model	1.59622155	2	.798110773	Prob > F	=	0.0000
Residual	.070981736	21	.003380083	R-squared	=	0.9574
				Adj R-squared	=	0.9534
Total	1.66720328	23	.072487099	Root MSE	=	.05814

lny	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
lnk	.2330537	.0635298	3.67	0.001	.1009363	.3651711
lnl	.807278	.1450762	5.56	0.000	.5055755	1.108981
_cons	-.1773099	.4342933	-0.41	0.687	-1.080472	.7258525

lnk (资本对数)与 *lnl* (劳动力对数)的系数分别为 0.233 与 0.807，且拟合优度R^2高达 0.957。

根据上表的回归结果，可得样本回归平面：

$$\widehat{\ln y_t} = -0.177 + 0.233 \ln k_t + 0.807 \ln l_t \qquad (5.5)$$

$\widehat{\ln y_t}$ 为 $\ln y_t$ 的拟合值或预测值。

由于 $\hat{\beta} = 0.233 = \dfrac{\partial \ln y_t}{\partial \ln k_t} = \dfrac{\partial y_t / y_t}{\partial k_t / k_t} \approx \dfrac{\Delta y_t / y_t}{\Delta k_t / k_t}$，故当 k_t 增加 1%

时，$\dfrac{\Delta y_t}{y_t} \approx \hat{\beta} \cdot \dfrac{\Delta k_t}{k_t} = 0.233 \times 1\% = 0.233\%$。

故 $\hat{\beta} = 0.233$ 可解释为弹性(elasticity)。

在 Stata 中，做完 OLS 回归后，可用命令 predict 来计算拟合值与残差。

```
. predict lny1
(option xb assumed; fitted values)
```

此命令将 *lny* 的拟合值记为 *lny1*。如果要计算残差，并记为 *e*，可输入命令

```
. predict e,residual
```

其中，选择项“residual”表示计算残差(如果省略此选择项，则默认为计算拟合值)。将 *lny* 及其拟合值、残差同时列表。

```
. list lny lny1 e
```

	lny	lny1	e
1.	4.60517	4.613595	-.0084246
2.	4.61512	4.66875	-.0536295
3.	4.718499	4.721073	-.0025745
4.	4.804021	4.793554	.010467
5.	4.820282	4.843644	-.0233621
6.	4.804021	4.808474	-.0044528
7.	4.962845	4.88667	.0761749
8.	5.02388	4.957679	.0662019
9.	5.01728	5.005354	.0119254
10.	4.836282	4.91085	-.074568
11.	5.043425	5.04442	-.0009945
12.	5.068904	5.078644	-.0097398
13.	5.030438	5.093026	-.0625884
14.	5.17615	5.141634	.0345157
15.	5.214936	5.162277	.0526584
16.	5.129899	5.143401	-.0135028
17.	5.241747	5.190166	.0515813
18.	5.416101	5.351499	.0646015
19.	5.42495	5.438601	-.0136506
20.	5.407172	5.475536	-.0683641
21.	5.384495	5.459777	-.0752818
22.	5.442418	5.47152	-.0291027
23.	5.187386	5.25739	-.0700038
24.	5.480639	5.338525	.142114

将产出对数及其拟合值画在一起(结果参见图 5.2)：

```
. line lny lny1 year,lp(solid dash)
```

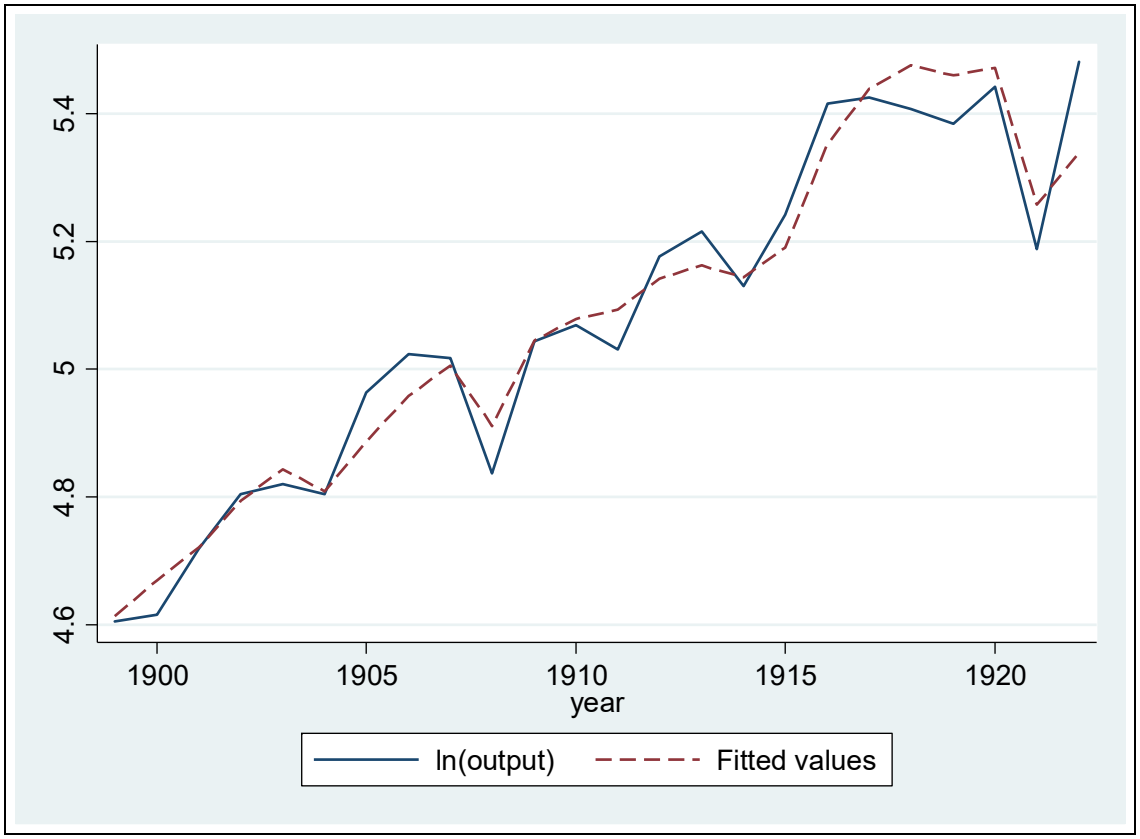

图 5.2 产出对数的实际值与预测值

5.2 多元线性回归模型

一般的多元回归模型可写为

$$y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_K x_{iK} + \varepsilon_i \quad (i = 1, \cdots, n) \qquad (5.6)$$

x_{i1}为个体 i 的第 1 个解释变量，x_{i2}为个体 i 的第 2 个解释变量，以此类推。

x_{ik}的第一个下标表示个体 i (共有n位个体，即样本容量为n)，而第二个下标表示第 k 个解释变量 (共有K个解释变量)。

在绝大多数情况下，回归方程都有常数项，故通常令$x_{i1} \equiv 1$ (恒等于 1)，则方程(5.6)简化为

$$y_i = \beta_1 + \beta_2 x_{i2} + \cdots + \beta_K x_{iK} + \varepsilon_i \qquad (5.7)$$

定义列向量$\mathbf{x}_i \equiv (1 \;\; x_{i2} \;\; \cdots \;\; x_{iK})'$(包含个体 i 的全部解释变量)

参数向量$\boldsymbol{\beta} \equiv (\beta_1 \;\; \beta_2 \;\; \cdots \;\; \beta_K)'$(包含全部回归系数)

则$\sum_{k=1}^{K} \beta_k x_{ik} = \mathbf{x}_i' \boldsymbol{\beta}$。故可将原模型(5.7)写为

$$y_i = \begin{pmatrix} 1 & x_{i2} & \cdots & x_{iK} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_K \end{pmatrix} + \varepsilon_i = \mathbf{x}_i'\boldsymbol{\beta} + \varepsilon_i \qquad (5.8)$$

上式对所有个体i都成立$(i = 1, \cdots, n)$，故有n个形如(5.8)的方程。将所有这n个方程都叠放在一起可得

$$\begin{pmatrix} y_1 = \mathbf{x}_1'\boldsymbol{\beta} + \varepsilon_1 \\ y_2 = \mathbf{x}_2'\boldsymbol{\beta} + \varepsilon_2 \\ \vdots \\ y_n = \mathbf{x}_n'\boldsymbol{\beta} + \varepsilon_n \end{pmatrix} \qquad (5.9)$$

将共同的参数向量$\boldsymbol{\beta}$向右边提出，经整理可得

$$\mathbf{y} \equiv \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \underbrace{\begin{pmatrix} \mathbf{x}_1' \\ \mathbf{x}_2' \\ \vdots \\ \mathbf{x}_n' \end{pmatrix}}_{\mathbf{X}} \boldsymbol{\beta} + \underbrace{\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}}_{\boldsymbol{\varepsilon}} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \qquad (5.10)$$

$\mathbf{y} \equiv \begin{pmatrix} y_1 & y_2 & \cdots & y_n \end{pmatrix}'$为被解释变量构成的列向量

$\boldsymbol{\varepsilon} \equiv \begin{pmatrix} \varepsilon_1 & \varepsilon_2 & \cdots & \varepsilon_n \end{pmatrix}'$为所有扰动项构成的列向量

$\mathbf{X}$为$n \times K$**数据矩阵**(data matrix)，其第 i 行包含个体 i 的全部解释变量，而第 k 列包含第 k 个解释变量的全部观测值，即

$$\mathbf{X} \equiv \begin{pmatrix} 1 & x_{12} & \cdots & x_{1K} \\ 1 & x_{22} & \cdots & x_{2K} \\ \cdots & \cdots & \cdots & \cdots \\ 1 & x_{n2} & \cdots & x_{nK} \end{pmatrix}_{n \times K} \tag{5.11}$$

5.3 OLS 估计量的推导

对于多元回归模型，OLS 估计量的最小化问题为

$$\min_{\hat{\beta}_1,\cdots,\hat{\beta}_K} \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \hat{\beta}_1 - \hat{\beta}_2 x_{i2} - \hat{\beta}_3 x_{i3} - \cdots - \hat{\beta}_K x_{iK})^2 \quad (5.12)$$

最小二乘法寻找使残差平方和(SSR) $\sum_{i=1}^{n} e_i^2$ 最小的 $(\hat{\beta}_1, \hat{\beta}_2, \cdots, \hat{\beta}_K)$。

在几何上，一元回归寻找最佳拟合的回归直线，使得观测值 y_i 到该回归直线的距离之平方和最小。

二元回归寻找最佳拟合的回归平面；而多元回归则寻找最佳拟合的回归超平面(hyperplane)。

此最小化问题的一阶条件为

$$\begin{cases} \dfrac{\partial}{\partial \hat{\beta}_1}\sum_{i=1}^{n} e_i^2 = -2\sum_{i=1}^{n}(y_i - \hat{\beta}_1 - \hat{\beta}_2 x_{i2} - \cdots - \hat{\beta}_K x_{iK}) = 0 \\ \dfrac{\partial}{\partial \hat{\beta}_2}\sum_{i=1}^{n} e_i^2 = -2\sum_{i=1}^{n}(y_i - \hat{\beta}_1 - \hat{\beta}_2 x_{i2} - \cdots - \hat{\beta}_K x_{iK}) x_{i2} = 0 \\ \vdots \\ \dfrac{\partial}{\partial \hat{\beta}_K}\sum_{i=1}^{n} e_i^2 = -2\sum_{i=1}^{n}(y_i - \hat{\beta}_1 - \hat{\beta}_2 x_{i2} - \cdots - \hat{\beta}_K x_{iK}) x_{iK} = 0 \end{cases} \tag{5.13}$$

消去方程左边的“–2”可得

$$\begin{cases} \sum_{i=1}^{n}(y_i-\hat{\beta}_1-\hat{\beta}_2x_{i2}-\cdots-\hat{\beta}_Kx_{iK})=0 \\ \sum_{i=1}^{n}x_{i2}(y_i-\hat{\beta}_1-\hat{\beta}_2x_{i2}-\cdots-\hat{\beta}_Kx_{iK})=0 \\ \vdots \\ \sum_{i=1}^{n}x_{iK}(y_i-\hat{\beta}_1-\hat{\beta}_2x_{i2}-\cdots-\hat{\beta}_Kx_{iK})=0 \end{cases} \tag{5.14}$$

这是一个包含 K 个未知数$(\hat{\beta}_1,\hat{\beta}_2,\cdots,\hat{\beta}_K)$与 K 个方程的联立方程组，称为**正规方程组**(normal equations)。

满足此正规方程组的$\hat{\boldsymbol{\beta}}\equiv(\hat{\beta}_1\ \ \hat{\beta}_2\cdots\hat{\beta}_K)'$称为 **OLS 估计量**(Ordinary Least Squares，简记 OLS)。

由于残差$e_i \equiv y_i - \hat{\beta}_1 - \hat{\beta}_2 x_{i2} - \cdots - \hat{\beta}_K x_{iK}$，正规方程组可写为

$$\begin{cases} \sum_{i=1}^{n} e_i = 0 \\ \sum_{i=1}^{n} x_{i2} e_i = 0 \\ \vdots \\ \sum_{i=1}^{n} x_{iK} e_i = 0 \end{cases} \tag{5.15}$$

上式每一方程都是乘积求和的形式，故可用向量内积表示。

比如，第 1 个方程可写为

$$\sum_{i=1}^{n} e_i = \begin{pmatrix} 1 & 1 & \cdots & 1 \end{pmatrix} \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix} = 0 \qquad (5.16)$$

而第 2 个方程可写为

$$\sum_{i=1}^{n} x_{i2} e_i = \begin{pmatrix} x_{12} & x_{22} & \cdots & x_{n2} \end{pmatrix} \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix} = 0 \qquad (5.17)$$

