中央财经大学
Central University of Finance and Economics

计量经济学

中国财政发展协同创新中心
陈怡心
cyx@cufe.edu.cn

第3章 多元线性回归

陈怡心
中央财经大学中国财政发展协同创新中心

二元线性回归

- 一元回归很可能遗漏了其他因素。
- 比如，在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。
- 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，其中年龄与工龄可视为“在职培训”(on the job training)的代理变量， 而在职培训是增加人力资本(human capital)的另一重要方式。

二元线性回归

- 首先考察比较简单的二元回归。

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

其中，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），而γ为给定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：

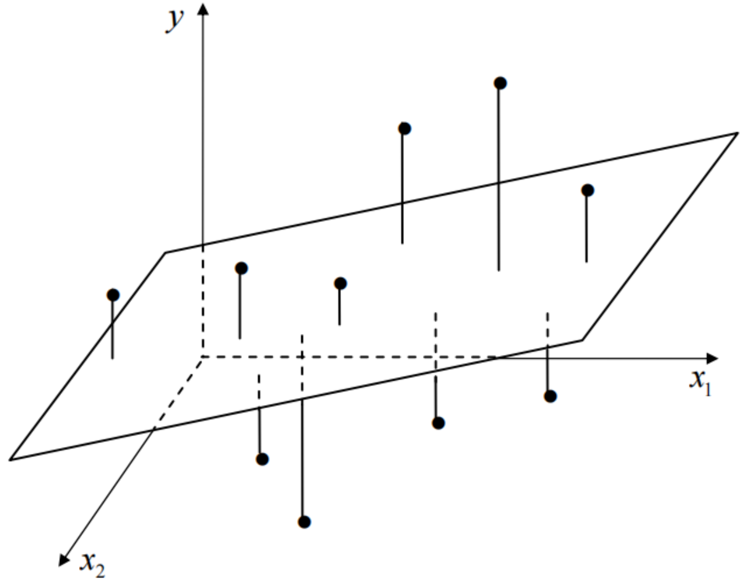
$$\min_{\hat{\alpha}, \hat{\beta}, \hat{\gamma}} \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} \left(y_i - \hat{\alpha} - \hat{\beta} x_{i1} - \hat{\gamma} x_{i2}\right)^2$$

二元线性回归

- 从几何上，这意味着寻找一个回归平面$\hat{y}_i \equiv \hat{\alpha} + \hat{\beta} x_{1i} + \hat{\gamma} x_{2i}$，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。
- 将目标函数分别对$\hat{\alpha}$，$\hat{\beta}$，$\hat{\gamma}$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。



多元线性回归模型

- 一般的多元线性回归模型可写为

$$y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + \varepsilon_i \quad (i = 1, \ldots, n)$$

- 其中，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。
- 一般地， 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

$$y_i = \beta_1 + \beta_2 x_{i2} + \cdots \beta_k x_{ik} + \varepsilon_i \quad (i = 1, \ldots, n)$$

- 多元线性回归模型可以更方便地以矩阵表示。如果不考虑扰动项 ε_i，此方程右边的主体部分为乘积之和，即$\sum_{k=1}^{k}\beta_k x_{\mathrm{ik}}$，其中 $x_{i1}\equiv 1$。
- 根据线性代数知识，乘积之和可写为两个向量的内积。
- 定义列向量$\boldsymbol{x_i}\equiv(1\ x_{i2}\ \cdots\ x_{ik})'$（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$\sum_{k=1}^{k}\beta_k x_{\mathrm{ik}}=\boldsymbol{x_i}'\boldsymbol{\beta}$。故可将原模型写为

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

- 上式对所有个体i都成立 $(i=1,\ldots,n)$，故有n个如上式的方程。将所有这n个方程叠放

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

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

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

- $\boldsymbol{y} \equiv (\, y_1 \;\; y_2 \;\; \cdots y_n)'$为被解释变量构成的列向量
- $\boldsymbol{\varepsilon} \equiv (\, \varepsilon_1 \;\; \varepsilon_2 \;\; \cdots \varepsilon_n)'$为所有扰动项构成的列向量
- ***X***为n×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

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

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

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

- 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(superplane)。

- 一阶条件为

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

- 消去方程左边的“- 2”可得

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

- 这是包含 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$\hat{\boldsymbol{\beta}} = (\hat{\beta}_1, \hat{\beta}_2, \cdots, \hat{\beta}_k)'$ 称为 OLS 估计量(OLS estimator)。

- 正规方程组可方便地用矩阵来表达。
- 由于残差$e_i \equiv y_i - \hat{\beta}_1 - \hat{\beta}_2 x_{i2} - \cdots - \hat{\beta}_k x_{ik}$，故正规方程组可写为

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

- 上式每一方程都是乘积求和的形式，可用向量内积表示

- 第1 个方程可写为

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

- 第2 个方程可写为

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

- 以此类推，第 K 个方程可写为

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

- 残差向量$\boldsymbol{e} \equiv (e_1 \quad e_2 \quad \cdots \quad e_n)'$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：

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

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

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

- 由于$\boldsymbol{X}'$的第 k 行包含第 k 个解释变量的全部观测值，根据$\boldsymbol{X}'\boldsymbol{e} = 0$也可看出，残差向量$\boldsymbol{e}$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

$$\boldsymbol{e} = \boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}$$

- 代入正规方程组可得

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

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

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

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

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

- 这就是多元回归的 OLS 估计量。

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

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

- 将所有个体的拟合值写为列向量$\hat{\boldsymbol{y}}$，推导可得

$$\hat{\boldsymbol{y}} \equiv (\hat{y}_1 \quad \hat{y}_2 \quad \cdots \quad \hat{y}_n)' = \boldsymbol{X}\hat{\boldsymbol{\beta}}$$

- 拟合值向量与残差向量正交，因为

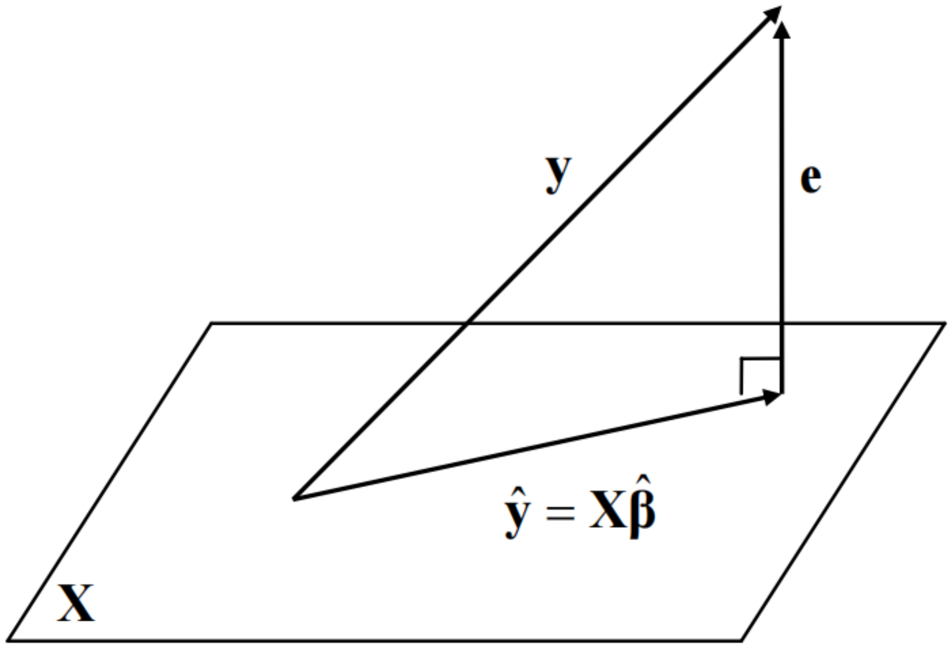
$$\hat{\boldsymbol{y}}'\boldsymbol{e} = (\boldsymbol{X}\hat{\boldsymbol{\beta}})'\boldsymbol{e} = \hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{e} = \hat{\boldsymbol{\beta}}' \cdot \boldsymbol{0} = 0$$

