CHAPITRE 2 Mutiple linear regression

$$\begin{cases} y_1 = \beta_0 + \beta_1 x_{11} + \cdots + \beta_k x_{1k} + \varepsilon_1 \quad (k \text{ var}) \\ \vdots \\ y_n = \beta_0 + \beta_1 x_{n1} + \cdots + \beta_k x_{nk} + \varepsilon_n. \end{cases}$$

$$y = X\beta + \varepsilon. \qquad y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \qquad X = \underbrace{\begin{bmatrix} 1 & x_{11} & \cdots & x_{1k} \\ \vdots & \vdots & & \vdots \\ 1 & x_{n1} & \cdots & x_{nk} \end{bmatrix}}_{n \times (k+1)}$$

$$\beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix} \qquad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

Ordinary least squares.

$$\min \; S(\beta) = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2} - \cdots - \beta_k x_{ik})^2$$

$$= (y - X\beta)'(y - X\beta) = (y' - \beta' X')(y - X\beta) = y'y - \beta' X' y - y' X\beta + \beta' X' X\beta = y'y - 2(X'y)'\beta + \beta' X' X\beta$$

matrix differentiation

1. Soit $c=(c_1,\dots c_K)'$, $\beta=(\beta_1,\dots\beta_K)'$

$$f(\beta)=c'\beta=c_1\beta_1+\cdots+c_K\beta_K$$

$$\frac{\partial f(\beta)}{\partial\beta}=\frac{\partial(c'\beta)}{\beta}=c$$

2. $$\frac{\partial f(\beta)}{\partial\beta}=\frac{\partial(\beta'A\beta)}{\partial\beta}=2A\beta.$$

Theorem: Estimators of ordinary least squares

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

$$\hat\beta=\beta+(X'X)^{-1}X'\cdot\varepsilon$$

Relation between ε and residuals e.

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

$$=[\underbrace{I-X(X'X)^{-1}X'}_{I-H}]y \qquad H=X(X'X)^{-1}X'$$

$$=\underline{(I-H)}\cdot y=\underline{M}\cdot y=M\cdot(X\beta+\varepsilon)=MX\beta+M\varepsilon.$$

$$MX = (I-H)X = X - HX = X - X\underbrace{(X'X)^{-1}X'X}_{I} = X - X = 0$$

$$e = M\varepsilon.$$

$\boxed{\begin{aligned} &\text{Conclusion:} \quad e = (I-H)y = My \\ &H = X(X'X)^{-1}X' \\ &e = M\varepsilon \\ &M = I-H. \end{aligned}}$

Theorem: $X' \cdot e = 0 \qquad \hat{y}' \cdot e = 0.$

Gauss-Markov

$E(\varepsilon_i) = 0, \quad Var(\varepsilon_i) = \sigma^2 \ \forall i, \quad E(\varepsilon_i \varepsilon_j) = 0. \ \forall i \neq j$

$E(\varepsilon) = 0. \quad E(\varepsilon\varepsilon') = \sigma^2 I = Cov(\varepsilon)$

$\underline{E(y) = X\beta}. \quad \underline{Cov(y) = \sigma^2 I}. \quad \underline{E(ee') = \sigma^2 M = \sigma^2 (I-H)}$

$H = X(X'X)^{-1}X'$

$Var(e_i) = \sigma^2 m_{ii} = \sigma^2 (1-h_{ii}), \quad h_{ii} \leq 1$

proof: (a) $y = X\beta + \varepsilon$. $E(y) = E(X\beta + \varepsilon) = X\beta + E(\varepsilon) = X\beta$.

(b) $\mathrm{Cov}(y) = E[(y - E(y))(y - E(y))'] = E[(y - X\beta)(y - X\beta)']$

$= E(\varepsilon\varepsilon') = \mathrm{Cov}(\varepsilon) = \sigma^2 I$.

(c) $e = M\varepsilon$, $E(e) = E(M\varepsilon) = 0$.

$E(ee') = \mathrm{Cov}(e) = \mathrm{Cov}(M\varepsilon) = M \underbrace{\mathrm{Cov}(\varepsilon)}_{\sigma^2 I} M' = \sigma^2 MM'$

M is symmetric and idempotent $MM' = M$.

$= \sigma^2 M = \sigma^2 (1 - H)$

(d) $M = \begin{bmatrix} m_{11} & \cdots & & \\ & \ddots & & \\ & & m_{ii} & \\ & & & \ddots \\ & & & & m_{nn} \end{bmatrix}$

Theorem: bias and variance of estimators.

