计量经济学

第 8 讲：回归模型 OLS 估计参数检验

授课人：刘　岩

中山大学商学院

2025 年 3 月 26 日

本讲内容

本节内容

小样本下单系数 t-检验的分布与标准误

- 原假设 $H_0 : \beta_k = \beta_{k,0}$，备择假设 $H_1 : \beta_k \neq \beta_{k,0}$，其中 $\beta_{k,0}$ 为回归模型 DGP 中系数 β_k 的假设真实值
- 原假设下，基于 OLS 估计值，可构造如下 t-统计量

$$t_k \equiv \frac{\hat{\beta}_k - \beta_{k,0}}{\sqrt{s^2[(\mathbf{X}^\intercal\mathbf{X})^{-1}]_{k+1,k+1}}} \sim t(N-K)\text{分布}$$

  其中，分母亦称为 OLS 估计系数 $\hat{\beta}_k$ 的标准误 (standard error)，记作 $s.e.(\hat{\beta}_k)$
- 注意，原假设 H_0 下 $\hat{\beta}_k$ 的分布为固定的 $t(N-K)$ 分布，只取决于样本量 N 和解释变量的个数 K，与作为模型未知量的 β 和 σ^2 无关

单系数 t-检验的计算步骤

给定实际样本取值 $\{Y_i, \boldsymbol{X}_i\}_{i=1}^N$，单系数估计值 $\hat{\beta}_k$ 的 t-检验相关计算步骤如下

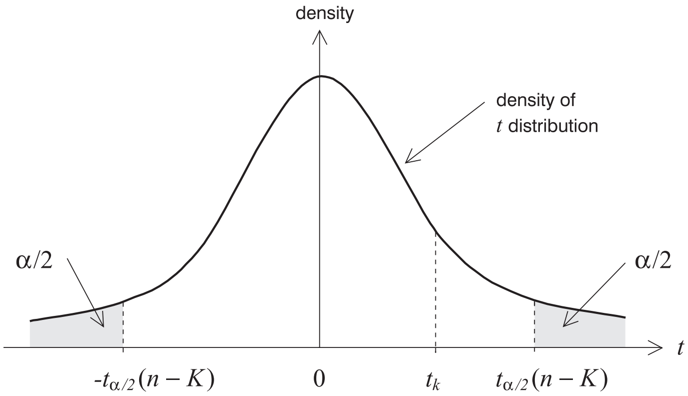
1. 根据样本，计算 $\hat{\beta}_k$ 的具体取值 b_k，矩阵 $\boldsymbol{Q}_N = \boldsymbol{X}^\intercal \boldsymbol{X}$，以及 $s^2 = \frac{\boldsymbol{e}^\intercal \boldsymbol{e}}{N-K}$ 的具体取值 $\bar{s}^2 = \frac{\bar{\boldsymbol{e}}^\intercal \bar{\boldsymbol{e}}}{N-K}$
   - 其中 $\bar{\boldsymbol{e}} = \boldsymbol{Y} - \boldsymbol{X}\boldsymbol{b}$，$\boldsymbol{b} = [b_0, \ldots, b_{K-1}]^\intercal$
2. 计算统计量 t_k 的具体取值

$$\bar{t}_k = \frac{b_k - \beta_{k,0}}{\sqrt{\bar{s}^2[(\boldsymbol{Q}_N)^{-1}]_{k+1,k+1}}} = \frac{b_k - \beta_{k,0}}{s.e.(\hat{\beta}_k)}$$

3. 计算 $t(N-K)$ 分布在 $\alpha = 1\%, 5\%$ 等显著性水平下的临界值 $t_{\alpha/2}(N-K)$，即右尾概率 $\mathbb{P}(t_k \geq t_{\alpha/2}(N-K)) = \frac{\alpha}{2}$，并按照如下规则拒绝或接受原假设
   - 若 $|\bar{t}_k| > t_{\alpha/2}(N-K)$，则在 α 显著性水平下，拒绝 H_0
   - 若 $|\bar{t}_k| < t_{\alpha/2}(N-K)$，则在 α 显著性水平下，接受 H_0

t-检验的拒绝域与接受域

给定显著性水平 α，自由度 $N-K$ 的 t-检验统计量 t 的拒绝域（拒绝 H_0）为 $|t| > t_{\alpha/2}(N-K)$，接受域为 $|t| \le t_{\alpha/2}(N-K)$，如下图所示



