韓國軍事運營分析學會誌
第 8 卷第 2 號, 1982. 12

종속자료에 대한 회귀분석

(Regression Analyis of Dependent Data)

정 규 련*

I. 서 론

일반적인 회귀분석 모형은 다음과 같이 표현할 수 있다.

$$Y = XB + e, \quad E(Y) = XB$$

단,

$$X = \begin{bmatrix} x_{10} & x_{11} & \cdots & x_{1k} \\ x_{20} & x_{21} & \cdots & x_{2k} \\ \vdots & \vdots & & \vdots \\ x_{n0} & x_{n1} & \cdots & x_{nk} \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1k} \\ 1 & x_{21} & \cdots & x_{2k} \\ \vdots & \vdots & & \vdots \\ 1 & x_{n1} & \cdots & x_{nk} \end{bmatrix},$$

$$B = \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_k \end{bmatrix}, \quad Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad e = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}$$

이러한 회귀분석 모형에서 오차가 독립 항등적으로 기대치 0, 분산이 σ^2 인 정규분포를 따르는 경우는 널리 연구되어 왔다. 이 논문에서는 표본이 특정한 종속관계에 있을 경우, 많은 결과들이 독립적인 경우와 같음을 보이고자 한다.

다음과 같은 모형을 생각해 보자.

$$Y = XB + e, \quad e \sim N(0, V)$$

단,

$$V = 1/2(A + A') + \alpha(I + E),$$

$$A = \begin{bmatrix} a_1 & a_1 & \cdots & a_1 \\ a_2 & a_2 & \cdots & a_2 \\ \vdots & \vdots & & \vdots \\ a_n & a_n & \cdots & a_n \end{bmatrix},$$

A'는 A의 전치이며 a_i와 α는 양의 상수, I는 $n \times n$ 단위행렬, E는 모든 원소가 1인 $n \times n$ 행렬이다. 이 모형에서 우리는 독립 항등적인 경우와 꼭 일치하지 않음을 견하게 될 것이다.

II. 기존모형의 개요

A. $\hat{B}$와 $\hat{\sigma}^2$의 성질

$$Y = XB + e, \quad e \sim N(0, \sigma^2 I)$$

이 모형에서 B와 σ^2은 미지이며, X는 계수(rank)가 $k+1$인 기지 행렬이다. 최우추정법으로 B와 σ^2을 추정하면,

$$\hat{B} = S^{-1}X'Y, \quad \text{단 } S^{-1} = (X'X)^{-1}$$

$$\hat{\sigma}^2 = \frac{(Y - X\hat{B})'(Y - X\hat{B})}{n-(k+1)} = \frac{Y'(I - XS^{-1}X')Y}{n-(k+1)}$$

이 경우 $\hat{B}$와 $\hat{\sigma}^2$은 다음과 같은 성질을 가진다.

(*) 육사 수학과

1. 일치성
2. 유효성
3. 불편성
4. 충분성
5. $\hat{B} \sim N(B, \sigma^2 S^{-1})$
6. 최소 분산 불편성
7. $\dfrac{(n-k-1)\hat{\sigma}^2}{\sigma^2} \sim \chi^2(n-k-1)$
8. $\hat{B}$와 $\hat{\sigma}^2$은 독립이다.

B. **結果 유도에 사용한 定理**

定理 1. $Y \sim N(\mu, \alpha^2 I)$이면, $\dfrac{Y'AY}{\sigma^2} \sim \chi^2(k, \lambda)$이기 위한 필요 충분조건은 A가 idempotent이다. 단,

$$\lambda = \frac{\mu' A \mu}{2\sigma^2}, \quad k = A\text{의 階數}$$

定理 2. $Y \sim N(\mu, V)$일 때, $Y'BY \sim \chi^2(k, \lambda)$이기 위한 필요 충분조건은 BV가 idempotent인 것이다. 단,

$$\lambda = \tfrac{1}{2}\mu' B \mu, \quad k = B\text{의 階數}$$

定理 3. $Y \sim N(\mu, V)$일 때, $Y'AY$와 $Y'BY$가 독립이기 위한 필요 충분조건은 $AVB = O$인 것이다.