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

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

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

- 拟合值 $\hat{\boldsymbol{y}}$可视为被解释变量 $\boldsymbol{y}$向解释变量超平面$\boldsymbol{X}$的投影(projection)。
- 由于拟合值为解释变量的线性组合，即 $\hat{\boldsymbol{y}} = \boldsymbol{X}\hat{\boldsymbol{\beta}}$ ，故拟合值向量$\hat{\boldsymbol{y}}$正好在超平$\boldsymbol{X}$上。
- 根据 OLS 的正交性，残差向量$\boldsymbol{e}$与$\hat{\boldsymbol{y}}$正交。



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)，故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

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

- 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，可定义拟合优度。

- **定义 拟合优度R^2为**

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

- 拟合优度的缺点是，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。
- 另一方面，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 。

- 引入校正拟合优度，对解释变量过多(模型不够简洁)进行惩罚。

- **定义 校正拟合优度(adjusted $\boldsymbol{R^2}$) $\bar{\boldsymbol{R}}^2$为**

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

- $\sum_{i=1}^{n} e_i^2$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的约束，故只有其中$(n-K)$个残差是(自由)独立的。
- 给定$\{e_1,\cdots,e_{n-k}\}$，即可根据正规方程组求解其余$\{e_{n-k+1},\cdots,e_n\}$。

- 类似地，$\sum_{i=1}^{n}(y_i - \bar{y})^2$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 $(n-1)$个离差是（自由）独立的。
- $\bar{R}^2$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)

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

古典线性回归模型的假定

- 为得到 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 1 线性假定(linearity)**
- 总体模型为

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

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

补充：回归方程的函数形式

- 连续型解释变量的边际效应：不变还是可变？
  ——交叉项
- 回归方程中变量的函数形式：线性还是对数？
  ——自然对数变换

补充：交叉项

- 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$x_{i2}x_{i3}$)
- 交叉项也称为互动项、交互项(interaction term)。
- 如果在线性模型中增加两个连续型解释变量的乘积作为新增解释变量，那么就会使得原有两个解释变量对y的边际效应是可变的。它允许某个变量的边际效应可以是自身以及其他解释变量的线性函数。

- 平方项，是一种特殊的交互项形式。

- 例(平方项)：考虑如下回归方程

$$\ln w_i = \beta_1 + \beta_2 s_i + \beta_3 s_i^2 + \varepsilon_i$$

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

$$\frac{\partial \ln w_i}{\partial s_i} = \beta_2 + 2\beta_3 s_i$$

- 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。
- 如果变量的边际效应不是常数，可考虑在回归方程中加入平方项。
- 只要将s^2也视为解释变量，则仍符合线性模型的假定。

- 例(互动项)：考虑如下生产函数方程

$$y_i = \beta_1 + \beta_2 k_i + \beta_3 l_i + \beta_4 k_i \times l_i + \varepsilon_i$$

- 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。

补充：自然对数变换

- 自然对数变换定义为那些定义为严格为正的变量：对于所有i，$y_i > 0$，那么$\ln y_i$才有定义；$x_{ij} > 0$，那么$\ln x_{ij}$才有定义。
- $\ln y_i$和$\ln x_{ij}$的微分都是相对的，或者表明其是y_i和x_{ij}的变化比例。

$$d\ln y_i = dy_i / y_i$$

- 它表明了对于所有$y_i > 0$，y_i的相对变化比例。

$$d\ln x_{ij} = dx_{ij} / x_{ij}$$

- 它表明了对于所有$x_{ij} > 0$，x_{ij}的相对变化比例。

补充：自然对数变换

- y_i和x_{ij}的变动百分比可以使用相对变动比例乘以100，从而有：

$$100(\mathrm{dln}y_i) = 100\left(\frac{\mathrm{d}y_i}{y_i}\right)$$

- 它表明了对于所有$y_i > 0$，y_i变动的百分比。

$$100(\mathrm{dln}x_{ij}) = 100\left(\frac{\mathrm{d}x_{ij}}{x_{ij}}\right)$$

- 它表明了对于所有$x_{ij} > 0$，x_{ij}变动的百分比。
- 从这个定义可以延伸到经济学上的弹性含义：y相对于x_j的弹性定义为：

$$\varepsilon_j = \frac{\mathrm{dln}y}{\mathrm{dln}x_j} = \frac{\mathrm{d}y/y}{\mathrm{d}x_j/x_j} = \frac{\mathrm{d}y}{\mathrm{d}x_j}\frac{x_j}{y}$$

- 例(函数形式)：经济学中常用的生产函数为 Cobb-Douglas 生产函数：

$$y_i = e^{\beta_1} k_i^{\beta_2} l_i^{\beta_3} e^{\varepsilon_i}$$

- e^{β_1}与e^{ε_i}分别为乘积形式的常数项与扰动项。两边取对数：

$$\ln y_i = \beta_1 + \beta_2 \ln k_i + \beta_3 \ln l_i + \varepsilon_i$$

- 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。
- 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 lnx)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 $(\boldsymbol{\beta_1} \cdots \boldsymbol{\beta_K})$的线性函数**(linear in parameters)。

补充：四种常见的函数设定形式：

- 模型1：线性线性模型

$$y_i = \beta_1 + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + u_i$$

- 所有变量均以线性形式出现，所有斜率系数直接取决于y和x_j测量单位，且是常数：

$$\beta_j = \frac{\partial y_i}{\partial x_{ij}} = \frac{y_i\text{的变动}}{x_{ij}\text{的变动}}$$

补充：四种常见的函数设定形式：

- 模型2：双对数模型

$$\ln y_i = \alpha_1 + \alpha_2 \ln x_{i2} + \cdots + \alpha_k \ln x_{ik} + u_i$$

- 所有变量均以对数形式出现，斜率系数不依赖于y_i和x_j的测量单位，且解释为弹性：

$$\alpha_j = \frac{\ln y_i}{\ln x_{ij}} = \frac{\partial y_i / y_i}{\partial x_{ij} / x_{ij}} = \frac{\%\Delta y_i}{\%\Delta x_{ij}}$$

- 这就是y_i相对于x_{ij}的偏弹性。

补充：四种常见的函数设定形式：

- 模型3：对数线性模型（半对数模型）

$$\mathrm{ln}y_i = \gamma_1 + \gamma_2 x_{i2} + \cdots + \gamma_k x_{ik} + u_i$$

- 被解释变量y_i以对数形式出现，但是解释变量x_j以线性形式进入。斜率系数解释为：

$$\gamma_j = \frac{\partial \mathrm{ln}y_i}{\partial x_{ij}} = \frac{\partial y_i / y_i}{\partial x_{ij}} = \frac{y_i\text{的相对变化}}{x_{ij}\text{的绝对变化}}$$

$$100\gamma_j = \frac{y_i\text{的百分比变化}}{x_{ij}\text{的绝对变化}} = \frac{\%\Delta y_i}{\Delta x_{ij}}$$

- 表示y相对于x_j的偏半弹性
- x增加100单位，导致了y增加了$100\gamma_j\%$。

补充：四种常见的函数设定形式：

- 模型4：线性对数模型（逆半对数模型）

$$y_i = \phi_1 + \phi_2 \ln x_{i2} + \cdots + \phi_k \ln x_{ik} + u_i$$

- 被解释变量以线性形式进入，但是解释变量以对数形式进入，其斜率系数解释为：

$$\phi_j = \frac{\partial y_i}{\partial \ln x_{ij}} = \frac{\partial y_i}{\partial x_{ij}/x_{ij}} = \frac{y_i\text{的绝对变化}}{x_{ij}\text{的相对变化}}$$

