计量经济学

第 5 讲：线性回归模型的 OLS 估计

授课人：刘　岩

中山大学商学院

2025 年 3 月 12 日

本讲内容

本节内容

单变量线性回归模型：基本术语

- 单一解释变量线性回归 (linear regression)：

$$Y_i = \beta X_i + \varepsilon_i, \quad i = 1, \ldots, N$$

- 如 i 为个人，Y_i 为其收入，X_i 为其教育水平，ε_i 为残差项 (residual term)
  - X 称为解释变量 (explanatory variable)，亦称为自变量 (independent variable)
    - 在计量经济学中，很少用自变量这个说法，因为几乎所有的变量，都没有独立性——即外生性 (exogeneity)
  - Y 称为结果变量 (outcome variable)，被解释变量，或因变量 (dependent variable)
    - 特别在因果推断中，标准称谓是结果变量
- ε 又称为误差项 (error term)，代表所有无法观测 (unobservable) 或未被观测到 (unobserved)，但对 Y 有影响的因素
- β 称为回归系数，简称系数 (coefficient)

单变量线性回归：系数估计

- 线性回归方程同时给出了一个数据生成过程 (DGP)：给定 $\{X_i\}$ 的分布、$\{\varepsilon_i\}$ 的分布及二者关系，回归方程确定了 $\{Y_i\}$ 与 $\{X_i, \varepsilon_i\}$ 的联合分布
- β 假设为未知 (unknown)，经济学家关心 β 的取值，计量经济学家的任务是通过样本数据 $\{X_i, Y_i\}_{i=1}^N$，估计 β 取值
- 可以使用的估计方法有很多种，最常见的是 一般最小二乘 (ordinary least square, OLS) 估计
- OLS 法：选择回归系数 β，最小化残差平方和

$$\min_{\beta} \sum_{i=1}^{N} (Y_i - \beta X_i)^2$$

该最小化问题的解称为 OLS 估计量，记为 $\hat{\beta}$

单变量线性回归：OLS 估计

- OLS 估计量的直接推导：最小化问题 1 阶条件（思考：2 阶条件如何?）

$$\sum_{i=1}^{N} X_i(Y_i - \beta X_i) = 0 \Rightarrow \hat{\beta}_N = \frac{\sum_{i=1}^{N} X_i Y_i}{\sum_{i=1}^{N} X_i^2}$$

- 进一步假设 $\{X_i, \varepsilon_i\}$ iid，$\mathbb{E}X_i = \mathbb{E}\varepsilon_i = 0$，且 X_i 与 ε_i 相互独立 $\forall i$，则上式中分子分母同时除以 N 可得

$$\hat{\beta}_N = \frac{\frac{1}{N}\sum_{i=1}^{N} X_i Y_i}{\frac{1}{N}\sum_{i=1}^{N} X_i^2} \xrightarrow{\text{a.s.}} \frac{\text{cov}(X_i, Y_i)}{\text{var}(X_i)} = \beta, \quad N \to \infty$$

其中，分子、分母的收敛由大数定律 (LLN) 保证，故 X_t 自回归方程系数的 OLS 估计值具有一致性

单变量线性回归：OLS 估计的矩阵形式

- 单变量线性回归 $Y_i = \beta X_i + \varepsilon_i, i = 1, \ldots, N$ 的向量形式：

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

  其中 $\boldsymbol{Y} = [Y_1, \ldots, Y_N]^{\mathsf{T}}$ 为被解释变量，$\boldsymbol{X} = [X_1, \ldots, X_N]^{\mathsf{T}}$ 解释变量，$\boldsymbol{\varepsilon} = [\varepsilon_1, \ldots, \varepsilon_N]^{\mathsf{T}}$ 残差项，注意矩阵写法中 $\boldsymbol{X}\beta$ 的顺序

- 最小化残差平方和可表示为

$$\min_{\beta} \boldsymbol{\varepsilon}^{\mathsf{T}}\boldsymbol{\varepsilon} = \min_{\beta} (\boldsymbol{Y} - \boldsymbol{X}\beta)^{\mathsf{T}}(\boldsymbol{Y} - \boldsymbol{X}\beta)$$

