BAB II
TINJAUAN PUSTAKA

2.1. Data *Outlier*

Data o*utlier* didefinisikan sebagai pengamatan yang menjadi bagian di dalam nilai yang diduga berasal dari populasi yang berbeda (Johnson dan Winchern, 2002), sedangkan menurut Hair dkk (1998) data o*utlier* adalah pengamatan yang dapat diidentifikasikan secara jelas yang berbeda dari pengamatan yang lain, namun data *outlier* dapat menunjukan karakteristik dari populasi. Apabila data *outlier* bermasalah tidak mewakili populasi dan bertentangan dari tujuan analisis maka secara serius dapat memberikan hasil uji statistik yang berbeda. Pemeriksaan atau identifikasi data *outlier* pada data harus dilakukan karena data *outlier* memberikan pengaruh pada ragam dan setelah data *outlier* teridentifikasi maka dapat diputuskan untuk mempertahankan atau menghapus data *outlier* tersebut.

Pendeteksian data *outlier* sangat penting dilakukan dalam prosedur statistika karena data *outlier* mampu mempengaruhi informasi data (Draper ,1981). Data *outlier* yang terdeteksi dari suatu contoh mungkin saja dihilangkan agar dapat dianalisis lebih lanjut. Tindakan tersebut dapat dilakukan apabila hanya terdapat satu data *outlier*, namun hal ini tidak mungkin dilakukan jika data *outlier* yang terdeteksi lebih dari satu, karena mungkin data *outlier* mengandung informasi yang penting. Data *outlier univariate* adalah data *outlier* yang muncul pada satu peubah, sedangkan data outlier *multivariate* muncul pada lebih dari satu peubah.

Identifikasi data *outlier* pada data peubah ganda didasarkan pada kuadrat jarak mahalanobis, didefinisikan sebagai berikut (Johnson dan Winchern, 2002):

$$MD_i{}^2 = [\boldsymbol{x_i} - \overline{\boldsymbol{x}}]_{(1\times p)}{}'\boldsymbol{S^{-1}}_{(p\times p)}[\boldsymbol{x_i} - \overline{\boldsymbol{x}}]_{(p\times 1)} \qquad (2.1)$$

di mana:
i = 1,2,...,n ; n= banyaknya objek
$\boldsymbol{x_i}$ = vektor objek ke-i
$\overline{\boldsymbol{x}}$ = vektor rata-rata
$\boldsymbol{S}$ = matriks *covariance*

Hipotesis yang melandasi pengujian data *outlier*:

H_0: pengamatan biasa
H_1: data *outlier*

$MD_i^2 \sim \chi^2_{\alpha/2(p)}$

H_0 ditolak jika $MD_i{}^2 > \chi^2{}_{\alpha/2(p)}$
Jika H_0 ditolak maka dapat dinyatakan bahwa pengamatan mengandung data *outlier*.

2.2. *Eigen Value* dan *Eigen Vector*

Menurut Leon (2001), dalam menentukan *eigen value* dan *eigen vector* dimisalkan $\boldsymbol{C}$ sebagai suatu matriks persegi. Skalar λ disebut sebagai suatu *eigen value* dengan menggunakan persamaan:

$$|\boldsymbol{C} - \lambda \boldsymbol{I}| = 0 \quad (2.2)$$

jika terdapat suatu vektor tak nol $\boldsymbol{a}$ sehingga

$$\boldsymbol{C}\boldsymbol{a} = \lambda\, \boldsymbol{a} \quad (2.3)$$

Vektor $\boldsymbol{a}$ disebut *eigen vector* atau vektor karakteristik matriks $\boldsymbol{C}$ yang berpadanan dengan *eigen value* λ.

2.3. Analisis Biplot

Analisis Biplot yang dikemukakan pertama kali oleh Gabriel pada tahun 1971 merupakan upaya untuk memberikan peragaan secara grafik dari matriks data $\boldsymbol{X}_{(n\times p)}$ dalam suatu plot dengan menumpangtindihkan vektor-vektor baris matriks $\boldsymbol{X}_{(n\times p)}$ (gambaran objek) dengan vektor-vektor yang mewakili kolom matriks $\boldsymbol{X}_{(n\times p)}$ (gambaran peubah). Adapun yang diperoleh gambaran dari analisis biplot tentang objek, misalnya kedekatan antar objek, gambaran tentang peubah dan keterkaitan antara objek dan peubah. Tampilan objek dalam Analisis Komponen Utama merupakan kasus khusus dari analisis biplot dan penghitungan dalam analisis biplot didasarkan pada Penguraian Nilai Singular (PNS) suatu matriks (Siswadi dan Suharjo, 1998). Kata "bi" dalam biplot menyatakan dua himpunan titik (yaitu baris dan kolom matriks tujuan) yang divisualisasikan oleh produk skalar, dan bukan menyatakan tampilan dua dimensi. Biplot dan geometrinya berlaku untuk ruang-ruang dimensi manapun, tetapi akan perlu mengurangi dimensi ketika matriks data memiliki dimensi tinggi sedangkan representasi memerlukan dimensi rendah, biasanya dua atau tiga.

Analisis biplot dapat dibangun dari suatu matriks data dengan masing-masing kolom mewakili peubah dan masing-masing baris mewakili objek.

$$\boldsymbol{X}_{(n\times p)} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{1p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}$$

Matriks $\boldsymbol{X}_{(n\times p)}$ adalah matriks yang memuat peubah-peubah yang akan diteliti sebanyak *p* dan objek penelitian sebanyak *n* (Mattjik dkk., 2011).

