10.4 Variasi Kanonik Sampel Dan Korelasi Kanonik Sampel

Sampel acak dari *n* observasi pada masing-masing variabel dari *(p + q)* variabel $X^{(1)}$, $X^{(2)}$ dapat digabungkan kedalam *((p + q) x n)* data matriks

$$X = \begin{bmatrix} X^{(1)} \\ \hdashline X^{(2)} \end{bmatrix} = \begin{bmatrix} x_{11}^{(1)} & x_{12}^{(1)} & \cdots & x_{1n}^{(1)} \\ x_{21}^{(1)} & x_{22}^{(1)} & \cdots & x_{2n}^{(1)} \\ \vdots & \vdots & \ddots & \vdots \\ x_{p1}^{(1)} & x_{p2}^{(1)} & \cdots & x_{pn}^{(1)} \\ \hdashline x_{11}^{(2)} & x_{12}^{(2)} & \cdots & x_{1n}^{(2)} \\ x_{21}^{(2)} & x_{22}^{(2)} & \cdots & x_{2n}^{(2)} \\ \vdots & \vdots & \ddots & \vdots \\ x_{p1}^{(2)} & x_{p2}^{(2)} & \cdots & x_{pn}^{(2)} \end{bmatrix} = \left[x_1, x_2, \ldots, x_n \right]$$

dimana $x_j = \begin{bmatrix} x_j^{(1)} \\ \hdashline x_j^{(2)} \end{bmatrix}$ **(10-25)**

Adapun vektor rata-rata sampelnya adalah

$$\underset{(p+q)\times 1}{\bar{x}} = \begin{bmatrix} \bar{x}^{(1)} \\ \hdashline \bar{x}^{(2)} \end{bmatrix}$$

dimana $\bar{x}^{(1)} = \frac{1}{n}\sum_{j=1}^{n} x_j^{(1)}$ dan $\bar{x}^{(2)} = \frac{1}{n}\sum_{j=1}^{n} x_j^{(2)}$ **(10-26)**

Dan matriks kovarian sampel dapat ditulis

$$\underset{(p+q)x(p+q)}{S} = \begin{bmatrix} \underset{(pxp)}{S_{11}} & \vdots & \underset{(pxq)}{S_{12}} \\ \cdots & & \cdots \\ \underset{(qxp)}{S_{21}} & \vdots & \underset{(qxq)}{S_{22}} \end{bmatrix} \quad \text{dimana} \quad S_{kl} = \frac{1}{n-1}\sum_{j=1}^{n}\left(x_j^{(k)} - \bar{x}^{(k)}\right)\left(x_j^{(l)} - \bar{x}^{(l)}\right)',$$

k,l = 1, 2 **(10-27)**

Kombinasi linear $\hat{U} = \hat{a}' x^{(1)}, \hat{V} = \hat{b}' x^{(2)}$ **(10-28)**

memiliki korelasi sampel $$r_{\hat{U},\hat{V}} = \frac{\hat{a}' S_{12}\, \hat{b}}{\sqrt{\hat{a}' S_{12}\, \hat{a}}\sqrt{\hat{b}' S_{12}\, \hat{b}}}$$ **(10-29)**

Pasangan pertama dari variasi kanonik sampel dalam kombinasi linear $\hat{U}_1$ dan $\hat{V}_1$ memiliki unit varian sampel yang memaksimumkan rasio (2-21).

Pada umumnya, *pasangan ke-k variasi kanonik sampel* adalah pasangan dari kombinasi linear $\hat{U}_k$ dan $\hat{V}_k$ yang memiliki unit varian sampel yang memaksimumkan (10-29) diantara kombinasi linear yang tidak berkorelasi dengan *k*-1 variasi kanonik sampel yang sebelumnya.

Korelasi sampel antara $\hat{U}_k$ dan $\hat{V}_k$ dinamakan korelasi kanonik sampel. Variasi sampel kanonik dan korelasi kanonik sampel dapat diperoleh dari matriks kovarian sampel S_{11}, $S_{12} = S_{21}'$, dan S_{22} dengan cara yang bersesuaian dengan persoalan yang dibahas dalam 2.1.