置信区间与 p-值

统计推断中，与检验显著性水平（拒绝域、接受域）密切相关的两个概念：置信区间与 p-值

- 置信区间：当 $\bar{t}_k$ 位于接受域中时，有如下不等式

$$-t_{\alpha/2}(N-K) \leq \frac{b_k - \beta_{k,0}}{s.e.(\hat{\beta}_k)} \leq t_{\alpha/2}(N-K) \Longleftrightarrow$$

$$\beta_{k,0} \in [b_k - s.e.(\hat{\beta}_k) \times t_{\alpha/2}(N-K), b_k + s.e.(\hat{\beta}_k) \times t_{\alpha/2}(N-K)]$$

最右的区间，称为估计值 $\hat{\beta}_k$ 置信度为 $1-\alpha$ 的置信区间 (confidence interval)

- 推断规则：当 H_0 下参数值 $\beta_{k,0} \in$ 置信区间时，接受 H_0；若否，则拒绝 H_0

置信区间与 p-值

- p-值：直接在原假设 t-统计量 t_k 所服从的分布 $t(N-K)$ 下，计算下述概率

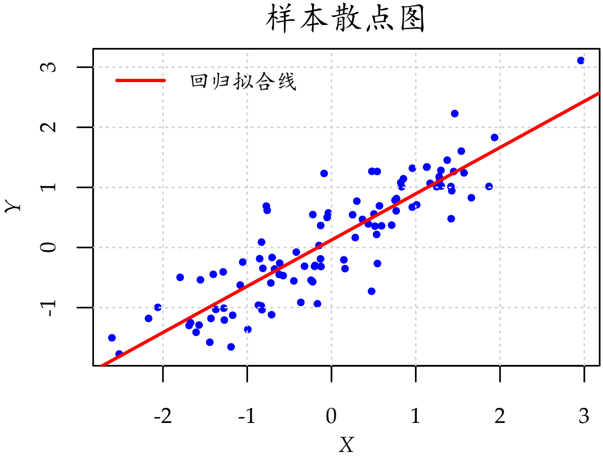
$$\mathbb{P}(-|\bar{t}_k| \le t_k \le |\bar{t}_k|) = 1 - 2\mathbb{P}(t_k > |\bar{t}_k|)$$

  并定义 $p = 1 - \mathbb{P}(-|\bar{t}_k| \le t_k \le |\bar{t}_k|) = 2\mathbb{P}(t_k > |\bar{t}_k|)$ 为 t-检验的 p-值
  - 推断规则：当 p-值大于检验所要求的显著性水平 α 时，接受 H_0；否则拒绝 H_0

注意 p-值不是越小越好，或者说，更小的 p-值，并不意味着原假设 H_0 正确的可能性越低；换言之，更小的 p-值，并不意味着假设检验第一类错误（冤枉好人）的可能性更低

第 6 讲中的数值模拟示例⋯

DGP: $\varepsilon_i \overset{\text{iid}}{\sim} \mathcal{N}(0, 0.5^2), X_i \overset{\text{iid}}{\sim} \mathcal{N}(0, 1), Y_i = 0.2 + 0.8X_i + \varepsilon_i, N = 100, K = 2$

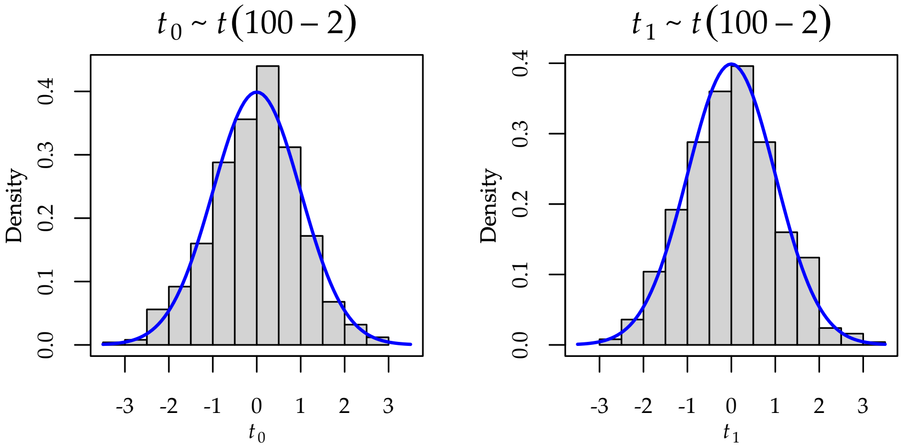


…估计结果

	估计值 b	标准误	t-值	显著性水平	p-值
$\hat{\beta}_0$	0.12484	0.04668	2.674	$***, \alpha = 1\%$	0.00877
$\hat{\beta}_1$	0.76955	0.04174	18.438	$***, \alpha = 1\%$	$< 2 \times 10^{-16}$