Analisis biplot adalah peragaan secara grafik dari baris dan kolom sebuah matriks data $\boldsymbol{X}^*_{(n\times p)}$dengan baris mewakili objek dan kolom mewakili peubah. Dalam setiap aplikasi, analisis biplot dimulai dengan mentransformasikan matriks $\boldsymbol{X}^*$ sebagai matriks data asal terhadap nilai rata - ratanya menjadi matriks $\boldsymbol{X}$ yang akan digambarkan.

$$\boldsymbol{X}^*_{(n\times p)} = \boldsymbol{X}_{(n\times p)} - (\boldsymbol{J}_{(n\times n)}\boldsymbol{X}_{(n\times p)})/n \qquad (2.4)$$

di mana:

$\boldsymbol{X}_{(n\times \mathrm{p})}$ = matriks data dengan *n* objek dan *p* peubah

$\boldsymbol{X}^*_{(n\times p)}$ = matriks data yang terkoreksi terhadap nilai tengah

$\boldsymbol{J}_{(n\times n)}$ = matriks berunsur bilangan satu

Biplot didefinisikan sebagai penguraian dari matriks tujuan ke dalam produk dari dua matriks, yang disebut matriks kiri dan kanan. Pernyataan tersebut dapat ditulis sebagai $\boldsymbol{X} = \boldsymbol{ULA'}$ dengan $\boldsymbol{X}$ adalah matriks tujuan, $\boldsymbol{U}$ adalah matriks kiri dan $\boldsymbol{A'}$ adalah matriks kanan (Hawkins, 2001).

Penguraian Nilai Singular dapat ditulis sebagai

$${}_{\mathrm{n}}\boldsymbol{X}_{\mathrm{p}} = {}_{n}\boldsymbol{U}_{r}\boldsymbol{L}_{r}\boldsymbol{A}_{p}' \qquad (2.5)$$

di mana:

$$\boldsymbol{L} = \begin{bmatrix} \sqrt{\lambda_1} & 0 & 0 & 0 \\ 0 & \sqrt{\lambda_2} & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & \sqrt{\lambda_{\mathrm{n}}} \end{bmatrix}$$

$$\boldsymbol{A} = [\boldsymbol{a_1}, \boldsymbol{a_2}, \ldots, \boldsymbol{a_r}]$$

$$\boldsymbol{U} = \left\{\frac{1}{\sqrt{\lambda_1}}\, \boldsymbol{X}^* \boldsymbol{a_1}, \frac{1}{\sqrt{\lambda_2}}\, \boldsymbol{X}^* \boldsymbol{a_2}, \dots, \frac{1}{\sqrt{\lambda_r}}\, \boldsymbol{X}^* \boldsymbol{a_r}\right\}$$

Matriks $\boldsymbol{U}$ dan $\boldsymbol{A}$ adalah matriks ortonormal kolom, dengan $\boldsymbol{U}'\boldsymbol{U} = \boldsymbol{A}'\boldsymbol{A} = \boldsymbol{I_r}$ (matriks identitas berdimensi *r*). Matriks $\boldsymbol{L}$ adalah matriks diagonal yang unsur diagonal-diagonalnya merupakan akar dari nilai *eigen* tak nol matriks $\boldsymbol{X}'\boldsymbol{X}$ atau matriks $\boldsymbol{X}\boldsymbol{X}'$. Matriks $\boldsymbol{A}$ adalah matriks yang kolom-kolomnya merupakan *eigen vector* yang berpadanan dengan *eigen value* tak nol dari matriks $\boldsymbol{X}'\boldsymbol{X}$, matriks $\boldsymbol{U}$ adalah matriks yang kolom-kolomnya merupakan *eigen vector* yang berpadanan dengan *eigen value* tak nol dari matriks $\boldsymbol{X}\boldsymbol{X}'$.

Mattjik (2011), misalkan $\boldsymbol{G} = \boldsymbol{U}\boldsymbol{L}^{\alpha}$ dan $\boldsymbol{H}' = \boldsymbol{L}^{1-\alpha}\boldsymbol{A}'$ dengan $0 \leq \alpha \leq 1$ persamaan (2.5) dijabarkan menjadi

$$\boldsymbol{X} = \boldsymbol{U}\boldsymbol{L}^{\alpha}\boldsymbol{L}^{1-\alpha}\boldsymbol{A}' = \boldsymbol{G}\boldsymbol{H}' \qquad (2.6)$$

Dari pendekatan matriks $\boldsymbol{X}$ pada dimensi dua diperoleh matriks $\boldsymbol{G}$ dan $\boldsymbol{H}$ sebagai berikut:

$$\boldsymbol{G}_{(n\times 2)} = \begin{bmatrix} g_{11} & g_{12} \\ \vdots & \vdots \\ g_{i1} & g_{i2} \\ \vdots & \vdots \\ g_{n1} & g_{n2} \end{bmatrix} \text{ dan } \boldsymbol{H}_{(p\times 2)} = \begin{bmatrix} h_{11} & h_{12} \\ \vdots & \vdots \\ h_{i1} & h_{i2} \\ \vdots & \vdots \\ h_{p1} & h_{p2} \end{bmatrix}$$

Matriks $\boldsymbol{G}$ adalah titik-titik koordinat dari *n* objek dan matriks $\boldsymbol{H}$ adalah titik-titik koordinat dari *p* peubah.

Untuk mendeskripsikan biplot perlu mengambil nilai α dalam mendefinisikan $\boldsymbol{G}$ dan $\boldsymbol{H}$. Pemilihan nilai α pada $\boldsymbol{G} = \boldsymbol{U}\boldsymbol{L}^{\alpha}$ dan $\boldsymbol{H}' = \boldsymbol{L}^{1-\alpha}A'$ bersifat sembarang dengan syarat $0 \leq \alpha \leq 1$. Pengambilan nilai α=0 dan α=1 berguna dalam interpretasi biplot.

