BAB III

PEREDUKSIAN RUANG INDIVIDU DENGAN ANALISIS KOMPONEN UTAMA

Analisis komponen utama adalah metode statistika multivariat yang bertujuan untuk mereduksi dimensi data dengan membentuk kombinasi linear–kombinasi linear dari variabel yang saling berkorelasi. Kombinasi linear yang terbentuk dinamakan komponen utama, di antara komponen utama tidak akan saling berkorelasi satu dengan yang lainnya. Dengan komponen utama tersebut data awal akan dapat direpresentasi secara maksimal namun dengan sesedikit mungkin komponen utama.

Komponen utama pertama adalah kombinasi linear dari variabel-variabel awal dengan variansi maksimum, komponen utama kedua adalah kombinasi linear yang mempunyai variansi maksimum di antara semua kombinasi linear yang tidak berkorelasi dengan komponen utama pertama, dan seterusnya. Pada dasarnya, analisis komponen utama terkait pada akar karakteristik dan vektor karakteristiknya. Koefisien pada komponen utama pertama berhubungan dengan nilai akar karakteristik terbesar, begitu pula dengan proporsi variansinya (Muirhead, 1982:380).

Jackson (1991:63) menyatakan bahwa terdapat tiga metode yang harus dipertimbangkan dalam pemilihan matriks yang digunakan untuk mendapatkan vektor karakteristik. Metode tersebut adalah sebagai berikut:

1. Semua variabel yang digunakan adalah variabel asli tidak dilakukan perubahan apapun.

2. Menggunakan matriks data terpusat, sehingga setiap vektor variabelnya menjadi $X - \bar{X}$, dengan demikian setiap variabel mempunyai rata-rata nol.
3. Dengan matriks data yang distandarkan, artinya setiap variabel dalam satuan standar. Sehingga setiap variabel mempunyai rata-rata nol dan variansi satu. Setiap variabel dinyatakan dengan $\frac{(X_i - \bar{X}_i)}{S_i}$.

Jika metode yang digunakan adalah matriks data terpusat yaitu dengan pengurangan rata-rata, maka matriksnya adalah matriks varians-kovarians, sedangkan jika data distandarkan maka yang digunakan adalah matriks korelasi.

Secara umum, matriks varians-kovarians lebih banyak digunakan, namun pada beberapa kasus, vektor karakteristik menjadi tidak tepat bila didasarkan pada matriks varians-kovarians. Kemungkinan penyebabnya adalah sebagai berikut:

1. Variabel awal menggunakan satuan yang berbeda, sehingga operasi *trace* dari matriks varians-kovarians menjadi tidak berarti. Ketika variabelnya dalam satuan yang berbeda, maka matriks data yang digunakan adalah matriks data yang distandarkan sehingga untuk mendapatkan vektor karakteristik digunakan matriks korelasi.
2. Variabel awal menggunakan satuan yang sama namun variansinya jauh berbeda. Jika kasusnya demikian, penggunaan matriks korelasi lebih tepat untuk digunakan.

Penggunaan matriks korelasi tersebar luas ke berbagai aplikasi, para pengguna jarang menggunakan matriks varians-kovarians dan meyakini bahwa penggunaanya tidak selamanya dapat digunakan untuk beberapa kasus. Walaupun demikian, ketika variabelnya dalam satuan ukuran yang sama dan besar

variansinya tidak jauh berbeda, maka matriks varians-kovarians lebih praktis untuk digunakan.

3.1 Pereduksian Ruang Variabel

Tujuan dari pereduksian ruang variabel dengan analisis komponen utama adalah mereduksi dimensi data yang terdiri dari variabel-variabel yang berkorelasi dengan jumlah yang banyak. Langkahnya adalah dengan mentransformasi variabel-variabel awal menjadi bentuk kombinasi linear yang tidak saling berkorelasi. Kombinasi linear tersebut dinamakan komponen utama, yang akan merepresentasikan keseluruhan dari variabel awal tanpa kehilangan banyak informasi.

Metode analisis komponen utama didasarkan pada hasil dari matriks *pxp* yang simetrik dan nonsingular, yaitu matriks varians kovarians V yang kemudian direduksi menjadi matriks diagonal L, dengan mengalikan V oleh matriks ortonormal U, sehingga persamaannya adalah sebagai berikut:

$$U^t V U = L \quad \text{(Jackson, 1991:7)} \qquad (3.1)$$

Diagonal dari elemen pada L adalah $\lambda_1, \lambda_2, \dots, \lambda_p$ yang kemudian disebut akar karakteristik atau nilai eigen dari V. Kolom-kolom dari U, $u_1, u_2, \dots, u_p$ disebut vektor karakteristik atau vektor eigen. Akar karakteristik dihasilkan dari solusi persamaan determinan yang disebut persamaan karakteristik.

$$|V - \lambda I| = 0 \qquad (3.2)$$

dengan I adalah matriks identitas.

Vektor karakteristik dihasilkan dari solusi persamaan

$$[V - \lambda I]t_i = 0 \tag{3.3}$$

dan

$$u_i = \frac{t_i}{\sqrt{t_i^t\, t_i}} \text{ (Jackson, 1991:8)} \tag{3.4}$$

untuk $i = 1,2, \dots, p$.

Langkah awal dalam analisis komponen utama adalah pada matriks varians kovarians (atau matriks korelasi). Misalkan untuk *p* variabel

$$V = \begin{bmatrix} s_2^1 & s_{12} & \cdots & s_{13} \\ s_{21} & s_2^2 & \cdots & s_{23} \\ \vdots & \vdots & \ddots & \vdots \\ s_{p1} & s_{p2} & \cdots & s_p^2 \end{bmatrix}$$

dengan s_i^2 variansi dari variabel ke-*i*, dan s_{ij} adalah kovarian dari variabel ke-*i* dengan variabel ke-*j*. Bila kovariansnya tidak sama dengan nol, ini mengindikasikan bahwa terdapat hubungan linear antara dua variabel. Besarnya hubungan yang digambarkan oleh koefisien korelasi adalah

$$r_{ij} = \frac{s_{ij}}{s_i s_j} \text{ (Jackson, 1991:11)} \tag{3.5}$$

Transformasi sumbu utama akan mentransformasi *p* variabel X_1, X_2, …, X_p yang berkorelasi menjadi *p* variabel baru Z_1, Z_2, …, Z_P yang tidak saling berkorelasi. Sumbu koordinat dari variabel baru tersebut digambarkan oleh vektor karakteristik $\overline{u}_i$, dengan transformasi

$$Z = U^t[X - \bar{X}] \text{ (Jackson, 1991:11)} \tag{3.6}$$

X adalah vektor *p* x 1 dari observasi pada variabel awal sedangkan $\bar{X}$ adalah vektor *p* x 1 sebagai rata-ratanya.