Akibat 10.2. Misalkan $\rho_1^{*2} \geq \rho_2^{*2} \geq \ldots \geq \rho_p^{*2}$ adalah *p* order nilai eigen dari $S_{11}^{-1/2} S_{12} S_{22}^{-1} S_{21} S_{11}^{-1/2}$ vektor eigen yang berkoresponden dengan $\hat{e}_1, \hat{e}_2, \ldots, \hat{e}_p$ dimana S_{kl} didefinisikan pada (2-19) dan $p \leq q$. Misalkan $\hat{f}_1, \hat{f}_2, \ldots, \hat{f}_q$ menjadi vektor eigen dari $S_{22}^{-1/2}S_{21}S_{11}^{-1}S_{12}S_{22}^{-1/2}$ dimana p yang pertama $\hat{f}$'s diperoleh dari $\hat{f}_k = \left(\frac{1}{\hat{\rho}_k^*}\right) S_{22}^{-1/2}S_{21}S_{11}^{-1/2}S\,\hat{e}_k, k = 1,2,\ldots,p.$ Pasangan variasi kanonik sampel ke-k adalah $\hat{U}_k = \underbrace{\hat{e}_k' S_{11}^{-1/2}}_{\hat{a}_k'} x^{(1)}$ dan $\hat{V}_k = \underbrace{\hat{f}_k' S_{22}^{-1/2}}_{\hat{b}_k'} x^{(2)}$ dimana $x^{(1)}$ dan $x^{(2)}$ adalah nilai variabel dari $X^{(1)}$ dan $X^{(2)}$ untuk unit ekperimen khusus. Variasi kanonik sampel pertama mempunyai korelasi sampel maksimum $r_{\hat{U}_1,\hat{V}_1} = \hat{\rho}_1^*$.

Untuk pasangan ke-k $r_{\hat{U}_k,\hat{V}_k} = \hat{\rho}_k^*$ dan korelasi ini merupakan kemungkinan terbesar diantara kombinasi linear yang tidak berkorelasi dengan k-1 variasi kanonik sampel sebelumnya. Jumlah $\hat{\rho}_1^*, \hat{\rho}_2^*, \ldots, \hat{\rho}_p^*$ adalah korelasi kanonik sampel. Jika $p > rank(S_{12}) = p_1$, maka korelasi kanonik sampel tak nol adalah $\hat{\rho}_1^*, \hat{\rho}_2^*, \ldots, \hat{\rho}_{p1}^*$.

Bukti. Bukti dari akibat 10.2 mengikuti bukti dari akibat 10.1 dengan S_{kl} menggantikan $\sum_{kl}$, k,l=1,2.

Variabel-variabel kanonik sampel mempunyai unit variansi sampel.

$$s_{\hat{U}_k,\hat{U}_k} = s_{\hat{V}_k,\hat{v}_k} = 1 \qquad \textbf{(10-30)}$$

dan sampel korelasi

$$r_{\hat{U}_k,\hat{U}_l} = r_{\hat{V}_k,\hat{v}_l} = 0, \qquad k \neq l \qquad \textbf{(10-31)}$$

$$r_{\hat{V}_k,\hat{v}_l} = 0, \qquad k \neq l$$

Penerapan dari $\hat{U}_k, \hat{V}_k$ sering dinantu dengan perhitungan sampel korelasi antara variable kanonik dan variable dalam himpunan $X^{(1)}$, $X^{(2)}$. Kami mendefinisikan matrik-matrik

$$\underset{p\times p}{\hat{A}} = [\hat{a}_1, \hat{a}_2, \ldots, \hat{a}_p]'; \qquad \underset{q\times q}{\hat{B}} = [\hat{b}_1, \hat{b}_2, \ldots, \hat{b}_q]' \qquad \textbf{(10-32)}$$

yang baris-barisnya adalah vektor-vektor koefisien dari variabel kanonik sampel. Analogi dengan persamaan **(10-17)**

$$\underset{(px1)}{\hat{U}} = \hat{A}_z\, z^{(1)}; \qquad \underset{(qx1)}{\hat{V}} = \hat{B}_z\, z^{(2)} \tag{10-33}$$

dan kita dapat definisikan

$R_{\hat{U},X^{(1)}}$ =matrik dari sampel korelasi pada $\hat{U}$ dengan $X^{(1)}$