- 统计软件汇报估计结构时，系数原假设取值默认为 0，即 $\beta_{0,0} = \beta_{1,0} = 0$
- 当 $N - K > 30$ 时，$t(N - K)$ 分布函数在数值上和，几乎无差异，此时可将 t_k 视作标准正态分布处理，相应的标准误 $s.e.(\hat{\beta}_k)$ 等于 $\hat{\beta}_k$ 的标准差

t_k 分布：500 次随机模拟



本节内容

多系数联合检验

- 一般情况，多系数线性联合假设检验 (joint hypothesis test)，可引入线性约束矩阵 (linear restriction matrix) $\boldsymbol{R}_{L\times K}$ 及约束目标向量 $\boldsymbol{r}_{L\times 1}$：

$$H_0 : \boldsymbol{R\beta} = \boldsymbol{r} \quad \Leftrightarrow \quad H_0 : \boldsymbol{R\beta} - \boldsymbol{r} = \boldsymbol{0}_{L\times 1}$$

  此时，估计向量仿射变换 $\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{r}$ 在 $\boldsymbol{X}$ 上的条件分布等于

$$\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{r}|\boldsymbol{X} \sim \mathcal{N}\left(\boldsymbol{0}, \sigma^2 \boldsymbol{R}\boldsymbol{Q}_N^{-1}\boldsymbol{R}^{\mathsf{T}}\right)$$

  注意，$L \leq K$，最多同时检验所有 K 个回归系数的取值

- 示例：$H_0 : \beta_0 = \beta_1 = \cdots = \beta_{K-1} = 0$，则 $\boldsymbol{R} = \boldsymbol{I}_K, \boldsymbol{r} = \boldsymbol{0}_{K\times 1}$

联合检验统计量：方差已知

- 若 σ^2 已知，则由 $\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{r}|\boldsymbol{X} \sim \mathcal{N}\left(\mathbf{0}, \sigma^2 \boldsymbol{R}\boldsymbol{Q}_N^{-1}\boldsymbol{R}^\intercal\right)$，可知下述统计量

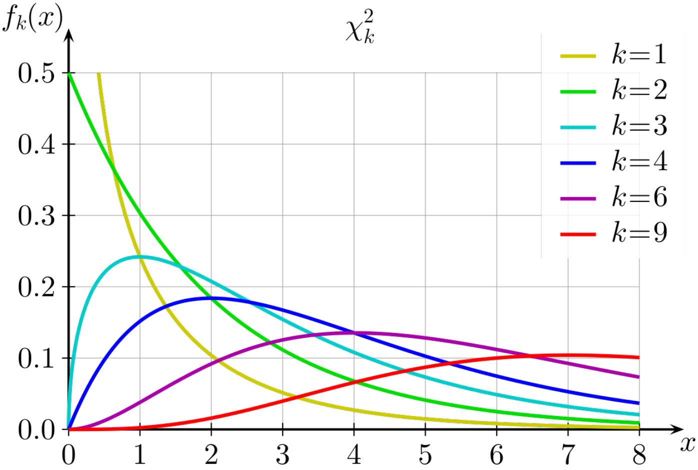
$$W_L \equiv (\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{r})^\intercal \left[\sigma^2 \boldsymbol{R}\boldsymbol{Q}_N^{-1}\boldsymbol{R}^\intercal\right]^{-1} (\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{r}) \sim \chi^2(L)$$

  服从自由度为 L 的 χ^2 分布，称为 Wald 统计量

- 推断规则：给定显著性水平 α，当样本计算得到的 W_L 统计量取值 $\overline{W}_L$ 大于 $\chi^2(L)$ 分布右尾临界值 $\chi^2_\alpha(L)$ 时，拒绝原假设
  - 类似于 t-检验，可计算 χ^2-检验的 p-值
  - 直观原理：$\overline{W}_L$ 越大，$\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{r}$ 的模长越长，其偏离 H_0 下零向量的幅度就越大