- 单一解释变量不易体现矩阵形式的优越性；多变量则不同

单变量线性回归：残差项方差的估计

- iid 假设下，残差项 ε_i 的方差 $\sigma_\varepsilon^2 = \text{var}(\varepsilon_i)$，同样假设为未知，需要从样本数据进行估计
- 按照 LLN 思路，可以基于样本方差 $\hat{\sigma}_\varepsilon^2 = \frac{1}{N}\sum_{i=1}^N \varepsilon_i^2$ 来估计 σ_ε^2；但唯一的观测值为 $\{X_i, Y_i\}$，而非 $\{\varepsilon_i\}$
- 间接估计思路：给定 β 的 OLS 一致估计量 $\hat{\beta}_N$，则大样本下

$$e_i \equiv Y_i - \hat{\beta}_N X_i \xrightarrow{\text{a.s.}} Y_i - \beta X_i = \varepsilon_i$$

因此，可利用残差项 ε_i 的样本估计值 e_i 来估计 σ_ε^2（后续会严格证明）：

$$\hat{\sigma}_e^2 = \frac{1}{N}\sum_{i=1}^N e_i^2 = \frac{1}{N}\sum_{i=1}^N (Y_i - \hat{\beta}_N X_i)^2 \xrightarrow{\text{a.s.}} \sigma_\varepsilon^2$$

本节内容

线性回归模型的一般形式

- 给定因变量 Y_i 与 K 个解释变量 $X_{0,i},\ldots,X_{K-1,i}$ 及其 N 个样本 $\{Y_i, X_{0,i},\ldots,X_{K-1,i}\}_{i=1}^{N}$
  - 其中，$X_{0,i}$ 通常假设为常数 1，即 $X_{0,i}\equiv 1$
- 并假设 Y_i 与 $X_{0,i},\ldots,X_{K-1,i}$ 之间满足线性关系：

$$\begin{aligned} Y_i &= \beta_0 X_{0,i} + \beta_1 X_{1,i} + \cdots + \beta_{K-1} X_{K-1,i} + \varepsilon_i, \quad i = 1,\ldots,N \\ &= \beta_0 + \beta_1 X_{1,i} + \cdots + \beta_{K-1} X_{K-1,i} + \varepsilon_i, \quad \text{若} X_{0,i}\equiv 1 \end{aligned}$$

  其中 ε_i 为残差项
- 上述线性回归模型的 OLS 估计，为如下最小化残差平方和问题的解

$$\min_{\{\beta_k\}_{k=0}^{K-1}} \sum_{i=1}^{N} (Y_i - \beta_0 X_{0,i} - \beta_1 X_{1,i} - \cdots - \beta_{K-1} X_{K-1,i})^2$$

- 用微积分求解极值问题的 1 阶最优条件来求解，可行，但会比较繁琐

线性回归模型的矩阵形式

- 多解释变量线性回归的一般写法

$$Y_i = \boldsymbol{X}_i^\top \boldsymbol{\beta} + \varepsilon_i, \quad i = 1, \ldots, N$$

其中 $\boldsymbol{X}_i^\top = [X_{0,i}, X_{1,i}, \ldots, X_{K-1,i}]$ 为 K 个回归变量

- 其矩阵形式为

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

其中

$$\boldsymbol{Y} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_N \end{bmatrix}_{N\times 1}, \boldsymbol{X} = \begin{bmatrix} \boldsymbol{X}_1^\top \\ \vdots \\ \boldsymbol{X}_N^\top \end{bmatrix}_{N\times K}, \boldsymbol{e} = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_N \end{bmatrix}_{N\times 1}, \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_{K-1} \end{bmatrix}_{K\times 1}$$

系数 OLS 估计的矩阵形式

- OLS 估计量为最小化残差平方和问题的解

$$\min_{\beta} \varepsilon^{\mathsf{T}}\varepsilon = \min_{\beta}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})^{\mathsf{T}}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})$$

结论 若 $\hat{\boldsymbol{\beta}}$ 满足 $\boldsymbol{X}^{\mathsf{T}}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}) = \mathbf{0}$（零向量），则 $\hat{\boldsymbol{\beta}}$ 为上述最小化问题的解；进一步，若 $\boldsymbol{X}^{\mathsf{T}}\boldsymbol{X}$ 为可逆矩阵，则 $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^{\mathsf{T}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathsf{T}}\boldsymbol{Y}$

- 只需要说明，$\forall \boldsymbol{\beta} \in \mathbb{R}^K$，下列不等式成立即可

$$(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})^{\mathsf{T}}(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}) \le (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})^{\mathsf{T}}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})$$

- 为此，只需考虑下式的展开

$$(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})^{\mathsf{T}}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}) = (\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}} + \boldsymbol{X}\hat{\boldsymbol{\beta}} - \boldsymbol{X}\boldsymbol{\beta})^{\mathsf{T}}(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}} + \boldsymbol{X}\hat{\boldsymbol{\beta}} - \boldsymbol{X}\boldsymbol{\beta})$$