Transformasi dari variabel X disebut komponen utama. Komponen utama ke-i mempunyai rata-rata nol dengan variansinya sebesar akar karakteristik ke-i yaitu λ_i. Komponen utama ke-i tersebut adalah

$$Z_i = U_i^t[X - \bar{X}] \text{ (Jackson, 1991:11)} \qquad (3.7)$$

Bila dalam kombinasi linear yang terbentuk, besarnya koefisien pada semua variabelnya hampir sama dan bertanda positif, maka hal ini mengindikasikan bahwa kombinasi linearnya diboboti rata oleh semua variabel didalamnya. Namun bila koefisien variabelnya berlawanan tanda, maka korelasi yang terjadi adalah korelasi negatif, artinya bila variabel yang satu nilainya semakin besar, variabel yang satunya akan semakin kecil.

Sifat umum dan komponen keragaman pada analisis komponen utama adalah:

1. Determinan dari matriks varians kovarians, $|V|$. Ini disebut *generalized variance*.
2. Jumlah variansi dari variabel:

   $s_1^2 + s_2^2 + \ ... + s_p^2 = Tr(V)$ (*trace* dari V)

Kegunaan sifat umum dan komponen keragaman pada analisis komponen utama tersebut adalah untuk mempertahankan nilai yaitu:

1. $|V| = |L| = \lambda_1 . \lambda_2 . \ldots . \lambda_p$

   Determinan dari matriks varians kovarians akan sama dengan hasil perkalian dari akar karakteristik yang merupakan determinan dari matriks diagonal L.

2. $Tr(V) = Tr(L)$

   Artinya jumlah dari variansi data sama dengan jumlah dari akar karakteristik.

Sifat variansi yang kedua akan digunakan untuk mengetahui proporsi variansi yang dijelaskan oleh komponen utama. Perbandingan dari masing-masing akar karakteristik dengan total karakteristik akan mengindikasikan proporsi dari variansi tersebut. Korelasi dari masing-masing komponen utama dengan setiap variabel awal yang terkait juga dapat diketahui. Untuk menentukan korelasi dari setiap komponen utama dengan setiap variabel awal adalah

$$r_{zx} = \frac{u_{ji}\sqrt{\lambda_i}}{s_j} \text{ (Jackson, 1991:14)} \qquad (3.8)$$

r_{zx} adalah korelasi antara komponen utama ke-i, Z_i, dengan variabel awal X_j.

3.2 Pereduksian Ruang Individu

Analisis komponen utama tidak hanya digunakan untuk mereduksi ruang variabel, ruang individu juga dapat direduksi dengan analisis komponen utama. Seperti halnya pereduksian pada ruang variabel, pereduksian ruang individu juga akan membentuk kombinasi linear-kombinasi linear dari individu yang saling berkorelasi. Artinya, pereduksian ruang individu dengan analisis komponen utama dapat dilakukan bila terdapat korelasi pada individunya. Sehingga pada akhirnya antara kombinasi linear yang terbentuk tidak akan terjadi korelasi. Kombinasi linear yang terbentuk selanjutnya dikatakan sebagai komponen utama.

Misalkan $X_{(pxn)}$ adalah matriks hasil pengukuran p buah variabel kuantitatif pada n individu, baris menyatakan variabel-variabel pengukuran, sedangkan kolom menyatakan individu-individu yang diukur dari variabel-variabel tersebut. Meskipun dalam menjelaskan informasi keseluruhan dibutuhkan sebanyak n individu, namun ada kalanya sebanyak n individu tersebut dapat

diwakili oleh *k* komponen utama. Sejumlah *k* komponen utama tersebut akan menggantikan *n* individu tanpa kehilangan banyak informasi.

$$X = \begin{bmatrix} x_1^1 & x_2^1 & \cdots & x_p^1 & \cdots & x_n^1 \\ x_1^2 & x_2^2 & \cdots & x_p^2 & \cdots & x_n^2 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ x_1^p & x_2^p & \cdots & x_p^p & \cdots & x_n^p \end{bmatrix}$$

x_i^j yang merupakan elemen baris ke-*j* dan kolom ke-*i*, adalah nilai pengukuran terhadap variabel ke-*j* pada individu ke-*i*, dengan *i* di $I = \{1,2,\dots,n\}$ dan $J = \{1,2,\dots,p\}$.

Urutan bilangan $(x_i^1, x_i^2, \dots, x_i^p)$ adalah urutan nilai pengukuran variabel pertama sampai dengan variabel ke-*p* pada individu ke-*i*, yang dapat dinyatakan dengan vektor

$$\bar{\boldsymbol{x}}_{\boldsymbol{i}} = \begin{bmatrix} x_i^1 \\ x_i^2 \\ \vdots \\ x_i^p \end{bmatrix} = \sum_{k=1}^{p} x_i^k \bar{\boldsymbol{e}}_{\boldsymbol{k}} \quad \text{di } E = R^p \tag{3.9}$$