- 上述检验一般性称为 Wald 检验

χ^2-分布的密度函数



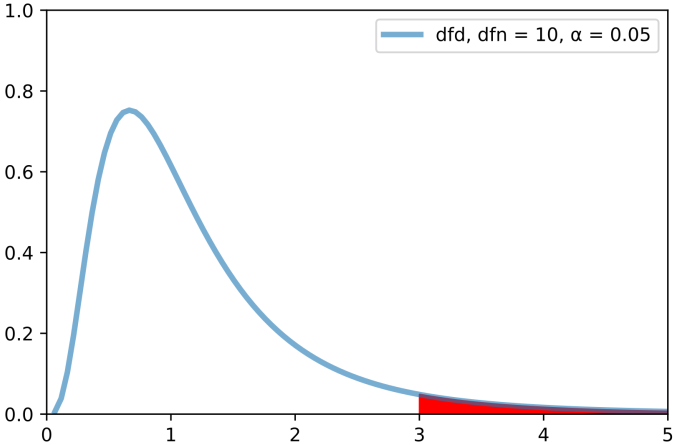
联合检验统计量：方差未知

- 在回归分析中，误差项方差 σ^2 未知，此时使用 F-检验，相应的检验统计量为

$$F \equiv \frac{(\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{r})^{\top}\left[\boldsymbol{R}\boldsymbol{Q}_N^{-1}\boldsymbol{R}^{\top}\right]^{-1}(\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{r})/L}{s^2} \sim F(L, N-K)$$

  服从自由度为 $(L, N-K)$ 的 F 分布 $F(L, N-K)$
  - $F(L, N-K)$ 分布定义为相互独立的两个随机变量 U, V 相除 $\frac{U/L}{V/(N-K)}$ 所得分布，其中 $U \sim \chi^2(L), V \sim \chi^2(N-K)$
  - 当自由度 $N-K \to \infty$ 时，$F(L, N-K)$ 分布收敛到 $L \times \chi^2(L)$ 分布
- 推断规则：与前述 Wald 检验类似，给定显著性水平 α，当样本计算所得 F 统计量取值 $\bar{F}$ 大于 $F(L, N-K)$ 分布右尾临界值 $F_\alpha(L, N-K)$ 时，拒绝原假设
  - 回归方程 OLS 估计后，统计软件自动汇报所有截距项外解释变量系数同时为 0 的 F-检验，如第 6 讲数值示例中，相应的 $\bar{F} = 340$，p-值为 2.2×10^{-16}

F-分布与拒绝域图示：$L = 10, N - K = 10$



本节内容

线性回归大样本理论基本假设

大样本理论关注 $N \to \infty$ 时，回归估计量的渐近分布 (asymptotic distribution)，相应的基本假设与小样本理论略有差异，此处仅介绍最简单的同方差 iid 情形假设

1. **iid 假设**：模型数据生成过程 $\{\boldsymbol{X}_i, \varepsilon_i\}$ 为 iid，即 $\forall i \neq j$，$[\boldsymbol{X}_i^{\top}, \varepsilon_i]$ 与 $[\boldsymbol{X}_j^{\top}, \varepsilon_j]$ 相互独立但同分布，这也意味着 $\{\boldsymbol{X}_i, \varepsilon_i, Y_i\}$ 为 iid 序列
2. **无多重共线性**：解释变量 2 阶交叉矩矩阵 $\boldsymbol{Q}_{XX} = \mathbb{E}[\boldsymbol{X}_i \boldsymbol{X}_i^{\top}]$ 满秩，即可逆
3. **当期外生性假设**：$\forall i$，$\mathbb{E}[\varepsilon_i | \boldsymbol{X}_i] = \mathbf{0}$，即 $\mathbb{E}[\boldsymbol{X}_i \varepsilon_i] = \mathbf{0}$，同期解释变量与残差无相关性
4. **同方差假设**：$\forall i$，$\mathbb{E}[\varepsilon_i^2 | \boldsymbol{X}_i] = \sigma^2$ 为常数

假设 1 保证可以使用 LLN 和 CLT，假设 3 保证估计系数一致性，假设 4 保证估计系数渐近正态分布

系数 OLS 估计的一致性

- 假设模型系数真实值为 $\boldsymbol{\beta}_0$，则有

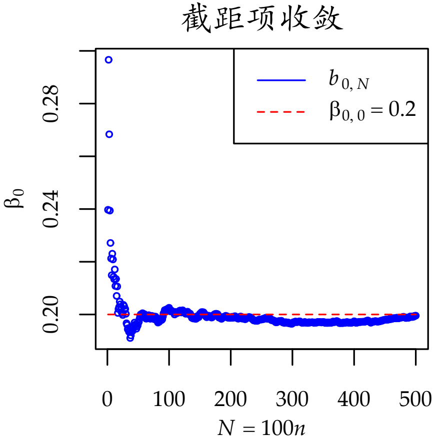
$$
\begin{aligned}
\hat{\boldsymbol{\beta}}_N &= \boldsymbol{\beta}_0 + (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{\varepsilon} = \boldsymbol{\beta}_0 + \left(\frac{1}{N} \boldsymbol{X}^\top \boldsymbol{X} \right)^{-1} \frac{1}{N} \boldsymbol{X}^\top \boldsymbol{\varepsilon} \\
&= \boldsymbol{\beta}_0 + \left(\frac{1}{N} \sum_{i=1}^{N} \boldsymbol{X}_i \boldsymbol{X}_i^\top \right)^{-1} \frac{1}{N} \sum_{i=1}^{N} \boldsymbol{X}_i \varepsilon_i
\end{aligned}
$$

