直交變換에 의한 回歸係數 및 重相關과 偏相關係數의 計算法

李　　鍾　　厚

Methods of Computations of Regression Coefficient, Multiple Correlation and Partial via Orthogonal Transformation.

Jong Hoo-Lee

〈目　　次〉

Abstract

The following topics are discussed in this article. First orthogonal transformation is introduced, and it is shown that the transformed probem is equivalent to the original problem.. Section 3 considers the distribution of estimators for the regression coefficients and the error variance. Finally we derive methods of computations of multiple correlation and partial via orthogonal transformation.

1. 序　　言

線型回歸問題에 있어서 一般的條件 아래서 直交變換을 利用하여 回歸係數와 重 및 偏相關係數의 推定値를 求하는 問題를 다음과 같이 생각한다.

2節에서 直交變換行列을 求하고 直交變換後의 分布와 原分布가 같다는 것을 밝힌다. 3節에서는 回歸係數의 推定量과 그의 分布 그리고 多變量分布의 重相關係數 및 偏相關係數의 算法을 提示한다.

2. 直交變換

回歸模型

$$\mathbf{y} = X\beta + \eta \qquad (2.1)$$

단 $\mathbf{y} = (y_1, \cdots\cdots, y_n)'$는 $n\times 1$의 從屬 rector 이고 $X = (\mathbf{x}_1, \cdots\cdots, \mathbf{x}_k)$는 $n\times k$의 資料行列, $\beta = (\beta_1, \cdots\cdots, \beta_k)'$는 $k\times 1$의 回歸係數 vector 그리고 $\eta = (\eta_1, \cdots\cdots, \eta_n)'$는 $n\times 1$의 誤差 vector 卽 $\eta \sim N(O, \sigma^2 I_n)$이다. 여기서 X의 階數는 k라 둔다.

(定理 2, 1) $\mathbf{x} = (x_1, \cdots\cdots, x_n)'$는 $n\times 1$의 零아닌 vector 라 한다. 여기서 $U = I - k\mathbf{x}\mathbf{x}'$ 단 $k = 2/\|\mathbf{x}\|^2$으로 두면 U는 對稱$(U = U')$이고 直交$(UU' = U'U = I)$, 그리고 $U^2 = I$ (involution) 이다. 이와 같이 定義된 行列 U를 elementary reflector 라 한다.

U를 elementary reflector 라 하고 $U\mathbf{x} = \alpha\ell_1$, 단 $\ell_1 = (1, 0, \cdots, 0)'$ $\alpha^2 = \|\mathbf{x}\|^2$ 따라서 $\alpha = \pm\|X\|$ 임을 要求하며, 다음 定理에서 이에 합당한 U를 구한다.

(定理 2, 2) $\mathbf{x}$는 零아닌 vector 이고 $\mathbf{x} \neq \alpha\ell_1$이라 한다. $\mathbf{z} = \mathbf{x} - \alpha\ell_1$, 단 $\alpha^2 = \|\mathbf{x}\|^2$ 그리고 $U = I - 2\mathbf{z}\mathbf{z}'/\|\mathbf{x}\|^2$으로 두면 U는 elementary reflector 이고 $U\mathbf{x} = \alpha\ell_1$이다.

證明, U는 $\mathbf{z} \neq 0$이므로 하나의 elementary reflector 이다. 또 $\|\mathbf{z}\|^2 = \|\mathbf{x}\|^2 - 2\alpha x_1 + \alpha^2 = 2(\|\mathbf{x}\|^2 - \alpha x_1)$이므로

$$\begin{aligned} U\mathbf{x} &= \mathbf{x} - 2\mathbf{z}\mathbf{z}'\mathbf{x}/\|\mathbf{z}\|^2 = \mathbf{x} - 2(\mathbf{x} - \alpha\ell_1)(\mathbf{x} - \alpha\ell_1)'\mathbf{x}/\|\mathbf{z}\|^2 \\ &= \mathbf{x} - 2(\mathbf{x} - \alpha\ell_1)(\|\mathbf{x}\|^2 - \alpha x_1)/[2(\|\mathbf{x}\|^2 - \alpha x_1)] \\ &= \mathbf{x} - (\mathbf{x} - \alpha\ell_1) = \alpha\ell_1. \end{aligned}$$

