6. Inference for regression parameters and models

对回归参数 $\beta \in \mathbb{R}^p$ 的检验：

$$H_0: A\beta = u \quad \text{v.s.} \quad H_1: A\beta \neq u$$

其中常数矩阵 $A \in \mathbb{R}^{m\times p}$ $rank(A) = m$，常数向量 $u \in \mathbb{R}^m$

Tips：通过设置 A 的每一行只有一个元素为 1，其他为 0，再令 $u=0$ 可以检验 β 的某些分量是否为 0 （A 第一列为 0，因为不检验截距项）

e.g. $p=3$ 时，令 $A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ $u=0 \Rightarrow \beta_2 = \beta_3 = 0$

Lemma 1 矩阵 $D = A(X^TX)^{-1}A^T$ 正定 $A \in \mathbb{R}^{m\times p}$ $rank(A) = m$

Proof $v^TDv = (A^Tv)(X^TX)^{-1}(A^Tv) \geq 0$ $(\because (X^TX)^{-1}$ 正定$)$ $\Rightarrow$ 只须证：$A^Tv \neq 0$，$v\neq 0$

$V = \{v: A^Tv = 0\}$ $rank(A^T) + dim(V) = m \Rightarrow dim(V) = 0 \Rightarrow v\neq 0, A^Tv \neq 0$

6.1 CLSE 约束的最小二乘估计（Constrained least squares estimation）

假设检验：H_0 成立，寻找反常事件，故要引入 H_0 的条件

Definition 1 我们称满足：

$$\hat{\beta}_c = \arg\min_b \|y - Xb\|^2$$

$$\text{s.t.} \begin{cases} Ab = u \\ b \in \mathbb{R}^p \end{cases}$$

的 $\hat{\beta}_c$ 为"满足 $A\beta = u$ 的约束最小二乘估计"，其解析解为

$$\hat{\beta}_c = \hat{\beta} - (X^TX)^{-1}A^TD^{-1}(A\hat{\beta} - u)$$

其中 $D = A(X^TX)^{-1}A^T$ 正定，$\hat{\beta}$ 是 OLSE 为 $\hat{\beta} = (X^TX)^{-1}X^Ty$

Proof: (Langrange) 构造: $F(b,\lambda) = \|y-Xb\|^2 + 2\lambda^T(Ab-u)$

其中 $\lambda \in \mathbb{R}^m$, $A \in \mathbb{R}^{m\times p}$, $b \in \mathbb{R}^p$, $u \in \mathbb{R}^m$, $X \in \mathbb{R}^{m\times p}$, $y \in \mathbb{R}^m$

求导 $\left.\frac{\partial F}{\partial b}\right|_{b=\hat{\beta}_c, \lambda=\hat{\lambda}_c} = -2X^T(y-X\hat{\beta}_c) + 2A^T\hat{\lambda}_c = 0$

求导 $\left.\frac{\partial F}{\partial \lambda}\right|_{b=\hat{\beta}_c, \lambda=\hat{\lambda}_c} = A\hat{\beta}_c - u = 0$

故: $\hat{\beta}_c = (X^TX)^{-1}X^Ty - (X^TX)^{-1}A^T\hat{\lambda}_c = \hat{\beta} - (X^TX)^{-1}A^T\hat{\lambda}_c$

$\therefore \quad u = A\hat{\beta}_c = A\hat{\beta} - A(X^TX)^{-1}A^T\hat{\lambda}_c = A\hat{\beta} - D\hat{\lambda}_c$

$\therefore \quad \hat{\lambda}_c = D^{-1}(A\hat{\beta} - u)$ 代入 $\hat{\beta}_c$ 有

$$\hat{\beta}_c = \hat{\beta} - (X^TX)^{-1}A^TD^{-1}(A\hat{\beta} - u)$$

下面证明唯一性:

Theorem 1 $\hat{\beta}_c = \hat{\beta} - (X^TX)^{-1}A^TD^{-1}(A\hat{\beta} - u)$ 满足:

(a). $A\hat{\beta}_c = u$

(b). 对 $\forall b \in \mathbb{R}^p$, $b \neq \hat{\beta}_c$, 满足 $Ab = u$, 均有

$$\|y-Xb\|^2 > \|y-X\hat{\beta}_c\|^2$$

Proof: 对于 (a). $A\hat{\beta}_c = A\hat{\beta} - A(X^TX)^{-1}A^TD^{-1}(A\hat{\beta}-u)$

$= A\hat{\beta} - DD^{-1}(A\hat{\beta}-u)$

$= A\hat{\beta} - A\hat{\beta} + u$

$= u \quad \square$

对于 (b).

$\|y-Xb\|^2 = \|y - X\hat{\beta} + X\hat{\beta} - Xb\|^2$

$$= \|y-X\hat{\beta}\|^2 + \|X\hat{\beta}-Xb\|^2 + 2(y-X\hat{\beta})^T(X\hat{\beta}-Xb)$$

注意到： $(y-X\hat{\beta})^T X = y^T X - \hat{\beta}^T X^T X = y^T X - y^T X(X^T X)^{-1} X^T X = 0$

$$\therefore \text{原式} = \|y-X\hat{\beta}\|^2 + \|X\hat{\beta}-Xb\|^2$$

$$= \|y-X\hat{\beta}\|^2 + \|X(\hat{\beta}-\hat{\beta}_c+\hat{\beta}_c-b)\|^2$$

$$= \|y-X\hat{\beta}\|^2 + \|X(\hat{\beta}-\hat{\beta}_c)\|^2 + \|X(\hat{\beta}_c-b)\|^2 + 2(\hat{\beta}-\hat{\beta}_c)^T X^T X(\hat{\beta}_c-b)$$

注意到： $(\hat{\beta}-\hat{\beta}_c)^T X^T X(\hat{\beta}_c-b) = [(X^T X)^{-1} A^T \hat{\lambda}_c]^T X^T X(\hat{\beta}_c-b)$

$$= \hat{\lambda}_c^T A (X^T X)^{-1} X^T X(\hat{\beta}_c-b) = \hat{\lambda}_c^T A(\hat{\beta}_c-b) = \hat{\lambda}_c^T(u-u) = 0$$

$$\therefore \text{原式} = \|y-X\hat{\beta}\|^2 + \|X(\hat{\beta}-\hat{\beta}_c)\|^2 + \|X(\hat{\beta}_c-b)\|^2$$

即： $\|y-Xb\|^2 = \|y-X\hat{\beta}\|^2 + \|X(\hat{\beta}-\hat{\beta}_c)\|^2 + \|X(\hat{\beta}_c-b)\|^2 \quad \forall b\in\mathbb{R}^p \ \ Ab=u$

取 $b=\hat{\beta}_c$ 有：

$$\|y-X\hat{\beta}_c\|^2 = \|y-X\hat{\beta}\|^2 + \|X(\hat{\beta}-\hat{\beta}_c)\|^2$$

于是有： $\|y-Xb\|^2 - \|y-X\hat{\beta}_c\|^2 = \|X(\hat{\beta}_c-b)\|^2 \geq 0$

取等时： $X(\hat{\beta}_c-b)=0 \quad \because \operatorname{rank}(X)=p \quad \therefore \dim(\operatorname{Col}(X)) = p-p = 0$

其中 $\operatorname{Col}(X) = \{v : Xv=0\} \Rightarrow \hat{\beta}_c - b = 0$ 矛盾！

故当 $b\neq\hat{\beta}_c$ 时，有 $\|y-Xb\|^2 - \|y-X\hat{\beta}_c\|^2 > 0 \quad \square$

