BAB 10

VARIABEL KANONIK DAN KORELASI KANONIK

10.1 Variabel Kanonik dan Korelasi Kanonik

Kita akan tertarik dalam mengukur dari kumpulan antara dua kelompok variabel. Kelompok pertama dari p variabel diwakili oleh (p x 1) vektor acak $X^{(1)}$. Kelompok kedua dari q variabel diwakili oleh (q x 1) vektor acak $X^{(2)}$. Kita asumsi, dalam pengembangan teoritis, bahwa $X^{(1)}$ mewakili himpunan yang lebih kecil, sehingga $p \leq q$.

Misalkan untuk vektor acak $X^{(1)}$ dan $X^{(2)}$:

$$
\begin{array}{ll}
E\left(X^{(1)}\right) = \mu^{(1)}; & Cov\left(X^{(1)}\right) = \sum_{11} \\
E\left(X^{(2)}\right) = \mu^{(2)}; & Cov\left(X^{(2)}\right) = \sum_{22} \\
Cov\left(X^{(1)}, X^{(2)}\right) = \sum_{22} = \sum_{21}^{'} &
\end{array}
\tag{10-1}
$$

Vektor acaknya :

$$
\underset{((p+q)x1)}{X} = \begin{bmatrix} X^{(1)} \\ \hdashline X^{(2)} \end{bmatrix} = \begin{bmatrix} X_1^{(1)} \\ X_2^{(1)} \\ \vdots \\ X_p^{(1)} \\ \hdashline X_1^{(2)} \\ X_1^{(2)} \\ \vdots \\ X_q^{(2)} \end{bmatrix}
\tag{10-2}
$$

Vektor rata-ratanya :

$$
\underset{((p+q)x1)}{\mu} = E(X) = \begin{bmatrix} E\left(X^{(1)}\right) \\ \hdashline E\left(X^{(2)}\right) \end{bmatrix} = \begin{bmatrix} \mu^{(1)} \\ \hdashline \mu^{(2)} \end{bmatrix}
\tag{10-3}
$$

Dan matriks kovariannya : $\underset{(p+q)(p+q)}{\sum} = E(X-\mu)(X-\mu)'$

$$\sum_{(p+q)\times(p+q)} = \left[\begin{array}{c|c} E\left(X^{(1)}-\mu^{(1)}\right)\left(X^{(1)}-\mu^{(1)}\right)' & E\left(X^{(1)}-\mu^{(1)}\right)\left(X^{(2)}-\mu^{(2)}\right)' \\ \hline E\left(X^{(2)}-\mu^{(2)}\right)\left(X^{(1)}-\mu^{(1)}\right)' & E\left(X^{(2)}-\mu^{(2)}\right)\left(X^{(2)}-\mu^{(2)}\right)' \end{array}\right] \qquad \textbf{(10-4)}$$

Kovarian antara pasangan variabel-variabel dari himpunan berbeda yaitu satu variabel dari $X^{(1)}$, satu variabel dari $X^{(2)}$ yang termuat di Σ_{12} atau ekuivalen di Σ_{21}. pq elemen dari Σ_{12} mengukur kumpulan antara dua himpunan. Ketika p dan q relatif besar, menginterpretasikan elemen dari Σ_{12} secara bersamaan biasanya adalah percuma. Selain itu, sering bahwa kombinasi linear dari variabel itu menarik dan berguna untuk memprediksi atau membandingkan tujuan. Tugas pokok dari analisis korelasi kanonik adalah meringkaskan kumpulan antara himpunan $X^{(1)}$ dan $X^{(2)}$ dalam syarat-syarat yang sedikit berhati-hati memilih kovarian (atau korelasi) daripada kovarian pq di Σ_{12}. Kombinasi linear menyediakan ringkasan sederhana mengukur suatu himpunan dari variabel. Himpunan

$$\begin{gathered} U = a'X^{(1)} \\ \text{dan} \\ V = b'X^{(2)} \end{gathered} \qquad \textbf{(10-5)}$$

Untuk beberapa bagian dari koefisien vektor a dan b. Dengan menggunakan (10-5) dan kombinasi linear Z = CX dimana,

$$\begin{aligned} \mu_Z &= E(Z) = E(CX) = C\mu_X \\ \sum{}_Z &= Cov(Z) = Cov(CX) = C\sum{}_X C' \end{aligned}$$

Sehingga,

$$\begin{aligned} Var(U) &= a'Cov\left(X^{(1)}\right)a = a'\sum{}_{11} a \\ Var(V) &= b'Cov\left(X^{(2)}\right)b = b'\sum{}_{22} b \\ Cov(U,V) &= a'Cov\left(X^{(1)}, X^{(2)}\right)b = a'\sum{}_{12} b \end{aligned} \qquad \textbf{(10-6)}$$

Kemudian dapat dicari koefisien vektor a dan b sedemikian sehingga,

$$Corr(U,V) = \frac{a'\sum_{12} b}{\sqrt{a'\sum_{11} a}\sqrt{b'\sum_{22} b}} \qquad \textbf{(10-7)}$$

sebisa mungkin bernilai besar.

Definisi:

Bagian *pertama pasangan dari variabel kanonik* adalah bagian dari kombinasi linear U_1, V_1 yang mempunyai unit variansi, yang memaksimalkan korelasi (10-7); Bagian kedua dari variabel kanonik adalah bagian dari kombinasi linear U_2, V_2 yang mempunyai unit variansi, yang memaksimalkan korelasi (10-7) diantara semua pilihan yang tidak berkorelasi dengan bagian pertama dari variabel kanonik.