- x增加100%导致y增加$\widehat{\phi}_2$个单位。

- **假定 2 严格外生性(strict exogeneity)**

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

- 严格外生性意味着，在给定数据矩阵$\boldsymbol{X}$的情况下，扰动项ε_i的条件期望为 0。
- 因此，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，大样本 OLS 可放松。

- 均值独立仅要求$\mathrm{E}(\varepsilon_i \mid \boldsymbol{X}) = c$，其中 c为常数，不一定为 0。
- 当回归方程有常数项时，要求$\mathrm{E}(\varepsilon_i \mid \boldsymbol{X}) = 0$ 并不会带来过多限制，因为如果$\mathrm{E}(\varepsilon_i \mid \boldsymbol{X}) = c \neq 0$，总可以把 c归入常数项。

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

$$\mathrm{E}(\varepsilon_i) = \mathrm{E}_{\boldsymbol{X}} \underbrace{\mathrm{E}(\varepsilon_i \mid \boldsymbol{X})}_{=0} = \mathrm{E}_{\boldsymbol{X}}(0) = 0$$

- 上式使用了迭代期望定律。

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

- 在线性代数中，如果两个向量的内积为 0，则这两个向量正交。
- 在概率统计中，两个随机变量的正交定义有所不同。

- 定义 如果随机变量 x, y满 足$E(xy) = 0$ ，则称 x, y 正 交(orthogonal)。
- 根据此定义，可证明解释变量与扰动项正交，因为

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

- **假定 3 不存在“严格多重共线性”(strict multicollinearity)，即数据矩阵X满列秩(full column rank)。**
- 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。
- 换言之， $\boldsymbol{X}$ 中不存在多余的变量.

- 考虑以下一元回归模型：

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

- 其数据矩阵$\boldsymbol{X}$为

$$\boldsymbol{X}=\begin{pmatrix}1 & s_1\\ 1 & s_2\\ \vdots & \vdots\\ 1 & s_n\end{pmatrix}$$

- 数据矩阵$\boldsymbol{X}$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

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

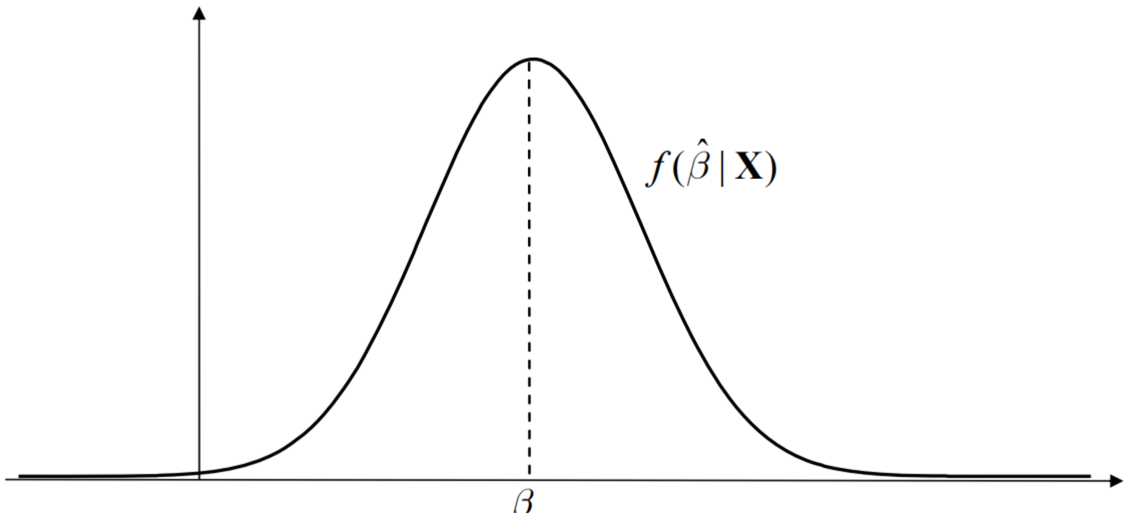
- $\bar{s}$与$\overline{lnw}$分别为 s与 lnw 的样本均值。

- 更一般地，对于多元回归，如果$\boldsymbol{X}$满列秩，则$\boldsymbol{X}'\boldsymbol{X}$为正定矩阵 (positive definite)， 故 $(\boldsymbol{X}'\boldsymbol{X})^{-1}$存在，可计算 OLS 估计量 $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$。
- 反之，如果$\boldsymbol{X}$不满列秩，则$(\boldsymbol{X}'\boldsymbol{X})^{-1}$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$\boldsymbol{X}$满列秩只是对数据的最低要求。
- 现实数据不容易出现严格多重共线性。

OLS的小样本性质

- OLS 估计量 $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$样本数据的函数，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 1-3 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)。
- OLS 估计量的表达式 $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$可知， $\hat{\boldsymbol{\beta}}$可视为 $\boldsymbol{y}$的线性组合(将 $[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}']$视为系数矩阵)，故为线性估计量。

- (2) **无偏性**： $\mathrm{E}(\hat{\boldsymbol{\beta}} \mid \boldsymbol{X}) = \boldsymbol{\beta}$ ，即$\hat{\boldsymbol{\beta}}$不会系统地高估或低估$\boldsymbol{\beta}$ 。



- 证明：抽样误差(sampling error)为

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

- 记 $\boldsymbol{A} \equiv (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$ 。给定解释变量$\boldsymbol{X}$，对上式两边求条件期望，根据严格外生性可得

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

- 移项可得$\mathrm{E}(\hat{\boldsymbol{\beta}} \mid \boldsymbol{X}) = \boldsymbol{\beta}$ 。
- 在此证明中，严格外生性不可或缺。
- 使用迭代期望定律，可进一步证明，无条件期望$\mathrm{E}(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$ ，因为

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

- (3) 估计量$\hat{\boldsymbol{\beta}}$的协方差矩阵
- 由于$\boldsymbol{\beta}$为常数，故$Var(\hat{\boldsymbol{\beta}}|\boldsymbol{X}) = Var(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}|\boldsymbol{X})$因此

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

- 在上式中，使用了协方差矩阵的夹心估计量表达式，而且$\boldsymbol{A} \equiv (\boldsymbol{X}'\boldsymbol{X})^{-\mathbf{1}}\boldsymbol{X}', \boldsymbol{A}' \equiv \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}$ (其中，$(\boldsymbol{X}'\boldsymbol{X})^{-1}$为对称矩阵)。
- 古典模型假定$Var(\boldsymbol{\varepsilon}|\boldsymbol{X}) = \sigma^2\boldsymbol{I}_n$ 称为“**球形扰动项**”。
- 在球形扰动项的假定下，表达式可大大简化：