우리들은 定理 2, 2의 方法에 의하여 直交行列 Q를 構成할 수 있다. $X = (\mathbf{x}_1, \cdots\cdots, \mathbf{x}_k)$, $k \leq n$, $\operatorname{rank} X = k$라 한다. $\mathbf{x}_1$이 $\alpha_1\ell_1$形이면 $U_1 = I$이고 $\mathbf{x}_1 \neq \alpha_1\ell_1$이면 U_1을 定理 2.2와 같이 둔다. 따라서 $U_1\mathbf{x}_1 = \alpha_1\ell_1$, $\mathbf{x}_j^{(1)} = U_1\mathbf{x}_j$, $j = 2, 3, \cdots\cdots, k$, 또 $X^{(1)} = [\alpha_1\ell_1, \mathbf{x}_2^{(1)}, \cdots\cdots, \mathbf{x}_k^{(1)}]$으로 두면 $X^{(1)}$은

$$X^{(j)} = \begin{bmatrix} R_{11} & X_{12}^{(j)} \\ O & X_{22}^{(j)} \end{bmatrix}, \qquad j = 1 \qquad (2.2)$$

인 形으로 表示된다. 단 $R_{11} = \alpha_1$이다. 우리들은 이 方法을 $(n-1)\times(k-1)$ 行列 $X_{12}^{(j)}$에 反復試行하여 elementery reflector [또는 單位行列 I_{n-1}] $\tilde{U}_2$를 얻는다. 그리고 $\tilde{U}_2$

$\times(X_{22}^{(j)}$의 첫째列$)=\alpha_2 \ell_2$, 단 $\ell_2=(1.0, \cdots\cdots, 0)'$ $(n-1)\times 1$ vector 이다. 그러므로

$$U_2 = \begin{bmatrix} I_1 & O \\ 0 & \widetilde{U}_2 \end{bmatrix}$$

으로 두면 $U_2 U_1 X = U_2 X^{(1)} = X^{(2)}$ 따라서 形(2.2)가 되며 R_{11}은 2×2 윗쪽 三角正則行列이다. 이 方法을 계속하면 直交行列 $U_j, j=1,\cdots\cdots,k$를 얻으면 $Q=U_k \cdots\cdots U_1$은 直交行列이고 다음과 같이 된다.

$$QX = \begin{array}{c} \begin{bmatrix} R \\ 0 \end{bmatrix} \end{array} \begin{array}{c} k \\ n-k \end{array} \tag{2.3}$$

단 R은 $k\times k$ 위로 三角正則行列이다. 따라서 다음의 定理를 얻는다.

(定理 2.3) $n\times n$ 直交行列 Q가 存在하여 $QQ'=Q'Q=I_n$, 그리고 式(2.3)이 成立한다.

이제부터 R의 (i,j) 元을 ξ_{ij}로 表示한다. R이 위로 三角正則行列이므로 $\xi_{ij}=0$, $i>j$, $\xi_{ii}\neq 0$, $i=1,2,\cdots\cdots,k$이다. (2.1)의 Q에 의한 變換을 $\mathbf{v}=Q\mathbf{y}$, $\epsilon=Q\eta$으로 둔다. $\mathbf{v}$와 ϵ을 첫째를 k次, 뒷부분을 $n-k$ 次의 列 vector 로 分割을 하면,

$$\begin{aligned} \mathbf{v} = \begin{bmatrix} \mathbf{v}_1 \\ \mathbf{v}_2 \end{bmatrix} &= \begin{bmatrix} R \\ 0 \end{bmatrix} \beta + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \end{bmatrix} \\ &= R_T \beta + \epsilon \end{aligned} \tag{2.4}$$

으로 表示된다. 變換된 模型에 對하여 다음 두 性質을 알 수 있다.

(定理 2.4), (1) 回歸模型(2.4)는 一般的假定 卽 $\epsilon\sim N(0, \sigma^2 I_n)$이며 誤差分散 σ^2은 (2.1)과 (2.4)가 다 같다.

(2) (2.1)로부터의 β의 通常最小自乘(OLS) 推定量은 (2.4)에 의한 通常最小自乘推定量과 一致한다.