$R_{\hat{V},X^{(2)}}$ =matrik dari sampel korelasi pada $\hat{V}$ dengan $X^{(2)}$

$R_{\hat{U},X^{(2)}}$ =matrik dari sampel korelasi pada $\hat{U}$ dengan $X^{(2)}$

$R_{\hat{V},X^{(1)}}$ =matrik dari sampel korelasi pada $\hat{V}$ dengan $X^{(1)}$

Sesuai dengan (10-19) kita mempunyai

$$\begin{aligned}
R_{\hat{U},X^{(1)}} &= \hat{A}S_{11}D_{11}^{-1/2} \\
R_{\hat{V},X^{(2)}} &= \hat{B}S_{22}D_{22}^{-1/2} \\
R_{\hat{U},X^{(2)}} &= \hat{A}S_{12}D_{22}^{-1/2} \\
R_{\hat{V},X^{(1)}} &= \hat{B}S_{21}D_{11}^{-1/2}
\end{aligned} \tag{10-34}$$

Dimana $D_{11}^{-1/2}$ adalah matrik diagonal berukuran *(pxp)* dengan elemen diagonal ke-i (sampel $\operatorname{var}(x_i^{(1)}))^{-1/2}$ dan $D_{22}^{-1/2}$ adalah matrik diagonal berukuran *(qxq)* dengan elemen diagonal ke-i (sampel $\operatorname{var}(x_i^{(2)}))^{-1/2}$.

Keterangan:

Jika observasi distandarisasikan, maka data matriks menjadi

$$Z = \left[\begin{array}{c} Z^{(1)} \\ \hdashline Z^{(2)} \end{array} \right] = \left[z_1, z_2, \ldots, z_n \right] \text{ dengan } z_j = \left[\begin{array}{c} z_j^{(1)} \\ \hdashline z_j^{(2)} \end{array} \right]$$

dan variasi kanonik sampel menjadi :

$$\underset{(px1)}{\hat{U}} = \hat{A}_z\, z^{(1)}; \qquad \underset{(qx1)}{\hat{V}} = \hat{B}_z\, z^{(2)} \qquad \textbf{(10-36)}$$

dimana $\hat{A}_z = \hat{A} D_{11}^{1/2}$ dan $\hat{B}_z = \hat{B} D_{22}^{1/2}$. Korelasi kanonik sampel tidak efektif dengan standardisasi. Sebagai catatan bahwa $D_{11}^{-1/2} = D_{22}^{-1/2} = I$ untuk observasi standard.

Contoh 10.4

Pada contoh 9.14, data terdiri dari ukuran-ukuran tulang dan tengkorak dari ayam White Leghorn. Dari contoh ini, ukuran tulang anak ayam untuk:

Kepala($X^{(1)}$): $\begin{cases} X_1^{(1)} = \text{panjang tengkorak} \\ X_2^{(1)} = \text{lebar tengkorak} \end{cases}$

Kaki ($X^{(2)}$): $\begin{cases} X_1^{(1)} = \text{panjang tulang paha} \\ X_2^{(2)} = \text{panjang tulang betis} \end{cases}$

mempunyai matrik sampel korelasi

$$R = \left[\begin{array}{c:c} R_{11} & R_{12} \\ \hdashline R_{21} & R_{22} \end{array}\right] = \left[\begin{array}{cc:cc} 1.0 & .505 & .569 & .602 \\ .505 & 1.0 & .422 & .467 \\ \hdashline .569 & .422 & 1.0 & .926 \\ .602 & .467 & .926 & 1.0 \end{array}\right]$$