定理 4. $Y \sim N(\mu, V)$일 때 $C'Y$와 $Y'AY$가 독립일 필요 충분조건은 $C'VA = O$인 것이다.

定理 5. (Hogg-Craig **定理**)

$Q = Q_1 + Q_2 + \cdots + Q_k$라 하자. 단, $Q, Q_1 \cdots Q_k$는 $N(\mu, \sigma^2)$에서 추출된 크기가 n인 확률표본으로 부터의 二次형식으로 된 확률변수들이다.

$\dfrac{Q}{\sigma^2} \sim \chi^2(r)$, $\dfrac{Q_i}{\sigma^2} \sim \chi^2(r_i)$, $(i = 1, 2, \cdots, k-1)$, Q_k는 非陰이라 하자. 그러면 $Q_1, Q_2, \cdots Q_k$는 相互 獨立이며

$$\frac{Q_k}{\sigma^2} \sim \chi^2\left(r_k = r - \sum_{i=1}^{k-1} r_i\right)\text{이다.}$$

定理 6. (Baldessari **定理**)

$Y \sim N(\mu, V)$이라 하고 $B_0, B_1, \cdots, B_k$는 $(n \times n)$ idempotent 행렬로서 $\sum B_j = I - \frac{1}{n}E$를 만족한다고 하자. 그리고 α는 陽數라 하자. 그러면 $Y'B_jY/\alpha$들이 獨立이면서 自由度가 r_j($r_j = B_j$의 階數)인 non-central Chi-square분포이기 위한 필요 충분조건은 共分散 行列 V가 다음과 같아야 한다.

$$V = \tfrac{1}{2}(A + A') + \alpha(I - E)$$

Ⅲ. 現 模型에서의 B와 $Y'(I - XS^{-1}X')Y$의 性質

$Y = XB + e$를 생각해 보자

여기서 $e \sim N(O, V)$, $V = \frac{1}{2}(A + A') + \alpha(I - E)$이다.

$$\hat{B} = S^{-1}X'Y, \quad \frac{(Y - X\hat{B})'(Y - X\hat{B})}{\alpha} = \frac{Y'(I - XS^{-1}X')Y}{\alpha} = \frac{Y'CY}{\alpha}$$

라 하자. 다음을 보이고자 한다.

A. $E(\hat{B}) = B$

B. $\dfrac{Y'CY}{\alpha} \sim \chi^2(n-k-1)$

C. $E(Y'CY) = \alpha(n-k-1)$

D. $\hat{B}$와 $Y'CY$는 縱屬이다.

A. $\hat{B}$의 기대치, $E(\hat{B})$

$$E(\hat{B}) = E(S^{-1}X'Y) = S^{-1}X'E(Y) = S^{-1}X'E(XB + e) = B$$

따라서 $\hat{B} = S^{-1}X'Y$는 B의 不偏 推定量이다.

B. $\dfrac{Y'CY}{\alpha}$의 분포

$\dfrac{Y'CY}{\alpha} \sim \chi^2(n-k-1)$이기 위한 필요 충분조건은 $1/\alpha CV$가 idempotent인 것이다.

$XS^{-1}X'E = E$, $XS^{-1}X'A' = A'$,

$AXS^{-1}X' = A$ 이므로

$$\frac{1}{\alpha}(I - XS^{-1}X')V(I - XS^{-1}X')$$

$$= \frac{1}{\alpha}(I - XS^{-1}X')\{\tfrac{1}{2}(A + A') + \alpha(I - E)\}(I - XS^{-1}X')$$

$$= (I - XS^{-1}X')$$

따라서 idempotent 이며 定理 2에 의하여 $\frac{Y'CY}{\alpha} \sim \chi^2(q, \lambda)$ 이다.