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

- **假定 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

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

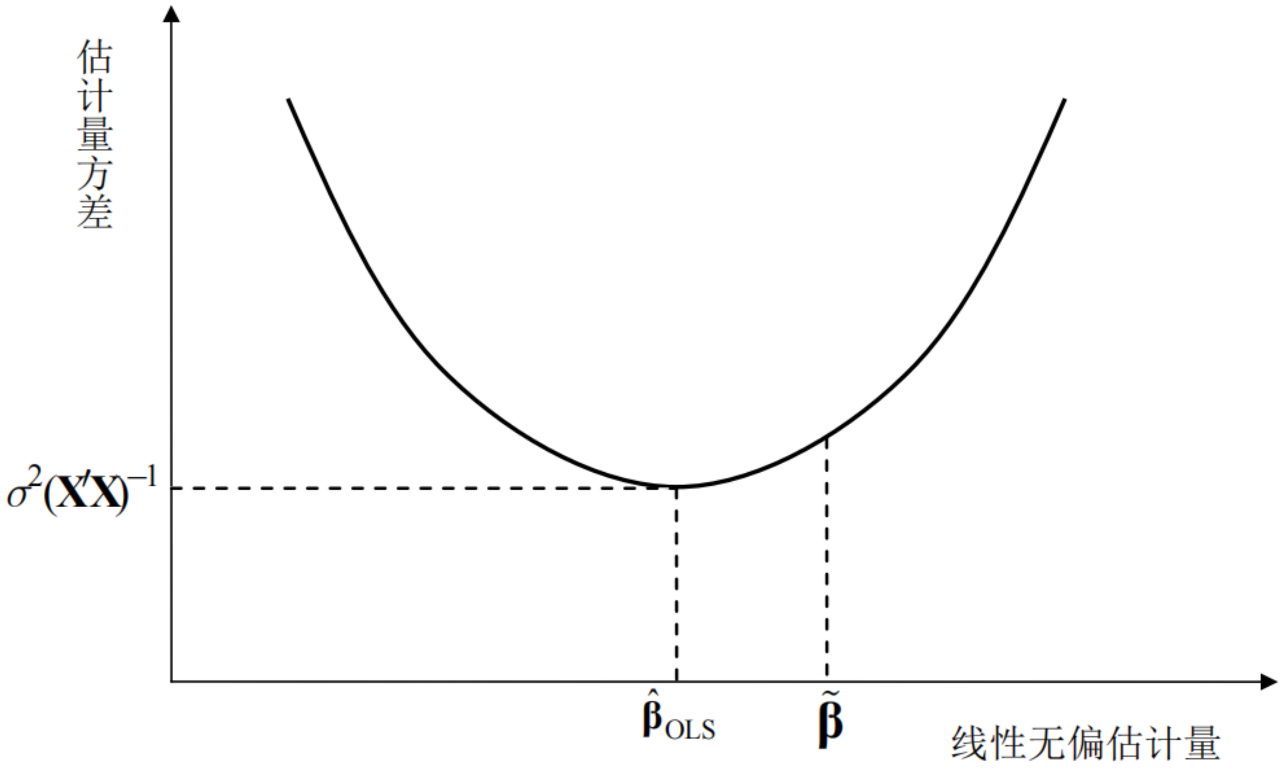
- 协方差矩阵$Var(\boldsymbol{\varepsilon}|\boldsymbol{X})$的主对角线元素都等于$\sigma^2$，即满足“条件同方差”(conditional homoskedasticity)，简称“同方差”。
- 如果不完全相等，则存在“条件异方差” (conditional heteroskedasticity)，简称“异方差”。
- 协方差矩阵$Var(\boldsymbol{\varepsilon}|\boldsymbol{X})$的非主对角线元素都为0，故不同个体的扰动项之间无“自相关”(autocorrelation)或“序列相关”(serial 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$\sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}$的关键(无偏性不依赖于球形扰动项)。
- 此表达式虽然简单，但只在同方差与无自相关的情况下才成立。
- 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)。
- 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。

高斯-马尔可夫定理(Gauss-Markov Theorem)

- 在假定 1-4 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。
- 严格来说，记 OLS 估计量为$\hat{\boldsymbol{\beta}}_{OLS}$，而任一线性无偏估计量为$\widetilde{\boldsymbol{\beta}}$则$[Var(\widetilde{\boldsymbol{\beta}}|\boldsymbol{X})-Var(\hat{\boldsymbol{\beta}}_{OLS}|\boldsymbol{X})]$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。
- 如果不满足球形扰动项(比如，存在异方差或自相关)，则高斯-马尔可夫定理不成立。



- （5）对扰动项方差的无偏估计
- 对于扰动项方差 $\sigma^2 = Var(\varepsilon_i)$，由于$\{\varepsilon_1, \cdots, \varepsilon_n\}$不可观测，$\{e_1, \cdots, e_n\}$将视为其实现值，得到对$\sigma^2$的估计：

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

- $(n-K)$为自由度。
- 为什么除以$(n-K)$而不除以n？
- 虽然有n个残差$\{e_1, e_2 \cdots, e_n\}$，但随机变量$\{e_1, e_2 \cdots, e_n\}$须满足K个正规方程$\boldsymbol{X}'\boldsymbol{e} = \boldsymbol{0}$，故只有其中$(n-K)$个残差是(自由)独立的。
- 经过校正后，才是“无偏估计”(unbiased estimator)，即$E(s^2) = \sigma^2$。

- 如果样本容量n很大，当$n \to \infty$时，则$\frac{n-K}{n} \to 1$是否进行“小样本校正”(small sample adjustment)并无多大差别。
- $s = \sqrt{s^2}$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 $\sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}$可用 $s^2(\boldsymbol{X}'\boldsymbol{X})^{-1}$来估计。
- 特别地，$\hat{\boldsymbol{\beta}}$的第k个分量$\hat{\beta}_k$的估计方差$s^2\,(\boldsymbol{X}'\boldsymbol{X})^{-1}_{kk}$，其中$(\boldsymbol{X}'\boldsymbol{X})^{-1}_{kk}$表示矩阵$(\boldsymbol{X}'\boldsymbol{X})^{-1}$的$(k,k)$元素，即主对角线上的第$k$个元素。

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

$$SE(\hat{\beta}_k) = \sqrt{s^2\ (\boldsymbol{X}'\boldsymbol{X})^{-1}_{kk}}$$

- 为什么称为“标准误差”?
- $Var(\hat{\beta}_k) = Var(\underbrace{\hat{\beta}_k - \beta_k}_{sampling\ error})$，故$\sqrt{Var(\hat{\beta}_k)}$=$\sqrt{Var(\hat{\beta}_k - \beta_k)}$
- “抽样误差的标准差”，简称“标准误（差）”
- 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)。

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$n \to \infty$，也称“有限样本理论”(finite sample theory)。
- 大样本理论要求$n \to \infty$，适用于较大的样本容量。

对单个系数的t检验

- 小样本理论虽适用于各种样本容量，但不易推导统计量的分布，需对随机变量的概率分布作很强假定。

- **假定 5** 在给定$\boldsymbol{X}$的情况下，$\boldsymbol{\varepsilon}|\boldsymbol{X}$ 的条件分布为正态，

即$\boldsymbol{\varepsilon}|\boldsymbol{X} \sim N(\mathbf{0}, \sigma^2 \boldsymbol{I_n})$。

对单个系数的t检验

- 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

$$H_0\text{：}\ \beta_k = c$$

- c为给定常数，也称为“假想值”(hypothesized value)。
- 通常$c = 0$，即检验变量x_{ik}的系数是否显著地不等于 0。

- 假设检验是一种概率意义上的反证法。
- 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，应拒绝原假设，接受“替代假设”(alternative hypothesis，也称“备择假设”)：

$$H_1\text{：}\ \beta_k \neq c$$

- 替代假设“H_1：$\beta_k \neq c$”也称“双边替代假设”(two-sided alternative hypothesis)，因为它既包括$\beta_k > c$，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，则应倾向于拒绝原假设。
- 使用此原理的这类统计检验称为“沃尔德检验”(Wald test)。
- 在衡量距离时，由于绝对距离依赖于变量单位，故需以标准差为基准来考虑相对距离。
- 由于扰动项$\boldsymbol{\varepsilon}|\boldsymbol{X} \sim N(\boldsymbol{0}, \sigma^2 \boldsymbol{I}_n)$，而$\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta} = \boldsymbol{A}\boldsymbol{\varepsilon}$为$\boldsymbol{\varepsilon}$的线性函数（其中$A \equiv (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$，故$\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}|\boldsymbol{X}$也服从正态分布。
- 由于$E\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}|\boldsymbol{X}\right) = 0$，$Var\left(\widehat{\boldsymbol{\beta}}|\boldsymbol{X}\right) = \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}$,故
$$\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}|\boldsymbol{X} \sim N\left(\boldsymbol{0}, \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}\right)$$
- 单独考虑上式的第k个分量：
$$\hat{\beta}_k - \beta_k|\boldsymbol{X} \sim N\left(0, \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}_{kk}\right)$$
- 其中，$(\boldsymbol{X}'\boldsymbol{X})^{-1}_{kk}$为矩阵$(\boldsymbol{X}'\boldsymbol{X})^{-1}$的$(k, k)$元素（即主对角线上第k个元素），而$\sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}_{kk}$为$\hat{\beta}_k$的方差。