$\{\bar{\boldsymbol{e}}_{\boldsymbol{1}}, \bar{\boldsymbol{e}}_{\boldsymbol{2}}, \dots, \bar{\boldsymbol{e}}_{\boldsymbol{p}}\}$ menyatakan basis kanonik dari ruang vektor individu *E*. Jadi, $\bar{\boldsymbol{x}}_{\boldsymbol{i}}$ menggambarkan vektor individu ke-*i* (*i* = 1, 2, ..., *n*) di *E*. Sedangkan urutan bilangan $\left(x_1^j, x_2^j, \dots, x_n^j\right)$ merupakan hasil pengukuran variabel ke-*j* terhadap individu pertama sampai dengan individu ke-*n* dan dapat dinyatakan sebagai

$$\bar{\boldsymbol{x}}^{\boldsymbol{j}} = \begin{bmatrix} x_1^j \\ x_2^j \\ \vdots \\ x_n^j \end{bmatrix} = \sum_{k=1}^{n} x_k^j \bar{\boldsymbol{f}}_{\boldsymbol{k}} \quad \text{di } F = R^n \tag{3.10}$$

$\{\bar{\boldsymbol{f}}_{\boldsymbol{1}}, \bar{\boldsymbol{f}}_{\boldsymbol{2}}, \dots, \bar{\boldsymbol{f}}_{\boldsymbol{n}}\}$ menyatakan basis kanonik dari ruang vektor variabel *F*. Artinya, $\bar{\boldsymbol{x}}^{\boldsymbol{j}}$ menggambarkan vektor variabel ke-*j* (*j* = 1, 2, ..., *p*) di *F*.

Pada E akan terdapat awan titik-titik individu $\{\bar{x}_i ;\ i = 1,2,\dots,n\}$ dan pada F akan terdapat awan titik-titik variabel $\{\bar{x}^j ;\ j = 1,2,\dots,p\}$. E^* dan F^* adalah ruang dual dari E dan F dengan $\{\bar{e}_1^*, \bar{e}_2^*, \dots, \bar{e}_p^*\}$ dan $\{\bar{f}_1^*, \bar{f}_2^*, \dots, \bar{f}_n^*\}$ adalah basis-basis dualnya.

Berdasarkan definisi basis dual, akan diperoleh:

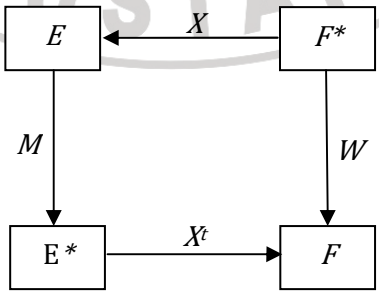
- $\bar{e}_j^*(\bar{x}_i) = \bar{e}_j^*\left(\sum_{k=1}^{p} x_i^k \bar{e}_k\right) = \langle \bar{e}_j^*, \bar{x}_i \rangle = x_i^j$ (3.11)
- $\bar{f}_i^*(\bar{x}^j) = \bar{f}_i^*\left(\sum_{k=1}^{n} x_k^j \bar{f}_k\right) = \langle \bar{f}_i^*, \bar{x}^j \rangle = x_i^j$ (3.12)

Sehingga dapat disimpulkan bahwa nilai $\bar{e}_j^*$ ada pada vektor individu ke-i atau dengan kata lain $\bar{e}_j^*$ menggambarkan variabel ke-j di E*. Sedangkan nilai $\bar{f}_i^*$ ada pada vektor variabel ke-j. Jadi, $\bar{f}_i^*$ menyatakan individu ke-i di F*.

Misalkan E ruang euclid dengan metrik M yang berperan mengukur kedekatan antara individu. Dengan memandang M sebagai isomorfisma dari E pada E^*, kemudian metrik W akan diterapkan untuk F^* sedemikian sehingga

$$\|\bar{x}_i - \bar{x}_k\|_M = \left\|\bar{f}_i^* - \bar{f}_k^*\right\|_W \quad (3.13)$$

dengan $X(\bar{f}_i^*) = \bar{x}_i;\ i = 1,2,\dots,n$. Mekanisme tersebut dapat disajkan dalam diagram dual berikut:



Gambar 1.1 Diagram Dual

Secara umum, untuk setiap $\bar{\boldsymbol{a}}$ dan $\bar{\boldsymbol{b}}$ di F^* dengan W, didefinisikan menjadi

$$\|X(\bar{\boldsymbol{a}}) - X(\bar{\boldsymbol{b}})\|_M = \|\bar{\boldsymbol{a}} - \bar{\boldsymbol{b}}\|_W \tag{3.14}$$

yang berarti pula bahwa untuk setiap $\bar{\boldsymbol{a}}$ di F^* berlaku:

$$\|X(\bar{\boldsymbol{a}})\|_M = \|\bar{\boldsymbol{a}}\|_W$$

Teorema 3.2.1

Jika untuk setiap $\bar{\boldsymbol{a}}$ di F^* berlaku $\|X(\bar{\boldsymbol{a}})\|_M = \|\bar{\boldsymbol{a}}\|_W$ maka diagram dual berlaku komutatif, artinya $W = X^t M X$.

Bukti:

Karena $\|X(\bar{\boldsymbol{a}})\|_M = \|\bar{\boldsymbol{a}}\|_W$ berlaku untuk setiap $\bar{\boldsymbol{a}}$ di F^*, maka untuk setiap pasangan (i,k) berlaku $M(\bar{\boldsymbol{x}}_{\boldsymbol{i}}, \bar{\boldsymbol{x}}_{\boldsymbol{k}}) = W(\bar{\boldsymbol{f}}_{\boldsymbol{i}}^*, \bar{\boldsymbol{f}}_{\boldsymbol{k}}^*)$, akan tetapi

$M(\bar{\boldsymbol{x}}_{\boldsymbol{i}}, \bar{\boldsymbol{x}}_{\boldsymbol{k}}) = \langle M(\bar{\boldsymbol{x}}_{\boldsymbol{i}}), \bar{\boldsymbol{x}}_{\boldsymbol{k}} \rangle = \langle MX(\bar{\boldsymbol{f}}_{\boldsymbol{i}}^*), X(\bar{\boldsymbol{f}}_{\boldsymbol{k}}^*) \rangle = \langle X^t MX(\bar{\boldsymbol{f}}_{\boldsymbol{i}}^*), (\bar{\boldsymbol{f}}_{\boldsymbol{k}}^*) \rangle$

dan $W(\bar{\boldsymbol{f}}_{\boldsymbol{i}}^*, \bar{\boldsymbol{f}}_{\boldsymbol{k}}^*) = \langle W(\bar{\boldsymbol{f}}_{\boldsymbol{i}}^*), \bar{\boldsymbol{f}}_{\boldsymbol{k}}^* \rangle$

Jadi untuk setiap pasangan (i,k) berlaku:

$X^t MX\left(\bar{\boldsymbol{f}}_{\boldsymbol{i}}^*, \bar{\boldsymbol{f}}_{\boldsymbol{k}}^*\right) = W\left(\bar{\boldsymbol{f}}_{\boldsymbol{i}}^*, \bar{\boldsymbol{f}}_{\boldsymbol{k}}^*\right)$, dengan kata lain $W = X^t MX$.