證明

1. $\epsilon=Q\eta$이므로 $\mathrm{var}(\epsilon)=E[Q\eta\eta'Q']=\sigma^2 QQ'=\sigma^2 I_n$
2. $\|\mathbf{v}-R_T\boldsymbol{\beta}\|^2=(\mathbf{y}-X\boldsymbol{\beta})'Q'Q(\mathbf{y}-X\boldsymbol{\beta})=\|\mathbf{y}-X\boldsymbol{\beta}\|^2$

이므로 두 最小自乘값은 一致한다.

이 簡單한 結果는 (2.4)에서 誘導한 推定과 (2.1)에서 誘導된 推定이 같음을 保證한다. 特히 $\epsilon_1\sim N(\mathbf{0}, \sigma^2 I_k)$, $\epsilon_2\sim N(\mathbf{0}, \sigma^2 I_{n-k})$이고 또 ϵ_1과 ϵ_2는 서로 獨立이다.

그림 A는 從屬變數 vector $\mathbf{y} = (y_1, y_2, y_3)$과 두 說明變數 vectors $\mathbf{x}_1 = (x_{11}, x_{12}, x_{13})$ 및 $\mathbf{x}_2 = (x_{21}, x_{22}, x_{23})$이며 各各 3回 測定한 값이다. Q를 곱하면 單純한 軸의 回轉이다.

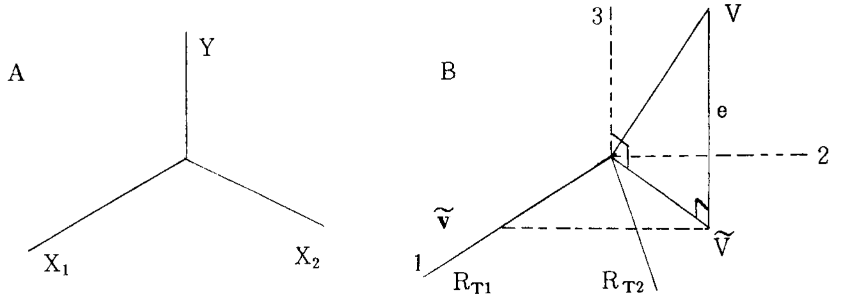

그림 A, $y = (y_1, y_2, y_3)'$, $X_1 = (x_{11}, x_{12}, x_{13})'$, $X_2 = (x_{21}, x_{22}, x_{23})'$, 그림 B, $V = (V_1, V_2, V_3)'$, $\mathbf{e} = (0, 0, V_3)'$, $\tilde{V} = (V_1, 0, 0)'$, $\hat{V} = (V_1, V_2, 0)'$, $R_{T1} = (\xi_{21}\ 0, 0)'$, $R_{T2} = (\xi_{21}, \xi_{22}, 0)'$

첫째 軸은 $\mathbf{x}_1$과 平行하고 $\mathbf{x}_2$는 첫째 軸과 둘째 軸이 決定하는 平面위에 둔다. 세째 軸은 이 平面에 垂直이다.

새 基底 $\mathbf{x}_1$은 $R_{T1} = \{\xi_{11}, 0, 0\}'$, $\mathbf{x}_2$는 $R_{T2} = \{\xi_{21}, \xi_{22}, 0\}'$ 그리고 $\mathbf{y}$는 $\mathbf{v} = (v_1\ v_2, v_3)'$으로 된다.

3. 推　　定

이제 回歸係數의 推定量 $\hat{\boldsymbol{\beta}}$의 여러 가지 性質을 調査해 보자.

(定理 3.1) $\hat{\boldsymbol{\beta}}$을 $\boldsymbol{\beta}$에 對한 OLS의 推定量이라 하고 (2.4)의 殘差 vector $\mathbf{e}$를

$$\mathbf{e} = \begin{bmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \end{bmatrix} = \mathbf{v} - R_T \hat{\boldsymbol{\beta}} \qquad (3.1)$$

으로 表示하면 다음의 結果를 얻는다.

1. $\hat{\boldsymbol{\beta}} = R^{-1}\mathbf{v}_1$ 이고 $\hat{\boldsymbol{\beta}} \sim N(\boldsymbol{\beta}, \sigma^2 (R'R)^{-1})$이다.