- 在原假设“H_0：$\beta_k = c$”成立的情况下：

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

- 如果σ^2已知，则标准化的统计量服从标准正态分布：

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

- 但通常σ^2未知，称为“厌恶参数”(nuisance parameter)。

- 合格的“检验统计量”(test statistic)须满足两个条件：

  （1）能根据样本数据算出。

  （2）它的概率分布已知。

- 对于统计量z_k，虽已知其分布为标准正态，但不知道σ^2，故无法根据数据计算统计量z_k的样本观测值。
- 以σ^2的估计量s^2替代σ^2，可得以下t统计量(t-statistic)。

- **定理(t 统计量的分布)** 在假定 1-5 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：

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

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

- 更一般地，t统计量的通用公式为

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

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

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。

　因此，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)$t_{\alpha/2}(n-K)$

  $t_{\alpha/2}(n-K)$的定义为

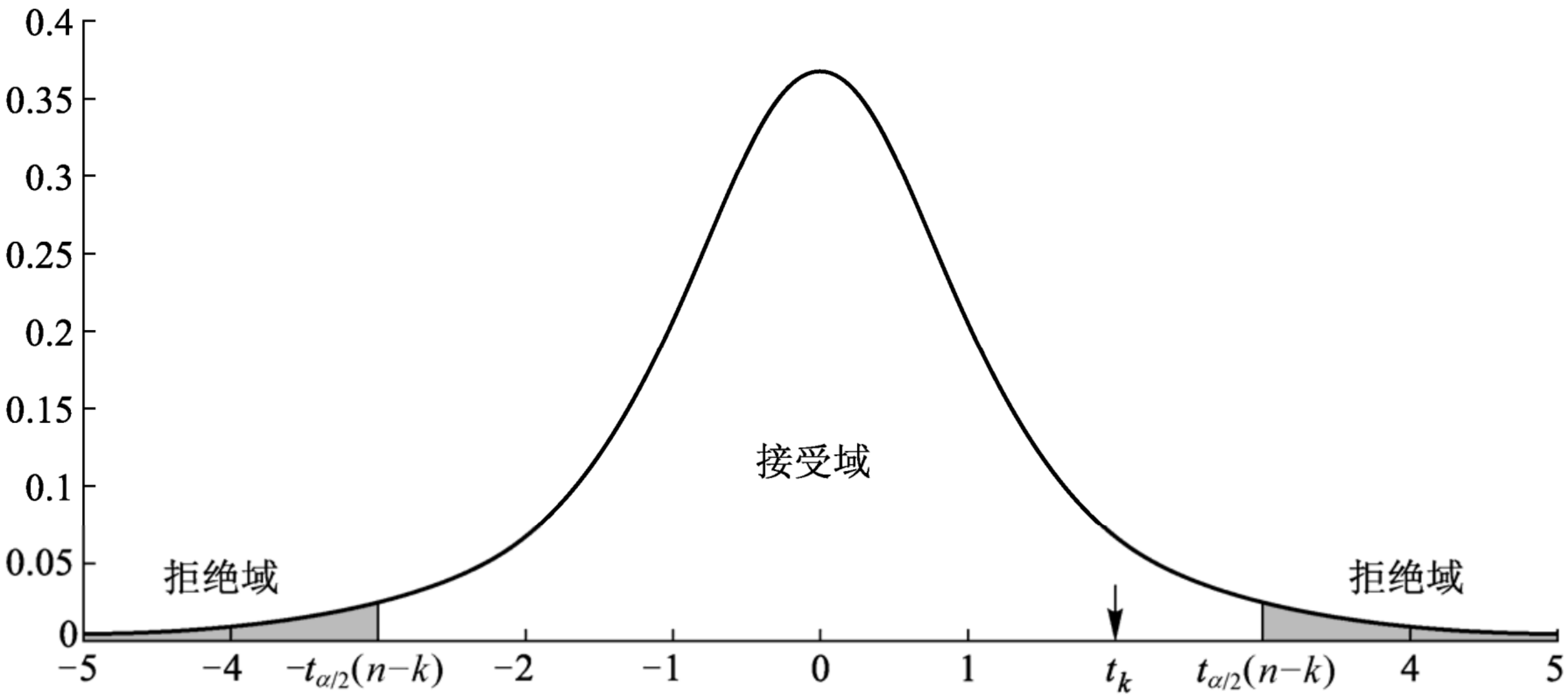
$$\mathrm{P}\left\{T > t_{\alpha/2}(n-K)\right\} = \mathrm{P}\left\{T < -t_{\alpha/2}(n-K)\right\} = \alpha/2$$

  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$\alpha/2$。

  通常取$\alpha = 5\%$，则$\alpha/2 = 2.5\%$。有时也使用$\alpha = 1\%$或$\alpha = 10\%$。

- 第三步：如果$|t_k| \geq t_{\alpha/2}(n-K)$，则 落入$t_k$ “拒绝域”(rejection region)，故拒绝H_0。反之，如果$|t_k| < t_{\alpha/2}(n-K)$，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“双边检验”(two-sided test)。



- t检验临界值表

续表

n'	$P(2)$:	0.50	0.20	0.10	0.05	0.02	0.01	0.005	0.002	0.001
	$P(1)$:	0.25	0.10	0.05	0.025	0.01	0.005	0.0025	0.001	0.0005
26		0.684	1.315	1.706	2.056	2.479	2.779	3.067	3.435	3.707
27		0.684	1.314	1.703	2.052	2.473	2.771	3.057	3.421	3.690
28		0.683	1.313	1.701	2.048	2.467	2.763	3.047	3.408	3.674
29		0.683	1.311	1.699	2.045	2.462	2.756	3.038	3.396	3.659
30		0.683	1.310	1.697	2.042	2.457	2.750	3.030	3.385	3.646
31		0.682	1.309	1.696	2.040	2.453	2.744	3.022	3.375	3.633
32		0.682	1.309	1.694	2.037	2.449	2.738	3.015	3.365	3.622
33		0.682	1.308	1.692	2.035	2.445	2.733	3.008	3.356	3.611
34		0.682	1.307	1.691	2.032	2.441	2.728	3.002	3.348	3.601
35		0.682	1.306	1.690	2.030	2.438	2.724	2.996	3.340	3.591
36		0.681	1.306	1.688	2.028	2.434	2.719	2.990	3.333	3.582
37		0.681	1.305	1.687	2.026	2.431	2.715	2.985	3.326	3.574
38		0.681	1.304	1.686	2.024	2.429	2.712	2.980	3.319	3.566
39		0.681	1.304	1.685	2.023	2.426	2.708	2.976	3.313	3.558
40		0.681	1.303	1.684	2.021	2.423	2.704	2.971	3.307	3.551
50		0.679	1.299	1.676	2.009	2.403	2.678	2.937	3.261	3.496
60		0.679	1.296	1.671	2.000	2.390	2.660	2.915	3.232	3.460
70		0.678	1.294	1.667	1.994	2.381	2.648	2.899	3.211	3.435
80		0.678	1.292	1.664	1.990	2.374	2.639	2.887	3.195	3.416
90		0.677	1.291	1.662	1.987	2.368	2.632	2.878	3.183	3.402
100		0.677	1.290	1.660	1.984	2.364	2.626	2.871	3.174	3.390
200		0.676	1.286	1.653	1.972	2.345	2.601	2.839	3.131	3.340
500		0.675	1.283	1.648	1.965	2.334	2.586	2.820	3.107	3.310
1000		0.675	1.282	1.646	1.962	2.330	2.581	2.813	3.098	3.300
∞		0.6745	1.2816	1.6449	1.9600	2.3263	2.5758	2.8070	3.0902	3.2905