여기서

$$q = \text{rank}\ \frac{(I - XS^{-1}X')}{\alpha}$$

$$= \text{rank}\ (I - XS^{-1}X')$$

$$= \text{trace}\ (I - XS^{-1}X')$$

$$= \text{trace}\ (I) - \text{trace}\ (X'XS^{-1})$$

$$= n - k - 1$$

$$\lambda = \tfrac{1}{2}D\ .. \frac{(I - XS^{-1}X')}{\alpha}XB = 0$$

그러므로

$$\frac{Y'CY}{\alpha} \sim \chi^2(n - k - 1)$$ 이다.

C. $Y'CY$의 기대치

위의 結果로 부터 $E(Y'CY) = \alpha(n-k-1)$ 임을 바로 알 수 있으며 또한 $\frac{Y'CY}{n-k-1}$ 는 α의 不偏 推定量임을 알 수 있다.

D. $\hat{B}$와 $Y'CY$는 縱屬이다

$$S^{-1}X'V(I - XS^{-1}X') = \tfrac{1}{2}(S^{-1}X'A' - S^{-1}X'A'XS^{-1}X') \neq 0$$

定理 4에 의하여 縱屬이다. 이 點이 獨立恒等的인 境遇와 相異하다. 이 點을 극복하기 위하여 다음 節에서 새로운 變形된 模型을 개발한다.

IV. 新 模型에서의 $\hat{\bar{B}}_1$와 $\bar{Y}'(I - \bar{X}\bar{S}^{-1}\bar{X}')\bar{Y}$의 性質

$\bar{Y} = \bar{X}B_1 + \bar{e}$를 생각하자.

여기서 $\bar{Y} = \{I - (\frac{1}{n})E\}Y$

$$\bar{X} = \{I - (\tfrac{1}{n})E\}X_1$$

$$B_1 = (b_1, b_2, \cdots, b_k)'$$

$$\bar{e} = \{I - (\tfrac{1}{n})E\}e$$

$$X_1 = \begin{bmatrix} x_{11} & \cdots & x_{1k} \\ \vdots & & \vdots \\ x_{n1} & \cdots & x_{nk} \end{bmatrix}$$

그러면

$$V(\bar{e}) = \alpha\{I - (\tfrac{1}{n})E\},$$ 즉

$$\bar{e} \sim N(O, \alpha(I - (\tfrac{1}{n})E))$$ 이다.

그리고

$$\bar{Y} \sim N(\bar{X}B_1, \alpha(I - \tfrac{1}{n}E))$$ 이다.

$\bar{S} = \bar{X}'\bar{X}$라 할 때 $\hat{B}_1 = \bar{S}^{-1}\bar{X}'\bar{Y}$라 하고 $\frac{1}{\alpha}\bar{Y}'(I - \bar{X}\bar{S}^{-1}\bar{X}')\bar{Y} = \frac{1}{\alpha}\bar{Y}'\bar{C}\bar{Y}$ 라 하자.

다음을 보이고자 한다.

A. $\hat{B}_1$의 기대치

B. $\frac{1}{\alpha}\bar{Y}'\bar{C}\bar{Y}$의 分布

C. $\bar{Y}'\bar{C}\bar{Y}$의 기대치

D. $\hat{B}_1$와 $\bar{Y}'\bar{C}\bar{Y}$의 獨立性

A. $\hat{B}_1$의 기대치, $E(\hat{B}_1)$

$$E(\hat{B}_1) = E(\bar{S}^{-1}\bar{X}'\bar{Y})$$

$$= \bar{S}^{-1}\bar{X}'E(\bar{Y})$$

$$= \bar{S}^{-1}\bar{X}'E(\bar{X}B_1 + \bar{e}) = B_1$$

따라서 $\hat{B}_1$은 B_1의 不偏推定量이다.

B. $\frac{1}{\alpha}\bar{Y}'\bar{C}\bar{Y}$의 분포

$\bar{C}\alpha(I - \frac{1}{n}E)$는 idempotent 이므로 定理 2에 의하여

$$\frac{1}{\alpha}\bar{Y}'\bar{C}\bar{Y} \sim \chi^2(q, \lambda)$$ 이다.

