

《线性回归》 —线性回归(9)

杨 瑛

清华大学 数学科学系

Email: yangying@mail.tsinghua.edu.cn

Tel: 62796887

2019.04.16

主要内容：线性回归模型中的置信区间和置信域

1 线性回归模型中的置信区间和置信域

- 联合推断
- 置信域
- 假设检验与CI
- 回归曲面的置信带
- 响应的预测区间和预测带
- 进一步的问题

引论

- ♠ 这一章主要研究检验线性模型参数的线性组合的双边置信区间，参数的联合置信域，预测区间等问题。
- ♠ 解决问题CI的方法是《统计推断》中一般CI理论和方法的具体应用。
- ♠ CI的构造方法有：
枢轴法、假设检验的拒绝域翻转，Bonferroni法。
- ♠ 考虑线性模型

$$\mathbf{Y} = \mathbf{X}\theta + \epsilon, \epsilon \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_n). \quad (1)$$

其中 \mathbf{X} 是秩为 p 的 $n \times p$ 设计矩阵， θ 是 $p \times 1$ 的未知参数， $\sigma^2 > 0$ 未知。

联合推断

♠ 感兴趣的问题:

- ✓ 确定 k 个线性组合: $\mathbf{a}_j^T \theta$ ($j = 1, \dots, k$)的双边置信区间,
- ✓ \mathbf{a}_j 的典型形式: $\mathbf{a}_1^T = (1, 0, \dots, 0)$, $\mathbf{a}_2^T = (0, 1, \dots, 0)$,
 $\mathbf{a}_p^T = (0, \dots, 0, 1)$ 等。

♠ 如何解决这个问题?

- ✓ 解决方案: 利用理论结果
 (Seber and Lee (2003), THEOREM 3.5):
 - $\hat{\theta} \sim N(\theta, \sigma^2(\mathbf{X}^T \mathbf{X})^{-1})$.
 - $(\hat{\theta} - \theta)\mathbf{X}^T \mathbf{X}(\hat{\theta} - \theta)/\sigma^2 \sim \chi_p^2$.
 - $\hat{\theta}$ 与 S^2 独立
 - $\text{RSS}/\sigma^2 \sim (n-p)S^2/\sigma^2 \sim \chi_{n-p}^2$
 - $\mathbf{a}^T \hat{\theta} \sim N(\mathbf{a}^T \theta, \sigma^2 \mathbf{a}^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{a})$, \mathbf{a} 是已知的列向量。

- ♠ 【课堂讨论】 如何利用这些理论结果, 得到
- ♠ ✓ $\mathbf{a}_j^T \theta$ ($j = 1, 2, \dots$) 的CI.
 $\mathbf{a}_j^T \theta, j = 1, 2, \dots$, 的联合置信域(Bonferroni t -区间)
- ♠ 【进阶内容】
 - ✓ Maximum Modulus t - intervals [LRA p. 122-123]
 - ✓ Scheffé S-Methods [LRA p. 123-124]
- ♠ 你的想法?
- ♠ 如何比较这些方法的优劣? 【理论的? 数值的?】什么标准? 【讨论】

置信域

♠ θ 的置信域:

$$1 - \alpha = \text{pr} \left[(\theta - \hat{\theta})' \mathbf{X}' \mathbf{X} (\theta - \hat{\theta}) \leq p S^2 F_{p, n-p}^{\alpha} \right] \quad (2)$$

这个结果是如何得到? 【讨论】

假设检验与CI

- ♠ 思考假设检验与CI之间的关系
- ♠ 举例说明【黑板】

回归曲面的置信带

♠ 感兴趣的问题：回归曲面

$$f(x_1, x_2, \dots, x_{p-1}) = \theta_0 + \theta_1 x_1 + \dots + \theta_{p-1} x_{p-1} = \mathbf{x}'\boldsymbol{\theta} \quad (3)$$

的置信水平为 $(1 - \alpha)\%$ 的置信域是什么？

♠ 或者给定 $\mathbf{x}_0 = (1, x_{01}, x_{02}, \dots, x_{0,p-1})'$ 之后，对应的 \mathbf{Y}_0 的置信区间。

♠ 解决方案：标准发放可以解决之！【略！】

响应的预测区间和预测带

- ♠ 以前已讲过简单线性模型的预测问题
- ♠ 如何推广到一般的多重线性模型(1)? 【讨论，黑板】

进一步的问题：

- ♠ 如何得到线性模型(1)中 θ 和 σ^2 的联合置信域？
- ♠ 【讨论，黑板】

作业：

- (1) 阅读Seber and Lee (2003)的Chapter 5.
- (2) 完成p.136-137中的Ex 1-7.
- (3) 利用Seber and Lee (2003), THEOREM 3.5 【或者ppt第四页】中的结果构造线性模型参数 θ 和 σ^2 的置信水平恰为 $1 - \alpha$ 的两种不同的联合置信区域，以及Bonferroni方法构造的 θ 和 σ^2 的置信水平至少或者恰好为 $1 - \alpha$ 的置信域。并从理论和数值方面比较这几种置信域优劣。进一步的问题：是否存在某种意义下 θ 和 σ^2 的置信水平恰为 $1 - \alpha$ 的最优置信域。最优的标准需要自己定义。