及 $\hat{\boldsymbol{\beta}}$ 所满足的性质即可

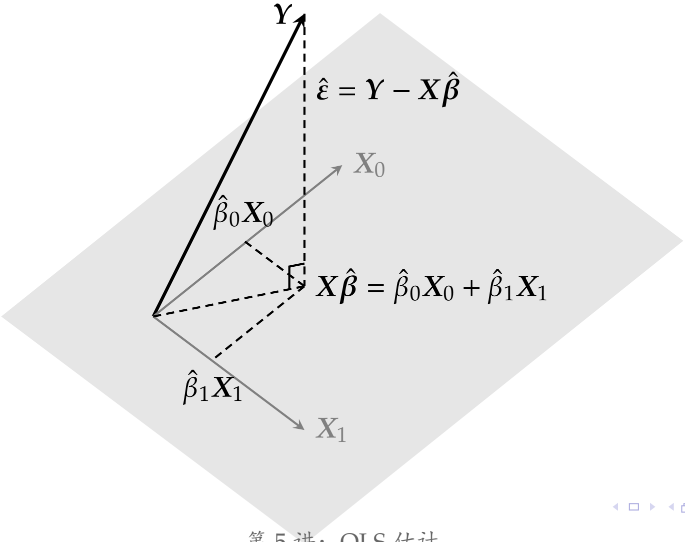
系数 OLS 估计矩阵表达式的证明

$$
\begin{aligned}
(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta})^{\top}(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta}) &= (\boldsymbol{Y}-\boldsymbol{X}\hat{\boldsymbol{\beta}}+\boldsymbol{X}\hat{\boldsymbol{\beta}}-\boldsymbol{X}\boldsymbol{\beta})^{\top}(\boldsymbol{Y}-\boldsymbol{X}\hat{\boldsymbol{\beta}}+\boldsymbol{X}\hat{\boldsymbol{\beta}}-\boldsymbol{X}\boldsymbol{\beta}) \\
&= (\boldsymbol{Y}-\boldsymbol{X}\hat{\boldsymbol{\beta}})^{\top}(\boldsymbol{Y}-\boldsymbol{X}\hat{\boldsymbol{\beta}}) + (\boldsymbol{Y}-\boldsymbol{X}\hat{\boldsymbol{\beta}})^{\top}(\boldsymbol{X}\hat{\boldsymbol{\beta}}-\boldsymbol{X}\boldsymbol{\beta}) \\
&\quad + (\boldsymbol{X}\hat{\boldsymbol{\beta}}-\boldsymbol{X}\boldsymbol{\beta})^{\top}(\boldsymbol{Y}-\boldsymbol{X}\hat{\boldsymbol{\beta}}) + (\boldsymbol{X}\hat{\boldsymbol{\beta}}-\boldsymbol{X}\boldsymbol{\beta})^{\top}(\boldsymbol{X}\hat{\boldsymbol{\beta}}-\boldsymbol{X}\boldsymbol{\beta}) \\
&= (\boldsymbol{Y}-\boldsymbol{X}\hat{\boldsymbol{\beta}})^{\top}(\boldsymbol{Y}-\boldsymbol{X}\hat{\boldsymbol{\beta}}) + 2\underbrace{(\boldsymbol{Y}-\boldsymbol{X}\hat{\boldsymbol{\beta}})^{\top}\boldsymbol{X}}_{\equiv 0}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}) \\
&\quad + \underbrace{(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})^{\top}\boldsymbol{X}^{\top}\boldsymbol{X}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})}_{\geq 0} \\
&\geq (\boldsymbol{Y}-\boldsymbol{X}\hat{\boldsymbol{\beta}})^{\top}(\boldsymbol{Y}-\boldsymbol{X}\hat{\boldsymbol{\beta}})
\end{aligned}
$$

系数 OLS 估计的几何视角

- 将 $\boldsymbol{X}$ 看做 $\mathbb{R}^N$ 中的 K 个列向量，则对任意 $\boldsymbol{\beta}$，$\boldsymbol{X\beta}$ 为这 K 个向量张成的子空间 $\mathcal{X}$ 中的一个向量
- $(\boldsymbol{Y}-\boldsymbol{X\beta})^\intercal(\boldsymbol{Y}-\boldsymbol{X\beta})$ 为 $\boldsymbol{Y}$ 到 $\boldsymbol{X\beta}$ 的距离
- 最小二乘估计就等价于寻找 $\boldsymbol{\beta}$ 使得 $\boldsymbol{Y}$ 到 $\mathcal{X}$ 距离最短
- 这又等价于使得残差向量 $\boldsymbol{\varepsilon}=\boldsymbol{Y}-\boldsymbol{X\beta}$ 与子空间 $\mathcal{X}$ 垂直，即 $\boldsymbol{X}^\intercal\boldsymbol{\varepsilon}=\boldsymbol{0}$
- OLS 估计可看做被解释变量对解释变量的投影 (projection)，$\boldsymbol{X}(\boldsymbol{X}^\intercal\boldsymbol{X})^{-1}\boldsymbol{X}^\intercal$ 称为投影矩阵