Bila pada teorema 3.2.1 didefinisikan M adalah matriks diagonal, dengan entri-entri pada setiap diagonalnya sebesar $\frac{1}{p-1}$, maka W adalah matriks varians-kovarians yang kemudian didefinisikan oleh Jackson (1991:190). Sehingga pada matriks data $X_{(pxn)}$ dengan $p < n$, matriks varians-kovarians untuk pereduksian variabel diperoleh dari perkalian matriks $XX^t/(n-1)$, sedangkan untuk pereduksian individu diperoleh dari $X^t X/(p-1)$. Sebelum menghitung matriks varians-kovarians untuk pereduksian ruang individu, setiap vektor individunya

dikurangi dengan vektor rata-ratanya. Sehingga rata-rata setiap vektor individunya sama dengan nol.

3.2.1 Penyajian Individu

Analisis komponen utama berusaha mereduksi ruang individu *p* menjadi berdimensi *k*, dengan *k*<*p*. Sehingga interpretasi dapat dilakukan pada ruang individu berdimensi *k*, tanpa kehilangan banyak informasi. Dengan melihat kesamaan karakteristik dari variabelnya, individu akan disajikan dalam kelompok-kelompok yang terdiri dari individu-individu yang mirip satu sama lain. Tujuannya adalah untuk menyajikan individu dalam kelompok-kelompok yang terdiri atas individu-individu yang saling berdekatan.

Pada dasarnya, pereduksian ruang individu dengan analisis komponen utama ini tidak akan cukup berarti bila pada individu–individu tidak mempunyai korelasi. Banyaknya maksimum komponen utama dari kombinasi linear dari individu tersebut sama dengan banyaknya individu awal. Bila pada sebuah pereduksian ruang individu, banyaknya komponen utama sama dengan individu awal, maka analisis komponen utama menjadi tidak berarti karena tidak didapatkan ruang individu dengan dimensi yang lebih kecil.

Misalkan $X_{(pxn)}$ adalah matriks data yang terdiri dari *p* variabel dan *n* individu. Maka terdapat awan titik-titik individu $\{\overline{x}_i\ ;\ i = 1,2,\dots,n\}$ di $\mathrm{E} = \mathrm{R}^p$. Misalkan terhadap individu ke-*i*, artinya terhadap setiap vektor $\overline{x}_i$ pada awan titik-titik individu tersebut diberikan bobot sebesar p_i, dengan nilai p_i lebih dari nol, dan $\sum_{i=1}^{n} p_i = 1$.

Vektor mean atau pusat gravitasi dari awan individu tersebut dinyatakan dengan vektor $\bar{g}$, dan didefinisikan dengan:

$$\bar{\boldsymbol{g}} = \sum_{i=1}^{n} p_i \bar{\boldsymbol{x}}_{\boldsymbol{i}} \tag{3.15}$$

Sedangkan elemen ke-*j* yang merupakan mean sampel untuk variabel ke-*j* adalah

$$\bar{\boldsymbol{g}}_{\boldsymbol{j}} = \sum_{i=1}^{n} p_i x_i^j \tag{3.16}$$

Khususnya jika dilakukan pembobotan yang sama untuk setiap individu, $p_i = \frac{1}{n}$; untuk setiap i ($i = 1,2,\dots,n$), maka

$$\bar{\boldsymbol{g}} = \frac{1}{n}\sum_{i=1}^{n} \bar{\boldsymbol{x}}_{\boldsymbol{i}} \text{ dan } \bar{\boldsymbol{g}}_{\boldsymbol{j}} = \frac{1}{n}\sum_{i=1}^{n} x_i^j$$

Definisi 3.2.1.1

Momen inersia individu $\bar{\boldsymbol{x}}_{\boldsymbol{i}}$ yang berbobot p_i terhadap suatu $\bar{\boldsymbol{a}}$ di *E* adalah bobot dikalikan dengan kuadrat jarak atau $p_i \|\bar{\boldsymbol{x}}_{\boldsymbol{i}} - \bar{\boldsymbol{a}}\|_M^2$.

Definisi 3.2.1.2

Momen inersia awan individu $\{\bar{x}_i$; i $= 1,2,\dots,n\}$, dengan $\bar{\boldsymbol{x}}_{\boldsymbol{i}}$ berbobot p_i, terhadap suatu $\bar{\boldsymbol{a}}$ di *E* adalah

$$I_{\bar{a}} = \sum_{i=1}^{n} p_i \|\bar{\boldsymbol{x}}_{\boldsymbol{i}} - \bar{\boldsymbol{a}}\|_M^2$$

Teorema 3.2.1.3

Untuk setiap $\bar{\boldsymbol{a}}$ di *E* berlaku:

$$I_{\bar{a}} = I_{\bar{g}} + \|\bar{\boldsymbol{g}} - \bar{\boldsymbol{a}}\|_M^2$$

Bukti:

Karena $\bar{\boldsymbol{x}}_{\boldsymbol{i}} - \bar{\boldsymbol{a}} = (\bar{x}_i - \bar{\boldsymbol{g}}) + (\bar{\boldsymbol{g}} - \bar{\boldsymbol{a}})$, maka

$$\|\bar{\boldsymbol{x}}_{\boldsymbol{i}} - \bar{\boldsymbol{a}}\|_M^2 = \|\bar{\boldsymbol{x}}_{\boldsymbol{i}} - \bar{\boldsymbol{g}}\|_M^2 + \|\bar{\boldsymbol{g}} - \bar{\boldsymbol{a}}\|_M^2 + 2M(\bar{\boldsymbol{x}}_{\boldsymbol{i}} - \bar{\boldsymbol{g}}, \bar{\boldsymbol{g}} - \bar{a})$$