以此类推，第 K 个方程可写为

$$\sum_{i=1}^{n} x_{iK} e_i = \begin{pmatrix} x_{1K} & x_{2K} & \cdots & x_{nK} \end{pmatrix} \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix} = 0 \qquad (5.18)$$

由此可知，残差向量$\mathbf{e} \equiv \begin{pmatrix} e_1 & e_2 & \cdots & e_n \end{pmatrix}'$与每个解释变量都正交，这是 OLS 估计量的一大特征。

将以上内积以矩阵形式表示可得

$$\underbrace{\begin{pmatrix} 1 & 1 & \cdots & 1 \\ x_{12} & x_{22} & \cdots & x_{n2} \\ \cdots & \cdots & \cdots & \cdots \\ x_{1K} & x_{2K} & \cdots & x_{nK} \end{pmatrix}}_{\mathbf{X}'} \underbrace{\begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix}}_{\mathbf{e}} = \underbrace{\begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}}_{\mathbf{0}} \qquad (5.19)$$

$\mathbf{X}'$为数据矩阵$\mathbf{X}$的转置。正规方程组可简洁地写为

$$\mathbf{X}'\mathbf{e} = \mathbf{0} \qquad (5.20)$$

由于$\mathbf{X}'$的第 k 行包含第 k 个解释变量的全部观测值，故根据$\mathbf{X}'\mathbf{e} = \mathbf{0}$也可看出，残差向量$\mathbf{e}$与每个解释变量都正交。

从$e_i \equiv y_i - (\hat{\beta}_1 + \hat{\beta}_2 x_{i2} + \cdots + \hat{\beta}_K x_{iK})$出发，可将残差向量写为(参见习题)

$$\mathbf{e} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} \qquad (5.21)$$

将上式代入正规方程组(5.20)可得

$$\mathbf{X}'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) = \mathbf{0} \qquad (5.22)$$

乘开来，并移项可知，最小二乘估计量$\hat{\boldsymbol{\beta}}$满足：

$$(\mathbf{X}'\mathbf{X})_{K\times K}\hat{\boldsymbol{\beta}}_{K\times 1} = \mathbf{X}'_{K\times n}\mathbf{y}_{n\times 1} \qquad (5.23)$$

假设$(\mathbf{X}'\mathbf{X})^{-1}$存在，可求解 OLS 估计量：

$$\hat{\boldsymbol{\beta}} \equiv (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \qquad (5.24)$$

5.4 OLS 的几何解释

定义被解释变量 y_i 的**拟合值**(fitted value)或**预测值**(predicted value)为

$$\hat{y}_i \equiv \hat{\beta}_1 + \hat{\beta}_2 x_{i2} + \cdots + \hat{\beta}_K x_{iK} \quad (i = 1, \cdots, n) \qquad (5.25)$$

将所有个体的拟合值写为列向量$\hat{\mathbf{y}}$，并参照方程(5.10)同样的推导可得

$$\hat{\mathbf{y}} \equiv \begin{pmatrix} \hat{y}_1 & \hat{y}_2 & \cdots & \hat{y}_n \end{pmatrix}' = \mathbf{X}\hat{\boldsymbol{\beta}} \qquad (5.26)$$

拟合值向量与残差向量正交，因为

$$\hat{\mathbf{y}}'\mathbf{e} = (\mathbf{X}\hat{\boldsymbol{\beta}})'\mathbf{e} = \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{e} = \hat{\boldsymbol{\beta}}' \cdot \mathbf{0} = 0 \qquad (5.27)$$

由于$\mathbf{e} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{y} - \hat{\mathbf{y}}$，故

$$\mathbf{y} = \hat{\mathbf{y}} + \mathbf{e}$$

被解释变量$\mathbf{y}$可分解为相互正交的拟合值$\hat{\mathbf{y}}$与残差$\mathbf{e}$之和。

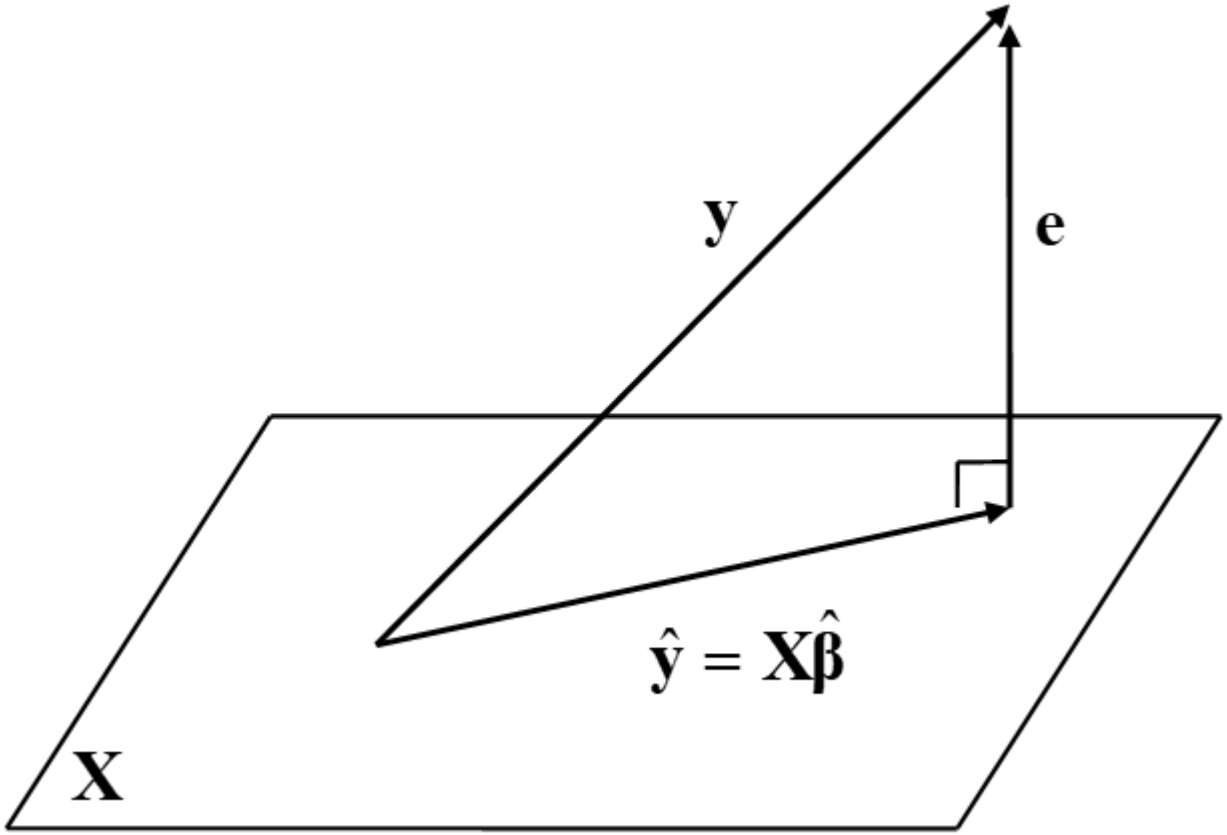

图 5.3 最小二乘法的正交性

在图 5.3 中，拟合值$\hat{\mathbf{y}}$可视为被解释变量$\mathbf{y}$向解释变量超平面$\mathbf{X}$的**线性投影**(linear projection)。

由于拟合值为解释变量的线性组合，即$\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$，故拟合值向量$\hat{\mathbf{y}}$正好在超平面$\mathbf{X}$上。

根据 OLS 的正交性，残差向量$\mathbf{e}$与$\hat{\mathbf{y}}$正交。

图 5.3 可视为 OLS 的几何解释。

5.5 拟合优度

对于多元回归，在回归方程有常数项的情况下，由于 OLS 的正交性，平方和分解公式依然成立 (证明方法与一元回归相同)。

仍可将被解释变量的离差平方和$\sum_{i=1}^{n}(y_i-\bar{y})^2$分解如下：

$$\underbrace{\sum_{i=1}^{n}(y_i-\bar{y})^2}_{\text{TSS}}=\underbrace{\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2}_{\text{ESS}}+\underbrace{\sum_{i=1}^{n}e_i^2}_{\text{RSS}} \qquad (5.28)$$

ESS 为模型可解释的部分，而 RSS 为模型不可解释的部分。

根据平方和分解公式(5.28)，可定义拟合优度。

定义 拟合优度R^2为

$$0 \le R^2 \equiv \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} = 1 - \frac{\sum_{i=1}^{n} e_i^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \le 1 \qquad (5.29)$$

拟合优度R^2的缺点是，如果增加解释变量的数目，则R^2只增不减，因为至少可让新增解释变量的系数为 0 而保持R^2不变。

通过最优地选择新增解释变量的系数(以及已有解释变量的系数)，通常可以提高R^2。

引入如下校正拟合优度，对解释变量过多(即模型不够简洁)进行惩罚。

定义 校正拟合优度(adjusted R^2) $\bar{R}^2$ 为

$$\bar{R}^2 \equiv 1 - \frac{\sum_{i=1}^{n} e_i^2 \Big/ (n-K)}{\sum_{i=1}^{n} (y_i - \bar{y})^2 \Big/ (n-1)} \qquad (5.30)$$

$\sum_{i=1}^{n} e_i^2$ 的**自由度**(degree of freedom)为 $(n-K)$。

虽然 $\sum_{i=1}^{n} e_i^2$ 由 n 个随机变量 $\{e_1, \cdots, e_n\}$ 所构成，但 $\{e_1, \cdots, e_n\}$ 受由 K 个方程组成的正规方程组(5.19)的约束，故只有其中 $(n-K)$ 个残差可自由取值。

给定$\{e_1,\cdots,e_{n-K}\}$，即可根据正规方程组求解其余$\{e_{n-K+1},\cdots,e_n\}$。

$\sum_{i=1}^{n}(y_i-\bar{y})^2$的自由度为$(n-1)$。

虽然$\sum_{i=1}^{n}(y_i-\bar{y})^2$由$n$个离差$\{(y_1-\bar{y}),\cdots,(y_n-\bar{y})\}$所构成，但这些离差之和必然为 0，即$\sum_{i=1}^{n}(y_i-\bar{y})=0$，故只有其中$(n-1)$个离差可自由取值。

定义式(5.30)通过调整自由度，达到对模型过于复杂的惩罚。

$\bar{R}^2$的缺点是，它可能为负值。

无论R^2还是$\bar{R}^2$，只反映拟合程度的好坏，除此并无太多意义。

评估回归方程是否显著，应使用 F 检验(R^2与 F 统计量也有联系)。

如果回归模型无常数项，则仍须使用**非中心** R^2 (uncentered R^2)：

$$R_{uc}^2 \equiv \frac{\sum_{i=1}^{n} \hat{y}_i^2}{\sum_{i=1}^{n} y_i^2} \qquad (5.31)$$

5.6 古典线性回归模型的假定

为了得到 OLS 估计量的良好性质，**古典线性回归模型** (Classical Linear Regression Model) 作了如下假定。

假定 5.1 线性假定(linearity)。总体模型为

$$y_i = \beta_1 + \beta_2 x_{i2} + \cdots + \beta_K x_{iK} + \varepsilon_i \quad (i = 1, \cdots, n) \qquad (5.32)$$

线性假设的含义是每个解释变量对y_i的边际效应为常数，比如$\dfrac{\partial y_i}{\partial x_{i2}} = \beta_2$(忽略扰动项$\varepsilon_i$)。

如果边际效应可变，可加入平方项(比如x_{i2}^2)或交叉项(比如$x_{i2}x_{i3}$)。交叉项也称为交互项(interaction term)。

例(平方项)：考虑如下回归方程，

$$\ln w_i = \beta_1 + \beta_2 s_i + \beta_3 s_i^2 + \varepsilon_i \qquad (5.33)$$

其中，$\ln w_i$为工资对数，s_i为教育年限。则教育年限对工资对数的边际效应为(忽略扰动项)

$$\frac{\partial \ln w_i}{\partial s_i} = \beta_2 + 2\beta_3 s_i \qquad (5.34)$$

如果$\beta_3 < 0$，则存在教育投资回报率递减。

反之，如果$\beta_3 > 0$，则存在教育投资回报率递增。

只要将s^2也视为解释变量(根据s可算出s^2的取值)，则仍然符合线性模型的假定。

例(交互项)：考虑如下生产函数方程，

$$y_i = \beta_1 + \beta_2 k_i + \beta_3 l_i + \beta_4 k_i \times l_i + \varepsilon_i \qquad (5.35)$$

其中，y为产出，k为资本，l为劳动力，而$k \times l$为资本与劳动力的互动项。

劳动力的边际产出为$\frac{\partial y_i}{\partial l_i} = \beta_3 + \beta_4 k_i$(忽略扰动项)。

如果$\beta_4 > 0$，则说明资本与劳动力是互补的，即随着资本上升，劳动力的边际产出也增加。

只要将$k \times l$也视为解释变量，则依然符合线性模型的假定。

例(函数形式)：经济学中常用的生产函数并非方程(5.35)，而是Cobb-Douglas 生产函数：

$$y_i = e^{\beta_1} k_i^{\beta_2} l_i^{\beta_3} e^{\varepsilon_i} \qquad (5.36)$$

其中，e^{β_1}与e^{ε_i}分别为乘积形式的常数项与扰动项。

将上式两边同时取对数可得

$$\ln y_i = \beta_1 + \beta_2 \ln k_i + \beta_3 \ln l_i + \varepsilon_i \qquad (5.37)$$