$E(\hat{\beta}) = \beta$. $\mathrm{Cov}(\hat{\beta}) = \sigma^2 (X'X)^{-1}$

$\mathrm{Cov}(\hat{y}) = \sigma^2 H$. $\hat{y} = X\hat{\beta}$

$\mathrm{var}(\hat{y}_i) = \sigma^2 h_{ii}$ $\hat{y}_i = x_i' \hat{\beta}$ avec $x_i' = (1, x_{i1}, \ldots x_{ik})$

proof: (a) $E(\hat{\beta}) = E(\beta + (X'X)^{-1}X'\varepsilon)$

$$= \beta + (X'X)^{-1}X' \cdot E(\varepsilon)$$

$$= \beta.$$

(b) $\mathrm{cov}(\hat{\beta}) = \mathrm{cov}(\beta + A\varepsilon)$ avec $A = (X'X)^{-1}X'$

$\Rightarrow \mathrm{cov}(\hat{\beta}) = A\,\mathrm{cov}(\varepsilon)A' = A \cdot \sigma^2 I A' = \sigma^2 AA'$

$= \sigma^2 (X'X)^{-1}X' \cdot (X \cdot (X'X)^{-1}) = \sigma^2 (X'X)^{-1}$

(c) $\mathrm{cov}(\hat{y}) = \mathrm{cov}(X\hat{\beta}) = X\,\mathrm{cov}(\hat{\beta})X' =$

$X \cdot \sigma^2 (X'X)^{-1} \cdot X' = \sigma^2 \cdot X(X'X)^{-1}X' = \sigma^2 H.$

(d) $\mathrm{var}(\hat{y}_i) = \mathrm{var}(x_i' \cdot \hat{\beta}) = x_i'\,\mathrm{cov}(\hat{\beta}) \cdot x_i$

$$= \sigma^2 x_i' (X'X)^{-1} x_i$$

Theorem: If $tr[(X'X)^{-1}] \to 0$,

when $n \to \infty$, then $\hat{\beta}$ is a convergent estimator of β.

Theorem. $S^2 = \frac{e'e}{n-k-1} = \frac{1}{n-k-1}\sum_{i=1}^{n} e_i^2$ is an unbaised and convergent estimator of σ^2.

$k+1$: number of regression parameters.

$E(s^2) = \sigma^2$ and $\lim_{n\to\infty} P(|s^2-\sigma^2| \geq \delta) = 0$.

$$K = \frac{(n-k-1)s^2}{\sigma^2} \sim \chi^2_{(n-k-1)}, \quad E(K) = n-k-1.$$

$$var(K) = 2(n-k-1)$$

P.S Chebyshev's inequality.

$$Pr(|X-\mu| \geq k\sigma) \leq \frac{1}{k^2}$$

$$\rightarrow Pr(|s^2-\sigma^2| \geq \delta) \leq \frac{var(s^2)}{\delta^2}$$

Relation between SSE, SST, SSR.

$$e_i = y_i - \hat{y}_i, \qquad e_i = (y_i - \bar{y}) - (\hat{y}_i - \bar{y})$$

$$\sum e_i \hat{y}_i = 0.$$

$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} y_i^2 - \sum_{i=1}^{n} \hat{y}_i^2$$

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

$$R^2 = \frac{\sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} = 1 - \frac{SSE}{SST}$$

$$R^2_{adjusted} = 1 - \frac{SSE / (n-k-1)}{SST / (n-1)}$$

Gauss-Markov Theorem.

estimate linear function $L\beta$ or $\ell\beta$.

L: matrix, ℓ: vector.

$y = X\beta + \varepsilon$, $\hat{\beta} = (X'X)^{-1}X'y$ estimator of β.

estimator $\ell'\hat{\beta}$ is the best unbaised linear estimator of $\ell'\beta$.

Least squares under linear Constraints

$$y = X\beta + \varepsilon$$

$$\text{s.t.} \quad C\beta - d = 0$$

estimator of least squares under constraints:

$$\hat{\beta}_C = \hat{\beta} + (X'X)^{-1}C'\left[C(X'X)^{-1}C'\right]^{-1}(d - C\hat{\beta})$$

$$\hat{\beta} = \beta + (X'X)^{-1}X'\varepsilon.$$

proof: $\min \quad S(\beta) = (y - X\beta)'(y - X\beta)$

$$= y'y - 2(X'y)'\beta + \beta'X'X\beta.$$

s.t $\quad C\beta - d = 0$, d vector $m \times 1$.

Lagrange: $\quad Q = S(\beta) + \lambda'(d - C\beta)$

$$= y'y - 2(X'y)'\beta + \beta'X'X\beta + d'\lambda - (C\beta)'\lambda.$$

$\lambda = (\lambda_1, \dots \lambda_m)$

$$\begin{cases} \dfrac{\partial Q}{\partial \lambda} = d - C\beta \\[2ex] \dfrac{\partial Q}{\partial \beta} = 2(X'X)\beta - 2X'y - C'\lambda \end{cases}$$

$$\Rightarrow \begin{cases} d = C\hat{\beta}_c \\[1ex] 2(X'X)\hat{\beta}_c - 2X'y - C'\hat{\lambda} = 0 \end{cases}$$

$$\Rightarrow \hat{\beta}_c = \hat{\beta} + \frac{1}{2}(X'X)^{-1}C'\hat{\lambda}$$

$$\hat{\lambda} = 2[C(X'X)^{-1}C']^{-1}(d - C\hat{\beta})$$