注：表上图中的阴影部分表示概率P，$P(2)$是双侧的概率，$P(1)$是单侧的概率，n'是自由度。以后附表同此。

计算p值

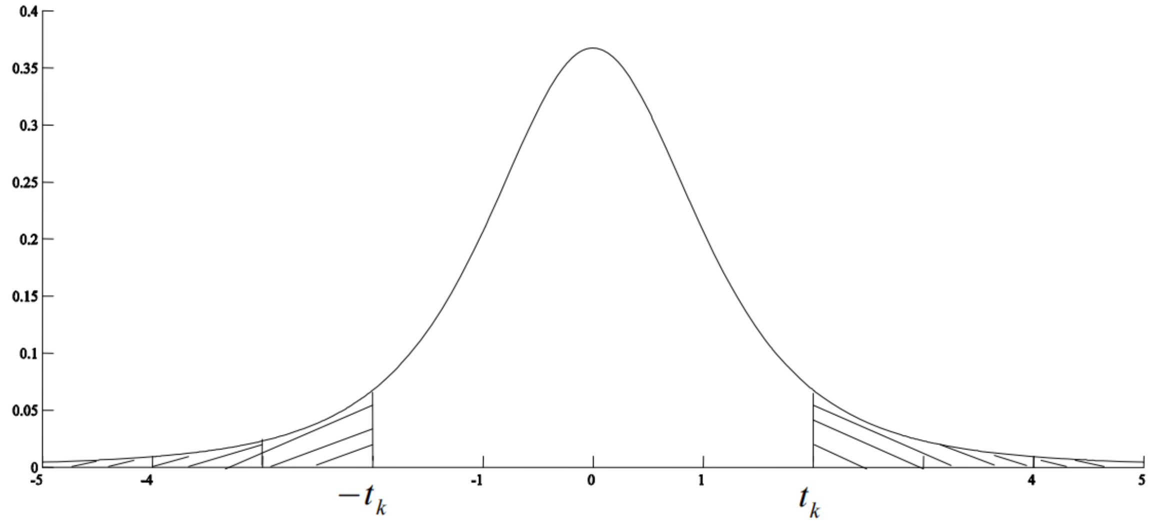
- 假设检验的基本逻辑就是概率意义上的反证法。
- 观测到样本数据的发生概率究竟小到何种程度，可通过下面的p值来度量。
- 在双边t检验中，给定t统计量的样本观测值t_k，此假设检验问题

的p值(probability value，即 p-value)为

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

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。



- 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。
- **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。
- 如选定显著性水平为 5%，只要 p值比 0.05 小，即可拒绝原假设。
- 比如, p值 = 0.03，则可在 5%的显著性水平上拒绝原假设。
- 进一步,“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，操作简便。

计算置信区间

- 有时点估计还不够，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α =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 $t_k \equiv \frac{\hat{\beta}_k - c}{SE(\hat{\beta}_k)} \sim t(n-K)$，故$t$统计量落入接受域的概率为 $(1-\alpha)$

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

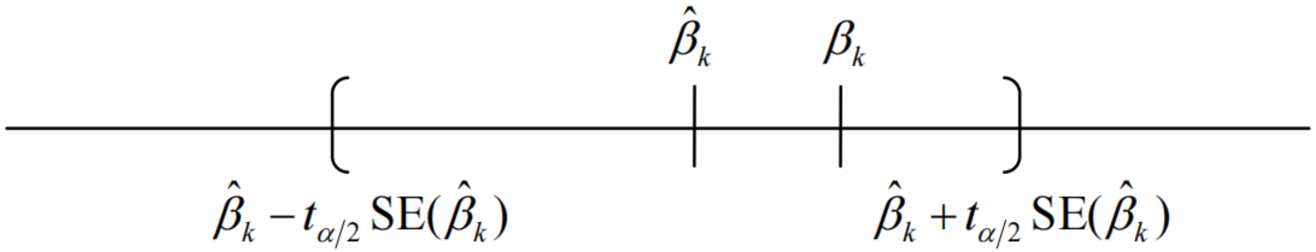
其中，$t_{\alpha/2}$为显著性水平为α的临界值。将不等式变形可得

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

由此可知，β_k的置信区间为

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

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



- 标准误SE（$\hat{\beta}_k$）越大，对β_k的估计越不准确，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

- 有时也进行单边检验(one-sided test)。
- 考虑原假设为H_0：$\beta_k = c$，而替代假设为H_1：$\beta_k > c$
- 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：

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

- 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。

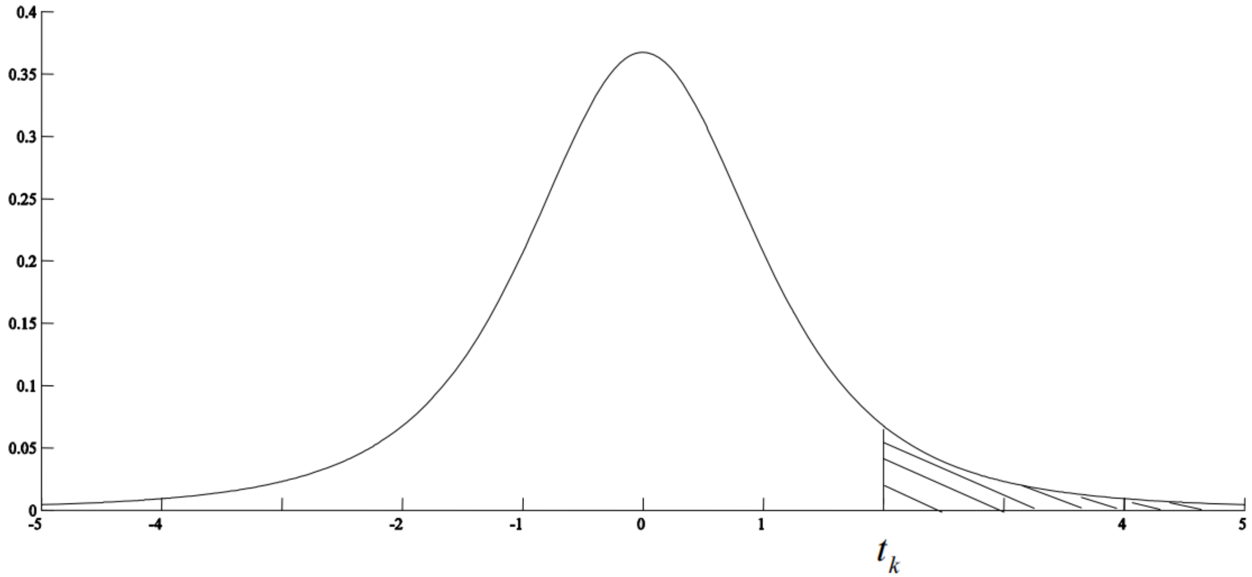
- 拒绝域只在概率分布的最右边一侧。
- 给定显著性水平α后，计算的临界值为$t_{\alpha/2}(n-K)$，使得取值大于此临界值的概率为α：