Sedangkan,

$$\sum_{i=1}^{n} p_i M(\bar{\boldsymbol{x}}_{\boldsymbol{i}} - \bar{\boldsymbol{g}}, \bar{\boldsymbol{g}} - \bar{\boldsymbol{a}}) = M\left(\sum_{i=1}^{n} p_i (\bar{\boldsymbol{x}}_{\boldsymbol{i}} - \bar{\boldsymbol{g}}, \bar{\boldsymbol{g}} - \bar{\boldsymbol{a}})\right)$$

$$= M\left(\sum_{i=1}^{n} p_i \bar{\boldsymbol{x}}_{\boldsymbol{i}} - \bar{\boldsymbol{g}} \sum_{i=1}^{n} p_i , \bar{\boldsymbol{g}} - \bar{\boldsymbol{a}}\right)$$

$$= M(\bar{\boldsymbol{g}} - \bar{\boldsymbol{g}}, \bar{\boldsymbol{g}} - \bar{\boldsymbol{a}}), \text{karena} \sum_{i=1}^{n} p_i \bar{\boldsymbol{x}}_{\boldsymbol{i}} = \bar{g}, \sum_{i=1}^{n} p_i = 1$$

$$= M(\bar{\mathbf{0}}, \bar{\boldsymbol{g}} - \bar{\boldsymbol{a}}) = 0$$

Jadi,

$$I_{\bar{a}} = \sum_{i=1}^{n} p_i \|\bar{\boldsymbol{x}}_{\boldsymbol{i}} - \bar{\boldsymbol{a}}\|_M^2$$

$$= \sum_{i=1}^{n} p_i \left(\|\bar{\boldsymbol{x}}_{\boldsymbol{i}} - \bar{\boldsymbol{g}}\|_M^2 + \|\bar{\boldsymbol{g}} - \bar{\boldsymbol{a}}\|_M^2\right)$$

$$= \sum_{i=1}^{n} p_i \|\bar{\boldsymbol{x}}_{\boldsymbol{i}} - \bar{\boldsymbol{g}}\|_M^2 + \|\bar{\boldsymbol{g}} - \bar{\boldsymbol{a}}\|_M^2 \sum_{i=1}^{n} p_i$$

$$= I_{\bar{g}} + \|\bar{\boldsymbol{g}} - \bar{\boldsymbol{a}}\|_M^2$$

Teorema tersebut kemudian dinamakan Teorema *Huyghens*, yang menyimpulkan bahwa vektor mean $\bar{\boldsymbol{g}}$ adalah vektor yang meminimumkan $I_{\bar{a}}$, artinya $I_{\bar{a}}$ akan minimum bila $\bar{\boldsymbol{g}} = \bar{\boldsymbol{a}}$.

Teorema 3.2.1.4

Momen inersia awan individu di E terhadap $\bar{\boldsymbol{g}}$, yakni $I_{\bar{g}}$ memenuhi:

$$I_{\bar{g}} = Tr(VM)$$

Bukti:

$$I_{\bar{g}} = \sum_{i=1}^{n} p_i \|\bar{\boldsymbol{x}}_{\boldsymbol{i}} - \bar{\boldsymbol{g}}\|_M^2 = \sum_{i=1}^{n} p_i \|\bar{\boldsymbol{x}}_{\boldsymbol{i}}\|_M^2 \quad \text{(karena } X \text{ terpusat)}$$

$$= \sum_{i=1}^{n} p_i \; \bar{\boldsymbol{x}}_{\boldsymbol{i}}^{\boldsymbol{t}} \, M \, \bar{\boldsymbol{x}}_{\boldsymbol{i}}$$

$$= \sum_{i=1}^{n} p_i \; Tr(\bar{\boldsymbol{x}}_{\boldsymbol{i}}^{\boldsymbol{t}} \, M \, \bar{\boldsymbol{x}}_{\boldsymbol{i}}) \quad \text{(karena } \bar{\boldsymbol{x}}_{\boldsymbol{i}}^{\boldsymbol{t}} \, M \, \bar{\boldsymbol{x}}_{\boldsymbol{i}} \text{ adalah bilangan riil)}$$

$$= \sum_{i=1}^{n} p_i \; Tr(\bar{\boldsymbol{x}}_{\boldsymbol{i}} \, \bar{\boldsymbol{x}}_{\boldsymbol{i}}^{\boldsymbol{t}} M)$$

$$= Tr\left(\sum_{i=1}^{n} (p_i \, \bar{\boldsymbol{x}}_{\boldsymbol{i}} \, \bar{\boldsymbol{x}}_{\boldsymbol{i}}^{\boldsymbol{t}}) \, M \right)$$

$$= Tr \; (VM)$$

Misalkan W adalah ruang bagian dari E, dan $W^{\perp}$ adalah M-ortogonal dari W maka $E = W \oplus W^{\perp}$, untuk setiap $i = 1,2, \dots, n$ kemudian dituliskan

$$\bar{\boldsymbol{x}}_{\boldsymbol{i}} = \bar{\boldsymbol{\alpha}}_{\boldsymbol{i}} + \bar{\boldsymbol{\beta}}_{\boldsymbol{i}} \tag{3.17}$$

dengan $\bar{\boldsymbol{\alpha}}_{\boldsymbol{i}}$ di W dan $\bar{\boldsymbol{\beta}}_{\boldsymbol{i}}$ di $W^{\perp}$. Jadi, $\bar{\boldsymbol{\alpha}}_{\boldsymbol{i}}$ adalah proyeksi M-ortogonal dari $\bar{\boldsymbol{x}}_{\boldsymbol{i}}$ pada W. Momen inersia awan individu $\{\bar{x}_i \; ; \; \mathrm{i} = 1,2, \dots, n\}$ terhadap ruang bagian W:

$$I_W = \sum_{i=1}^{n} p_i \|\bar{\boldsymbol{\beta}}_{\boldsymbol{i}}\|_M^2 \tag{3.18}$$

- $I_W = 0$ jika dan hanya jika $\{\bar{x}_i \; ; \; \mathrm{i} = 1,2, \dots, n\} \subset W$.