Analisis korelasi kanonik dari himpunan variabel kepala dan leg menggunakan R menghasilkan dua korelasi kanonik dan pasangan variabel yang bersesuaian:

$$\hat{\rho}_1^* = 0.631 \qquad \begin{array}{l} \hat{U}_1 = 0.781\, z_1^{(1)} + 0.345\, z_2^{(1)} \\ \hat{V}_1 = 0.060\, z_1^{(2)} + 0.944\, z_2^{(2)} \end{array}$$

$$\hat{\rho}_2^* = 0.057 \qquad \begin{array}{l} \hat{U}_2 = -0.856\, z_1^{(1)} + 1.106\, z_2^{(1)} \\ \hat{V}_2 = -2.648\, z_1^{(2)} + 2.475\, z_2^{(2)} \end{array}$$

Disini $z_i^{(1)}$, i=1,2 dan $z_i^{(2)}$, i=1,2 adalah nilai-nilai terstandarisasi untuk himpunan 1 dan 2,respectively. Hasil di atas diperoleh dari software statistika SAS.

Contoh 10.5

Sebagai bagian dari studi besar dari pengaruh struktur organisasi pada "kepuasan pekerjaan", Dunham[3] menyelidiki kadar hubungan kepuasan pekerjaan dengan karakteristik pekerjaan.

Dengan menggunakan instrumen penelitian, Dunham obtained ukuran-ukuran dari p=5 karakteristik pekerjaan dan q=7 variabel kepuasan pekerjaan untuk n=784 eksekutif dari cabang perusahaan dagang besar.

Apakah ukuran-ukuran kepuasan pekerjaan berasosiasi dengan karakteristik pekerjaan?jawabannya mungkin dapat diperoleh desain pekerjaan. Variabel-variabel karakteristik pekerjaan asli, $X^{(1)}$, dan variabel-variabel kepuasan pekerjaan, $X^{(2)}$, didefinisikan sebagai:

$$\mathbf{X}^{(1)} = \begin{bmatrix} X_1^{(1)} \\ X_2^{(1)} \\ X_3^{(1)} \\ X_4^{(1)} \\ X_5^{(1)} \end{bmatrix} = \begin{bmatrix} \text{feedback} \\ \text{task significance} \\ \text{task variety} \\ \text{task identity} \\ \text{autonomy} \end{bmatrix}$$

$$\mathbf{X}^{(2)} = \begin{bmatrix} X_1^{(2)} \\ X_2^{(2)} \\ X_3^{(2)} \\ X_4^{(2)} \\ X_5^{(2)} \\ X_6^{(2)} \\ X_7^{(2)} \end{bmatrix} = \begin{bmatrix} \text{supervisor satisfaction} \\ \text{career-future satisfaction} \\ \text{financial satisfaction} \\ \text{workload satisfaction} \\ \text{company identification} \\ \text{kind-of-work satisfaction} \\ \text{general satisfaction} \end{bmatrix}$$

Matrik sampel korelasi berdasarkan pada 784 responden adalah:

$$\mathbf{R} = \left[\begin{array}{c|c} \mathbf{R}_{11} & \mathbf{R}_{12} \\ \hline \mathbf{R}_{21} & \mathbf{R}_{22} \end{array}\right]$$

$$= \left[\begin{array}{ccccc|ccccccc}
1.0 & & & & & .33 & .32 & .20 & .19 & .30 & .37 & .21 \\
.49 & 1.0 & & & & .30 & .21 & .16 & .08 & .27 & .35 & .20 \\
.53 & .57 & 1.0 & & & .31 & .23 & .14 & .07 & .24 & .37 & .18 \\
.49 & .46 & .48 & 1.0 & & .24 & .22 & .12 & .19 & .21 & .29 & .16 \\
.51 & .53 & .57 & .57 & 1.0 & .38 & .32 & .17 & .23 & .32 & .36 & .27 \\
\hline
.33 & .30 & .31 & .24 & .38 & 1.0 & & & & & & \\
.32 & .21 & .23 & .22 & .32 & .43 & 1.0 & & & & & \\
.20 & .16 & .14 & .12 & .17 & .27 & .33 & 1.0 & & & & \\
.19 & .08 & .07 & .19 & .23 & .24 & .26 & .25 & 1.0 & & & \\
.30 & .27 & .24 & .21 & .32 & .34 & .54 & .46 & .28 & 1.0 & & \\
.37 & .35 & .37 & .29 & .36 & .37 & .32 & .29 & .30 & .35 & 1.0 & \\
.21 & .20 & .18 & .16 & .27 & .40 & .58 & .45 & .27 & .59 & .31 & 1.0
\end{array}\right]$$