β_2为产出的资本弹性，即资本每增加 1%，产出平均增加百分之几；

β_3为产出的劳动力弹性，即劳动力每增加 1%，产出平均增加百分之几。

只要将$\ln y_i$视为被解释变量，而将$\ln k_i$与$\ln l_i$视为解释变量，则仍符合线性模型的假定。

只要将回归方程中变量的高次项(比如x^2)或函数(比如$\ln x$)都作为变量来看待，则依然满足线性假定。

线性假定的本质要求是，回归函数是参数$(\beta_1 \cdots \beta_K)$的线性函数(linear in parameters)。

假定 5.2 严格外生性(strict exogeneity)要求

$$\mathrm{E}(\varepsilon_i \mid \mathbf{X}) = \mathrm{E}(\varepsilon_i \mid \mathbf{x}_1, \cdots, \mathbf{x}_n) = 0 \quad (i = 1, \cdots, n)$$

严格外生性意味着，在给定数据矩阵$\mathbf{X}$的情况下，扰动项ε_i的条件期望为 0。

ε_i均值独立于(mean-independent)所有解释变量的观测数据，而不仅仅是同一观测数据$\mathbf{x}_i$中的解释变量。

这意味着，ε_i与所有个体的解释变量都不相关，即$\mathrm{Cov}(\varepsilon_i, x_{jk}) = 0,\ \forall\ j, k$。

此假定很强，在第 6 章大样本 OLS 可放松。

均值独立仅要求$\mathrm{E}(\varepsilon_i \mid \mathbf{X}) = c$，其中 c 为常数，不一定为 0。

当回归方程有常数项时，要求$\mathrm{E}(\varepsilon_i \mid \mathbf{X}) = 0$并不会带来过多限制，因为如果$\mathrm{E}(\varepsilon_i \mid \mathbf{X}) = c \neq 0$，总可以把 c 归入常数项。

从$\mathrm{E}(\varepsilon_i \mid \mathbf{X}) = 0$出发，可证明扰动项的无条件期望也为0，因为

$$\mathrm{E}(\varepsilon_i) = \mathrm{E}_{\mathbf{X}} \underbrace{\mathrm{E}(\varepsilon_i \mid \mathbf{X})}_{=0} = \mathrm{E}_{\mathbf{X}}(0) = 0 \qquad (5.38)$$

从$\mathrm{Cov}(\varepsilon_i, x_{jk}) = 0$出发，可证明扰动项与解释变量“正交”。

在线性代数中，如两个向量的内积为0，则这两个向量正交。

这与在概率统计中，两个随机变量正交的定义有所不同。

定义 如果随机变量x, y满足$\mathrm{E}(xy) = 0$，则称x, y**正交**(orthogonal)。

解释变量与扰动项正交，因为

$$0 = \mathrm{Cov}(x_{jk}, \varepsilon_i) = \mathrm{E}(x_{jk}\varepsilon_i) - \mathrm{E}(x_{jk})\underbrace{\mathrm{E}(\varepsilon_i)}_{=0} = \mathrm{E}(x_{jk}\varepsilon_i) \qquad (5.39)$$

假设数据$(y_i, x_{i1}, \cdots, x_{iK})$为 iid 的随机样本，则严格外生性可简化为$\mathrm{E}(\varepsilon_i \mid \mathbf{x}_1, \cdots, \mathbf{x}_n) = \mathrm{E}(\varepsilon_i \mid \mathbf{x}_i) = 0$(由于不同个体相互独立，故$\varepsilon_i$与所有$\mathbf{x}_j$都相互独立，只要$i \neq j$)。

假定$\mathrm{E}(\varepsilon_i \mid \mathbf{x}_i) = 0$意味着$\mathbf{x}_i$与$\varepsilon_i$不相关，仿佛在$\varepsilon_i$取值之前，$\mathbf{x}_i$的取值已经确定，故称$\mathbf{x}_i$为**前定变量**(predetermined variable)。

对于时间序列，$\mathrm{E}(\varepsilon_t \mid \mathbf{x}_t) = 0$(将下标变为 t，表示时间)，故$\mathbf{x}_t$与ε_t不相关，称$\mathbf{x}_t$为**同期外生**(contemporaneously exogenous)。

随机样本的假设虽然容易理解，但可能过强，因为不同个体之间可能存在相关性(比如，时间序列通常存在自相关)。

假定 5.3 不存在**严格多重共线性**(strict multicolinearity)，即数据矩阵**X满列秩**(full column rank)，故列秩等于列数。

数据矩阵的各列向量为线性无关，即不存在某个解释变量为另一解释变量的倍数，或可由其他解释变量线性表出的情形。

X中不存在多余的变量。

考虑一元回归模型：

$$\ln w_i = \beta_1 + \beta_2 s_i + \varepsilon_i \quad (i = 1, \cdots, n) \qquad (5.40)$$

其数据矩阵$\mathbf{X}$为

$$\mathbf{X}=\begin{pmatrix}1 & s_1\\ 1 & s_2\\ \vdots & \vdots\\ 1 & s_n\end{pmatrix} \qquad (5.41)$$

数据矩阵$\mathbf{X}$满列秩要求，解释变量s_i不是常数项的固定倍数，即s_i应有变动，不能是常数。

如果所有个体的教育年限s_i都相同，则无法定义 OLS 估计量

$$\hat{\beta}=\frac{\sum_{i=1}^{n}(s_i-\bar{s})(\ln w_i-\overline{\ln w})}{\sum_{i=1}^{n}(s_i-\bar{s})^2} \qquad (5.42)$$

其中，$\bar{s}$与$\overline{\ln w}$分别为s与lnw的样本均值。

因为上式分母$\sum_{i=1}^{n}(s_i-\bar{s})^2=0$，故无法估计教育年限对工资对数的作用。

对于多元回归，如果$\mathbf{X}$满列秩，则$\mathbf{X}'\mathbf{X}$为**正定矩阵**(positive definite)，故$(\mathbf{X}'\mathbf{X})^{-1}$存在，可以计算OLS估计量$\hat{\boldsymbol{\beta}}=(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$。

反之，如果$\mathbf{X}$不满列秩，则$(\mathbf{X}'\mathbf{X})^{-1}$不存在，无法定义OLS估计量。

此时，称$\boldsymbol{\beta}$**不可识别**(unidentified)。

数据矩阵$\mathbf{X}$满列秩只是对数据的最低要求。

在现实数据中，并不容易出现严格多重共线性。

即使出现，Stata 也会自动识别，并去掉多余的变量。

5.7 OLS 的小样本性质

OLS 估计量$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$为样本数据的函数，故也是随机变量，其分布称为**抽样分布**(sampling distribution)。

在古典线性回归模型的假定 5.1-5.3 之下，OLS 估计量具有以下良好性质。

(1) 线性性：OLS 估计量$\hat{\boldsymbol{\beta}}$为**线性估计量**(linear estimator)。

从 OLS 估计量的表达式$\hat{\boldsymbol{\beta}}=(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$可知，$\hat{\boldsymbol{\beta}}$可视为$\mathbf{y}$的线性组合(将$[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']$视为系数矩阵)，故为线性估计量。

(2) 无偏性：$\mathrm{E}(\hat{\boldsymbol{\beta}}\,|\,\mathbf{X})=\boldsymbol{\beta}$，即$\hat{\boldsymbol{\beta}}$不会系统地高估或低估$\boldsymbol{\beta}$，参见图 5.4。

证明：抽样误差(sampling error)为

$$\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}=(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}-\boldsymbol{\beta}=(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\varepsilon})-\boldsymbol{\beta}=\underbrace{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'}_{\mathbf{A}}\boldsymbol{\varepsilon}\equiv\mathbf{A}\boldsymbol{\varepsilon} \tag{5.43}$$

其中，记$\mathbf{A} \equiv (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$。

给定解释变量$\mathbf{X}$，上式两边求条件期望，根据严格外生性可得

$$\mathrm{E}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} \mid \mathbf{X}) = \mathrm{E}(\mathbf{A}\boldsymbol{\varepsilon} \mid \mathbf{X}) = \mathbf{A}\underbrace{\mathrm{E}(\boldsymbol{\varepsilon} \mid \mathbf{X})}_{=\mathbf{0}} = \mathbf{0} \quad (5.44)$$

移项可得，$\mathrm{E}(\hat{\boldsymbol{\beta}} \mid \mathbf{X}) = \boldsymbol{\beta}$。

在此证明中，严格外生性不可或缺。

进一步，无条件期望$\mathrm{E}(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$，因为

$$\mathrm{E}(\hat{\boldsymbol{\beta}}) = \mathrm{E}_{\mathbf{X}}\mathrm{E}(\hat{\boldsymbol{\beta}} \mid \mathbf{X}) = \mathrm{E}_{\mathbf{X}}(\boldsymbol{\beta}) = \boldsymbol{\beta} \quad (5.45)$$

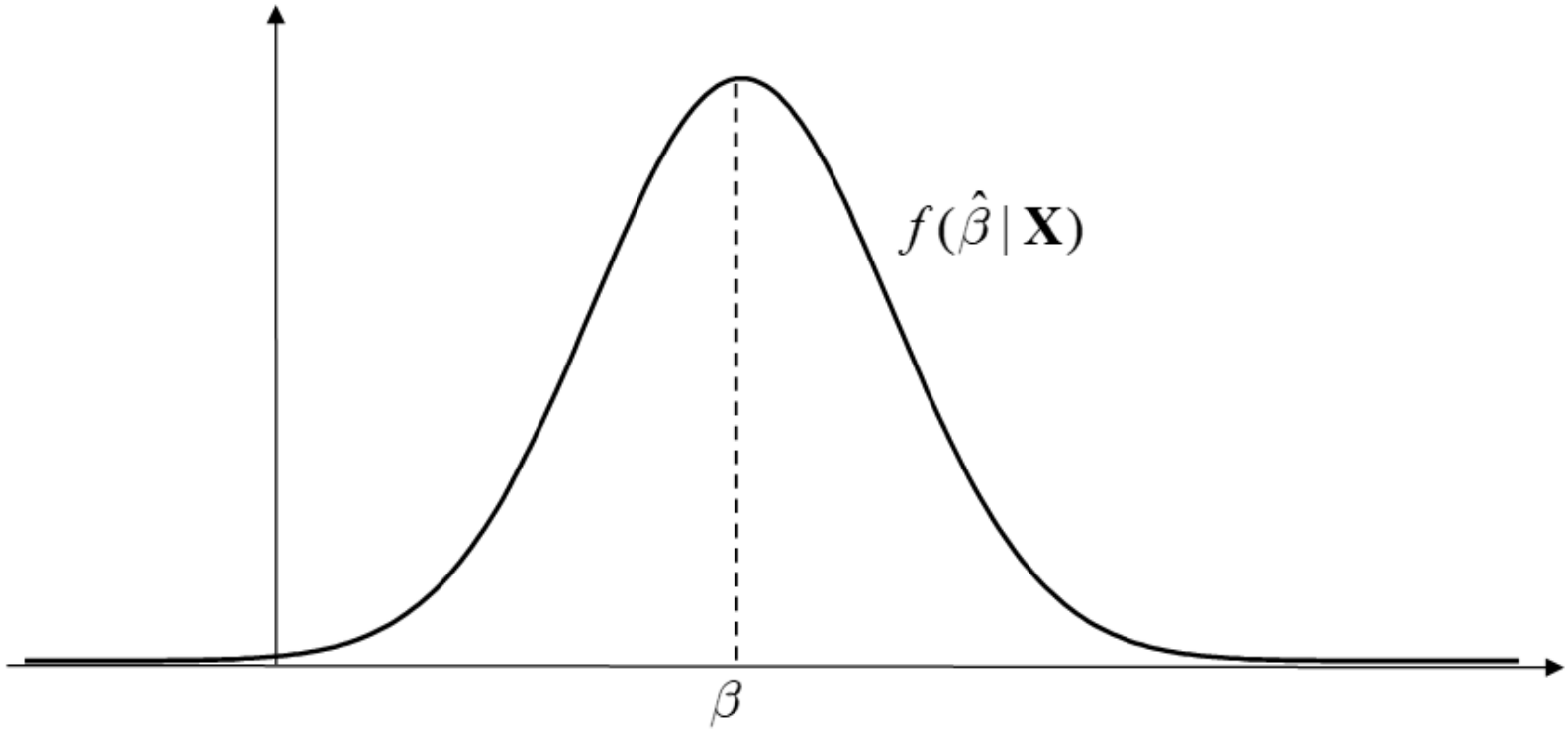

图 5.4 OLS 估计量$\hat{\beta}$的无偏性

(3) 估计量$\hat{\boldsymbol{\beta}}$的协方差矩阵

由于$\boldsymbol{\beta}$为常数，故$\mathrm{Var}(\hat{\boldsymbol{\beta}} \mid \mathbf{X}) = \mathrm{Var}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} \mid \mathbf{X})$，因此

$$\begin{aligned}\mathrm{Var}(\hat{\boldsymbol{\beta}} \mid \mathbf{X}) &= \mathrm{Var}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} \mid \mathbf{X}) = \mathrm{Var}(\mathbf{A}\boldsymbol{\varepsilon} \mid \mathbf{X}) \\ &= \mathbf{A}\,\mathrm{Var}(\boldsymbol{\varepsilon} \mid \mathbf{X})\mathbf{A}' = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathrm{Var}(\boldsymbol{\varepsilon} \mid \mathbf{X})\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\end{aligned} \tag{5.46}$$

上式使用了协方差矩阵的夹心估计量表达式，且$\mathbf{A} \equiv (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$，$\mathbf{A}' \equiv \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$。