Jika α = 0, Hasil pemfaktoran ini disebut biplot GH atau CMP (*Column Metric Preserving*) biplot. Penyajian ini mempertahankan jarak antar kolom dan digunakan untuk mempresentasikan ragam dan hubungan antar ragam. Dengan α = 0 didapat $\boldsymbol{G} = \boldsymbol{U}\boldsymbol{L}^0$ dan $\boldsymbol{H}' = \boldsymbol{L}\boldsymbol{A}'$ sehingga:

$$\begin{aligned} \boldsymbol{X}'\boldsymbol{X} &= (\boldsymbol{G}\boldsymbol{H}')'(\boldsymbol{G}\boldsymbol{H}') \\ &= \boldsymbol{H}\boldsymbol{G}'\boldsymbol{G}\boldsymbol{H}' \\ &= \boldsymbol{H}\boldsymbol{U}'\boldsymbol{U}\boldsymbol{H}' \\ &= \boldsymbol{H}\boldsymbol{H}' \end{aligned} \qquad (2.7)$$

Matriks $\boldsymbol{U}$ ortonormal dan $\boldsymbol{X'X}$ = **(n-1)$\boldsymbol{S}$** dengan *n* adalah banyaknya objek pengamatan dan $\boldsymbol{S}$ adalah matriks *covariance* dari matriks **X** maka $\boldsymbol{HH'}$=(*n*-1)$\boldsymbol{S}$**.** Hasil kali $\boldsymbol{h_i}'\,\boldsymbol{h_i}$ adalah akan sama dengan *(n-1)* kali *covariance* s_{jk} antara peubah ke-*j* dan peubah ke-*k*. Selanjutnya untuk mengetahui keragaman peubah digunakan matriks $\boldsymbol{H}$.

$$\boldsymbol{HH'} = \begin{bmatrix} h_{11} & h_{12} \\ \vdots & \vdots \\ h_{i1} & h_{i2} \\ \vdots & \vdots \\ h_{p1} & h_{p2} \end{bmatrix} \begin{bmatrix} h_{11} & \dots & h_{i1} & \dots & h_{p1} \\ h_{12} & \dots & h_{i2} & \dots & h_{p2} \end{bmatrix}$$

$$= \begin{bmatrix} h_{11}^{\mathbf{2}} + h_{12}^{\mathbf{2}} & \dots & h_{11}h_{p1} + h_{12}h_{p2} \\ \dots & \dots & \dots \\ h_{11}h_{p1} + h_{12}h_{p2} & \dots & h_{p1}^{\mathbf{2}} + h_{p2}^{\mathbf{2}} \end{bmatrix} \quad (2.8)$$

Diagonal utama pada matriks $\boldsymbol{HH'}$: $h_{11}^{\mathbf{2}} + h_{12}^{\mathbf{2}}$, ... , $h_{j1}^{\mathbf{2}} + h_{j2}^{\mathbf{2}}$, ..., $h_{p1}^{\mathbf{2}} + h_{p2}^{\mathbf{2}}$ menggambarkan keragaman dari peubah. Sedangkan, j=1,2,…,*p* menyatakan panjang vektor peubah (dengan jarak Euclid dari titik O(0,0)). Sehingga dapat disimpulkan bahwa panjang vektor peubah sebanding dengan keragaman dari peubah.

Nilai korelasi antar peubah dapat dinyatakan sebagai berikut:

$$\cos\theta = \frac{\boldsymbol{h_i}'\boldsymbol{h_j}}{\|\boldsymbol{h_{ii}}\|\|\boldsymbol{h_{jj}}\|} = \frac{s_{ij}}{\sqrt{s_{ii}}\sqrt{s_{jj}}} = r_{ij} \quad (2.9)$$

di mana θ adalah sudut antara vektor $\boldsymbol{h_i}$ dan $\boldsymbol{h_j}$**.**

Artinya, *cosinus* sudut antara vektor $\boldsymbol{h_i}$ dan $\boldsymbol{h_j}$ merupakan korelasi antara peubah ke-*i* dengan peubah ke-*j*. Bila sudut antara kedua vektor tersebut mendekati nol maka makin besar korelasi positif antara kedua peubah. Bila sudut antara kedua vektor mendekati π, maka makin besar pula korelasi negatif antara kedua peubah tersebut. Korelasi sama dengan 1, jika $\theta = 0$. Jika θ mendekati $\pi/2$ maka makin kecil korelasi antara kedua peubah tersebut dan korelasi sama dengan nol jika $= \pi/2$.

Jika α=1, hasil pemfaktorannya disebut JK atau RMP (*Row Metric Preserving*) biplot. Penyajian dalam kasus ini adalah jarak antar pasangan baris dan digunakan untuk mempelajari objek. Dengan α=1 maka ***G=UL*** dan $\boldsymbol{H'}$**=*A*** sehingga diperoleh :

$$\begin{aligned} XX' &= (GH')(GH')' \\ &= GH'HG' \\ &= GA'AG' \\ &= GG' \end{aligned} \qquad (2.10)$$

Pada keadaan ini, jarak *euclid* antara $\boldsymbol{g_i}$dan $\boldsymbol{g_j}$ akan sama dengan jarak Euclid antara objek pengamatan $\boldsymbol{x_i}$ dan $\boldsymbol{x_j}$.Vektor baris ke-i sama dengan skor komponen utama untuk responden ke-i dari hasil analisis komponen utama. Untuk $\boldsymbol{G = UL}$ maka unsur ke-*k* dari $\boldsymbol{g_i}$ adalah $u_{ik}\sqrt{\lambda_k}$. Hasil tersebut sama dengan z_{ik} yang merupakan skor komponen utama ke-k dari objek ke-i .

Jika α =1/2 memberikan skala atau bobot yang sama untuk baris dan kolom. Hasil pemfaktorannya disebut *symmetric* biplot. Hal ini biasanya sigunakan untuk interpretasikan interaksi dalam percobaan dua faktor. Biplot ini digunakan sebagai alat diagnose untuk mengidentifikasi model sederhana yang dapat menjelaskan data. Dari penyajian biplot dengan α = 1/2, akan diketahui bentuk hubungan antara peubah dan objek yang akan diteliti bersama.