2. $\mathbf{c}$가 零아닌 $k \times 1$ vector 이면, $\mathbf{c}'\hat{\boldsymbol{\beta}}$는 $\mathbf{c}'\boldsymbol{\beta}$에 對한 最小分散線型不偏推定量이다. ($\eta$에 對한 正規性의 假定은 必要가 없다)

3. $\mathbf{e} = (\mathbf{0}', \mathbf{e}_2')'$: 誤差自乘合(ESS)은 $ESS = \mathbf{e}'\mathbf{e} = \epsilon_2'\epsilon_2$에 의하여 주어진다. 그리고 $ESS/\sigma^2 \sim \chi^2(n-k)$ 이고 $\hat{\boldsymbol{\beta}}$와 獨立이다. 단 $\chi^2(n)$은 自由度 n의 中心 χ^2 分布이다.

4. $S^2 = ESS/(n-k)$는 σ^2의 不偏推定量이고 $\hat{\boldsymbol{\beta}}$와 獨立이다.

證明

1. $ESS = \| \mathbf{v} - R_T \beta \|^2 = (\mathbf{v}_1 - R\boldsymbol{\beta})' \ (\mathbf{v}_1 - R\boldsymbol{\beta}) + \mathbf{v}_2'\mathbf{v}_2$ (3.2)

으므로 $\beta = R^{-1}\ _1$일 때 (3.2)는 最小가 된다. 따라서 $\mathbf{e}_1 = \mathbf{v}_1 - R\hat{\boldsymbol{\beta}} = 0$

$$\hat{\boldsymbol{\beta}} = R^{-1}\mathbf{v}_1 = R^{-1}(R\boldsymbol{\beta} + \epsilon_1) = \boldsymbol{\beta} + R^{-1}\epsilon_1 \qquad (3.3)$$

따라서 $\hat{\boldsymbol{\beta}} \sim N(\beta, \sigma^2(R'R)^{-1})$이며 $\mathbf{e}_1 = \mathbf{0}$, $\mathbf{e}_2 = \epsilon_2$ 이다.

2. 任意의 線型推定量은 $g_1'\mathbf{v}_1 + g_2'\mathbf{v}_2 = g_1'(R\beta + \epsilon_1) + g_2'\epsilon_2$ 으로 쓸 수 있다.

g_1, g_2는 各各 $k \times 1$, $(n-k) \times 1$의 定 vector 이다.

$E[g'(R\beta + \epsilon_1) + g_2'\epsilon_2] = g_1' R\beta$에서 不偏性을 充足시키려면 $g_1' = \mathbf{c}'R^{-1}$가 必要條件이다. 따라서 $g'\mathbf{v}_1 = \mathbf{c}'\boldsymbol{\beta}$이다. 또 $var(\mathbf{c}'\hat{\boldsymbol{\beta}} + g_2'\epsilon_2) = var(\mathbf{c}'\hat{\boldsymbol{\beta}}) + g_2' g_2 \sigma^2$

여기서 $g_2 = 0$일 때 最小가 된다.

萬一 $E\eta = 0$, $var(\eta) = \sigma^2 I$이면 $E\epsilon = 0$, $var(\epsilon) = \sigma^2 I$이므로 앞 結果에서 正規性은 必要하지 않는다.

3. 1에서 $ESS = \epsilon_2'\epsilon_2 \sim \sigma^2\chi^2(n-k)$, 定義로부터 χ^2은 r.v.이고 $\hat{\boldsymbol{\beta}}$은 ϵ_1만의 函數이므로 $\hat{\beta}$과 ESS 는 서로 獨立이다.

4. 3에서 明白하다.

直交(Q)回轉變換은 說明變數의 첫째 **k**行을 分離시키고, 나머지 行의 從屬變數는 $\boldsymbol{\beta}$와 關係없이 誤差項과 같다.

問題 (2.1)과 變換後의 問題(2.4)는 一致함을 다음과 같이 쉽게 알 수 있다.

$R'R = R_T'R_T = R_T'Q'QR_T = X'X$, 그리고 $R'\mathbf{v}_1 = R_T'\mathbf{v} = R_T'Q'Q\mathbf{v} = X'\mathbf{y}$이므로

$$\hat{\boldsymbol{\beta}} = R^{-1}\mathbf{v}_1 = (R'R)^{-1}R'\mathbf{v}_1 = (X'X)^{-1}X'\mathbf{y}$$

또, $var(\hat{\boldsymbol{\beta}}) = (R'R)^{-1}\sigma^2 = (X'X)^{-1}\sigma^2$

이것은 一般的인 경우의 結果이다.

