김은영

제3장 중선형회귀모형

3.3 회귀직선의 적합도와 분산분석

< 중선형회귀에서의 분산분석표(ANOVA table) >

요인	제곱합	자유도	평균제곱합	F비
회귀	SSR	$p-1$	$MSR=\dfrac{SSR}{(p-1)}$	$F_0=\dfrac{MSR}{MSE}=\dfrac{SSR/(p-1)}{SSE/(n-p)}$
오차	SSE	$n-p$	$MSE=\dfrac{SSE}{(n-p)}$	
전체	SST	$n-1$		

여기서, $SST=\sum_{i=1}^{n}\left(y_i-\bar{y}\right)^2$, $SSR=\sum_{i=1}^{n}\left(\hat{y}_i-\bar{y}\right)^2$, $SSE=\sum_{i=1}^{n}\left(y_i-\hat{y}_i\right)^2$

■ 결정계수 $R^2=\dfrac{SSR}{SST}=1-\dfrac{SSE}{SST}$

[장점]

▸ 그 의미가 명확하게 해석이 쉽다.

▸ 특히 두 개 이상의 모형들을 비교하고자 할 때 사용한다.

[단점]

▸ 회귀모형에 포함된 설명변수의 개수가 많아질수록 SSR이 증가하므로, 결정계수의 값도 항상 증가한다. 따라서 결정계수를 모형선택의 기준으로 할 경우 설명변수의 개수가 많은 모형을 선택하는 경향이 있을 수 있으므로 적합하지 않다.

→ 수정된 결정계수

3.4 이차형식과 제곱합의 분포

3.4.1 제곱합의 이차형식

■ 1과 J의 유용한 관계식들

▸ $1=(1,\ \cdots,1)^t$

▸ $J=\begin{pmatrix} 1 & \cdots & 1 \\ \vdots & & \vdots \\ 1 & \cdots & 1 \end{pmatrix}$

▸ $11^t=J$

▸ $1^t1=n$

▸ $J \bullet 1=\begin{pmatrix} 1 & \cdots & 1 \\ \vdots & & \vdots \\ 1 & \cdots & 1 \end{pmatrix}\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}=\begin{pmatrix} n \\ \vdots \\ n \end{pmatrix}=n \bullet 1$

▸ $1^t \bullet J=(1 \cdots 1)\begin{pmatrix} 1 & \cdots & 1 \\ \vdots & & \vdots \\ 1 & \cdots & 1 \end{pmatrix}=(n \cdots n)=n \bullet 1^t$

▸ $J^2=\begin{pmatrix} 1 & \cdots & 1 \\ \vdots & & \vdots \\ 1 & \cdots & 1 \end{pmatrix}\begin{pmatrix} 1 & \cdots & 1 \\ \vdots & & \vdots \\ 1 & \cdots & 1 \end{pmatrix}=\begin{pmatrix} n & \cdots & n \\ \vdots & & \vdots \\ n & \cdots & n \end{pmatrix}=n \bullet J$

1. SST

$$SST=\sum_{i=1}^{n}\left(y_i-\overline{y}\right)^2=\sum_{i=1}^{n}\left(y_i^2-2\overline{y}y_i+\overline{y}^2\right)=\sum_{i=1}^{n}y_i^2-2\overline{y}\sum_{i=1}^{n}y_i+n\overline{y}^2=\sum_{i=1}^{n}y_i^2-n\overline{y}^2$$

$$SST=y^ty-n\left(\frac{1^ty}{n}\right)^2=y^ty-\frac{1}{n}(y^t1)(1^ty)=y^ty-\frac{1}{n}y^t11^ty=y^t\left(I-\frac{1}{n}J\right)y=y^tTy$$

여기서, $T=I-\dfrac{1}{n}J$

$$T^2=\left(I-\frac{1}{n}J\right)\left(I-\frac{1}{n}J\right)=I^2-\frac{1}{n}J-\frac{1}{n}J+\frac{1}{n^2}JJ=I^2-\frac{2}{n}J+\frac{1}{n}J=I-\frac{1}{n}J$$

$\rightarrow\ T=I-\dfrac{1}{n}J$: 멱등행렬

$$rank(T)=tr(T)=tr\left(I-\frac{1}{n}J\right)=tr(I)-\frac{1}{n}tr(J)=tr(I)-\frac{1}{n}n=n-1$$

2. SSE

$$SSE = e^t e = [(I-H)y]^t[(I-H)y] = \left[y^t(I-H)^t\right][(I-H)y] = y^t(I-H)y = y^t Ey$$

여기서, $E = I - H$: 멱등행렬

$$rank(E) = tr(I-H) = tr(I) - tr(H) = n - p$$

여기서, $tr(H) = tr\left[X\left(X^tX\right)^{-1}X^t\right] = tr\left[\left(X^tX\right)^{-1}X^tX\right] = tr\left(I_p\right) = p$

$*$ $tr(AB) = tr(BA)$

3. SSR

$$SSR = SST - SSE = y^t Ty - y^t Ey = y^t(T-E)y$$

$$= y^t\left[\left\{I - \frac{1}{n}J\right\} - \{I-H\}\right]y = y^t\left(H - \frac{1}{n}J\right)y = y^t Ry$$

여기서, $R = H - \dfrac{1}{n}J$: 멱등행렬

$$R^2 = \left(H - \frac{1}{n}J\right)\left(H - \frac{1}{n}J\right) = H - \frac{2}{n}JH + \frac{1}{n^2}J^2$$ 여기서, $JH = J$[1]

$$= H - \frac{2}{n}J + \frac{1}{n}J = H - \frac{1}{n}J$$

$$rank(R) = tr\left(H - \frac{1}{n}J\right) = tr(H) - \frac{1}{n}tr(J) = p - \frac{1}{n}n = p - 1$$

$SST = y^t Ty$	$T = I - \frac{1}{n}J$: 멱등행렬
	$rank(T) = n-1$
$SSR = y^t Ry$	$R = H - \frac{1}{n}J$: 멱등행렬
	$rank(R) = p-1$
$SSE = y^t Ey$	$E = I - H$: 멱등행렬
	$rank(E) = n-p$

1) $X^t = X^t$

$X^tX\left(X^tX\right)^{-1}X^t = X^t$ 여기서, $X\left(X^tX\right)^{-1}X^t = H$

$X^tH = X^t$

양변의 첫 번째 열벡터를 생각해 보자.

$1^tX\left(X^tX\right)^{-1}X^t = 1^t$

$11^tX\left(X^tX\right)^{-1}X^t = 11^t$

$\rightarrow$ $JH = J$