CANONICAL VARIATE COEFFICIENTS AND CANONICAL CORRELATIONS

	Standardized variables							Standardized variables						
	$z_1^{(1)}$	$z_2^{(1)}$	$z_3^{(1)}$	$z_4^{(1)}$	$z_5^{(1)}$	$\hat{\rho}_1^*$		$z_1^{(2)}$	$z_2^{(2)}$	$z_3^{(2)}$	$z_4^{(2)}$	$z_5^{(2)}$	$z_6^{(2)}$	$z_7^{(2)}$
$\hat{\mathbf{a}}_1'$:	.42	.21	.17	−.02	.44	.55	$\hat{\mathbf{b}}_1'$:	.42	.22	−.03	.01	.29	.52	−.12
$\hat{\mathbf{a}}_2'$:	−.30	.65	.85	−.29	−.81	.23	$\hat{\mathbf{b}}_2'$:	.03	−.42	.08	−.91	.14	.59	−.02
$\hat{\mathbf{a}}_3'$:	−.86	.47	−.19	−.49	.95	.12	$\hat{\mathbf{b}}_3'$:	.58	−.76	−.41	−.07	.19	−.43	.92
$\hat{\mathbf{a}}_4'$:	.76	−.06	−.12	−1.14	−.25	.08	$\hat{\mathbf{b}}_4'$:	.23	.49	.52	−.47	.34	−.69	−.37
$\hat{\mathbf{a}}_5'$:	.27	1.01	−1.04	.16	.32	.05	$\hat{\mathbf{b}}_5'$:	−.52	−.63	.41	.21	.76	.02	.10

Sebagai contoh, pasangan variate kanonik sampel pertama adalah

$$\hat{U}_1 = .42z_1^{(1)} + .21z_2^{(1)} + .17z_3^{(1)} - .02z_4^{(1)} + .44z_5^{(1)}$$

$$\hat{V}_1 = .42z_1^{(2)} + .22z_2^{(2)} - .03z_2^{(2)} + .01z_4^{(2)} + .29z_5^{(2)} + .52z_6^{(2)} - .12z_7^{(2)}$$

Dengan sampel korelasi kanonik $\hat{\rho}_1^* = .55$

$\mathbf{X}^{(1)}$ variables	Sample canonical variates $\hat{U}_1$	$\hat{V}_1$	$\mathbf{X}^{(2)}$ variables	Sample canonical variates $\hat{U}_1$	$\hat{V}_1$
1. Feedback	.83	.46	1. Supervisor satisfaction	.42	.75
2. Task significance	.74	.41	2. Career-future satisfaction	.35	.65
3. Task variety	.75	.42	3. Financial satisfaction	.21	.39
4. Task identity	.62	.34	4. Workload satisfaction	.21	.37
5. Autonomy	.85	.48	5. Company identification	.36	.65
			6. Kind-of-work satisfaction	.44	.80
			7. General satisfaction	.28	.50

Kelima variabel karakteristik pekerjaan kira-kira memiliki korelasi yang sama dengan variasi kanonik pertama $\hat{U}_1$.

Sampel korelasi antara $\hat{U}_1$ dan $\hat{V}_1$ adalah $\hat{\rho}_1^* = .55$. Ini menunjukkan overlap antara karakteristik pekerjaan dan kepuasan pekerjaan.