古典模型对扰动项协方差矩阵$\mathrm{Var}(\boldsymbol{\varepsilon} \mid \mathbf{X})$作了最简单的假定，即$\mathrm{Var}(\boldsymbol{\varepsilon} \mid \mathbf{X}) = \sigma^2\mathbf{I}_n$，与单位矩阵成正比，称为“球形扰动项”。

在球形扰动项的假定下，表达式(5.46)可大大简化：

$$\begin{aligned}\mathrm{Var}(\hat{\boldsymbol{\beta}} \mid \mathbf{X}) &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\sigma^2\mathbf{I}_n)\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\ &= \sigma^2(\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{X})(\mathbf{X}'\mathbf{X})^{-1} = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\end{aligned} \tag{5.47}$$

假定 5.4 球型扰动项(spherical disturbance)，即扰动项满足“同方差”、“无自相关”的性质，故扰动项$\boldsymbol{\varepsilon}$的协方差矩阵可写为

$$\mathrm{Var}(\boldsymbol{\varepsilon} \mid \mathbf{X}) = \sigma^2\mathbf{I}_n = \begin{pmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma^2 \end{pmatrix} \tag{5.48}$$

其中，$\mathbf{I}_n$为n阶单位矩阵。

协方差矩阵$\mathrm{Var}(\boldsymbol{\varepsilon}\,|\,\mathbf{X})$的主对角线元素都等于$\sigma^2$，即满足**条件同方差**(conditional homoskedasticity)，简称**同方差**。

如果不完全相等，则存在**条件异方差**(conditional heteroskedasticity)，简称**异方差**。

协方差矩阵$\mathrm{Var}(\boldsymbol{\varepsilon}\,|\,\mathbf{X})$的非主对角线元素都为0，故不同个体的扰动项之间无**自相关**(autocorrelation)或**序列相关**(serial correlation)；反之，则存在自相关。

球型扰动项假定是证明协方差表达式$\sigma^2(\mathbf{X}'\mathbf{X})^{-1}$的关键(无偏性则不依赖于球形扰动项)。

如果存在条件异方差，则方差表达式有所不同，应使用“稳健标准误差”(robust standard error)，参见第 7 章。

引入球形扰动项假定的另一好处是，可以证明 OLS 估计量在某种范围内是最有效率的估计量，即方差最小。

(4) **高斯−马尔可夫定理**(Gauss-Markov Theorem)：

在假定 5.1-5.4 之下，最小二乘法是**最佳线性无偏估计**(Best Linear Unbiased Estimator，简记 BLUE)，即在所有线性的无偏估计中，最小二乘法的方差最小，参见图 5.5。

记 OLS 估计量为$\hat{\boldsymbol{\beta}}_{\text{OLS}}$，而任一线性无偏估计量为$\tilde{\boldsymbol{\beta}}$，则$[\text{Var}(\tilde{\boldsymbol{\beta}} \mid \mathbf{X}) - \text{Var}(\hat{\boldsymbol{\beta}}_{\text{OLS}} \mid \mathbf{X})]$为半正定矩阵。

高斯-马尔可夫定理的核心假设是球形扰动项；若不满足球形扰动项(存在异方差或自相关)，则高斯-马尔可夫定理不成立。

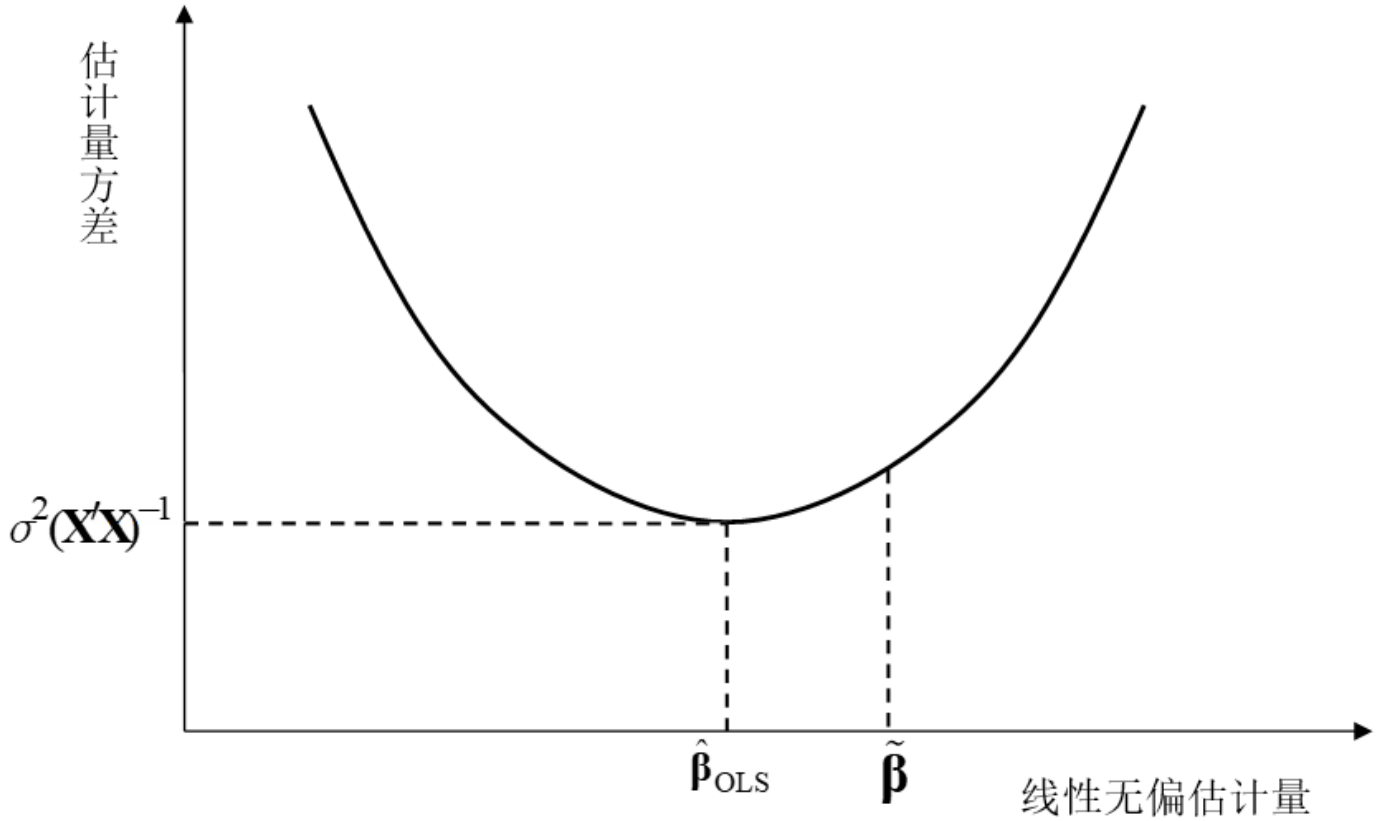

图 5.5 OLS 估计量为 BLUE

(5) 对扰动项方差的无偏估计

对于扰动项方差$\sigma^2 = \mathrm{Var}(\varepsilon_i)$，由于$\{\varepsilon_1, \cdots, \varepsilon_n\}$不可观测，将$\{e_1, \cdots, e_n\}$视为其实现值，可得到对$\sigma^2$的估计：

$$s^2 \equiv \frac{1}{n-K}\sum_{i=1}^{n} e_i^2 \qquad (5.49)$$

其中，$(n-K)$为自由度。为什么除以$(n-K)$而不除以n？

虽然共有n个残差$\{e_1, e_2, \cdots, e_n\}$，随机变量$\{e_1, e_2, \cdots, e_n\}$必须满足$K$个正规方程$\mathbf{X}'\mathbf{e} = \mathbf{0}$，故只有其中$(n-K)$个残差可自由取值。

经过校正后，才是**无偏估计**(unbiased estimator)，即$\mathrm{E}(s^2)=\sigma^2$。

如果样本容量n很大，当$n\to\infty$时，则$\dfrac{n-K}{n}\to 1$，是否进行**小样本校正**(small sample adjustment)并无多大差别。

称$s=\sqrt{s^2}$为**回归方程的标准误差**(standard error of the regression)，简称**回归方程的标准误**，它衡量回归方程扰动项的波动幅度。

OLS估计量$\hat{\boldsymbol{\beta}}$的协方差矩阵$\sigma^2(\mathbf{X}'\mathbf{X})^{-1}$可用$s^2(\mathbf{X}'\mathbf{X})^{-1}$来估计。

$\hat{\boldsymbol{\beta}}$的第k个分量$\hat{\beta}_k$的估计方差为$s^2(\mathbf{X}'\mathbf{X})^{-1}_{kk}$，其中$(\mathbf{X}'\mathbf{X})^{-1}_{kk}$表示矩阵$(\mathbf{X}'\mathbf{X})^{-1}$的$(k,k)$元素，即主对角线上的第$k$个元素。

称$\sqrt{s^2(\mathbf{X}'\mathbf{X})^{-1}_{kk}}$为 OLS 估计量$\hat{\beta}_k$的**标准误差**(standard error)，简称**标准误**，记为$\mathrm{SE}(\hat{\beta}_k)$，即

$$\mathrm{SE}(\hat{\beta}_k) \equiv \sqrt{s^2(\mathbf{X}'\mathbf{X})^{-1}_{kk}} \qquad (5.50)$$

更一般地，称对某统计量的标准差之估计值(estimated standard deviation)为该统计量的“标准误”(standard error)，作为对统计量估计误差的度量。

得到参数的点估计(point estimate)后，还须给出相应的标准误，才能知道此点估计的准确程度(或不确定性)。

5.8 对单个系数的 t 检验

计量经济学中的**统计推断**(statistical inference)方法，可分为两大类，即**小样本理论**(small sample theory)与**大样本理论**(large sample theory)。

无论样本容量是多少，小样本理论都成立，不需要让样本容量 $n \to \infty$，故也称**有限样本理论**(finite sample theory)。

反之，大样本理论要求 $n \to \infty$，适用于较大的样本容量。

小样本理论适用于各种样本容量，但不易推导其统计量的分布，需对随机变量概率分布作很强的具体假定，比如正态分布。

假定 5.5 在给定$\mathbf{X}$的情况下，$\boldsymbol{\varepsilon}|\mathbf{X}$的条件分布为正态，即$\boldsymbol{\varepsilon}|\mathbf{X} \sim N(\mathbf{0}, \sigma^2\mathbf{I}_n)$。

首先考虑最简单的假设检验(hypothesis testing)，即对单个回归系数β_k进行检验。

需要检验的**原假设**(null hypothesis，也称“零假设”)为

$$H_0 : \beta_k = c \qquad (5.51)$$

其中，c为给定常数，也称为**假想值**(hypothesized value)。

通常$c=0$，此时即检验变量x_{ik}的系数是否显著地不等于 0。

假设检验的实质是一种概率意义上的反证法，即首先假设原假设成立，然后看在原假设成立的前提下，是否导致不太可能发生的“小概率事件”在一次抽样的样本中出现。

如果小概率事件竟然在一次抽样实验中被观测到，则说明原假设不可信，应该拒绝原假设，而接受**替代假设**(alternative hypothesis，也称“备择假设”)：

$$H_1:\beta_k \neq c \qquad (5.52)$$

替代假设“$H_1:\beta_k \neq c$”也称为**双边替代假设**(two-sided alternative hypothesis)，它既包括$\beta_k > c$，也包括$\beta_k < c$的情形。

这类检验称为**双边检验**(two-sided test)。

如果未知参数β_k的估计值$\hat{\beta}_k$离c较远，则倾向于拒绝原假设。

使用此原理的这类统计检验称为**沃尔德检验**(Wald test)。

在衡量距离远近时，由于绝对距离依赖于变量的单位，故需要以标准差为基准来考虑相对距离。

由于扰动项$\boldsymbol{\varepsilon}|\mathbf{X} \sim N(\mathbf{0}, \sigma^2\mathbf{I}_n)$，而$\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} = \mathbf{A}\boldsymbol{\varepsilon}$为$\boldsymbol{\varepsilon}$的线性函数(其中$\mathbf{A} \equiv (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$)，故$(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \,|\, \mathbf{X}$也服从正态分布。

由于$\mathrm{E}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\,|\,\mathbf{X})=\mathbf{0}$，$\mathrm{Var}(\hat{\boldsymbol{\beta}}\,|\,\mathbf{X})=\sigma^2(\mathbf{X}'\mathbf{X})^{-1}$，故

$$(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})\,|\,\mathbf{X}\sim N\left(\mathbf{0},\sigma^2(\mathbf{X}'\mathbf{X})^{-1}\right) \qquad (5.53)$$

单独考虑上式的第k个分量，则有

$$(\hat{\beta}_k-\beta_k)\,|\,\mathbf{X}\sim N\left(0,\sigma^2(\mathbf{X}'\mathbf{X})^{-1}_{kk}\right) \qquad (5.54)$$

其中，$(\mathbf{X}'\mathbf{X})^{-1}_{kk}$为矩阵$(\mathbf{X}'\mathbf{X})^{-1}$的$(k,k)$元素(即主对角线上第$k$个元素)，而$\sigma^2(\mathbf{X}'\mathbf{X})^{-1}_{kk}$为$\hat{\beta}_k$的方差。

在原假设“$H_0 : \beta_k = c$”成立的情况下，上式可写为

$$(\hat{\beta}_k - c) \mid \mathbf{X} \sim N\left(0, \sigma^2 (\mathbf{X}'\mathbf{X})^{-1}_{kk}\right) \quad (5.55)$$

如果σ^2已知，则标准化的统计量服从标准正态分布：

$$z_k \equiv \frac{\hat{\beta}_k - c}{\sqrt{\sigma^2 (\mathbf{X}'\mathbf{X})^{-1}_{kk}}} \sim N(0,1) \quad (5.56)$$