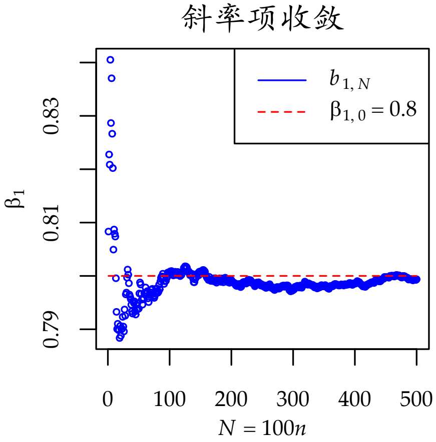
- 由 LLN 可知，$\frac{1}{N} \sum_{i=1}^{N} \boldsymbol{X}_i \boldsymbol{X}_i^\top \xrightarrow{\text{a.s.}} \boldsymbol{Q}_{XX}$，$\frac{1}{N} \sum_{i=1}^{N} \boldsymbol{X}_i \varepsilon_i \xrightarrow{\text{a.s.}} \mathbb{E}[\boldsymbol{X}_i \varepsilon_i] = \mathbf{0}$，故

$$
\hat{\boldsymbol{\beta}}_N = \boldsymbol{\beta}_0 + \left(\frac{1}{N} \sum_{i=1}^{N} \boldsymbol{X}_i \boldsymbol{X}_i^\top \right)^{-1} \frac{1}{N} \sum_{i=1}^{N} \boldsymbol{X}_i \varepsilon_i \xrightarrow{\text{a.s.}} \boldsymbol{\beta}_0 + \boldsymbol{Q}_{XX}^{-1} \mathbf{0} = \boldsymbol{\beta}_0
$$

大样本 OLS 估计收敛示例：500 组回归，每组增加 100 样本

DGP: $\varepsilon_i \overset{\text{iid}}{\sim} \mathcal{N}(0, 0.5^2), X_i \overset{\text{iid}}{\sim} \mathcal{N}(0,1), Y_i = 0.2 + 0.8X_i + \varepsilon_i, N = 100n, n = 1, \ldots, 500$





系数 OLS 估计的渐近分布

- 为分析 $\hat{\boldsymbol{\beta}}_N$ 的渐近分布，考虑

$$\sqrt{N}(\hat{\boldsymbol{\beta}}_N - \boldsymbol{\beta}_0) = \underbrace{\left(\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{X}_i\boldsymbol{X}_i^{\intercal}\right)^{-1}}_{\xrightarrow{\text{a.s.}}\boldsymbol{Q}_{XX}^{-1}\text{ by LLN}} \times \underbrace{\sqrt{N}\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{X}_i\varepsilon_i}_{\xrightarrow{\text{d}}\mathcal{N}(\boldsymbol{0},\sigma^2\boldsymbol{Q}_{XX})\text{ by CLT}}$$

- 注意 $\mathbb{E}[\boldsymbol{X}_i\varepsilon_i \cdot \varepsilon_i\boldsymbol{X}_i^{\intercal}] = \mathbb{E}\mathbb{E}[\varepsilon_i^2\boldsymbol{X}_i\boldsymbol{X}_i^{\intercal}|\boldsymbol{X}_i] = \sigma^2\mathbb{E}[\boldsymbol{X}_i\boldsymbol{X}_i^{\intercal}] = \sigma^2\boldsymbol{Q}_{XX}$
- 在 $\boldsymbol{X}_i\varepsilon_i$ 为 iid 序列假设下，由 CLT 可得

$$\sqrt{N}\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{X}_i\varepsilon_i \xrightarrow{\text{d}} \mathcal{N}(\boldsymbol{0},\sigma^2\boldsymbol{Q}_{XX})$$

- 再由 $N^{-1}\sum_i \boldsymbol{X}_i\boldsymbol{X}_i^{\intercal} \xrightarrow{\text{a.s.}} \boldsymbol{Q}_{XX}$ 可得：$\sqrt{N}(\hat{\boldsymbol{\beta}}_N - \boldsymbol{\beta}_0) \xrightarrow{\text{d}} \mathcal{N}(\boldsymbol{0},\sigma^2\boldsymbol{Q}_{XX}^{-1})$