回歸模型 (2.1) $\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\eta}$에 對해서 殘差分散最小의 뜻으로 X에 의한 **y**의 最良線型推推定量을

$$\mathbf{y}^* = X\hat{\boldsymbol{\beta}} \qquad (3.4)$$

이라하면 이를 回歸平面이라 하며, $\mathbf{y}^*$와 **y**間의 相關係數가 **y**와 X사이의 重相關係數이다. 이를 $\rho_{y(x)}$로 表示되면

$$\rho_{y(x)} = \frac{E[(\mathbf{y}-\overline{\mathbf{y}})(\mathbf{y}^*-\overline{\mathbf{y}}^*)]}{\sqrt{\mathrm{var}(\mathbf{y})\ \ \mathrm{var}(\mathbf{y}^*)}} = \sqrt{\boldsymbol{\lambda}'\Lambda^{-1}\boldsymbol{\lambda}/\lambda_0}$$

이다. 단 $\boldsymbol{\lambda} = E[(\mathbf{y}-E(\mathbf{y}))(\mathbf{y}^*-E(\mathbf{y}^*)]$, $E[(X-E(X))(X-E(X))'] = \Lambda$, $\mathrm{var}(\mathbf{y}) = \lambda_0$ $\overline{\mathbf{y}} = E(\mathbf{y})$, $\overline{\mathbf{y}}^* = E(\mathbf{y}^*)$.

標本資料를 $\mathbf{x}_i = (x_{i1}, \cdots\cdots, x_{in})'$, $i=1, 2, \cdots\cdots, k$, $Y = (y_1, \cdots\cdots, y_n)'$ $Z = (z_1, \cdots\cdots, z_n)'$, $X = (\mathbf{x}_1, \cdots\cdots, \mathbf{x}_k)$로 두고, Y와 X의 重相關係數와 Y와 Z의 偏相關係數를 생각해 보자.

먼저 Y의 回歸模型 $y_i = X\boldsymbol{\beta} = \beta_1 + \beta_2 x_{2i} + \cdots\cdots + \beta_k x_{ki}$, $i=1, \cdots\cdots, n$에서 $\Sigma(y_i - \beta_1 - \beta_2 x_{2i} \cdots\cdots - \beta_k x_{ki})^2$을 最小로 하는 β_i의 값을 $\hat{\beta}_i$로 두었다. $e_i = y_i - y_i^* = y_i - \hat{\beta}_1 - \hat{\beta}_2 x_{2i} - \cdots\cdots - \hat{\beta}_k x_{ki}$ 로 두면 正規方程式은 $\Sigma e_i = 0$, $\Sigma x_{2i} e_i = 0$, $\cdots\cdots$, $\Sigma x_{ki} e_i = 0$으로 주어지며 $\hat{\beta}$은 이들 正規方程式에서 求해진다. 따라서 $\Sigma y_i = \Sigma y_i^*$, $\overline{y} = \overline{y}^*$, $\Sigma e_i \overline{y} = 0$이다. 이로부터

$$\Sigma(y_i - y_i^*) y_i^* = \Sigma e_i (\hat{\beta}_1 + \hat{\beta}_2 x_{2i} + \cdots\cdots + \hat{\beta}_k x_{ki}) = 0$$

이다. 그러므로 다음 式을 얻는다.

$$\begin{aligned}\Sigma(y_i - \overline{y})(y_i^* - \overline{y}) &= \Sigma(y_i - y_i^* + y_i^* - \overline{y})(y_i^* - \overline{y}) \\ &= \Sigma(y_i - y_i^*)(y_i^* - \overline{y}) + \Sigma(y_i^* - \overline{y})^2 = \Sigma(y_i^* - \overline{y})^2\end{aligned}$$

y의 總自乘合을 $TSS = \Sigma(y_i - \overline{y})^2$, 殘差 卽 誤差의 自乘合을 $ESS = \Sigma(y_i - y_i^*)^2$, 回歸에 의한 自乘合을 $SSR = \Sigma(y_i^* - \overline{y})^2$ 으로 表示한다.