6.2 假设检验

$$\hat{\beta}_c = \hat{\beta} - (X^TX)^{-1}A^T D^{-1}(A\hat{\beta} - u)$$

(1) SST，SSR 和 SSE

Propersity

$$\sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 + \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

$$SST = SSR + SSE$$

Rm: 证明见第7周作业，核心：

$$\left.\frac{\partial Q}{\partial b}\right|_{b=\hat{\beta}} = \sum_{i=1}^{n} x_i e_i = X^T(y - X\hat{\beta}) = 0 \in \mathbb{R}^p$$

利用矩阵形式：

$$SSE = \|y - \hat{y}\|^2 = \|y - X\hat{\beta}\|^2$$

(2) 约束 $A\beta = u$ 下的 SSE_c

$$SSE_c = \|y - X\hat{\beta}_c\|^2$$

(3) 检验 H_0 的统计量构造： $\frac{SSE}{SSE_c}$

若假设 H_0 不成立，则 $A\beta = u$ 的约束条件会显著损害拟合效果，使 SSE_c 大于 SSE，所以自然地我们去检验

$$\frac{SSE}{SSE_c}$$

Very Important

REMINDER

Theorem 2 添加正态假设 $\varepsilon \sim N(0, \sigma^2 I)$，我们有：

(1) $\frac{SSE}{\sigma^2} \sim \chi^2(n-p)$

(2) 若 $H_0: A\beta = u$ 成立，则 $\frac{SSE_c - SSE}{\sigma^2} \sim \chi^2(m)$

(3) 若 $H_0: A\beta = u$ 成立，则 $SSE \perp SSE_c - SSE$

(4) 若 $H_0: A\beta = u$ 成立，则

$$\frac{(SSE_c - SSE)/m}{SSE/(n-p)} \sim F(m, n-p)$$

Proof (1) $SSE = \|y-\hat{y}\|^2 = \|e\|^2 = e^T e$ $\quad \|y - X\hat{\beta}_c\|^2$

由作业题： $\dfrac{e^T e}{\sigma^2} \sim \chi^2(n-p)$ $\quad = \|y - X\hat{\beta}\|^2 + \|X(\hat{\beta} - \hat{\beta}_c)\|^2$

Proof (2) 由之前的证明唯一性：

$$\|y - X\hat{\beta}_c\|^2 = \|y - X\hat{\beta}\|^2 + \|X(\hat{\beta} - \hat{\beta}_c)\|^2$$

$\therefore \quad SSE_c - SSE = \|X(\hat{\beta} - \hat{\beta}_c)\|^2 \quad \Leftarrow ☆$

$$= \|X[\hat{\beta} - (\hat{\beta} - (X^TX)^{-1}A^T D^{-1}(A\hat{\beta} - u)]\|^2$$

$$= \{X(X^TX)^{-1}A^T D^{-1}(A\hat{\beta} - u)\}^T \{X(X^TX)^{-1}A^T D^{-1}(A\hat{\beta} - u)\}$$

$$= (A\hat{\beta} - u)^T D^{-1} A (X^TX)^{-1} X^T \cdot X(X^TX)^{-1} A^T D^{-1} (A\hat{\beta} - u)$$

$$= (A\hat{\beta} - u)^T D^{-1} A (X^TX)^{-1} A^T D^{-1} (A\hat{\beta} - u)$$

$$= (A\hat{\beta} - u)^T D^{-1} D D^{-1} (A\hat{\beta} - u)$$

$$= (A\hat{\beta} - u)^T D^{-1} (A\hat{\beta} - u) \qquad D = A(X^TX)^{-1}A^T \in \mathbb{R}^{m\times m}$$

$\because \hat{\beta} \sim N(\beta, (X^TX)^{-1}\sigma^2) \Rightarrow (A\hat{\beta} - u) \sim N(A\beta - u, \sigma^2 D)$ 而当 H_0 成立时，