Perlu dipahami sebelumnya bahwa biplot adalah upaya membuat gambar di ruang berdimensi banyak menjadi gambar di ruang berdimensi dua. Pereduksian dimensi ini mengakibatkan menurunnya informasi yang terkandung dalam biplot. Biplot yang mampu memberikan informasi sebesar 70% dari seluruh informasi dianggap cukup (Matjjik, 2011).

Sebagai ilustrasi analisis biplot klasik adalah sebagai berikut: Analisis biplot dapat dibangun dari suatu matriks data dengan masing-masing kolom mewakili peubah (*p*) dan masing-masing baris mewakili objek (*n*), dengan menggunakan data harga barang dalam satuan seribu rupiad di 4 pasar, *n*=4 (A, B, C, D) dan *p*=4 (X1, X2, X3, X4). Data disajikan pada Tabel 2.3:

Tabel 2.3. Data Ilustrasi Biplot Klasik

Objek	X1	X2	X3	X4
A	46,0	21,2	18,0	16,7
B	54,4	21,1	17,3	13,3
C	48,6	20,1	18,2	16,0
D	61,0	21,2	18,0	19,2

Setelah ditransformasi sesuai dengan persamaan (2.4), diperoleh matriks sebagai berikut:

$$\boldsymbol{X}^{*} = \begin{array}{c|cccc} & X1 & X2 & X3 & X4 \\ A & 20,53 & 27,88 & 22,88 & 31,15 \\ B & -4,28 & -5,43 & -5,63 & -8,65 \\ C & -7,48 & -9,23 & -7,53 & -11,85 \\ D & -8,78 & -13,23 & -9,73 & -10,65 \end{array}$$

Matriks ($\boldsymbol{A}$) merupakan *eigen vector* yang berpadanan dengan *eigen value* dari matriks ***X'X***. Berkut ini merupakan perhitungan untuk memperoleh matriks $\boldsymbol{A}$:

$$\boldsymbol{X'X} = \begin{bmatrix} 20,53 & -4,28 & -7,48 & -8,78 \\ 27,88 & -5,43 & -9,23 & -13,23 \\ 22,88 & -5,63 & -7,53 & -9,73 \\ 31,15 & -8,65 & -11,85 & -10,65 \end{bmatrix} \begin{bmatrix} 20,53 & 27,88 & 22,88 & 31,15 \\ -4,28 & -5,43 & -5,63 & -8,65 \\ -7,48 & -9,23 & -7,53 & -11,85 \\ -8,78 & -13,23 & -9,73 & -10,65 \end{bmatrix}$$

$$= \begin{bmatrix} 572,43 & 780,33 & 635,14 & 858,37 \\ 780,33 & 1066,45 & 866,19 & 1165,40 \\ 635,14 & 866,19 & 706,10 & 953,96 \\ 858,37 & 1165,40 & 953,96 & 1298,99 \end{bmatrix}$$

Mencari Nilai Eigen:

$$|\boldsymbol{X'X} - \boldsymbol{\lambda I}| = \boldsymbol{0}$$

$$\left|\begin{bmatrix} 572,43 & 780,33 & 635,14 & 858,37 \\ 780,33 & 1066,45 & 866,19 & 1165,40 \\ 635,14 & 866,19 & 706,10 & 953,96 \\ 858,37 & 1165,40 & 953,96 & 1298,99 \end{bmatrix} - \begin{bmatrix} \boldsymbol{\lambda}_1 & 0 & 0 & 0 \\ 0 & \boldsymbol{\lambda}_2 & 0 & 0 \\ 0 & 0 & \boldsymbol{\lambda}_3 & 0 \\ 0 & 0 & 0 & \boldsymbol{\lambda}_4 \end{bmatrix}\right| = 0$$

$$\begin{vmatrix} 572,43 - \boldsymbol{\lambda}_1 & 780,33 & 635,14 & 858,37 \\ 780,33 & 1066,45 - \boldsymbol{\lambda}_2 & 866,19 & 1165,40 \\ 635,14 & 866,19 & 706,10 - \boldsymbol{\lambda}_3 & 953,96 \\ 858,37 & 1165,40 & 953,96 & 1298,99 - \boldsymbol{\lambda}_4 \end{vmatrix} = 0$$

$\lambda_1 = 3631,25$

$\lambda_2 = 12,11$

$\lambda_3 = 0,62$

$\lambda_4 = 0,00$

Karena analisis biplot yang diterapkan analisis biplot dua dimensi, maka *eigen value* yang digunakan merupakan merupakan dua *eigen value* yang terbesar yaitu 3631,25 dan 12,11, sehingga diperoleh matriks ***L*** adalah sebagai berikut:

$$\boldsymbol{L} = \begin{bmatrix} \sqrt{3631,25} & 0 \\ 0 & \sqrt{12,11} \end{bmatrix}$$

Eigen vector pertama matriks kanan ➔ $\boldsymbol{X'Xa_1 = a_1\lambda_1}$

$$\begin{bmatrix} 572,43 & 780,33 & 635,14 & 858,37 \\ 780,33 & 1066,45 & 866,19 & 1165,40 \\ 635,14 & 866,19 & 706,10 & 953,96 \\ 858,37 & 1165,40 & 953,96 & 1298,99 \end{bmatrix} \begin{bmatrix} a_{11} \\ a_{12} \\ a_{13} \\ a_{14} \end{bmatrix} = 3631,25 \begin{bmatrix} a_{11} \\ a_{12} \\ a_{13} \\ a_{14} \end{bmatrix}$$

Sehingga diperoleh vektor $\boldsymbol{a}_1$ sebagai berikut:

$$\begin{bmatrix} a_{11} \\ a_{12} \\ a_{13} \\ a_{14} \end{bmatrix} = \begin{bmatrix} 0,4 \\ 0,54 \\ 0,44 \\ 0,60 \end{bmatrix}$$