- $I_{\bar{g}} = I_W + I_{W^\perp}$, karena

$$I_{\bar{g}} = \sum_{i=1}^{n} p_i \|\bar{\boldsymbol{x}}_i\|_M^2 = \sum_{i=1}^{n} p_i \|\bar{\boldsymbol{\alpha}}_i\|_M^2 + \sum_{i=1}^{n} p_i \|\bar{\boldsymbol{\beta}}_i\|_M^2 \qquad (3.19)$$

Teorema 3.2.1.5

Misalkan W ruang bagian dari E, jika $W = W_1 \oplus W_2$ dengan $W_1 \perp W_2$, maka:

$I_{W^\perp} = I_{W_1^\perp} + I_{W_2^\perp}$

Bukti:

$E = W \oplus W^\perp$. Maka $E = W_1 \oplus W_2 \oplus W^\perp$.

Sehingga untuk setiap $i = 1,2, \dots, n$, dari persamaan (3.17) $\bar{\boldsymbol{x}}_i = \bar{\boldsymbol{\alpha}}_i + \bar{\boldsymbol{\beta}}_i$ dengan $\bar{\boldsymbol{\alpha}}_i$ di W dan $\bar{\boldsymbol{\beta}}_i$ di $W^\perp$, sedangkan $\bar{\boldsymbol{\alpha}}_i = \bar{\boldsymbol{\gamma}}_i + \bar{\boldsymbol{\delta}}_i$ dengan $\bar{\boldsymbol{\gamma}}_i$ di W_1 dan $\bar{\boldsymbol{\delta}}_i$ di W_2.

Berdasarkan dalil Pythagoras,

$$I_W = \sum_{i=1}^{n} p_i \|\bar{\boldsymbol{\alpha}}_i\|_M^2 = \sum_{i=1}^{n} p_i \left(\|\bar{\boldsymbol{\gamma}}_i\|_M^2 + \|\bar{\boldsymbol{\delta}}_i\|_M^2 \right)$$

$$= I_{W_1^\perp} + I_{W_2^\perp}$$

karena $\bar{\boldsymbol{\gamma}}_i$ dan $\bar{\boldsymbol{\delta}}_i$ merupakan proyeksi M-ortogonal dari $\bar{\boldsymbol{x}}_i$ masing–masing pada W_1 dan W_2.

Akibat:

$I_{\bar{g}} = I_W + I_{W^\perp}$, maka $I_W = I_{\bar{g}} - I_{W_1^\perp} - I_{W_2^\perp}$

Analisis komponen utama berusaha mereduksi dimensi ruang individu $(E = R^p)$ menjadi berdimensi k ($k < p$). Ini dilakukan untuk membentuk kelompok-kelompok individu bila ruang individunya berada pada ruang vektor yang berdimensi p ($p > 3$). Pembentukan kelompok-kelompok individu tersebut akan didapat melalui bidang P. Bidang P yang dibangun oleh $\bar{u}_1$ dan $\bar{u}_2$

dinamakan bidang utama sedangkan sumbu Δ_{u_i} yang dibangun oleh $\bar{u}_i$ adalah sumbu utama ke-*i*.

$$\mathrm{P} = \Delta_{u_1} \oplus \Delta_{u_2}$$

Bila kualitas penyajian di P, artinya bagian inersia global yang diterangkan oleh P cukup baik, maka dengan memproyeksikan awan individu $\{\bar{\boldsymbol{x}}_{\boldsymbol{i}}\,;\, i = 1,2,\dots,n\}$, dapat dilakukan analisis terhadap individu secara visual melalui P. Sehingga pengelompokkan individu–individu yang berdekatan dapat dilakukan dengan melihat awan proyeksi individu di P.

Misalkan $\bar{\boldsymbol{\alpha}}_{\boldsymbol{i}}$ adalah proyeksi dari $\bar{\boldsymbol{x}}_{\boldsymbol{i}}$ pada P, maka

$$\bar{\boldsymbol{\alpha}}_{\boldsymbol{i}} = c_i^1\, \bar{\boldsymbol{u}}_{\boldsymbol{1}} + c_i^2\, \bar{\boldsymbol{u}}_{\boldsymbol{2}} \tag{3.20}$$

untuk setiap $i = 1,2,\dots,n$. Misalkan $\bar{\boldsymbol{\beta}}_{\boldsymbol{j}}$ adalah proyeksi *M*-ortogonal dari $\bar{\boldsymbol{e}}_{\boldsymbol{j}}$ pada P, maka proyeksi sumbu Δ_{e_j} di P dibangun oleh $\bar{\boldsymbol{\beta}}_{\boldsymbol{j}}\,; j = 1,2,\dots,p$. Untuk mengetahui kordinat dari $\bar{\boldsymbol{\beta}}_{\boldsymbol{j}}$:

$$\bar{\boldsymbol{\beta}}_{\boldsymbol{j}} = \beta_j^1\, \bar{\boldsymbol{u}}_{\boldsymbol{1}} + \beta_j^2\, \bar{\boldsymbol{u}}_{\boldsymbol{2}} \tag{3.21}$$

Jadi, $\beta_j^1 = M(\bar{\boldsymbol{e}}_{\boldsymbol{j}}, \bar{\boldsymbol{u}}_{\boldsymbol{1}})$ dan $\beta_j^2 = M(\bar{\boldsymbol{e}}_{\boldsymbol{j}}, \bar{\boldsymbol{u}}_{\boldsymbol{2}})$.