$A\beta - u = 0 \Rightarrow (A\hat{\beta} - u) \sim N(0, \sigma^2 D)$ 而 D 可逆，

而由 Capter 4 (Proposition 4) 知

$$\frac{SSE_c - SSE}{\sigma^2} = \left(\frac{A\hat{\beta} - u}{\sigma}\right)^T \cdot D^{-1} \cdot \left(\frac{A\hat{\beta} - u}{\sigma}\right) \sim \chi^2(m)$$

Proof (3): SSE 是 e 的函数， $SSE_c - SSE$ 是 $\hat{\beta}$ 的函数，

而 $\hat{\beta} \perp e \Rightarrow SSE \perp SSE_c - SSE$

Proof (4): 由 F-distribution 定义可证：

Rm: $\hat{\sigma}^2 = \dfrac{SSE}{n-p} = \dfrac{e^T e}{n-p} = \dfrac{1}{n-p}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$

Remark:

若 $H_0: A\beta = u$ 成立，则

$$\frac{(A\hat{\beta}-u)^T D^{-1}(A\hat{\beta}-u)/m}{SSE/(n-p)} \sim F(m, n-p)$$

(4) 特例：对单个 β_j 检验

令 $A = [0,0,\cdots,1,\cdots,0,0] \in \mathbb{R}^{1\times p}$ $\quad m=1$ $\quad u=0$，就得到了对 β_j 检验：

$$H_0: \beta_j = 0 \quad v.s. \quad H_1: \beta_j \neq 0$$

因为此时 $A\hat{\beta} - u = \hat{\beta}_j \qquad D = \{(X^TX)^{-1}\}_{[j,j]}$

于是，F-test 变为：

若 $H_0: \beta_j = 0$ 成立，$\dfrac{\hat{\beta}_j^2}{\{(X^TX)^{-1}\}_{[j,j]} \cdot SSE/(n-p)} \sim F(1, n-p)$

↑

① 可以判断 "β_j 与 0"

② 进一步，单变量可判断 "β_j 的正负"

↓

由 $t^2(a) = F(1,a)$ 可知：

若 $H_0: \beta_j = 0$ 成立，$\dfrac{\hat{\beta}_j}{\sqrt{\{(X^TX)^{-1}\}_{[j,j]} \cdot SSE/(n-p)}} \sim t(n-p)$

6.3 拟合优度 Goodness of fit

(1) 决定系数：Coefficient of Determination

$$R^2 = \frac{SSR}{SST}$$

(2) y与所有自变量的复相关系数

$$R = \left(\frac{SSR}{SST}\right)^{\frac{1}{2}}$$

(3) 对整体自变量的检验：

利用 $\frac{SSR}{SSE}$（等价的）作检验，假设只能对 $\beta_j\ (j\neq 1)$ 整体检验：

$H_0: \beta_j = 0\ (j=2,\cdots,p)$，则

$$\frac{(n-p)/(p-1)}{R^{-2}-1} = \frac{SSR/(p-1)}{SSE/(n-p)} \sim F(p-1, n-p)$$

Proof: $\because \frac{(SSE_c - SSE)/(m)}{SSE/(n-p)} \sim F(m, n-p)$

取 $A = \begin{bmatrix} 0 & 1 & & \\ \vdots & & \ddots & \\ 0 & & & 1 \end{bmatrix} \in R^{(p-1)\times p}$ $u=0 \Rightarrow \text{rank}(A) = m = p-1$, $H_0: A\beta = u = 0$

$\therefore \frac{(SSE_c - SSE)/(p-1)}{SSE/(n-p)} \sim F(p-1, n-p)$

$\because A\beta = 0 \Rightarrow \beta_1 \neq 0 \quad \beta_j = 0 \Rightarrow y_i = \beta_1 + \varepsilon_i$ $\Rightarrow$ 预测 $\hat{y}_i = \bar{y}$