Eigen vector kedua matriks kanan ➔ ***X'Xa₂= a₂λ₂***

$$\begin{bmatrix} 572,43 & 780,33 & 635,14 & 858,37 \\ 780,33 & 1066,45 & 866,19 & 1165,40 \\ 635,14 & 866,19 & 706,10 & 953,96 \\ 858,37 & 1165,40 & 953,96 & 1298,99 \end{bmatrix} \begin{bmatrix} a_{21} \\ a_{22} \\ a_{23} \\ a_{24} \end{bmatrix} = 12,11 \begin{bmatrix} a_{21} \\ a_{22} \\ a_{23} \\ a_{24} \end{bmatrix}$$

$$\begin{bmatrix} a_{21} \\ a_{22} \\ a_{23} \\ a_{24} \end{bmatrix} = \begin{bmatrix} -0,15 \\ -0,64 \\ -0,10 \\ 0,75 \end{bmatrix}$$

Matriks kiri ($\boldsymbol{U}$) merupakan *eigen vector* yang berpadanan dengan *eigen value* dari matriks $\boldsymbol{XX'}$. Berkut ini merupakan perhitungan untuk memperoleh matriks ***U***

$$\boldsymbol{XX'} = \begin{bmatrix} 20,53 & 27,88 & 22,88 & 31,15 \\ -4,28 & -5,43 & -5,63 & -8,65 \\ -7,48 & -9,23 & -7,53 & -11,85 \\ -8,78 & -13,23 & -9,73 & -10,65 \end{bmatrix} \begin{bmatrix} 20,53 & -4,28 & -7,48 & -8,78 \\ 27,88 & -5,43 & -9,23 & -13,23 \\ 22,88 & -5,63 & -7,53 & -9,73 \\ 31,15 & -8,65 & -11,85 & -10,65 \end{bmatrix}$$

$$= \begin{bmatrix} 2691,88 & -637,09 & -951,83 & -1102,96 \\ -637,09 & 154,17 & 226,83 & 256,08 \\ -951,83 & 226,83 & 338,02 & 386,98 \\ -1102,96 & 256,08 & 386,98 & 459,90 \end{bmatrix}$$

Eigen vector pertama matriks kiri➔ ***XX'u₁= u₁λ₁***

$$\begin{bmatrix} 2691,88 & -637,086 & -951,833 & -1102,96 \\ -637,09 & 154,169 & 226,832 & 256,08 \\ -951,83 & 226,832 & 338,024 & 386,98 \\ -1102,96 & 256,084 & 386,977 & 459,90 \end{bmatrix} \begin{bmatrix} u_{11} \\ u_{12} \\ u_{13} \\ u_{14} \end{bmatrix} = 3631,25 \begin{bmatrix} u_{11} \\ u_{12} \\ u_{13} \\ u_{14} \end{bmatrix}$$

$$\begin{bmatrix} u_{11} \\ u_{12} \\ u_{13} \\ u_{14} \end{bmatrix} = \begin{bmatrix} 0,86 \\ -0,20 \\ -0,30 \\ -0,35 \end{bmatrix}$$

Eigen vector kedua matriks kiri➔ ***XX'u₂= u₂λ₂***

$$\begin{bmatrix} 2691,88 & -637,09 & -951,83 & -1102,96 \\ -637,09 & 154,17 & 226,83 & 256,08 \\ -951,83 & 226,83 & 338,02 & 386,98 \\ -1102,96 & 256,08 & 386,98 & 459,90 \end{bmatrix} \begin{bmatrix} u_{21} \\ u_{22} \\ u_{23} \\ u_{24} \end{bmatrix} = 12,11 \begin{bmatrix} u_{21} \\ u_{22} \\ u_{23} \\ u_{24} \end{bmatrix}$$

$$\begin{bmatrix} u_{21} \\ u_{22} \\ u_{23} \\ u_{24} \end{bmatrix} = \begin{bmatrix} 0{,}08 \\ -0{,}53 \\ -0{,}33 \\ 0{,}78 \end{bmatrix}$$

Menentukan koordianat objek➔ $\boldsymbol{G=UL^{\alpha}}$

$\alpha = 1/2$

$\boldsymbol{G} = \boldsymbol{UL}^{\alpha} = \boldsymbol{UL}^{1/2}$

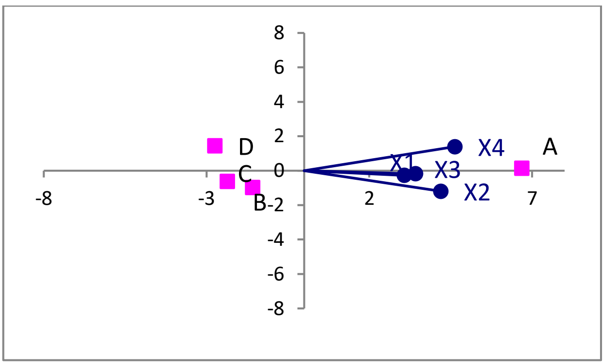
$$\begin{bmatrix} 0{,}86 & 0{,}08 \\ -0{,}20 & -0{,}53 \\ -0{,}30 & -0{,}33 \\ -0{,}35 & 0{,}78 \end{bmatrix} \begin{bmatrix} 3631{,}25^{1/4} & 0 \\ 0 & 12{,}11^{1/4} \end{bmatrix} = \begin{array}{cc} \boldsymbol{g_1} & \boldsymbol{g_2} \end{array} \begin{bmatrix} 6{,}68 & 0{,}15 \\ -1{,}55 & -0{,}99 \\ -2{,}33 & -0{,}62 \\ -2{,}72 & 1{,}46 \end{bmatrix}$$

$$\boldsymbol{H'} = \boldsymbol{L^{1-\alpha}A'} = \boldsymbol{L^{1/2}A'}$$

$$\boldsymbol{H'} = \begin{bmatrix} 3631{,}25^{1/4} & 0 \\ 0 & 12{,}11^{1/4} \end{bmatrix} \begin{bmatrix} 0{,}4 & 0{,}54 & 0{,}44 & 0{,}60 \\ -0{,}15 & -0{,}64 & -0{,}10 & 0{,}75 \end{bmatrix}$$

$$\boldsymbol{H'} = \begin{array}{c} \boldsymbol{h_1} \\ \boldsymbol{h_2} \end{array} \begin{bmatrix} 3{,}11 & 4{,}19 & 3{,}42 & 4{,}66 \\ -0{,}28 & -1{,}19 & -0{,}19 & -0{,}40 \end{bmatrix}$$