$$\begin{aligned}\Sigma(y_i - \overline{y})^2 &= \Sigma(y_i - y_i^* + y_i^* - \overline{y})^2 \\ &= \Sigma(y_i - y_i^*)^2 + \Sigma(y_i^* - \overline{y})^2 + 2\Sigma(y_i - y_i^*)(y_i^* - \overline{y}) \\ &= \Sigma(y_i - y_i^*)^2 + \Sigma(y_i^* - \overline{y})^2\end{aligned}$$

따라서 $TSS = ESS + SSR$ 이다.

(定理 3.2) 標本資料 (X, Y)에서 Y와 X의 重相關係數는

$$r_{y(x)} = 1 - \frac{ESS}{TSS} = 1 - \frac{\mathbf{v}_2'\mathbf{v}_2}{\sum_2^n \mathbf{v}_i^2}$$

이고 $TSS = \sum_2^n \mathbf{v}_i^2$, $\mathbf{v}_1 = \sqrt{n}\ \overline{y}$ 이다.

證明 (定理 3.1)에서 $ESS = (\mathbf{v}_1 - R\beta)'(\mathbf{v}_1 - R\beta) + \mathbf{v}_2'\mathbf{v}_2$는 $\hat{\beta} = R^{-1}\mathbf{v}_1$일 때 最小가

가 된다. 그러므로 $ESS = \mathbf{v}_2'\mathbf{v}_2 = \Sigma(y_i - y_i^*)^2$이며, 다음을 얻는다.

$$r^2_{y(x)} = \frac{[\Sigma(y_i - \bar{y})(y_i^* - \bar{y})]^2}{\mathrm{var}(y)\ \mathrm{var}(y^*)} = \frac{[\Sigma(y_i^* - \bar{y})^2]^2}{\Sigma(y_i - \bar{y})^2\,\Sigma(y_i^* - \bar{y})^2}$$

$$= \frac{\Sigma(y_i^* - \bar{y})^2}{\Sigma(y_i - \bar{y})^2} = 1 - \frac{ESS}{TSS}$$

直交變換Q의 構成에서 第1行의 成分은 모두 $\frac{1}{\sqrt{n}}$이다. 따라서 $TSS = \sum_{2}^{n} v_i^2$, $v_1 = \sqrt{n}\ \bar{y}$ 이다.

(定理 3.3) 標本資料(X, Y, Z)에서, X에 관한 Y, Z 의 偏相關係數 $r_{yz(x)}$ 는

$$r^2_{yz(x)} = \frac{(\mathbf{v}_2'\mathbf{w}_2)^2}{ESS(y)\ ESS(z)} = \frac{(\mathbf{w}_2'\mathbf{w}_2)^2}{(\mathbf{v}_2'\mathbf{v}_2)(\mathbf{w}_2'\mathbf{w}_2)}$$

이다. 단 $\mathbf{v}_2 = Q(Y - X\hat{\beta})$, $\hat{\beta} = R^{-1}\mathbf{v}_1$, $\mathbf{w}_2 = Q(Z - X\hat{\alpha})$, $\hat{\alpha} = R^{-1}\mathbf{w}_1$

證明, $ESS(\mathbf{y}) = n\,\mathrm{var}(\mathbf{y} - \mathbf{y}^*) = \Sigma(y_i - y_i^*)^2$

$$= (Y - X\hat{\beta})'(Y - X\hat{\beta}) = (Y - X\hat{\beta})'Q'Q(Y - X\hat{\beta})$$

$$= \begin{bmatrix} \mathbf{v}_1 - R\hat{\beta} \\ \mathbf{v}_2 \end{bmatrix}' \begin{bmatrix} \mathbf{v}_1 - R\hat{\beta} \\ \mathbf{v}_2 \end{bmatrix} = \begin{bmatrix} 0 \\ \mathbf{v}_2 \end{bmatrix}' \begin{bmatrix} 0 \\ \mathbf{v}_2 \end{bmatrix} = \mathbf{v}_2'\mathbf{v}_2$$