无回归参数 $\Rightarrow SSR_c = 0 \Rightarrow SST = SSE_c$

$$\therefore \frac{SSR/(p-1)}{SSE/(n-p)} = \frac{(SST-SSE)/(p-1)}{SSE/(n-p)}$$

$$= \frac{(SSE_c - SSE)/(p-1)}{SSE/(n-p)} \sim F(p-1, n-p)$$

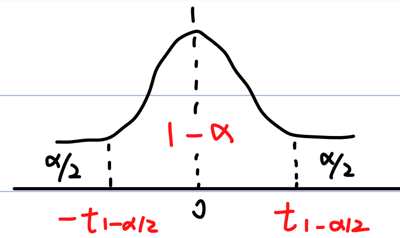
$SST = \sum(y_i - \bar{y})^2 \quad SSR = \sum(\hat{y}_i - \bar{y})^2 \quad SSE = \sum(\hat{y}_i - y_i)^2$

6.4 区间估计 Confidence Intervals

(1) 对单变量参数估计 $u = \beta_j$

$$\frac{\hat{\beta}_j - \beta_j}{\sqrt{\{(X^TX)^{-1}\}_{[j,j]} \cdot SSE/(n-p)}} \sim t(n-p)$$

$$P\left(\left|\frac{\hat{\beta}_j - \beta_j}{\sqrt{\{(X^TX)^{-1}\}_{[j,j]} \cdot SSE/(n-p)}}\right| \leq t_{1-\alpha/2}\right) \geq 1-\alpha$$



(2) β_j 的 $(1-\alpha)$ 置信区间 CI

$$\left(\hat{\beta}_j - t_{1-\alpha/2}(n-p) \cdot \sqrt{\frac{\{(X^TX)^{-1}\}_{[j,j]} \cdot SSE}{n-p}},\ \hat{\beta}_j - t_{1-\alpha/2}(n-p) \cdot \sqrt{\frac{\{(X^TX)^{-1}\}_{[j,j]} \cdot SSE}{n-p}}\right)$$

7. 中心化 &. 标准化 Centralization &. Standardization

由 $X(y - X\hat{\beta}) = \sum_{i=1}^{n} x_i^T e_i = 0 \in \mathbb{R}^p$，取第一行有 $\sum_{i=1}^{n}(y_i - x_i^T\hat{\beta}) = 0$

$$\bar{y} = \bar{x}^T \cdot \hat{\beta} = \hat{\beta}_1 + \bar{x}_2 \cdot \hat{\beta}_2 + \cdots + \bar{x}_p \cdot \hat{\beta}_p$$

拟合一定过均值点

(1) 中心化：

$$\hat{y}_i - \bar{y} = (x_{i2} - \bar{x}_2) \cdot \hat{\beta}_2 + \cdots + (x_{ip} - \bar{x}_p) \cdot \hat{\beta}_p \quad (i = 1, 2, \cdots, n)$$

(2) 标准化：

$$x_{ij}^* = \frac{x_{ij} - \bar{x}_j}{\hat{\sigma}_{x_j}}$$

$i = 1, 2, \cdots, n$ $\quad j = 2, 3, \cdots, p$

$$y_i^* = \frac{y_i - \bar{y}}{\hat{\sigma}_y}$$

就有：

$$y_i^* = \beta_2^* x_{i2}^* + \beta_3^* x_{i3}^* + \cdots + \beta_p^* x_{ip}^*$$

$$D = A(X^T X)^{-1} A^T$$

$$\hat{\beta} = (X^T X)^{-1} X^T y$$

$$\hat{\beta}_c = \hat{\beta} - (X^T X)^{-1} A^T D^{-1} (A\hat{\beta} - u)$$

$$\|X(\hat{\beta}_c - \hat{\beta})\|^2 = \frac{(A\hat{\beta} - u)^T}{\sigma} D^{-1} \frac{(A\hat{\beta} - u)}{\sigma} \sim \chi^2(m)$$