Karena komponen ke-*j* dari $\bar{e}_j$ berharga satu dan komponen lainnya nol, maka:

$$\beta_j^1 = \bar{\boldsymbol{e}}_{\boldsymbol{j}}{}^t M \bar{\boldsymbol{u}}_{\boldsymbol{1}}$$

$$= (0,\dots,0,1,0,\dots,0) \begin{bmatrix} m_{11} & m_{12} & \cdots & m_{1p} \\ m_{21} & m_{22} & \cdots & m_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ m_{p1} & m_{p2} & \cdots & m_{pp} \end{bmatrix} \begin{bmatrix} u_1^1 \\ u_1^2 \\ \vdots \\ u_1^p \end{bmatrix}$$

$$= \sum_{k=1}^{p} m_{jk}\, \tilde{\mathrm{u}}_1^k \tag{3.22}$$

Dengan cara yang sama didapat

$$\beta_j^2 = \sum_{k=1}^{p} m_{jk}\, u_2^k \tag{3.23}$$

Secara umum, dengan menuliskan

$$\bar{\boldsymbol{e}}_{\boldsymbol{j}} = \beta_j^1 \bar{\boldsymbol{u}}_{\boldsymbol{1}} + \beta_j^2 \bar{\boldsymbol{u}}_{\boldsymbol{2}} + \ldots + \beta_j^p \bar{\boldsymbol{u}}_{\boldsymbol{p}} \tag{3.24}$$

Kordinat $\bar{\boldsymbol{e}}_{\boldsymbol{j}}$ pada $\bar{\boldsymbol{u}}_{\boldsymbol{i}}$ adalah

$$\beta_j^i = \sum_{k=1}^{p} m_{jk}\, u_i^k \tag{3.25}$$

Dalam hal ini $M = I_p$ (metrik euclid klasik), maka

$$\beta_j^i = u_i^j \tag{3.26}$$

Jadi vektor ke-*j* dari vektor karakteristik $\bar{\boldsymbol{u}}_{\boldsymbol{i}}$ sama dengan kordinat $\bar{\boldsymbol{e}}_{\boldsymbol{j}}$ pada $\bar{\boldsymbol{u}}_{\boldsymbol{i}}$.

Bila kualitas penyajian di P kurang memuaskan, maka penyajiannya dapat dilakukan pada ruang bagian berdimensi tiga.

$$\mathrm{P} = \Delta_{u_1} \oplus \Delta_{u_2} \oplus \Delta_{u_3}$$

Pada dasarnya sama dengan penyajian di P, hanya saja data disajikan pada bidang – bidang berikut:

$\mathrm{P}_1 = \Delta_{u_1} \oplus \Delta_{u_2}$

$\mathrm{P}_2 = \Delta_{u_1} \oplus \Delta_{u_3}$

$\mathrm{P}_3 = \Delta_{u_2} \oplus \Delta_{u_3}$

Bila kualitas pada ruang bagian berdimensi tiga belum cukup optimal, maka bidang - bidangnya akan semakin banyak, hingga kualitas penyajiannya memadai.

3.2.2 Kualitas Global

Pada prinsipnya, komponen-komponen utama akan disajikan melalui bidang P. Komponen utama yang dihasilkan harus dapat menjelaskan total variansi. Kualitas komponen-komponen utama tersebut dinamakan kualitas global.

Bila sebagian besar (80% - 90%) dari persentasi kualitas penyajian individu untuk *n* yang besar dapat dijelaskan oleh satu, dua, atau tiga kombinasi linear dari individu-individu tersebut, maka komponen utama tersebut dapat menggantikan *n* indvidu awal tanpa kehilangan banyak informasi.

Karena P $=\Delta_{u_1} \oplus \Delta_{\tau_2}$, berdasarkan **akibat Teorema 3.2.1.5**,

$$I_p = Tr(VM) - I_{\Delta_{u_1}^{\perp}} - I_{\Delta_{u_2}^{\perp}}$$

atau

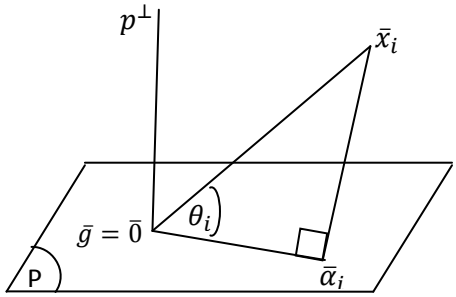
$$I_p = Tr(VM) - \lambda_1 - \lambda_2$$

Sehingga kualitas penyajian individu secara global di P ditunjukkan oleh besarnya

$$\frac{\lambda_1+\lambda_2}{Tr(VM)} \tag{3.27}$$

3.2.3 Kualitas Individual

Kualitas penyajian individu $\bar{\boldsymbol{x}}_{\boldsymbol{i}}$ oleh $\bar{\boldsymbol{\alpha}}_{\boldsymbol{i}}$ di P dapat diukur dengan membandingkan $\|\bar{\boldsymbol{x}}_{\boldsymbol{i}}\|_M = \|\bar{\boldsymbol{\alpha}}_{\boldsymbol{i}}\|_M$. Misalkan θ_i adalah sudut antara $\bar{\boldsymbol{x}}_{\boldsymbol{i}}$ dan $\bar{\boldsymbol{\alpha}}_{\boldsymbol{i}}$.



Gambar 3.2 Kualitas Individual

$$\cos\theta_i = \frac{\|\bar{\alpha}_i\|}{\|\bar{x}_i\|} \tag{3.28}$$

menyatakan kualitas penyajian $\bar{\boldsymbol{x}}_{\boldsymbol{i}}$ oleh $\bar{\boldsymbol{\alpha}}_{\boldsymbol{i}}$. Makin besar harga $\cos\theta_i$, makin bagus kualitasnya. Cos θ_i akan menyatakan alat ukur yang bagus, bila $\bar{\boldsymbol{x}}_{\boldsymbol{i}}$ cukup jauh dari $p^{\perp}$.

3.2.4 *Minimum Covariance Determinant*

Analisis komponen utama klasik didasarkan dari matriks varians kovarians dari data, oleh karena itu akan sangat sensitif dengan observasi yang berbeda dengan yang lainnya (pencilan). Akibatnya, komponen utama seringkali tertarik ke arah pencilan serta variansi dari observasi-observasi lainnya mungkin menjadi lebih besar. Pereduksian dimensi data menjadi kurang terpercaya bila pencilan tersebut dibiarkan begitu saja dalam data. *Minimum covariance determinant* (MCD) adalah salah satu metode untuk mendeteksi pencilan.