通常σ^2未知，称为**厌恶参数**(nuisance parameter)。

虽然我们对σ^2不感兴趣，但σ^2却出现在表达式(5.56)中。

合格的**检验统计量**(test statistic)必须满足两个条件。

首先，它应能够根据样本数据计算出来。

其次，它的概率分布是已知的。

对于上式的z_k统计量，虽然已知其分布为标准正态，但由于不知道σ^2，故无法根据数据计算z_k统计量的样本观测值。

以σ^2的估计量s^2来替代σ^2，即可得到以下t统计量(t-statistic)。

定理(*t* 统计量的分布) 在假定 5.1-5.5 均满足，且原假设“$H_0:\beta_k=c$”也成立的情况下，t统计量服从自由度为$(n-K)$的t分布：

$$t_k \equiv \frac{\hat{\beta}_k - c}{\mathrm{SE}(\hat{\beta}_k)} \sim t(n-K) \qquad (5.57)$$

其中，$\mathrm{SE}(\hat{\beta}_k) \equiv \sqrt{s^2(\mathbf{X}'\mathbf{X})^{-1}_{kk}}$为$\hat{\beta}_k$的标准误。

更一般地，t统计量的通用公式为

$$t \equiv \frac{\text{估计量} - \text{假想值}}{\text{估计量的标准误}} \qquad (5.58)$$

t统计量度量估计量($\hat{\beta}_k$)离假想值(c)的距离，并以估计量的标准误$\mathrm{SE}(\hat{\beta}_k)$作为距离的度量单位，即此距离为标准误的多少倍。

t检验的步骤

第一步：计算t统计量，记其具体取值为t_k。

如果H_0为真，则$|t_k|$很大的概率将很小(为小概率事件)，不应在抽样中观测到。如果$|t_k|$很大，则H_0较不可信。

第二步：计算**显著性水平**(significance level)为α的**临界值**(critical value)$t_{\alpha/2}(n-K)$，其中$t_{\alpha/2}(n-K)$的定义为

$$\mathrm{P}\left\{T > t_{\alpha/2}(n-K)\right\} = \mathrm{P}\left\{T < -t_{\alpha/2}(n-K)\right\} = \alpha/2 \qquad (5.59)$$

其中，随机变量$T \sim t(n-K)$。

随机变量T大于$t_{\alpha/2}(n-K)$，或小于$-t_{\alpha/2}(n-K)$的概率都是$\alpha/2$，参见图 5.6。

在实践中，通常取$\alpha = 5\%$，则$\alpha/2 = 2.5\%$。

有时也使用$\alpha = 1\%$或$\alpha = 10\%$。

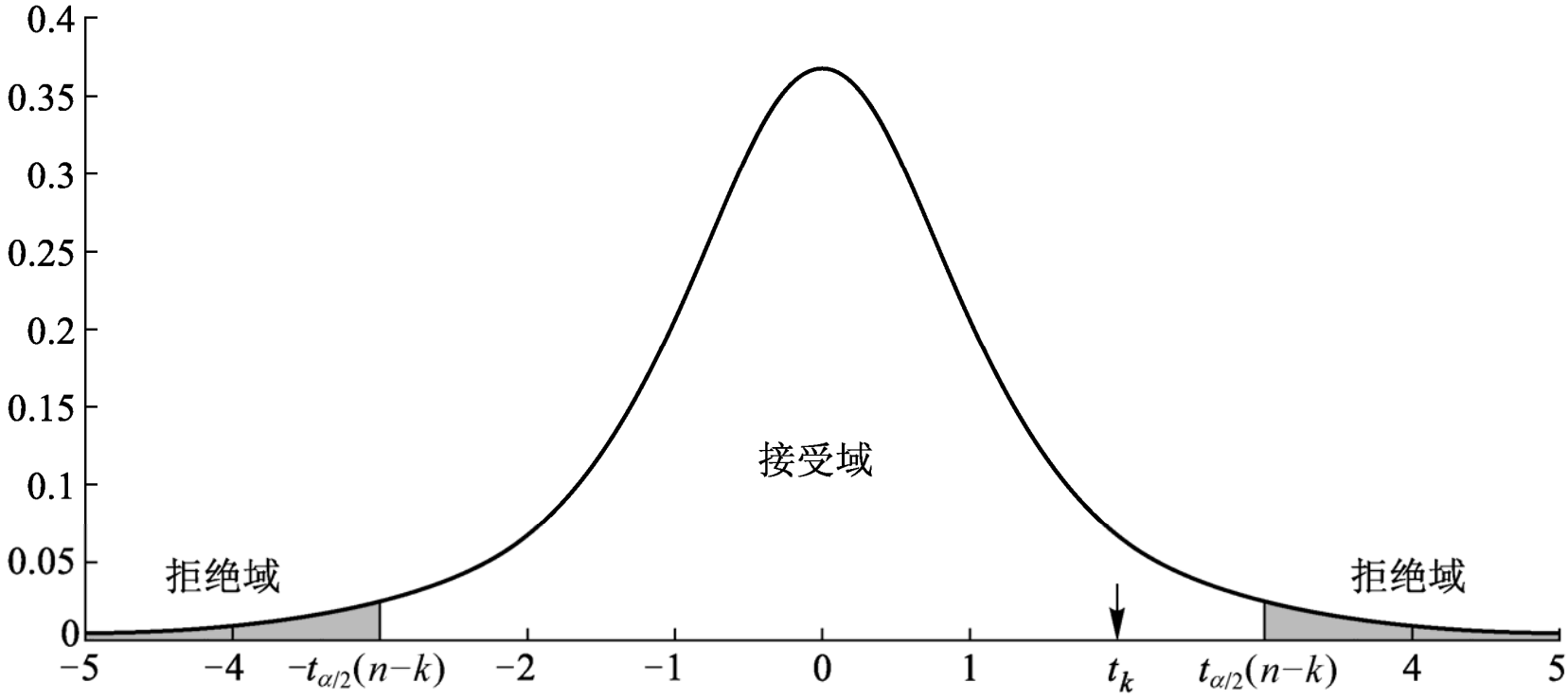

图 5.6 双边t检验的临界值与拒绝域

第三步：如果$\left|t_k\right| \geq t_{\alpha/2}(n-K)$，则$t_k$落入**拒绝域**(rejection region)，故拒绝H_0。

反之，如果$\left|t_k\right| < t_{\alpha/2}(n-K)$，则$t_k$落入**接受域**(acceptance region)，故接受H_0。

因为拒绝域分布在t分布两边，故称为**双边检验**(two-sided test)。

计算 p 值

假设检验的基本逻辑就是概率意义上的反证法，即如果在一次抽样中看到很不可能发生的小概率事件，则拒绝原假设。

至于观测到样本数据的发生概率究竟小到何种程度，可通过下面的 p 值来度量。

在双边 t 检验中，给定 t 统计量的样本观测值 t_k，此假设检验问题的 **p 值**(probability value，即 p-value)为

$$p\text{值} \equiv \mathrm{P}\left(\left|T\right| > \left|t_k\right|\right) \qquad (5.60)$$

其中，随机变量 $T \sim t(n-K)$。

给定t统计量t_k，则p值衡量比$|t_k|$更大的t分布两端的尾部概率，参见图 5.7。

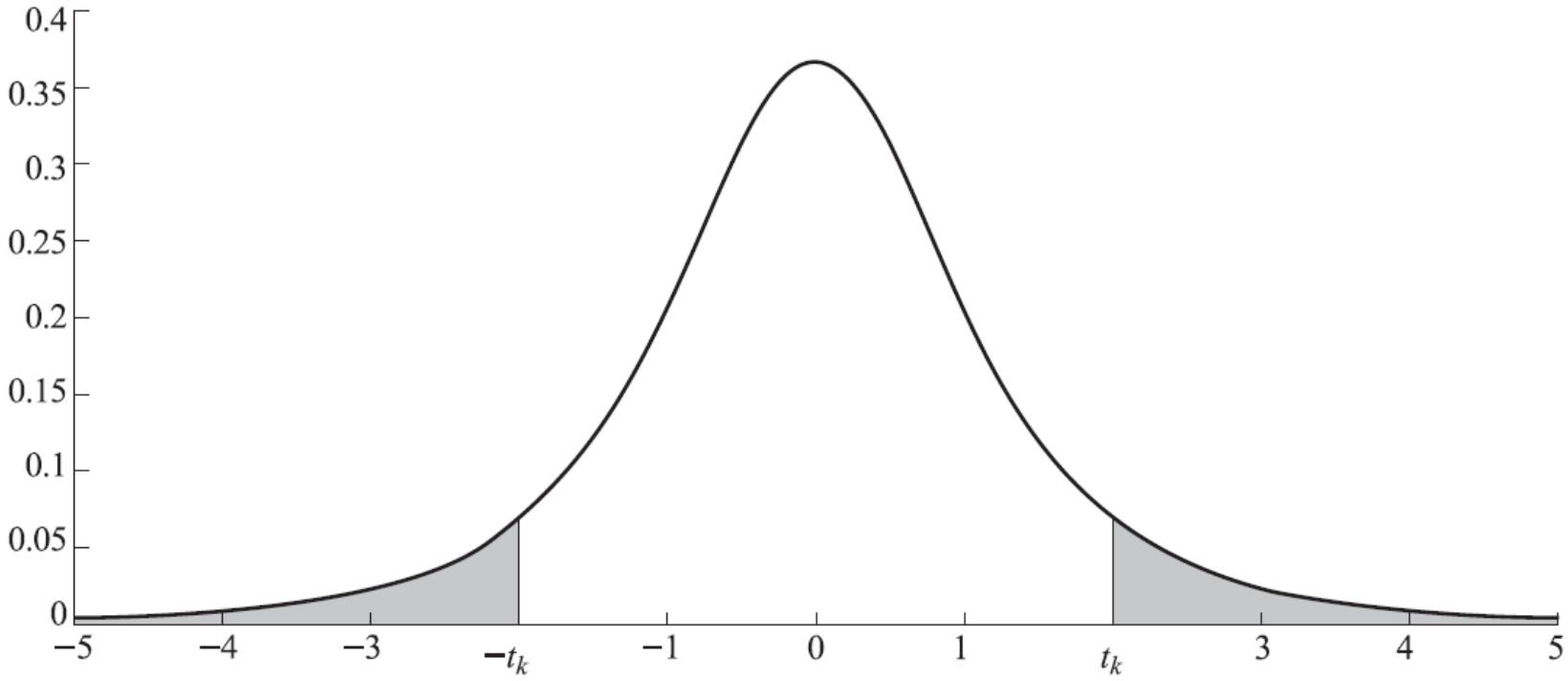

图 5.7 双边t检验的 p 值

如果 p 值为 0.05，则正好可以在 5%的显著性水平上拒绝原假设，但无法在 4.9%的显著性水平上拒绝原假设。

p 值的一般定义如下。

定义 称原假设可被拒绝的最小显著性水平为此假设检验问题的 **p 值**(probability value，即 p-value)。

p 值越小，则越倾向于拒绝原假设。

如果选定显著性水平为 5%，则只要 p 值比 0.05 小，即可拒绝原假设。

比如， p 值 = 0.03，可在 5%的显著性水平上拒绝原假设。

“p值= 0.03”还可“在 3%的显著性水平上拒绝原假设”。

使用p值进行假设检验一般比临界值更有信息量。

使用p值进行检验的另一好处是，只要将p值与 0.05 相比，即可得到检验结果，操作十分简便。

传统的检验需要将统计量与临界值相比，但临界值依分布与自由度而变。

计算置信区间

有时还希望进行区间估计，即参数最可能的取值范围。

假设**置信度**(confidence level)为$(1-\alpha)$(比如$\alpha=5\%$，则$1-\alpha=95\%$)，即要找到**置信区间**(confidence interval)，使得该区间覆盖真实参数β_k的概率为$(1-\alpha)$。

由于$t_k=\dfrac{\hat{\beta}_k-\beta_k}{\mathrm{SE}(\hat{\beta}_k)}\sim t(n-K)$，故$t$统计量落入接受域的概率为$(1-\alpha)$：

$$\mathrm{P}\left\{-t_{\alpha/2}<\frac{\hat{\beta}_k-\beta_k}{\mathrm{SE}(\hat{\beta}_k)}<t_{\alpha/2}\right\}=1-\alpha \qquad (5.61)$$

其中，$t_{\alpha/2}$为显著性水平为α的临界值。

将上式中的不等式变形可得

$$\mathrm{P}\left\{\hat{\beta}_k - t_{\alpha/2}\,\mathrm{SE}(\hat{\beta}_k) < \beta_k < \hat{\beta}_k + t_{\alpha/2}\,\mathrm{SE}(\hat{\beta}_k)\right\} = 1-\alpha \qquad (5.62)$$

由此可知，β_k的置信区间为

$$\left[\hat{\beta}_k - t_{\alpha/2}\,\mathrm{SE}(\hat{\beta}_k),\ \hat{\beta}_k + t_{\alpha/2}\,\mathrm{SE}(\hat{\beta}_k)\right] \qquad (5.63)$$

此置信区间以点估计$\hat{\beta}_k$为中心，区间半径为$t_{\alpha/2}\,\mathrm{SE}(\hat{\beta}_k)$，参见图 5.8。