Pada langkah ke-k:

Bagian *ke-k pasangan dari variabel kanonik* adalah bagian dari kombinasi linear U_k, V_k yang mempunyai unit variansi, yang memaksimalkan korelasi (10-7) diantara semua pilihan yang tidak berkorelasi dengan bagian k-1 sebelumnya dari pasangan variabel kanonik.

Korelasi antara bagian ke-k dari variabel kanonik dinamakan korelasi kanonik ke-k.

Akibat 10.1. Misalkan p ≤ q dan vektor acak $X^{(1)}$ dan $X^{(2)}$ mempunyai, $Cov(X^{(1)}) = \underset{(pxp)}{\sum_{11}}$, $Cov(X^{(2)}) = \underset{(qxq)}{\sum_{22}}$ dan $Cov(X^{(1)}, X^{(2)}) = \underset{(pxq)}{\sum_{12}}$ dimana $\sum$ mempunyai rank lengkap. Untuk koefisien vector $\underset{(px1)}{a}$ dan $\underset{(qx1)}{b}$, bentuk kombinasi linear U = a'$X^{(1)}$ dan V = b'$X^{(2)}$. Maka $\max_{a,b} Corr(U,V) = \rho_1^*$ diperoleh dengan kombinasi linear (variabel kanonik bagian pertama).

$U_1 = e_1' \sum_{11}^{-1/2} X^{(1)}$ dan $V_1 = f_1' \sum_{22}^{-1/2} X^{(2)}$, Bagian ke-k dari variabel kanonik, k = 2, 3, ..., p, $U_k = e_k^{'} \sum_{11}^{-1/2} X^{(1)}$ dan $V_k = f_k^{'} \sum_{22}^{-1/2} X^{(2)}$ memaksimumkan $Corr(U_k, V_k) = \rho_k^*$ diantara kombinasi linear yang tidak berkorelasi dengan variabel kanonik 1, 2, ..., k-1 sebelumnya. $\rho_1^{*2} \geq \rho_2^{*2} \geq \ldots \geq \rho_p^{*2}$ adalah nilai eigen dari $\sum_{11}^{-1/2} \sum_{12} \sum_{22}^{-1} \sum_{21} \sum_{11}^{-1/2}$ dan $e_1, e_2, ..., e_p$ adalah vektor eigen (p x 1). (Jumlah $\rho_1^{*2} \geq \rho_2^{*2} \geq \ldots \geq \rho_p^{*2}$ juga nilai eigen p paling besar dari matriks $\sum_{11}^{-1/2} \sum_{12} \sum_{22}^{-1} \sum_{21} \sum_{11}^{-1/2}$ yang bersesuaian dengan vektor eigen (q x 1), f_1, f_2, ..., f_p. Tiap f_i adalah proporsi untuk $\sum_{11}^{-1/2} \sum_{12} \sum_{22}^{-1} \sum_{21} \sum_{11}^{-1/2} e_i$). Variasi kanonik mempunyai sifat sebagai berikut:

$$
\begin{aligned}
& Var(U_k) = Var(V_k) = 1 & \\
& Cov(U_k, U_l) = Corr(U_k, U_l) = 0 & k \neq l \\
& Cov(V_k, V_l) = Corr(V_k, V_l) = 0 & k \neq l \\
& Cov(U_k, V_l) = Corr(U_k, V_l) = 0 & k \neq l
\end{aligned}
$$

untuk k, l = 1, 2, ..., p.

Jika variabel awal distandardisasikan dengan $Z^{(1)} = [Z_1^{(1)}, Z_2^{(1)}, \ldots, Z_p^{(1)}]^T$ dan $Z^{(2)} = [Z_1^{(2)}, Z_2^{(2)}, \ldots, Z_q^{(2)}]^T$ maka variabel kanonik berbentuk:

$$
\begin{aligned}
U_k &= a_k^{'} Z^{(1)} = e_k^{'} \rho_{11}^{-1/2} Z^{(1)} \\
V_k &= b_k^{'} Z^{(2)} = f_k^{'} \rho_{22}^{-1/2} Z^{(2)}
\end{aligned}
\qquad \textbf{(10-8)}
$$

Disini $Cov(Z^{(1)}) = \rho_{11}, Cov(Z^{(2)}) = \rho_{22}, Cov(Z^{(1)}, Z^{(2)}) = \rho_{12} = \rho_{21}$ dan e_k dan f_k adalah vektor-vektor eigen dari $\rho_{11}^{-1/2} \rho_{12} \rho_{22}^{-1} \rho_{21} \rho_{11}^{-1/2}$ dan $\rho_{22}^{-1/2} \rho_{21} \rho_{11}^{-1} \rho_{12} \rho_{22}^{-1/2}$ secara berurut. Korelasi kanonik ρ_k^* memenuhi,

$$
Corr(U_k, V_k) = \rho_k^*, \text{dimana } k = 1,2,..., p \qquad \textbf{(10-9)}
$$

dan $\rho_1^{*2} \geq \rho_2^{*2} \geq \ldots \geq \rho_p^{*2}$ adalah vektor eigen tak nol dari matriks $\rho_{11}^{-1/2} \rho_{12} \rho_{22}^{-1} \rho_{21} \rho_{11}^{-1/2}$ atau matriks $\rho_{22}^{-1/2} \rho_{21} \rho_{11}^{-1} \rho_{12} \rho_{22}^{-1/2}$.