여기서

$$q = \text{rank}\left(\frac{1}{\alpha}\right)(I - \bar{X}\bar{S}^{-1}\bar{X}')$$
$$= \text{rank}\,(I - \bar{X}\bar{S}^{-1}\bar{X}')$$
$$= \text{trace}\,(I - \bar{X}\bar{S}^{-1}\bar{X}')$$
$$= \text{trace}\,(I) - \text{trace}(\bar{X}'\bar{X}\bar{S}^{-1})$$
$$= n - k$$

$$\alpha = \tfrac{1}{2}B_1'\bar{X}'\left(\frac{1}{\alpha}\right)(I - \bar{X}\bar{S}^{-1}\bar{X}')$$
$$\bar{X}B_1 = 0$$

따라서 $\frac{1}{\alpha}\bar{Y}'\bar{C}\bar{Y} \sim \chi^2(n-k)$

C. Y'CY의 기대치

위의 結果로 부터 $E(\bar{Y}'\bar{C}\bar{Y}) = \alpha(n-k)$ 임과 $\frac{\bar{Y}'\bar{C}\bar{Y}}{n-k}$ 는 α 의 不偏推定量임을 바로 알 수 있다.

D. $B_1 = \bar{S}^{-1}\bar{X}\bar{Y}$ 와 $\bar{Y}'\bar{C}\bar{Y}$ 의 獨立性

$$\bar{S}^{-1}\bar{X}'\alpha\left(I - \frac{1}{n}E\right)(I - \bar{X}\bar{S}^{-1}\bar{X}')$$
$$= \alpha\left(\bar{S}^{-1}\bar{X}' - \frac{1}{n}\bar{S}^{-1}\bar{X}'E\right)(I - \bar{X}\bar{S}^{-1}\bar{X}')$$
$= 0$ ($\bar{X}'E = 0$ 임을 유의)

定理 4에 의하여 이들은 獨立이다.

V. 結 論

위에서 論한 結果로 부터 $\hat{B}_1$ 의 分散, $r'B_1$ 의 구간 추정, 하나의 미래 자료에 대한 구간 예측을 論하고자 한다.

A. $V(\hat{B}_1)$

$$V(\hat{B}_1) = E\{(\hat{B}_1 - B_1)(\hat{B}_1 - B_1)'\}$$
$$= E\{(\bar{S}^{-1}\bar{X}'\bar{Y} - B_1)(\bar{S}^{-1}\bar{X}'\bar{Y} - B_1)'\}$$
$$= \bar{S}^{-1}\bar{X}'V(\bar{e})\bar{X}\bar{S}^{-1}$$
$$= \alpha\bar{S}^{-1}\bar{X}'\left(I - \frac{1}{n}E\right)\bar{X}\bar{S}^{-1}$$
$$= \alpha\bar{S}^{-1}$$

따라서 $\hat{B}_1 \sim N(B_1,\ \alpha\bar{S}^{-1})$

B. $r'B_1$ 의 구간추정

$r'\hat{B}' \sim N(r'B_1,\ \alpha r'\bar{S}^{-1}r)$ 이므로
$\frac{r'\hat{B}_1 \sim r'B_1}{(\alpha r'\bar{S}^{-1}r)^{\frac{1}{2}}} \sim N(0,1)$ 이다.

또한

$\frac{\bar{Y}'\bar{C}\bar{Y}}{\alpha} \sim \chi^2(n-k)$ 이고 $\hat{B}_1$ 와 독립이므로

$$\frac{r'\hat{B}_1 - r'B_1}{(\alpha r'\bar{S}^{-1}r)^{\frac{1}{2}}}\left\{\frac{(n-k)}{\bar{Y}'\bar{C}\bar{Y}}\right\}^{\frac{1}{2}} \sim$$
$t(n-k)$ 이다.