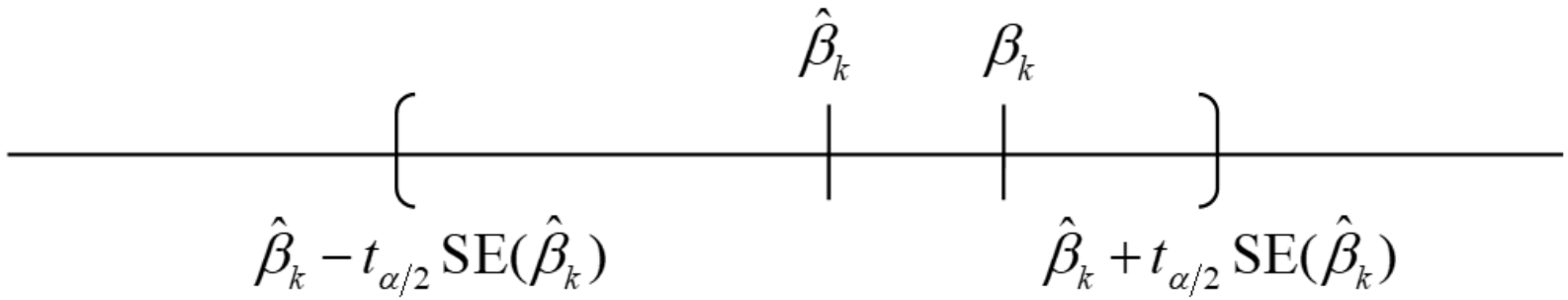

图 5.8 置信区间的示意图

标准误$\mathrm{SE}(\hat{\beta}_k)$越大，则对$\beta_k$的估计越不准确，故置信区间也越宽。

置信区间是随机区间，随着样本不同而不同。

如果置信度为 95%，抽样 100 次，得到 100 个置信区间，大约 95 个置信区间能覆盖到真实参数β_k。

单边检验

假设检验有时也进行**单边检验**(one-sided test)。

不失一般性，考虑原假设为$H_0:\beta_k=0$，而替代假设为$H_1:\beta_k>0$；比如，从理论上认为解释变量x_k对y的作用不可能为负。

仍可计算t统计量：

$$t_k \equiv \frac{\hat{\beta}_k}{\mathrm{SE}(\hat{\beta}_k)} \sim t(n-K) \qquad (5.64)$$

如果此t统计量很大，则倾向于拒绝原假设；而如果此t统计量很小(比如为负数)，则倾向于接受原假设。

因此，拒绝域只在概率分布的最右边一侧。

给定显著性水平α后，需要计算的临界值为$t_\alpha(n-K)$，使得取值大于此临界值的概率为α：

$$\mathrm{P}\left\{T > t_\alpha(n-K)\right\} = \alpha \qquad (5.65)$$

其中，随机变量$T \sim t(n-K)$。

如果要计算此单边检验的p值，则为比统计量t_k更大的右侧尾部概率(参见图 5.9)：

$$p\text{值} \equiv \mathrm{P}\left(T > t_k\right) \qquad (5.66)$$

由于拒绝域只在分布的右侧尾部，故也称**单边右侧检验**(one-sided right-tail test)。

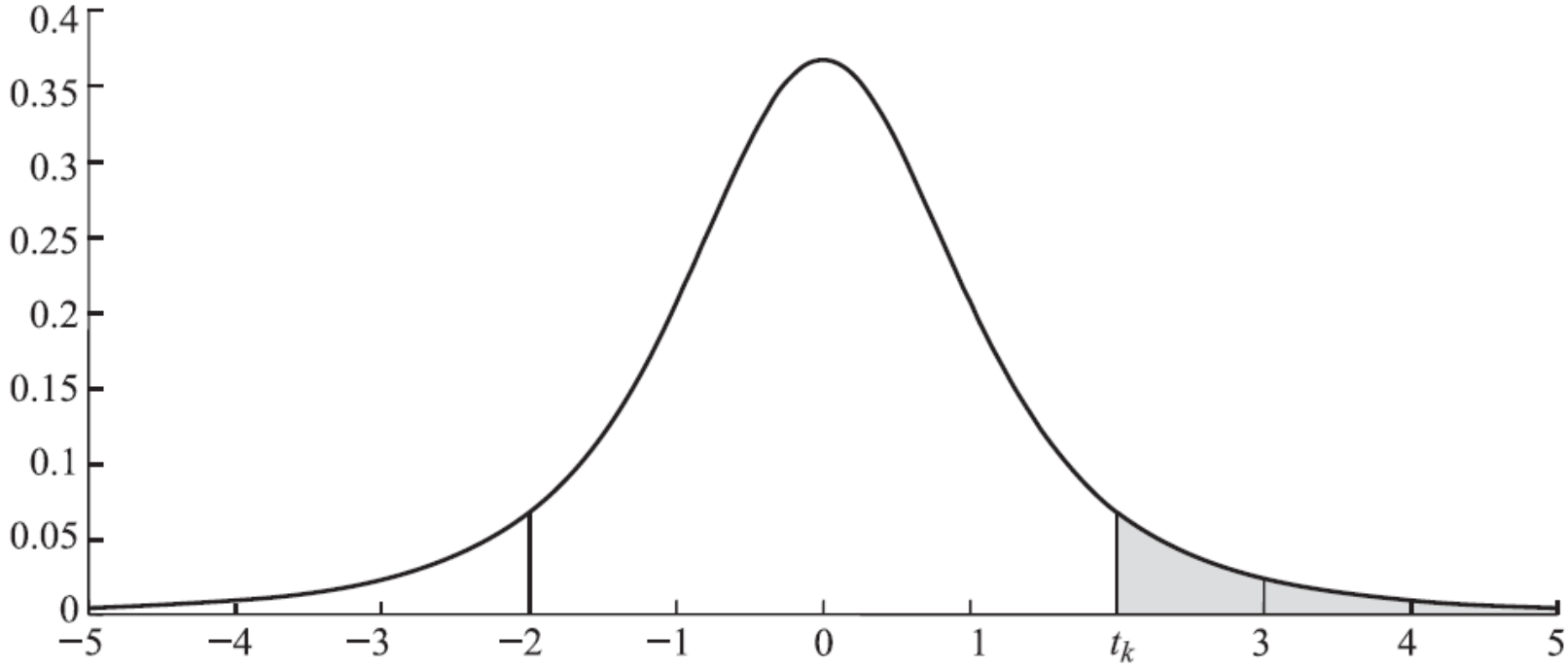

图 5.9 单边右侧t检验的p值

相应地，如果原假设为 $H_0:\beta_k=0$，而替代假设为 $H_1:\beta_k<0$，则为**单边左侧检验**(one-sided left-tail test)。

此时，t 统计量越小(比如为负数)，则越倾向于拒绝原假设；故拒绝域只在分布的左侧尾部。

因此，对于单边左侧检验，计算 p 值的公式为(参见图 5.10)：

$$p\text{值} \equiv \mathrm{P}\left(T<t_k\right) \qquad (5.67)$$

其中，随机变量 $T\sim t(n-K)$。

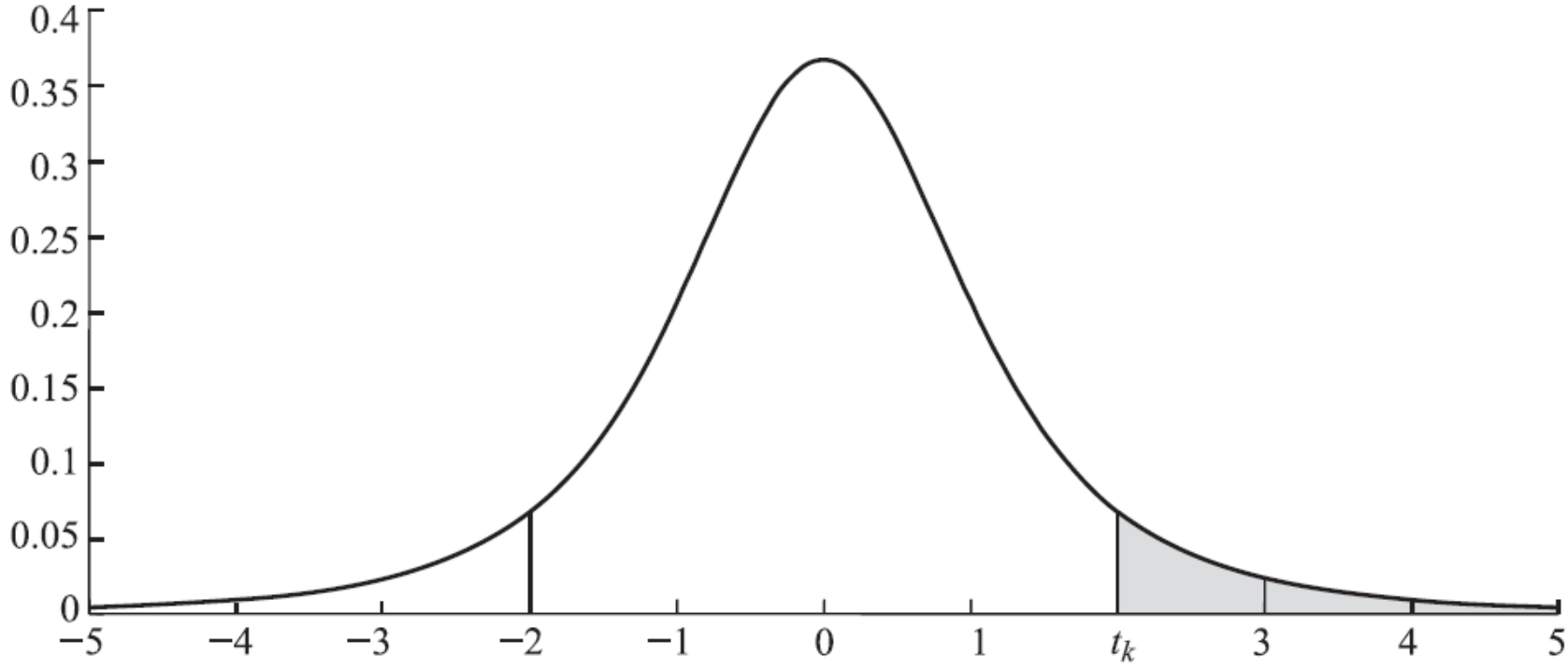

图 5.10 单边左侧t检验的p值

第 I 类错误与第 II 类错误

在进行假设检验时，可能犯以下两类错误。

定义　第 I 类错误(Type I error)指的是，虽然原假设为真，但却根据观测数据做出了拒绝原假设的错误判断，即“弃真”。

第I类错误的发生概率为

$$\mathrm{P}(\text{拒绝}H_0 \,|\, H_0) = \mathrm{P}(\text{检验统计量落入拒绝域} \,|\, H_0)=\alpha \qquad (5.68)$$

其中，α正是此检验的显著性水平。

定义　第 II 类错误(Type II error)指的是，虽然原假设为假(替代假设为真)，但却根据观测数据做出了接受原假设的错误判断，即“存伪”。

第II类错误的发生概率为

$$\mathrm{P}(\text{接受 } H_0 \,|\, H_1) = \mathrm{P}(\text{检验统计量落入接受域} \,|\, H_1) \qquad (5.69)$$

在β_k可能取值的参数空间中，通常H_0仅包含一个点(比如$\beta_k=0$)，故容易计算第I类错误的发生概率，即$\mathrm{P}(\text{拒绝} H_0 \,|\, H_0)=\alpha$。

反之，替代假设H_1则一般包括许多点(比如$\beta_k \neq 0$)，故不易计算第II类错误的发生概率。

第Ⅰ类错误与第Ⅱ类错误存在此消彼长的关系，即如果减少第Ⅰ类错误的发生概率，则第Ⅱ类错误的发生概率必然增加；反之亦然。

一般来说，如果同时减少第I类错误与第II类错误的发生概率，则必须增加样本容量。

在进行假设检验时，一般先指定可接受的发生第I类错误的最大概率，即显著性水平α(比如 5%)，而不指定第II类错误的发生概率(通常更难计算)。

定义 称“1 减去第Ⅱ类错误的发生概率”为统计检验的**功效**(power)，即

$$\text{功效} = 1 - \mathrm{P}(\text{接受}H_0 \,|\, H_1) = \mathrm{P}(\text{拒绝}H_0 \,|\, H_1) \qquad (5.70)$$

功效为在原假设为错误的情况下，拒绝原假设的概率。

在进行检验时，通常知道第I类错误的发生概率，而不知道第II类错误的发生概率。

如果拒绝原假设，则比较理直气壮，因为知道犯错概率(显著性水平)。

反之，如果接受原假设，则比较没有把握，因为通常不知犯错概率(可能较高)。

5.9 对线性假设的 *F* 检验

如想知道整个回归方程是否显著，即除常数项以外，所有解释变量的回归系数是否都为零。

这需要检验以下原假设：

$$H_0: \beta_2 = \cdots = \beta_K = 0 \qquad (5.71)$$

其中，β_1为常数项。

此原假设等价于对$(K-1)$个约束条件进行联合检验(joint test)：

$$H_0: \beta_2 = 0,\ \beta_3 = 0,\ \cdots,\ \beta_K = 0 \qquad (5.72)$$

例 对于模型 $y=\beta_1 x_1+\beta_2 x_2+\beta_3 x_3+\beta_4 x_4+\varepsilon$，检验以下两个约束：

$$H_0:\beta_2=\beta_3,\ \beta_4=0 \qquad (5.73)$$

将此原假设的两个约束写成向量形式，经整理可得

$$H_0:\begin{pmatrix}\beta_2-\beta_3\\ \beta_4\end{pmatrix}=\begin{pmatrix}0\\0\end{pmatrix} \qquad (5.74)$$

进一步，上式可写为

$$H_0: \begin{pmatrix} \beta_2 - \beta_3 \\ \beta_4 \end{pmatrix} = \underbrace{\begin{pmatrix} 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}}_{\mathbf{R}} \underbrace{\begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{pmatrix}}_{\boldsymbol{\beta}} = \underbrace{\begin{pmatrix} 0 \\ 0 \end{pmatrix}}_{\mathbf{r}} \tag{5.75}$$