Dengan menggunakan koordinat objek (***G***) dan koordinat peubah (***H***) diperoleh gambar biplot dengan $\boldsymbol{g_1}$merupakan sumbu X objek dan $\boldsymbol{g_2}$ sebagai sumbu Y objek. Demikian juga untuk koordinat peubah, $\boldsymbol{h_1}$ sebagai sumbu X dan $\boldsymbol{h_2}$ sebagai sumbu Y. Dengan menumpangtindihkan koordinat peubah dan koordinat objek diperoleh gambar biplot sebagaimana disajikan dalam Gambar 2.1



Gambar 2.1. Ilustrasi Biplot Klasik dengan α = ½

Adapun yang diperoleh gambaran dari analisis biplot adalah:

1. Kedekatan antar objek, yaitu objek mempunyai kemiripan dengan objek lain yang ditunjukkan dengan posisi objek-objek tersebut.
2. Keragaman peubah, yaitu dengan membandingkan panjang vektor peubah. Peubah dengan keragaman kecil digambarkan dengan vektor yang pendek, sebaliknya jika keragamannya besar digambarkan dengan vektor yang panjang.
3. Korelasi antar peubah, dalam hal ini peubah digambarkan sebagai vektor. Dua peubah yang berkorelasi positif digambarkan sebagai dua vektor dengan arah yang sama atau membentuk sudut lancip. Sedangkan dua peubah yang berkorelasi negatif digambarkan sebagai dua vektor dengan arah berlawanan atau membentuk sudut tumpul. Apabila sudut yang dibentuk siku-siku, maka dua peubah tersebut tidak saling berkorelasi.
4. Keterkaitan peubah dengan objek. Objek yang letaknya sepihak dengan arah vektor peubah, menunjukkan objek tersebut nilainya di atas rata-rata, jika berlawanan berarti nilainya di bawah rata-rata (Mattjik, 2011).

2.4. Analisis *Robust* Biplot

Pada analisis biplot dengan penguraian nilai singular biasa data tidak boleh mengandung data *outlier*, tapi terkadang didalam segugus data mungkin terdapat data yang salah atau data. Keberadaan data *outlier* pada suatu data dapat mengganggu proses analisis data, sehingga mengakibatkan ragam pada data menjadi besar dan memiliki rentang yang lebar. Akan tetapi, membuang data *outlier* bukanlah sikap yang bijaksana karena adakalanya dapat memberikan informasi yang penting. Oleh karena itu, ditawarkan metode baru untuk mengeksplorasi data tersebut dengan analisis biplot kekar. Metode pendugaan *robust* biplot dilakukan dengan menggunakan matriks *covariance* yang *robust* (Daigle, 1992).

Pada analisis *robust* biplot, analisis biplot dapat dibangkitkan dengan menggunakan matriks *covariance yang* robust. Analisis biplot dengan menggunakan matriks *covariance yang robust* dilakukan dengan menduga *eigen value* dan *eigen vector* kiri ($\boldsymbol{U}$) dan kanan ($\boldsymbol{A}$) sehingga hasil dugaan tersebut tahan terhadap data *outlier*

(Hawkins, 2001). Menurut Kurt (2011), *eigen value* yang *robust* diperoleh dengan persamaan:

$$|\boldsymbol{S}_{FMCD} - \boldsymbol{\lambda}_{FMCD}\boldsymbol{I}| = \boldsymbol{0} \qquad (2.11)$$

Menggunakan matriks *covariance* dan *eigen value* yang *robust, eigen vector* kanan diperoleh dengan persamaan:

$$\boldsymbol{S}_{FMCD}\boldsymbol{a} = \boldsymbol{\lambda}_{FMCD}\boldsymbol{a} \qquad (2.12)$$

Sedangkan matriks *eigen vector* kiri diperoleh dengan menggunakan persamaan:

$$\boldsymbol{u_i} = \frac{1}{\sqrt{\lambda_i}}\boldsymbol{X}^*\boldsymbol{a_i} \qquad (2.13)$$

Sama seperti biplot klasik, Penentuan matriks kanan ($\boldsymbol{A}$) dan matriks kiri ($\boldsymbol{U}$) menggunakan penguraian nilai singular. Analisis biplot menggunakan dua dimensi, maka *eigen value* yang digunakan untuk $\boldsymbol{A}$ dan $\boldsymbol{U}$ adalah dua *eigen value* terbesar.

2.5. *Fast Minimum Covariance Determinan* (*fast*-MCD)

Metode yang dapat digunakan untuk mencari matriks *covariance robust* adalah *fast Minimum Covariance Determinant (fast*-MCD). Metode *MCD* yang dikemukakan oleh Rosseeuw adalah salah satu penaksir *robust* dalam analisis peubah ganda, Penaksir *robust MCD* merupakan rata-rata dan *covariance* dari sebagian pengamatan yang meminimumkan determinan matriks *covariance*-nya.

Metode *MCD* akan mencari himpunan bagian dari $\boldsymbol{X}$ dengan sejumlah h elemen di mana { $(n + p + 1)/2 \leq h \leq n$ }. Misalkan himpunan bagian itu adalah $\boldsymbol{X}_h$, akan terdapat C_h^n kombinasi yang harus ditemukan untuk mendapatkan penaksir *MCD*. Untuk n kecil, penaksir *MCD* dapat dengan cepat dihitung dan ditemukan, tetapi jika n besar, maka akan banyak sekali kombinasi sub sampel dari $\boldsymbol{X}$ yang harus ditemukan dan dalam penghitungannya pun akan cukup memakan waktu. Untuk mengatasi keterbatasan ini, maka Rousseeuw dan van Driesseen (1999) menemukan suatu algoritma baru untuk metode *MCD*, yang dinamakan metode *fast-MCD*.