$Q\mathbf{z} = \mathbf{w} = \begin{bmatrix} \mathbf{w}_1 \\ \mathbf{w}_2 \end{bmatrix}$ 으로 두면

$$ESS(\mathbf{z}) = n\ \mathrm{var}(\mathbf{z} - \mathbf{z}^*) = (Z - X\hat{\alpha})'(Z - X\hat{\alpha}) = \begin{bmatrix} \mathbf{w}_1 - R\hat{\alpha} \\ \mathbf{w}_2 \end{bmatrix}' \begin{bmatrix} \mathbf{w}_1 - R\hat{\alpha} \\ \mathbf{w}_2 \end{bmatrix}$$

$$= \begin{bmatrix} \mathbf{0} \\ \mathbf{w}_2 \end{bmatrix}' \begin{bmatrix} \mathbf{0} \\ \mathbf{w}_2 \end{bmatrix} = \mathbf{w}_2'\mathbf{w}_2, \quad \text{단 } \hat{\alpha}' = R^{-1}\mathbf{w}_1$$

$$\Sigma(y_i - y_i^*)(z_i - z_i^*) = (Y - X\hat{\beta})'(Z - X\hat{\alpha}) = \begin{bmatrix} \mathbf{0} \\ \mathbf{v}_2 \end{bmatrix}' \begin{bmatrix} \mathbf{0} \\ \mathbf{w}_2 \end{bmatrix} = \mathbf{v}_2'\mathbf{w}_2$$

위의 結果를 偏相關係數의 定義式에 넣으면,

$$r^2_{yz(x)} = \frac{\mathbf{v}_2'\mathbf{w}_2}{(\mathbf{v}_2'\mathbf{v}_2)(\mathbf{w}_2'\mathbf{w}_2)}$$

이다.

以上에서 定理 2.3 의 直交變換 Q에 의하여 重相關係數와 偏相關係數를 셈할 수 있다.

例 다음과 같은 標本資料

$$(X, Y, Z) = (X_0, X_1, X_2, Y, Z) = \begin{bmatrix} 1 & 1 & 2 & 2.99 & 3 \\ 1 & 2 & 4 & 5.69 & 6 \\ 1 & 3 & 5 & 7.02 & 8 \\ 1 & 4 & 7 & 9.54 & 10 \\ 1 & 5 & 8 & 9.28 & 11 \end{bmatrix}$$

에서 回歸函數를

$$\begin{aligned} y_i &= \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \eta_i, & \eta_i &\sim i,i,d,N(0,1) \\ Z_i &= \alpha_0 + \alpha_1 x_{1i} + \alpha_2 x_{2i} + \epsilon_1, & \epsilon_i &\sim i,i,d,N(0,1) \end{aligned} \quad i = 1, \cdots\cdots, 5,$$

이라 하고 回歸係數 $\hat{\alpha}$, $\hat{\beta}$와 重相關係數 및 偏相關係數를 求해보자. 먼저 定理 2.2에 의하여 直交行列 Q를 셈하면

$$Q = \begin{bmatrix} 0.4472 & 0.4472 & 0.4472 & 0.4472 & 0.4472 \\ -0.6325 & -0.3162 & 0.0000 & 0.3162 & 0.6325 \\ -0.3650 & 0.5477 & -0.3655 & 0.5477 & -0.3649 \\ 0.3982 & -0.6171 & -0.1793 & 0.6170 & -0.2188 \\ 0.3290 & 0.1386 & -0.7964 & -0.1391 & 0.4679 \end{bmatrix}$$

이고 $Qy = \mathbf{v}$, $Q\mathbf{z} = \mathbf{w}$, $QX = R_T$는 다음과 같다.

$$\mathbf{v} = \begin{bmatrix} 15.438 \\ 5.196 \\ 1.298 \\ 0.276 \\ -0.804 \end{bmatrix} \quad \mathbf{w} = \begin{bmatrix} 16.994 \\ 6.325 \\ 0.730 \\ -0.179 \\ -0.797 \end{bmatrix} \quad R_T = \begin{bmatrix} 2.236 & 6.708 & 11.627 \\ 0 & 3.162 & 4.744 \\ 0 & 0 & 0.548 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

y의 $TSS = 5.196^2 + 1.298^2 + 0.276^2 + (-0.804)^2 = 29.40$

y의 $ESS = 0.276^2 + (-0.804)^2 = 0.723$, $S_y^2 = ESS/2 = 0.3615$ $\quad S_y = 0.601$

z 의 $TSS = 6.325^2 + 0.730^2 + (-0.179)^2 + (-0.797)^2 = 41.20$

z 의 $ESS = (-0.179)^2 + (-0.797)^2 = 0.667$, $S_z^2 = 0.677/2 = 0.3335$ $S_z = 0.577$

$$\mathbf{v}_2\mathbf{w}_2 = 0.276 \times (-0.179) + (-0.804) \times (-0.797) = 0.591$$