Definisi 3.2.4.1 MCD (Hardin dan Rocke, 2002:626)

Diketahui $X_{\sim} = \{x_1, x_2, \dots, x_n\}$ merupakan himpunan data dari *n* pengamatan dan p variabel dengan $n \geq p + 1$. Penaksir MCD merupakan pasangan $t \in R^p$ dan C

adalah matriks simetris definit positif berdimensi *pxp* dari suatu subsampel berukuran *h* pengamatan dengan $(n+p+1)/2 \leq h \leq n$ dengan

$$T_1 = \frac{1}{h}\sum_{i \in H_1} \boldsymbol{x_i} \tag{3.29}$$

$$C_1 = \frac{1}{h}\sum_{i \in H_1} (\boldsymbol{x_i} - T_1)(\boldsymbol{x_i} - T_1)^t \tag{3.30}$$

yang meminimumkan det(C).

Metode MCD mencari himpunan bagian dari X, sejumlah h elemen dengan h integer terkecil dari $(n+p+1)/2$. Tetapi, jika n besar, maka banyak sekali kombinasi subsampel yang harus ditemukan untuk mendapatkan penaksir MCD. Karena keterbatasan tersebut Rousseeuw dan Drissen membuat sebuah algoritma *Fast* MCD dengan teorema *C-Step*.

Teorema 3.2.4.2 *C-Step* (Rousseeuw dan Drissen, 1999:214)

Misalkan himpunan data $X_n = \{x_1, x_2, \ldots, x_n\}$ dari n pengamatan dengan p variabel. Misalkan $H_1 \subset \{x_1, x_2, \ldots, x_n\}$ dengan $|H_1| = h$, dan

$$T_1 = \frac{1}{h}\sum_{i \in H_1} \boldsymbol{x_i}$$

$$C_1 = \frac{1}{h}\sum_{i \in H_1} (\boldsymbol{x_i} - T_1)(\boldsymbol{x_i} - T_1)^t$$

Jika $\det(C_1) \neq 0$, maka definisi jarak relatifnya adalah

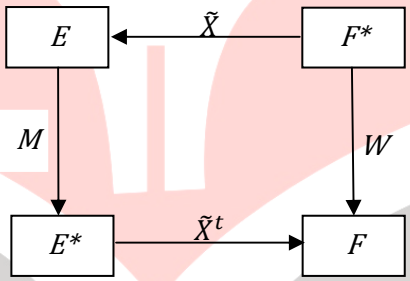
$$d_1(i) = \sqrt{(\boldsymbol{x_i} - T_1)^t C_1^{-1} (\boldsymbol{x_i} - T_1)} \quad \text{untuk } i = 1,2, \ldots, n. \tag{3.31}$$

Selanjutnya ambil H_2 sedemikian sehingga $\{d_1(i); i \in H_2\} = \{(d_1)_{1:n}, \ldots, (d_1)_{h:n}\}$, dengan $(d_1)_{1:n} \leq (d_1)_{2:n} \leq \cdots \leq (d_1)_{n:n}$ adalah urutan jarak dan hitung T_2 dan

C_2 berdasarkan himpunan H_2. Maka $\det(C_2) \leq \det(C_1)$ jika dan hanya jika $T_1 = T_2$ dan $C_1 = C_2$.

3.2.5 Pembobotan Pencilan

Putrasto (1996:12) mengungkapkan bahwa setiap pencilan akan diboboti, kemudian dibentuk matriks data baru $\tilde{X}$ dari matriks data X dengan pembobotan. Mekanisme diagram dual dari transformasi matriks data X menjadi matriks data dengan pembobotan pencilan $\tilde{X}$, nampak pada diagram dual berikut:



gambar 3.3 Diagram Dual dengan Pembobotan

Pembobotan pencilan tersebut dinyatakan dengan matriks diagonal, yaitu matriks Δ_{ii}. Setiap entri ke-ii yang merupakan pencilan diberi bobot satu sedangkan yang bukan merupakan pencilan diboboti nol. Kemudian matriks tersebut dikalikan dengan $\frac{1}{n_i}$, dengan n_i adalah banyaknya pencilan.

Definisi 3.2.5.1 Pembobotan Pencilan (Putrasto, 1996:15)

Misalkan X adalah matriks data asli, maka matriks pembobotan pencilan dinotasikan dengan $\tilde{X}$.

$$\tilde{X} = X\left(I_{nxn} - 1_n 1_n^t \frac{1}{n_i} \Delta\right)^t$$

dengan 1_n adalah matriks *nxn*, dengan entri pertama hingga entri ke-*n* pada vektor pertamanya bernilai satu. Kemudian matriks $\tilde{X}$ akan menggantikan matriks X, sehingga pembentukan matriks varians-kovariansnya tidak lagi dari X. Begitu pula dalam pembentukan kombinasi linearnya, penentuan akar karakteristiknya didapatkan dari matriks varians-kovarians dari $\tilde{X}$.

Berdasarkan uraian-uraian sebelumnya, maka dapat disimpulkan bahwa langkah-langkah dalam pereduksian ruang individu adalah sebagai berikut:

1. Menentukan matriks varians-kovarians dari X, yaitu dengan $X^tX/(p-1)$. Namun sebelumnya, setiap vektor individunya harus dikurangi dengan vektor rata-ratanya. Sehingga rata-rata setiap vektor individunya sama dengan nol.
2. Menentukan akar karakteristik dan vektor karakteristiknya.
3. Membentuk kombinasi linear dari vektor karakteristik yang ortonormal.
4. Menghitung proporsi komponen utama untuk menentukan banyaknya komponen yang akan diambil.

Untuk mendapatkan hasil yang lebih akurat, dapat dilakukan pendeteksian dan penanganan pencilan sebagai berikut:

1. Pendeteksian pencilan dengan menggunakan metode *minimum covariance determinant*.
2. Transformasi matriks data X menjadi matriks data dengan pembobotan pencilan $\tilde{X}$, yaitu dengan $\tilde{X} = X\left(I_{nxn} - 1_n 1_n^t \frac{1}{n_i} \Delta\right)^t$.
3. Menentukan matriks varians-kovarians dari matriks data $\tilde{X}$. Lakukan seperti langkah-langkah pada pereduksian individu biasa.