Matriks *covariance* sangat rentan terhadap data *outlier*, oleh karena itu diperlukan matriks *covariance* kekar dengan metode *fast*-MCD. Berikut ini adalah algoritma dari *fast*-MCD.

1. Ambil himpunan bagian dari matriks $\boldsymbol{X}$ secara acak**,** dimisalkan himpunan bagian tersebut sebagai H_1 dengan jumlah elemen sebanyak h.

di mana $h = \frac{n+p+1}{2}$

2. Hitung vektor rata-rata $\boldsymbol{t_1}$ dan matriks *covariance* $\boldsymbol{S_1}$ dari H_1 dengan menggunakan persamaan:
$$\boldsymbol{t_1}_{(p\times h)} = \frac{1}{h}\sum_{i\in H}\boldsymbol{x_i} \tag{2.14}$$
dan
$$\boldsymbol{S_1} = \frac{1}{h-1}\sum_{\boldsymbol{i=1}}^{\boldsymbol{h}}(\boldsymbol{x_i}-\boldsymbol{t_1})'_{(p\times h)}\,(\boldsymbol{x_i}-\boldsymbol{t_1})_{(h\times p)} \tag{2.15}$$
3. Hitung determinan dari matriks *covariance* $\boldsymbol{S}$
4. Kemudian hitung jarak relatif dari setiap pengamatan terhadap rata-rata $\bar{x}_h$ dan *covariance* $\boldsymbol{S}$ dengan rumus dari Persamaan:
$$d_1(i) = \sqrt{(\boldsymbol{x_i}-\boldsymbol{t_1})'\boldsymbol{S_1}^{-1}(\boldsymbol{x_i}-\boldsymbol{t_1})}, \text{ untuk i=1,2,…,}n \tag{2.16}$$
5. Urutkan pengamatan tersebut berdasarkan jarak mahalanobis, dari terkecil hingga terbesar.
6. Ambil elemen dari *h* pengamatan dengan jarak terkecil berdasarkan tahapan (5) untuk menjadi elemen himpunan bagian H_2 , ulangi tahapan (2) sampai tahapan (5) sehingga ditemukan himpunan bagian yang konvergen dan memiliki determinan matriks *covariance* yang terkecil yaitu: $|\boldsymbol{S_{n+1}}| < |\boldsymbol{S_n}|$
7. Berdasarkan anggota h tersebut, data selanjutnya diboboti:
$$w_i\begin{cases}1, \text{jika } (\boldsymbol{x_{ij}}-\boldsymbol{t})'\boldsymbol{S^{-1}}(\boldsymbol{x_{ij}}-\boldsymbol{t}) < x^2_{\alpha/2,p}\\ \qquad 0, \text{lainnya}\end{cases}$$
8. Berdasarkan pembobot di atas, penduga *fast*-MCD adalah:
$$\boldsymbol{\bar{x}}_{fMCD} = \frac{\sum_{i=1}^{n} w_i x_{ij}}{\sum_{i=1}^{n} w_i} \tag{2.17}$$
$$\boldsymbol{S}_{fMCD} = \frac{\sum_{i=1}^{n} w_i(\boldsymbol{x_{ij}}-\boldsymbol{\bar{x}_{MCD}})'(\boldsymbol{x_{ij}}-\boldsymbol{\bar{x}_{MCD}})}{(\sum_{i=1}^{n} w_i)-1} \tag{2.18}$$

Identifikasi pencilan dengan menggunakan persamaan (2.1) menjadi tidak maksimal jika terdapat lebih dari satu *outlier* karena adanya efek *masking* dan efek *swamping*. Efek *Masking* terjadi saat pencilan tidak terdeteksi sebagai *outlier* karena adanya pencilan lain yang berdekatan. Efek *Swaping* terjadi saat pengamatan baik terdeteksi sebagai *outlier*. Efek ini dapat diatasi dengan menggunakan jarak *robust* dengan persamaan:
$$RD_i^2 = (\boldsymbol{x_i}-\boldsymbol{\bar{x}_{fMCD}})'\boldsymbol{S_{fMCD}^{-1}}(\boldsymbol{x_i}-\boldsymbol{\bar{x}_{fMCD}}) \tag{2.19}$$

Hipotesis yang melandasi pengujian data *outlier*:
H_0: pengamatan biasa
H_1: data *outlier*
$RD_i^{\,2} \sim \chi^2{}_{\alpha/2(p)}$
H_0 ditolak jika $RD_i^{\,2} > \chi^2{}_{\alpha/2(p)}$
Jika H_0 ditolak maka dapat dinyatakan bahwa pengamatan mengandung data *outlier*.

Sebagai ilustrasi *robust biplot* sebagai berikut:

Data merupakan data yang mengandung *outlier* tentang beberapa hasil panen kedelai di New York yang dikelompokkan berdasarkan lingkungan dan telah terkoreksi terhadap rata-rata sebagaimana disajikanpada Tabel 2.2.