更一般地，考虑检验m个线性假设是否同时成立：

$$H_0: \underbrace{\mathbf{R}}_{m\times K} \underbrace{\boldsymbol{\beta}}_{K\times 1} = \underbrace{\mathbf{r}}_{m\times 1}$$

其中，$\mathbf{r}$为m维列向量($m < K$)，$\mathbf{R}$为$m \times K$维矩阵，而且$\text{rank}(\mathbf{R}) = m$，即$\mathbf{R}$满行秩，没有多余或自相矛盾的行或方程。

在上例中，$\mathbf{R}=\begin{pmatrix}0 & 1 & -1 & 0\\ 0 & 0 & 0 & 1\end{pmatrix}$，而$\mathbf{r}=\begin{pmatrix}0\\ 0\end{pmatrix}$。

根据沃尔德检验原理，由于$\hat{\boldsymbol{\beta}}$是$\boldsymbol{\beta}$的估计量，故如果H_0成立，则$(\mathbf{R}\hat{\boldsymbol{\beta}}-\mathbf{r})$应比较接近 $\mathbf{0}$(零向量)。

这种接近程度可用其二次型来衡量，比如

$$(\mathbf{R}\hat{\boldsymbol{\beta}}-\mathbf{r})'[\mathrm{Var}(\mathbf{R}\hat{\boldsymbol{\beta}}-\mathbf{r})]^{-1}(\mathbf{R}\hat{\boldsymbol{\beta}}-\mathbf{r}) \quad (5.76)$$

其中，$(\mathbf{R}\hat{\boldsymbol{\beta}}-\mathbf{r})$的协方差矩阵可写为

$$
\begin{aligned}
\mathrm{Var}(\mathbf{R}\hat{\boldsymbol{\beta}}-\mathbf{r}) &= \mathrm{Var}(\mathbf{R}\hat{\boldsymbol{\beta}}) && \text{(去掉常数，方差不变)} \\
&= \mathbf{R}\,\mathrm{Var}(\hat{\boldsymbol{\beta}})\mathbf{R}' && \text{(夹心估计量的公式)} \\
&= \sigma^2\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}' && \text{(因为}\mathrm{Var}(\hat{\boldsymbol{\beta}})=\sigma^2(\mathbf{X}'\mathbf{X})^{-1}\text{)}
\end{aligned} \tag{5.77}
$$

其中，σ^2可由s^2来估计，故有如下定理。

定理(*F* 统计量的分布) 在假定 5.1-5.5 均满足，且原假设“$H_0:\ \mathbf{R}\boldsymbol{\beta}=\mathbf{r}$”也成立的情况下，则 F 统计量服从自由度为$(m, n-K)$的F分布：

$$
F \equiv \frac{(\mathbf{R}\hat{\boldsymbol{\beta}}-\mathbf{r})'[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}(\mathbf{R}\hat{\boldsymbol{\beta}}-\mathbf{r})/m}{s^2} \sim F(m, n-K) \tag{5.78}
$$

F 检验的步骤

第一步：计算F统计量。如果H_0为真，则“F统计量很大”的概率将很小(为小概率事件)，不应在一次抽样中观测到。

如果F统计量很大，则H_0较不可信。

第二步：计算显著性水平为α的临界值$F_\alpha(m, n-K)$，其中$F_\alpha(m, n-K)$的定义为

$$\mathrm{P}\left\{\widetilde{F} > F_\alpha(m, n-K)\right\} = \alpha \qquad (5.79)$$

其中，随机变量$\widetilde{F} \sim F(m, n-K)$。$\widetilde{F}$大于临界值$F_\alpha(m, n-K)$的概率恰好为$\alpha$。

第三步：如果F统计量大于临界值$F_{\alpha}(m, n-K)$，即落入右边拒绝域，则拒绝H_0；反之，如果F统计量小于临界值，即落入接受域，则接受H_0，参见图 5.11。

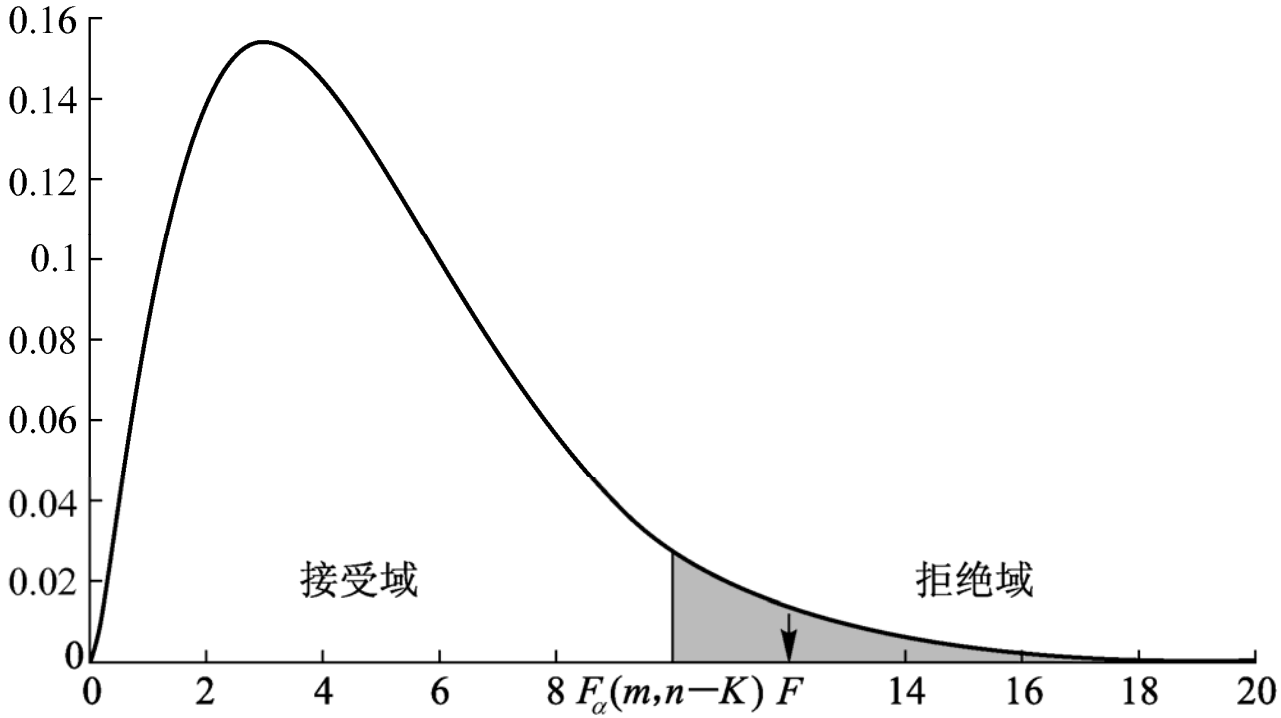

图 5.11 F检验

对于F检验，也可使用p值来进行。

给定F统计量的样本观测值，此假设检验问题的p值为比F统计量更大的F分布的右侧尾部概率，即

$$p\text{值} \equiv \mathrm{P}(\tilde{F} > F) \qquad (5.80)$$

其中，随机变量$\widetilde{F} \sim F(m, n-K)$，而$F$为$F$统计量的取值。

5.10 *F* 统计量的似然比原理表达式

在作假设检验时，如果接受原假设，则可将此原假设作为约束条件，代入最小二乘法的最优化问题。

使用约束条件下的最小二乘法，即**约束最小二乘法**(Restricted OLS 或 Constrained OLS)，可得到 *F* 统计量的另一方便表达式。考虑以下约束极值问题：

$$\begin{aligned} &\min_{\hat{\boldsymbol{\beta}}} \ \mathrm{SSR}(\hat{\boldsymbol{\beta}}) \\ &s.t. \ \mathbf{R}\hat{\boldsymbol{\beta}} = \mathbf{r} \end{aligned} \tag{5.81}$$

$\mathrm{SSR}(\hat{\boldsymbol{\beta}})$为残差平方和，是$\hat{\boldsymbol{\beta}}$的函数；而$\hat{\boldsymbol{\beta}}$还须满足约束条件$\mathbf{R}\hat{\boldsymbol{\beta}} = \mathbf{r}$。

若$H_0:\mathbf{R}\boldsymbol{\beta}=\mathbf{r}$正确，加上此约束不应使残差平方和增大很多。

记无约束的残差平方和SSR，有约束的残差平方和为SSR^*。

在H_0正确的情况下，$(\text{SSR}^*-\text{SSR})$不应很大。

由此可构造如下F统计量。通过求解此约束极值问题，可以证明：

$$F=\frac{(\text{SSR}^*-\text{SSR})/m}{\text{SSR}/(n-K)} \qquad (5.82)$$

其中，m为约束条件个数(即矩阵$\mathbf{R}$的行数)，n为样本容量，而K为参数个数(即$\boldsymbol{\beta}$的维度)。

此 F 统计量表达式有时更易计算。

这种通过比较“条件极值”与“无条件极值”而进行的检验，统称为**似然比检验**(Likelihood Ratio test，简记 LR)。

F 统计量的似然比表达式(5.82)，也可以通过拟合优度来表示。

记 R^2 为无约束回归的拟合优度，R_*^2 为约束回归的拟合优度，则

$$F = \frac{(R^2 - R_*^2)/m}{(1-R^2)/(n-K)} \qquad (5.83)$$

如果去掉约束条件后拟合优度上升越多，即 $(R^2 - R_*^2)$ 越大，则越应该拒绝约束条件成立的原假设。

证明：在方程(5.82)右边的分子分母同时除以被解释变量的离差平方和$\mathrm{TSS}\equiv\sum_{i=1}^{n}(y_i-\bar{y})^2$可得

$$F=\frac{\left.\dfrac{(\mathrm{SSR}^*-\mathrm{SSR})}{\mathrm{TSS}}\right/m}{\left.\dfrac{\mathrm{SSR}}{\mathrm{TSS}}\right/(n-K)} \qquad (5.84)$$

由于$\dfrac{\mathrm{SSR}}{\mathrm{TSS}}=1-R^2$，而$\dfrac{\mathrm{SSR}^*}{\mathrm{TSS}}=1-R_*^2$，故

$$F=\frac{\left[(1-R_*^2)-(1-R^2)\right]/m}{(1-R^2)/(n-K)}=\frac{(R^2-R_*^2)/m}{(1-R^2)/(n-K)} \qquad (5.85)$$

考虑一个特殊情形，即检验整个回归方程的显著性。

命题 对于线性回归方程$y_i = \beta_1 + \beta_2 x_{i2} + \cdots + \beta_K x_{iK} + \varepsilon_i$，检验原假设“$H_0 : \beta_2 = \cdots = \beta_K = 0$”的$F$统计量等于$\dfrac{R^2/(K-1)}{(1-R^2)/(n-K)}$。

证明：对于$H_0 : \beta_2 = \cdots = \beta_K = 0$，共有$(K-1)$个约束，故在表达式(5.83)中，$m = (K-1)$。

另一方面，当原假设成立时，$y_i = \beta_1 + \varepsilon_i$，故约束回归只是对常数项回归，因此$R_*^2 = 0$(参见第 4 章习题)。

将$m = (K-1)$与$R_*^2 = 0$代入表达式(5.85)即得证。

此命题表明了F统计量与拟合优度R^2之间的关系。

R^2并非决定F统计量的唯一因素；F统计量还取决于样本容量n，以及与解释变量个数K。

5.11 预测

有时也用计量模型进行预测(prediction 或 forecasting)，即给定解释向量$\mathbf{x}_0$的(未来)取值，预测被解释变量y_0的取值。

假设计量模型对所有观测值都成立(包括外推到未来的观测值)，

$$y_0 = \mathbf{x}_0'\boldsymbol{\beta} + \varepsilon_0 \qquad (5.86)$$

记$\hat{\boldsymbol{\beta}}$为$\boldsymbol{\beta}$的 OLS 估计量，对y_0作点预测为

$$\hat{y}_0 \equiv \mathbf{x}_0'\hat{\boldsymbol{\beta}} \tag{5.87}$$

预测误差(prediction error)$(\hat{y}_0 - y_0)$可写为

$$\hat{y}_0 - y_0 = \mathbf{x}_0'\hat{\boldsymbol{\beta}} - (\mathbf{x}_0'\boldsymbol{\beta} + \varepsilon_0) = \mathbf{x}_0'(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) - \varepsilon_0 \tag{5.88}$$

$\hat{y}_0$为**无偏预测**(unbiased predictor)，即用$\hat{y}_0$作为y_0的预测值不会系统地高估或低估y_0，因为

$$\mathrm{E}(\hat{y}_0 - y_0) = \mathbf{x}_0' \mathrm{E}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) - \mathrm{E}(\varepsilon_0) = 0 \tag{5.89}$$

其中，由于$\hat{\boldsymbol{\beta}}$是$\boldsymbol{\beta}$的无偏估计，故$\mathrm{E}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = 0$。

有时候，还希望知道此预测的置信区间。

计算预测误差$(\hat{y}_0 - y_0)$的方差为：

$$\begin{aligned}
\mathrm{Var}(\hat{y}_0 - y_0) &= \mathrm{Var}[\mathbf{x}_0'(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) - \varepsilon_0] \\
&= \mathrm{Var}(\varepsilon_0) + \mathrm{Var}[\mathbf{x}_0'(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})] \\
&= \sigma^2 + \mathrm{Var}[\mathbf{x}_0'(\hat{\boldsymbol{\beta}})] \qquad \text{(去掉常数，方差不变)} \\
&= \sigma^2 + \mathbf{x}_0' \,\mathrm{Var}(\hat{\boldsymbol{\beta}})\mathbf{x}_0 \qquad \text{(夹心估计量的公式)} \\
&= \sigma^2 + \sigma^2 \mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0 \qquad \text{(因为}\mathrm{Var}(\hat{\boldsymbol{\beta}})=\sigma^2(\mathbf{X}'\mathbf{X})^{-1}\text{)}
\end{aligned}$$