OLS 估计的几何视角



$\boldsymbol{Y}$
$\hat{\varepsilon} = \boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}$
X_0
$\hat{\beta}_0 \boldsymbol{X}_0$
$\boldsymbol{X}\hat{\boldsymbol{\beta}} = \hat{\beta}_0 \boldsymbol{X}_0 + \hat{\beta}_1 \boldsymbol{X}_1$
$\hat{\beta}_1 \boldsymbol{X}_1$
X_1

计算示例

- 考虑 $Y_i = \beta_0 + \beta_1 X_i, i = 1,2,3$，其中 $Y_i = 1,1,3$，$X_i = 1,2,3$，则相应地有

$$\boldsymbol{X} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix}, \qquad \boldsymbol{Y} = \begin{bmatrix} 1 \\ 1 \\ 3 \end{bmatrix}$$

- 由 $\boldsymbol{\beta} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{Y}$，可知

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \left(\begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \right)^{-1} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 3 \end{bmatrix} = \begin{bmatrix} -\frac{1}{3} \\ 1 \end{bmatrix}$$

“回归”概念的提出：人口学家、统计学家，Galton 1886

205 对父母身高与 930 个子女身高数据集：相对父母身高，子女身高向均值回归

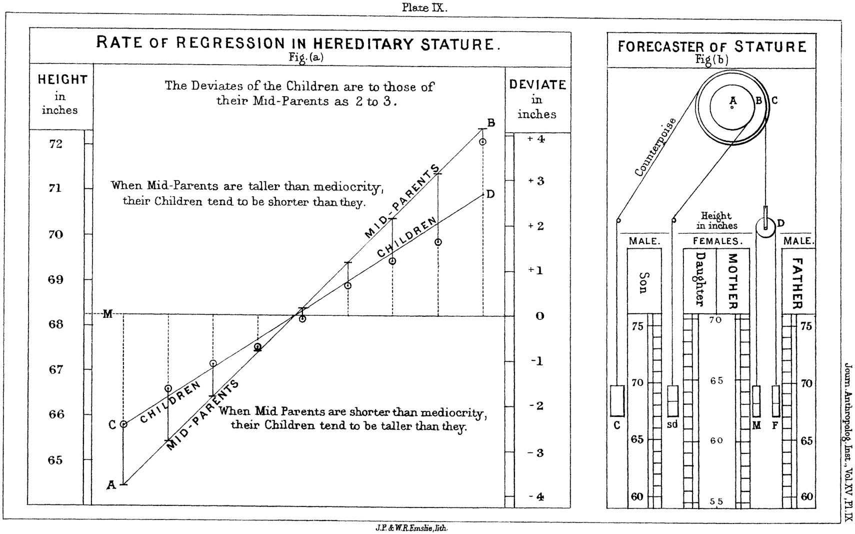
- Galton, F. 1886. Regression towards Mediocrity in Hereditary Stature. *Journal of the Anthropological Institute of Great Britain and Ireland*, vol.15: 246--63.

REGRESSION *towards* MEDIOCRITY *in* HEREDITARY STATURE.
By FRANCIS GALTON, F.R.S., &c.

[WITH PLATES IX AND X.]

THIS memoir contains the data upon which the remarks on the Law of Regression were founded, that I made in my Presidential Address to Section H, at Aberdeen. That address, which will appear in due course in the Journal of the British Association, has already been published in "Nature," September 24th. I reproduce here the portion of it which bears upon regression, together with some amplification where brevity had rendered it obscure, and I have added copies of the diagrams suspended at the meeting, without which the letterpress is necessarily difficult to follow. My object is to place beyond doubt the existence of a simple and far-reaching law that governs the hereditary transmission of, I believe, every one of those simple qualities which all possess, though in unequal degrees. I once before ventured to draw attention to this law on far more slender evidence than I now possess.

Galton 图示：向均值回归 regression to the mean



Galton 图示：二元正态分布拟合，密度函数等高线为椭圆