Tabel 2.2. Data Ilustrasi *Robust* Biplot

	X1	X2	X3	X4
A	-1,7652	-1,9017	-1,2149	-1,4880
B	-0,8390	-1,0932	-0,1427	-0,1617
C	-1,1410	-0,8732	-0,9054	-1,1479
D	0,3813	0,4808	0,3198	0,4356
E	0,2875	0,1261	0,3221	0,4200
F	1,0470	1,1328	0,6041	0,1773
G	2,0297	2,1285	1,0168	1,7646

dengan menggunakan algoritma *fast Minimum Covariance Determinant*, diperoleh matriks *covariance robust* $\boldsymbol{S_{fMCD}}$ sebagai berikut:

$$\boldsymbol{S_{fMCD}} = \begin{bmatrix} 1{,}7610 & 1{,}8327 & 1{,}0347 & 1{,}3544 \\ 1{,}8327 & 1{,}9391 & 1{,}0500 & 1{,}3713 \\ 1{,}0347 & 1{,}0500 & 0{,}6535 & 0{,}8410 \\ 1{,}3544 & 1{,}3713 & 0{,}8410 & 1{,}1782 \end{bmatrix}$$

Berdasarakan matriks *covariance* robust, matriks kiri ($\boldsymbol{A}$) diperoleh dari *eigen vector* yang berpadanan dengan *eigen value* dari matriks *covariance robust*, Oleh karena itu, terlebih dahulu dicari *eigen value* dengan persamaan (2,11)

Eigen value yang diperoleh dari matriks *covariance* adalah
λ_1 = 5.3442
λ_2 = 0,1527
λ_3 = 0,0307
λ_4 = 0,0042

Dari *eigen value*, dapat ditentukan matriks singular $\boldsymbol{L}$ sebagai berikut:

$$\boldsymbol{L} = \begin{bmatrix} \sqrt{5,3442} & 0 \\ 0 & \sqrt{0,1527} \end{bmatrix}$$

Matriks $\boldsymbol{A}$ merupakan *eigen vector* yang berpadanan dengan *eigen value* diperoleh dengan menggunakan persamaan (2.12) sebagai berikut:

$$\boldsymbol{A} = \begin{bmatrix} 0,5726 & -0,1917 \\ 0,5945 & -0,5555 \\ 0,3401 & 0,2965 \\ 0,4505 & 0,7528 \end{bmatrix}$$

Matriks $\boldsymbol{U}$ diperoleh dengan persamaan (2.13), yaitu:

$$\boldsymbol{U} = \begin{bmatrix} -0,5695 & -0,0894 \\ -0,2210 & 0,6312 \\ -0,3528 & -0,4480 \\ 0,1429 & 0,0863 \\ 0,0951 & 0,2994 \\ 0,2752 & -0,5405 \\ 0,6302 & 0,0611 \end{bmatrix}$$

α=1/2

G=ULᵅ=UL^(1/2)

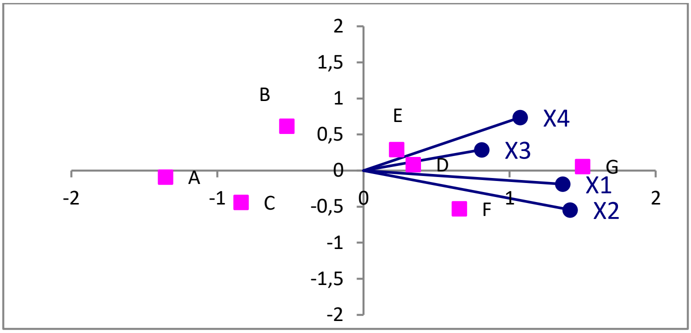
$$\begin{bmatrix} -0,5695 & -0,0894 \\ -0,2210 & 0,6312 \\ -0,3528 & -0,4480 \\ 0,1429 & 0,0863 \\ 0,0951 & 0,2994 \\ 0,2752 & -0,5405 \\ 0,6302 & 0,0611 \end{bmatrix} \begin{bmatrix} 5,4622^{1/2} & 0 \\ 0 & 1,5040^{1/2} \end{bmatrix} = \begin{array}{cc} \boldsymbol{g_1} & \boldsymbol{g_2} \\ \end{array} \begin{bmatrix} -1,3553 & -0,5572 \\ -0,5260 & -0,2163 \\ -0,8395 & -0,3451 \\ 0,3400 & 0,1398 \\ 0,2262 & 0,0930 \\ 0,6549 & 0,2692 \\ 1,4996 & 0,6165 \end{bmatrix}$$

$$\boldsymbol{H'} = \boldsymbol{L^{1-\alpha}A'} = \boldsymbol{L^{1/2}A'}$$

$$\boldsymbol{H'} = \begin{bmatrix} 5,4622^{1/2} & 0 \\ 0 & 1,5041^{1/2} \end{bmatrix} \begin{bmatrix} 0,5726 & 0,5945 & 0,3402 & 0,4505 \\ -0,1917 & -0,5555 & 0,2965 & 0,7528 \end{bmatrix}$$

$$\boldsymbol{H'} = \begin{array}{c} \boldsymbol{h_1} \\ \boldsymbol{h_2} \end{array} \begin{bmatrix} 1,3626 & 1,4147 & 0,8095 & 1,0721 \\ -0,1875 & -0,5435 & 0,2901 & 0,7365 \end{bmatrix}$$

Dengan menggunakan koordinat objek ($\boldsymbol{G}$) dan koordinat peubah ($\boldsymbol{H}$) diperoleh gambar biplot dengan $\boldsymbol{g}_1$merupakan sumbu X objek dan $\boldsymbol{g}_2$ sebagai sumbu Y objek. Demikian juga untuk koordinat peubah, $\boldsymbol{h_1}$ sebagai sumbu X dan $\boldsymbol{h_2}$ sebagai sumbu Y. Dengan menumpangtindihkan koordinat peubah dan koordinat objek diperoleh gambar biplot sebagaimana disjikan dalam Gambar 2.1



Gambar 2.2. Ilustrasi *robust* biplot α = 1/2

2.6. Pemeriksaan Kesesuaian Biplot

Gabriel (1971) dalam Mattjik (2002), mengemukakan ukuran matriks ***X*** dalam biplot dalam bentuk:

$$\rho^2 = \frac{\lambda_1 + \lambda_2}{\sum \lambda_j} \qquad (2.20)$$

di mana λ_1 adalah *eigen value* terbesar pertama, λ_2 adalah *eigen value* terbesar kedua.Jika nilai statistik uji semakin mendekati satu berarti biplot yang diperoleh dari matriks pendekatan berpangkat dua akan memberikan penyajian data yang semakin baik mengenai informasi-informasi yang terdapat pada data sebenarnya.