Y 와 X 의 重相關係數

$$r_{y\,(x)}^2 = 1 - \frac{ESS}{TSS} = 1 - \frac{0.723}{29.40} = 0.975.$$

Z 와 X 의 重相關係數 $r_{z\,(x)}^2$

$$r_{z\,(x)}^2 = 1 - \frac{0.667}{41.20} = 0.984$$

또 Y 와 X, Z 와 X의 重相關係數를 公式에 의하여 셈하면 $r_{y\,(x)}^2 = \lambda' \Sigma^{-1} \lambda / var(y) = 0.975$ $r_{z\,(x)}^2 = \xi' \Sigma^{-1} \xi / var(z) = 0.984$, 단 $\lambda = (X - \overline{X})'(Y - \overline{Y})$, $\xi = (X - \overline{X})'(Z - \overline{Z})$. $\Sigma = X'X$,

$\hat{\alpha} = R^{-1}\mathbf{w}_1$에서 $\hat{\alpha}_3 = 0.7245/0.548 = 1.3221$

$\hat{\alpha}_2 = (6.325 - 4.744 \times 1.3221)/\ 3.162 = 0.0185$

$\hat{\alpha}_1 = (16.9941 - 11.627 \times 1.3221 - 6.708 \times 0.0185)/2.326 = 0.6704$

$$R'R = S = X'X = \begin{bmatrix} 5 & 15 & 26 \\ 15 & 55 & 93 \\ 26 & 93 & 158 \end{bmatrix}, \quad (R'R)^{-1} = \begin{bmatrix} 2.733 & 3.2 & -2.333 \\ 3.2 & 7.6 & -5 \\ -2.333 & -5 & 3.333 \end{bmatrix}$$

$$var(\hat{\alpha}) = S_z^2 (R'R)^{-1} = \begin{bmatrix} 0.9116 & 1.0672 & -0.7782 \\ 1.0672 & 2.5346 & -1.6675 \\ -0.7782 & -1.6675 & 1.1117 \end{bmatrix}$$

$\hat{\beta} = R^{-1}\mathbf{v}_1$ 에서 $\hat{\beta}_3 = 1.2974/0.548 = 2.368$

$\hat{\beta}_2 = (5.196 - 4.743 \times 2.369)/\ 3.162 = -1.910$

$\hat{\beta}_1 = [15.438 - 11.628 \times 2.369 - 6.708 \times (-1.910)]/2.236 = 0.315$

$$var(\hat{\beta}) = S_y^2 (R'R)^{-1} = \begin{bmatrix} 0.912 & 1.067 & -0.778 \\ 1.067 & 2.535 & -1.668 \\ -0.778 & -1.668 & 1.112 \end{bmatrix}$$

$$R^{-1} = \begin{bmatrix} 0.4772 & -0.9487 & -1.2789 \\ 0 & 0.3162 & -2.7388 \\ 0 & 0 & 1.8258 \end{bmatrix}$$

參 考 文 獻

1. Graig F. Ansley (1985) Quick Proofs of Some Regression Theorems via the QR Algorithm. The American Statistician vol. 39, no 1.
2. Graybill F. A. (1961) Introduction to Matrices with Application in Statistics. Wadsworth Publishing Co.
3. Graybill F. A. (1976) Theory and Application of the Linear Model North Scituate MA: Duxbury Press.
4. Searle. S. R. (1971) Linear Models. New York, John Wiley.
5. 朴聖炫, (1981) 回歸分析, 大英社.
6. 北川敏男, (1966) 多變量解析, 共立社.
7. 中村正一, (1979) 例解多變量解析入門 日刊工業新聞社.