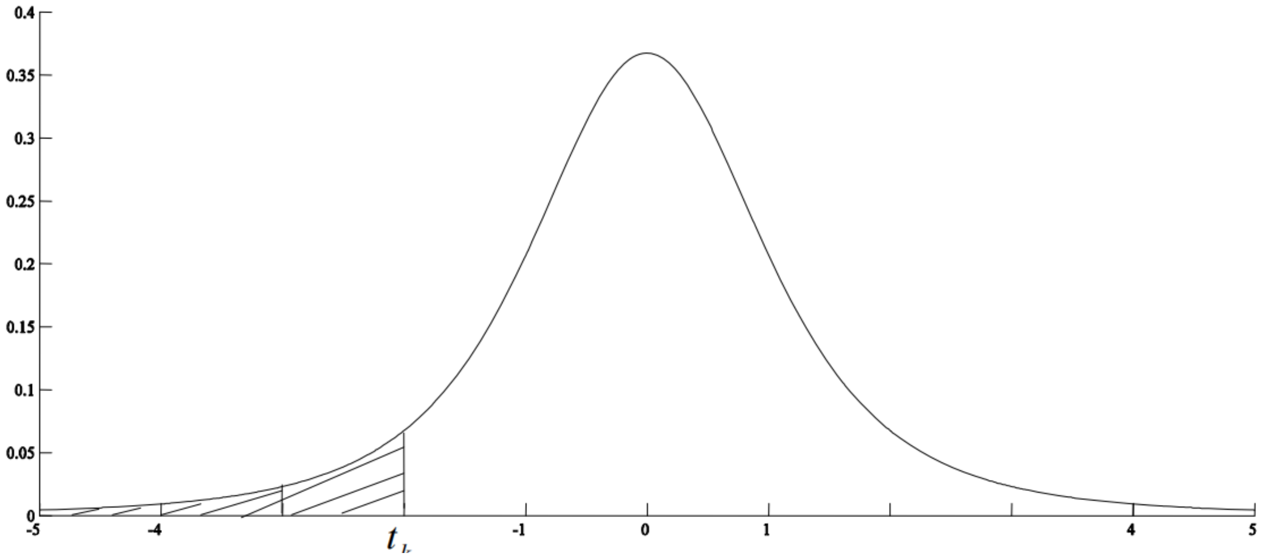
$$\mathrm{P}\left\{T > t_{\alpha}(n-K)\right\} = \alpha$$

- 其中，随机变量$T \sim t(n-K)$
- 如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。

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

- 拒绝域只在分布的右侧尾部，故称“单边右侧检验”(one-sided right-tail test)。


0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
-5
-4
-2
-1
0
1
4
5
t_k


0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
-5
-4
-1
0
1
4
5

- 如果原假设H_0：$\beta_k = c$，而替代假设为H_1：$\beta_k < c$,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。

- 对于单边左侧检验，计算 p值的公式为

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

- 其中，随机变量$T \sim t(n-K)$

t检验举例

- PRE： $y_i = \beta_1 + \beta_2 x_i + u_i, i = 1, \ldots, N (N = 74)$

- β_2的估计结果：

$$\hat{\beta}_2 = 2.044063，\ se(\hat{\beta}_2) = 0.3768341$$

- 假设检验： 在5%显著性水平下， $H_0: \beta_2 = 0 \text{ vs. } H_1: \beta_2 \neq 0$

t检验举例

- 方法1： t检验统计量法
- (1)使用t检验统计量和临界值$t_{\hat{\beta}_2} = \frac{\hat{\beta}_2}{se(\hat{\beta}_2)} = \frac{2.044063}{0.3768341} = 5.424$
- (2)感兴趣的临界值：

$$t_{1-\frac{\alpha}{2}}(N-k) = t_{1-0.025}(74-2) = t_{0.975}(72) = |t_{0.025}(72)|$$

- 查t分布表：

$$t_{0.975}(60) = 2.000,\quad t_{0.975}(120) = 1.980,$$

- 从而$t_{0.025}(72)$的取值位于2.000和1.980之间。
- (3)将$t_{\hat{\beta}_2} = 5.424$与临界值$t_{0.975}(72)$进行对比：

$$t_{\hat{\beta}_2} = 5.424 > 2.000 > t_{0.975}(72)$$

- (4)拒绝原假设。

t检验举例

- 方法2：为β_2计算95%水平（$= 1 - \alpha$）的置信区间。
- (1)当$\left|t_{\hat{\beta}_2}\right| \le t_{0.975}(72)$时，我们保留原假设，这就意味着：

$$\hat{\beta}_2 - t_{0.975}(72) \cdot se\left(\hat{\beta}_2\right) \le \beta_2 \le \hat{\beta}_2 + t_{0.975}(72) \cdot se(\hat{\beta}_2)$$

- (2)代入估计值，于是：

$$2.044063 - t_{0.975}(72) \cdot 0.3768341 \le \beta_2 \le 2.044063 + t_{0.975}(72) \cdot 0.3768341$$
$$1.292858 \le \beta_2 \le 2.795268$$

- (3)95%的置信区间是[1.292858,2.795268]，原假设$\beta_2 = 0$位于置信区间以外.
- (4)拒绝原假设。

t检验举例

- 方法3：计算p值
- (1)t检验统计量：

$$t_{\hat{\beta}_2} = 5.424$$

- (2)在t分布表中来查找这一数字$t_{1-\frac{\alpha}{2}}(72) = 5.424$所在分位点，或者使用统计软件包如Stata计算p值：$\mathrm{p-value} = 0.000$

- (3)由于$\mathrm{p-value} = 0.000 < 0.05 =$显著性水平，那么拒绝原假设。

第Ⅰ类错误和第Ⅱ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

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

- 其中,α正是检验的显著性水平。

- **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

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

- 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β_k=0)，故容易计算第Ⅰ类错误的发生概率，即P(拒绝$H_0 | H_0$)=α。
- 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。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。
- 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(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”或“势”(power)，即

$$功效 = 1 - \mathrm{P}(接受H_0 \mid H_1) = \mathrm{P}(拒绝H_0 \mid H_1)$$

- 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)。
- 如果接受原假设，则比较没有把握，因为通常不知犯错概率。

对线性假设的F检验

- 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。
- 检验以下原假设：

$$H_0: \beta_2 = \cdots = \beta_K = 0$$

- 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)：

$$H_0: \beta_2 = 0,\ \beta_3 = 0,\ \cdots,\ \beta_K = 0$$

- 例 对于模型$y=\beta_1x_1+\beta_2x_2+\beta_3x_3+\beta_4x_4+\varepsilon$,检验以下两个约束:

$$H_0: \beta_2=\beta_3,\ \beta_4=0$$

- 写成向量形式:

$$H_0: \begin{pmatrix} \beta_2-\beta_3 \\ \beta_4 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

- 上式可写为

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

- 更一般地，考虑检验m个线性假设是否同时成立：

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

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

$$\boldsymbol{R} = \begin{pmatrix} 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}，\ 而\ \boldsymbol{r} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

- 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（$\boldsymbol{R}\hat{\boldsymbol{\beta}}-\boldsymbol{r}$）应比较接近 0(零向量)。
- 这种接近程度可用其二次型来衡量，比如

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

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

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

- 其中，σ^2可由s^2来估计。

- 为何使用协方差矩阵的逆矩阵作为二次型的矩阵?
- 回顾： **定理(t 统计量的分布)** 在假定 1-5 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$(n-K)$ t分布：

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

- 将此t统计量平方可得：

$$t_k{}^2 \equiv \frac{(\hat{\beta}_k - c)^2}{\widehat{Var(\hat{\beta}_k)}} = (\hat{\beta}_k - c)\left[\widehat{Var(\hat{\beta}_k)}\right]^{-1}(\hat{\beta}_k - c) \sim F(1, n-K)$$

- **定理(F 统计量的分布)**
- 在假定 1-5 均满足，且原假设“$H_0: \boldsymbol{R\beta} = \boldsymbol{r}$ ”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：