따라서

$$\Pr\left(r'\hat{B}_1 - mt_{\frac{\alpha_1}{2}} \le r'B_1 \le r'\hat{B}_1 + mt_{\frac{\alpha_1}{2}}\right) = 1 - \alpha_1$$

여기서 $m = \left\{\frac{(r'\bar{S}^{-1}r)\,\bar{Y}'\bar{C}\bar{Y}}{n-k}\right\}^{\frac{1}{2}}$

C. 하나의 미래자료에 대한 구간예측

y_{n+1} 을 하나의 미래 표본이라 하고 $Y \sim N(XB, V)$ 라 가정하자. 여기서 X는 $(n+1) \times (k+1)$ 行列, 階數는 $k+1$, 또 V는 앞에서 定義한 것과 같으며 $(n+1) \times (n+1)$ 行列이다.

$$y^*_{n+1} = y_{n+1} - \bar{y}_{n+1},\ \left(\bar{y}_{n+1} = \frac{1}{n+1}\sum_{i=1}^{n+1} y_i\right)$$

이라 정의하면

$$y^*_{n+1} = \frac{n}{n+1}(y_{n+1} - \bar{y}_n)$$
$$= \frac{n}{n+1}(B_1'X_{n+1} + e_{n+1})$$

여기서

$$X_{n+1} = (x_{(n+1)1},\ x_{(n+1)2},\ \cdots,\ x_{(n+1)k})'$$

그러면

$$E(y^*_{n+1}) = \frac{n}{n+1} B'_1 X_{n+1}$$ 이며

$$V(y^*_{n+1}) = \left(\frac{n}{n+1}\right)^2 \alpha$$ 이다.

다음, $\hat{y}^*_{n+1} = \frac{n}{n+1} \hat{B}'_1 X_{n+1}$ 라 하면

$$E(\hat{y}^*_{n+1}) = \frac{n}{n+1} B'_1 X_{n+1}$$ 이며

$$V(\hat{y}^*_{n+1}) = \alpha\left(\frac{n}{n+1}\right)^2 X'_{n+1} \bar{S}^{-1} X_{n+1}$$ 이다.

$Z = y^*_{n+1} - \hat{y}^*_{n+1}$ 이라 하면

$$Z \sim N\left\{0,\ \alpha\left(\frac{n}{n+1}\right)\left(1+\frac{n}{n+1}\right) X'_{n+1} \bar{S}^{-1} X_{n+1}\right\}$$ 이며

$Z' = Z / \{V(Z)\}^{\frac{1}{2}} \sim N(0,1)$ 이다.

더우기

$\frac{1}{\alpha} \bar{Y}'\bar{C}\bar{Y} \sim \chi^2(n-k)$ 이고 Z' 과 독립이므로,

$$T = \frac{Z'}{\left(\frac{\bar{Y}'\bar{C}\bar{Y}}{\alpha(n-k)}\right)^{\frac{1}{2}}} \sim t(n-k)$$ 이다.

結局, y_{n+1}을 구하기 위한 下界 L과 上界 U는 다음과 같다. (100 r % 신뢰구간)

$$L = \frac{n+1}{n}\left(y^*_{n+1} - t_{\frac{r}{2}} d\right) + \bar{y}_n$$

$$U = \frac{n+1}{n}\left(y^*_{n+1} + t_{\frac{r}{2}} d\right) + \bar{y}_n$$

여기서

$$d = \left[\frac{\left\{\frac{n}{n+1} + \left(\frac{n}{n+1}\right)^2\right\} X'_{n+1} \bar{S}^{-1} X_{n+1} \bar{Y}'\bar{C}\bar{Y}}{n-k}\right]^{\frac{1}{2}}$$

參 考 文 獻

[1] Baldessari, B., "Analysis of Variance of Dependent Data", Statistica, Vol. 26, pp.895-903, 1966.

[2] Graybill, F.A., An Introduction to Linear Statistical Models, Vol. 1, McGraw-Hill, New York.

[3] Hogg, R.V. and Craig, A.T., Introduction to Mathematical Statistics, 3rd. ed., Macmillan, New York.

[4] Searle, S.R., Linear Models, Wiley, New York.