(5.90)

其中，假设ε_0与$\hat{\boldsymbol{\beta}}$不相关(估计$\hat{\boldsymbol{\beta}}$没用到ε_0的信息)。

预测误差的方差有两个来源：

(1) 抽样误差$\sigma^2\mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0$(不能精确知道参数$\boldsymbol{\beta}$)；

(2) y_0本身的不确定性(ε_0的方差σ^2)。

如果样本很大，则抽样误差将很小；但扰动项的方差σ^2始终存在。

将方程(5.90)中的σ^2用s^2来替代，并开平方，则可得到预测误差$(\hat{y}_0 - y_0)$的标准误：

$$\mathrm{SE}(\hat{y}_0 - y_0) = s\sqrt{1 + \mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0} \qquad (5.91)$$

由此可得 t 统计量：

$$\frac{\hat{y}_0 - y_0}{\mathrm{SE}(\hat{y}_0 - y_0)} \sim t\left(n-K\right) \qquad (5.92)$$

进一步，y_0的置信度为$(1-\alpha)$的置信区间为

$$\left(\hat{y}_0 - t_{\alpha/2}\mathrm{SE}(\hat{y}_0 - y_0),\ \hat{y}_0 + t_{\alpha/2}\mathrm{SE}(\hat{y}_0 - y_0)\right) \qquad (5.93)$$

其中，$t_{\alpha/2}$为显著性水平为α的$t(n-K)$分布的双边检验临界值。

5.12 多元回归的 Stata 实例

在 Stata 中进行多元回归的命令为

```
. regress y x1 x2 x3
```

其中，“y”为被解释变量，而“x1 x2 x3”为解释变量。

以数据集 grilic.dta 为例，该数据集包括 758 名美国年轻男子的数据。对以下方程进行多元回归估计：

$$\ln w = \beta_1 + \beta_2 s + \beta_3 expr + \beta_4 tenure + \beta_5 smsa + \beta_6 rns + \varepsilon \quad (5.94)$$

被解释变量为 *lnw* (工资对数)，主要解释变量包括 *s* (教育年限)、*expr* (工龄)、*tenure* (在现单位工作年限)、*smsa* (是否住在大城市)以及 *rns* (是否住在美国南方)。

为估计方程(5.94)，可输入如下命令

```
. use grilic.dta,clear

. reg lnw s expr tenure smsa rns
```

Source	SS	df	MS	Number of obs	=	758
				F(5, 752)	=	81.75
Model	49.0478814	5	9.80957628	Prob > F	=	0.0000
Residual	90.2382684	752	.119997697	R-squared	=	0.3521
				Adj R-squared	=	0.3478
Total	139.28615	757	.183997556	Root MSE	=	.34641

lnw	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
s	.102643	.0058488	17.55	0.000	.0911611	.114125
expr	.0381189	.0063268	6.02	0.000	.0256986	.0505392
tenure	.0356146	.0077424	4.60	0.000	.0204153	.0508138
smsa	.1396666	.0280821	4.97	0.000	.0845379	.1947954
rns	-.0840797	.0287973	-2.92	0.004	-.1406124	-.0275471
_cons	4.103675	.085097	48.22	0.000	3.936619	4.270731

“_cons”表示常数项，“R-squared”显示 $R^2=0.3521$，“Adj R-squared”显示 $\overline{R}^2=0.3478$。

表上方的回归结果显示，残差平方和$\sum_{i=1}^{n} e_i^2 = 90.24$，方程的标准误差(Root MSE)为$s = 0.34641$。

检验整个方程显著性的F统计量为 81.75，其对应的p值(Prob > F)为 0.0000，表明这个回归方程整体是高度显著的。

所有解释变量(包括常数项)的回归系数的p值(P>|t|)都小于 0.01，故均在 1%水平上显著，且符号与理论预期一致。

教育年限(*s*)的系数估计值为 0.103，即教育投资回报率为 10.3%。

工龄(*expr*)与在现单位工作年限(*tenure*)的回报率分别为 3.8%与 3.6%(可视为在职培训的回报率)，小于正规教育的回报率。

住在大城市的回报率高达 14.0%，甚至高于一年教育的回报率，说明了环境的重要性。

变量 *rns* 的系数为-0.084，表明在给定其他变量的情况下，南方居民的工资比北方居民低 8.4%。

常数项的估计值为 4.104，这意味着未受任何教育($s = 0$)、也无工作经验($expr = tenure = 0$)、不住在大城市($smsa = 0$)，且身在北方($rns = 0$)的年轻男子预期工资对数为 4.104。

如果要显示回归系数的协方差矩阵，可输入命令

```
. vce
```

“vce”表示“variance covariance matrix estimated”。

Covariance matrix of coefficients of regress model

e(V)	s	expr	tenure	smsa	rns	_cons
s	.00003421					
expr	8.660e-06	.00004003				
tenure	-3.997e-08	-.00001107	.00005994			
smsa	-.0000144	3.261e-06	-7.819e-06	.00078861		
rns	8.524e-06	7.334e-07	7.259e-06	.00012486	.00082928	
_cons	-.00046567	-.00016778	-.00008646	-.00038746	-.00043997	.0072415

上表中的主对角线元素为各回归系数的方差，而非主对角线元素则为相应的协方差。

为演示目的，加上选择项“noconstant”，进行无常数项回归：

```
. reg lnw s expr tenure smsa rns,noc
```

Source	SS	df	MS	Number of obs	=	758
				F(5, 753)	=	9902.73
Model	24282.9531	5	4856.59061	Prob > F	=	0.0000
Residual	369.293555	753	.490429688	R-squared	=	0.9850
				Adj R-squared	=	0.9849
Total	24652.2466	758	32.5227528	Root MSE	=	.70031

lnw	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
s	.3665333	.0041742	87.81	0.000	.3583389	.3747277
expr	.1331991	.0121535	10.96	0.000	.1093403	.1570578
tenure	.0846129	.0155168	5.45	0.000	.0541515	.1150743
smsa	.3592339	.0560206	6.41	0.000	.2492588	.4692089
rns	.1652489	.0572715	2.89	0.004	.0528181	.2776796

根据无常数项回归的估计，教育投资回报率高达每年 36.7%，显然不合理。

由于常数项很显著，故忽略常数项将导致估计偏差，得不到一致估计。

即使真实模型不包括常数项，在回归中加入常数项，也不会导致不一致的估计，故危害较小。

反之，如果真实模型包括常数项，但在回归时被忽略了，则可能导致严重的估计偏差。

一般建议在回归中包括常数项。

如果只对南方居民的子样本进行回归，可使用虚拟变量 *rns*：

```
. reg lnw s expr tenure smsa if rns
```

Source	SS	df	MS	Number of obs	=	204
				F(4, 199)	=	36.07
Model	17.603542	4	4.40088551	Prob > F	=	0.0000
Residual	24.2783596	199	.122001807	R-squared	=	0.4203
				Adj R-squared	=	0.4087
Total	41.8819016	203	.206314786	Root MSE	=	.34929

lnw	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
s	.1198242	.0113156	10.59	0.000	.0975103	.1421381
expr	.0451903	.0122572	3.69	0.000	.0210197	.069361
tenure	.0092643	.0156779	0.59	0.555	-.0216518	.0401804
smsa	.1746563	.0506762	3.45	0.001	.0747251	.2745876
_cons	3.806148	.1586202	24.00	0.000	3.493356	4.11894

如果只对北方居民的子样本进行回归，可使用命令：

```
. reg lnw s expr tenure smsa if ~rns
```

其中，“~”表示逻辑的“否”(not)运算；也可使用“!”，即“!rns”，效果相同。

Source	SS	df	MS	Number of obs	=	554
				F(4, 549)	=	62.45
Model	29.486457	4	7.37161426	Prob > F	=	0.0000
Residual	64.8019636	549	.118036364	R-squared	=	0.3127
				Adj R-squared	=	0.3077
Total	94.2884207	553	.170503473	Root MSE	=	.34356

lnw	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
s	.0944787	.0068365	13.82	0.000	.0810498	.1079076
expr	.0358675	.0073558	4.88	0.000	.0214184	.0503165
tenure	.0455117	.0088792	5.13	0.000	.0280703	.0629531
smsa	.1199364	.0337443	3.55	0.000	.0536526	.1862202
_cons	4.214014	.0995796	42.32	0.000	4.018411	4.409618

如果只对中学以上($s \geq 12$)的子样本进行回归，可输入命令：

```
. reg lnw s expr tenure smsa rns if s>=12
```

Source	SS	df	MS	Number of obs	=	679
				F(5, 673)	=	69.81
Model	41.8750434	5	8.37500867	Prob > F	=	0.0000
Residual	80.7410668	673	.119971867	R-squared	=	0.3415
				Adj R-squared	=	0.3366
Total	122.61611	678	.18084972	Root MSE	=	.34637

lnw	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
s	.1077261	.0066792	16.13	0.000	.0946115	.1208408
expr	.0344524	.0071189	4.84	0.000	.0204745	.0484304
tenure	.0363033	.0082594	4.40	0.000	.0200859	.0525206
smsa	.1583146	.0298248	5.31	0.000	.0997537	.2168754
rns	-.074063	.0308884	-2.40	0.017	-.1347123	-.0134137
_cons	4.015335	.098159	40.91	0.000	3.8226	4.20807

如果只对中学以上($s \geq 12$)且在南方居住的子样本进行回归，可输入命令：

```
. reg lnw s expr tenure smsa if s>=12 & rns
```

Source	SS	df	MS	Number of obs	=	174
				F(4, 169)	=	32.17
Model	15.404067	4	3.85101675	Prob > F	=	0.0000
Residual	20.2300414	169	.119704387	R-squared	=	0.4323
				Adj R-squared	=	0.4188
Total	35.6341084	173	.205977505	Root MSE	=	.34598

lnw	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
s	.1269124	.0131847	9.63	0.000	.1008845	.1529404
expr	.0226531	.0156062	1.45	0.148	-.0081551	.0534613
tenure	.0146869	.0182079	0.81	0.421	-.0212573	.0506312
smsa	.2136309	.0548788	3.89	0.000	.1052947	.3219671
_cons	3.699026	.1873691	19.74	0.000	3.329141	4.068912

回到最初估计的全样本：

```
. quietly reg lnw s expr tenure smsa rns
```

其中，前缀“quietly”表示不汇报回归结果。

如果要计算被解释变量的拟合值($\widehat{\ln w}$)，并将其记为 *lnw1*，可使用命令：

```
. predict lnw1
```

如果要计算残差，并将其记为 *e*，可输入命令：

```
. predict e,residual
```

选择项“residual”表示计算残差，默认计算拟合值。

对于回归方程
$\ln w = \beta_1 + \beta_2 s + \beta_3 expr + \beta_4 tenure + \beta_5 smsa + \beta_6 rns + \varepsilon$，考虑检验教育投资回报率是否为 10%，即检验原假设“$H_0 : \beta_2 = 0.1$”，可使用命令：

```
. test s=0.1
```

此命令检验的原假设为，变量 s 的系数等于 0.1。

```
( 1)  s = .1

       F(  1,   752) =    0.20
            Prob > F =    0.6515
```

由于t分布的平方为F分布，故 Stata 统一汇报F统计量及其p值。

p值 $= 0.6515$，故无法拒绝原假设。

对于单个系数的检验，手工计算t统计量也十分方便。

根据公式(5.58)可得

$$t \equiv \frac{\text{估计量} - \text{假想值}}{\text{估计量的标准误}} = \frac{0.102643 - 0.1}{0.0058488} = 0.45188757 \sim t(n-K) = t(752) \tag{5.95}$$

由于默认为双边检验，故可计算此t统计量对应的p值如下：

```
. dis ttail(752,0.45188757)*2
.65148029
```

“ttail(752,0.45188757)”表示自由度为 752 的 t 分布比 0.45188757 更大的右侧尾部概率，正好是反向的累积分布函数。

手工计算的t统计量的p值，与 Stata 汇报的F统计量的p值完全相同。

如果要进行单边检验，比如原假设仍为$H_0:\beta_2=0.1$，而替代假设为$H_1:\beta_2>0.1$，则拒绝域在t分布的右侧尾部。

相应的t统计量仍为 0.45188757，但在计算p值时，只须计算大于此t统计量的右侧尾部概率即可：

```
. dis ttail(752,0.45188757)
.32574014
```

由于 p 值仍高达 0.3257，故仍可接受原假设。

考虑检验 *expr* 与 *tenure* 的系数是否相等，即检验 $H_0:\beta_3=\beta_4$，可输入命令：

```
. test expr=tenure
```

```
 ( 1)  expr - tenure = 0

       F(  1,   752) =    0.05
            Prob > F =    0.8208
```

由于 p 值 $=0.8208$，可以轻松地接受原假设。

为演示目的，考虑检验工龄回报率与现单位年限回报率之和是否等于教育回报率，即“$H_0:\beta_3+\beta_4=\beta_2$”，可使用命令：

```
. test expr+tenure=s
```

```
 ( 1)  - s + expr + tenure = 0

       F(  1,   752) =    8.82
            Prob > F =    0.0031
```

由于p值$=0.0031$，故可在1%的显著性水平上拒绝原假设，即认为$\beta_3+\beta_4\neq\beta_2$。