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

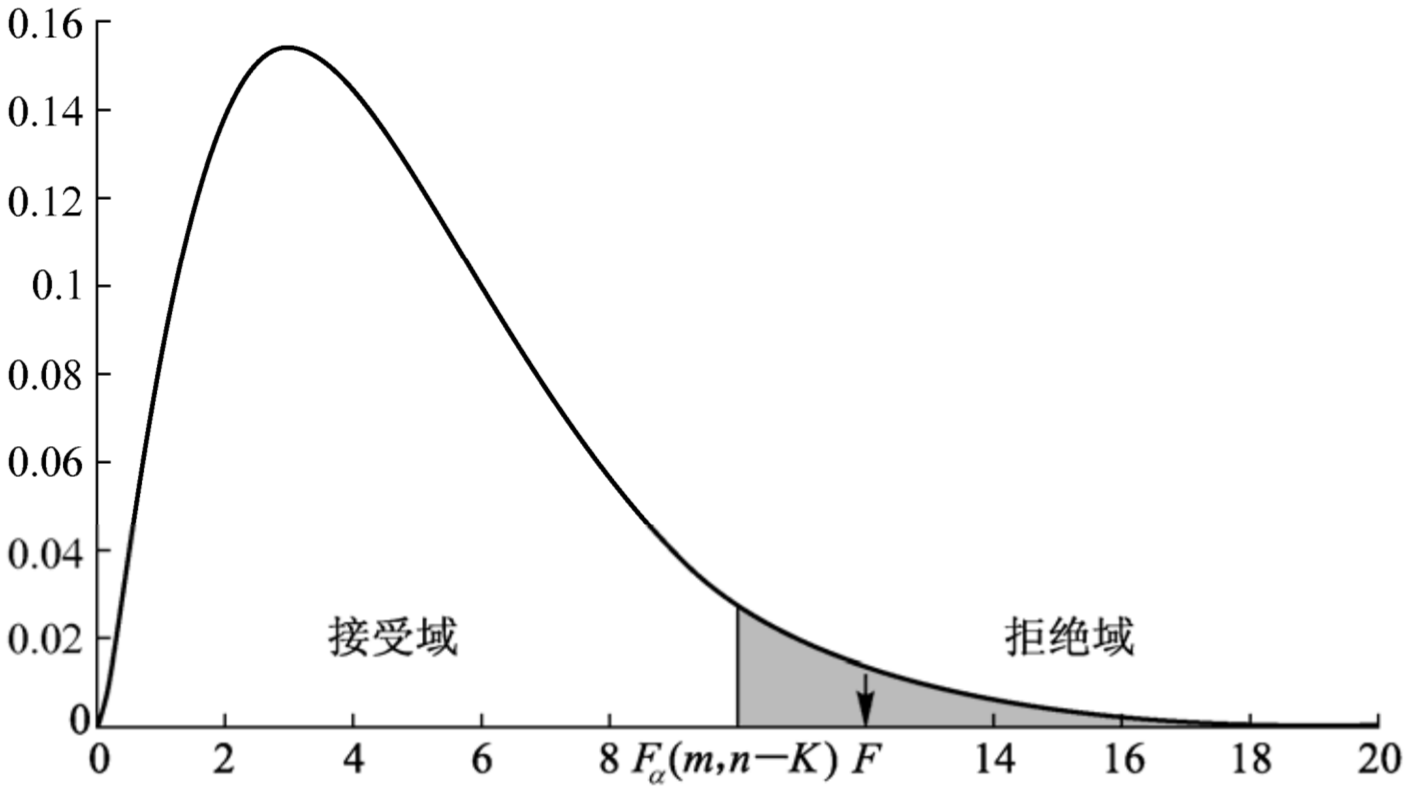
F检验的步骤

- 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

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

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

- 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。



- 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

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

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

F统计量的似然比原理表达式

- 在做假设检验时，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。
- 考虑约束极值问题：

$$\min_{\hat{\boldsymbol{\beta}}} \ \mathrm{SSR}(\hat{\boldsymbol{\beta}})$$

$$s.t. \ \ \boldsymbol{R}\hat{\boldsymbol{\beta}} = \boldsymbol{r}$$

- 其中，$SSR(\hat{\beta})$为残差平方和，是$\widehat{\boldsymbol{\beta}}$的函数；而$\widehat{\boldsymbol{\beta}}$还须满足约束条件$\boldsymbol{R}\widehat{\boldsymbol{\beta}} - \boldsymbol{r}$（s.t.表示受约束，即subject to）

- $\widehat{\boldsymbol{\beta}}$并不能任意取值，而只能在所有满足$\boldsymbol{R}\widehat{\boldsymbol{\beta}} = \boldsymbol{r}$的子集中，选择使残差平方和$SSR(\hat{\beta})$最小化的$\widehat{\boldsymbol{\beta}}$。
- 如果H_0: $\boldsymbol{R}\widehat{\boldsymbol{\beta}} = \boldsymbol{r}$正确，则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，而有约束的残差平方和SSR*。
- 在H_0正确的情况下，$(SSR^* - SSR)$不应很大。
- 构造如下F统计量。通过求解此约束极值问题，可以证明：

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

- 其中，m为约束条件个数（即矩阵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容易计算。
- 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，LR)。

- F统计量的似然比表达式，也可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

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

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

- 证明：方程右边的分子分母同时除以TSS $\equiv \sum_{i=1}^{n}(y_i-\bar{y})^2$

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

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

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

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"$y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots \beta_k x_{ik} + \varepsilon_i$"，检验原假设"$H_0$：$\beta_2 = \cdots = \beta_k = 0$"的$F$统计量等于

$$\frac{R^2/(K-1)}{(1-R^2)/(n-K)}$$

- 证明：对于H_0：$\beta_2 = \cdots = \beta_k = 0$，共有$(K-1)$个约束，故在表达式$F = \frac{(R^2-R_*^2)/m}{(1-R^2)/(n-K)}$中，$m = (K-1)$。
- 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$R_*^2 = 0$。
- 将$m = (K-1)$与$R_*^2 = 0$代入表达式$\frac{(R^2-R_*^2)/m}{(1-R^2)/(n-K)}$即得证。
- R^2并非决定F统计量的唯一因素；还取决于样本容量n与解释变量K的个数

预测

- 有时也用计量模型进行预测(prediction 或 forecasting)，即给定解释向量$\boldsymbol{x}_0$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），

$$y_0 = \boldsymbol{x}_0'\boldsymbol{\beta} + \varepsilon_0$$

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

$$\hat{y}_0 = \boldsymbol{x}_0'\widehat{\boldsymbol{\beta}}$$

- 例 使用一阶自回归（First Order Autoregression）进行预测：

$$y_t = \alpha + \beta y_{t-1} + \varepsilon_t$$

进行OLS估计可得

$$\hat{y}_t = \alpha + \beta \hat{y}_{t-1} + \varepsilon_t$$

- 假设y_{t-1}已知（比如，去年的通货膨胀率），则可预测y_t（今年的通货膨胀率)

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

$$\hat{y}_0 - y_0 = \boldsymbol{x}_0'\widehat{\boldsymbol{\beta}} - (\boldsymbol{x}_0'\boldsymbol{\beta} + \varepsilon_0) = \boldsymbol{x}_0'\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) - \varepsilon_0$$

- $\hat{y}_0$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

$$E(\hat{y}_0 - y_0) = \boldsymbol{x}_0' E\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) - E(\varepsilon_0) = 0$$

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

- 有时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

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

- 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，即抽样误差$\sigma^2 \boldsymbol{x}_0'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{x}_0$(不能精确知道参数$\boldsymbol{\beta}$），以及$y_0$本身的不确定性（$\varepsilon_0$的方差$\sigma^2$）。
- 如果样本很大，则抽样误差将很小；但扰动项方差σ^2始终存在。

- 将方程中的σ^2用s^2替代，可得预测误差$(\hat{y}_0 - y_0)$的标准误：

$$SE(\hat{y}_0 - y_0) = \sqrt{s^2 + s^2 \boldsymbol{x}_0'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{x}_0} = s\sqrt{1 + \boldsymbol{x}_0'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{x}_0}$$

由此可得t统计量：

$$\frac{\hat{y}_0 - y_0}{SE(\hat{y}_0 - y_0)} \sim t(n-K)$$

y_0的置信度为$(1-\alpha)$的置信区间为

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

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。

谢谢！