BAB II

LANDASAN TEORI

2.1 Kualitas

2.1.1 Konsep Kualitas

Kualitas produk merupakan salah satu faktor penting yang harus diperhatikan oleh perusahaan industri agar perusahaan industri tersebut mampu bersaing dan bertahan di bidangnya. Ada banyak pendapat para ahli kualitas yang mendefinisikan kualitas itu sendiri, beberapa diantaranya adalah sebagai berikut:

- Menurut Goetsch Davis (Erch, 2007) kualitas merupakan suatu kondisi dinamis yang berhubungan dengan produk, jasa, manusia, proses, dan lingkungan yang memenuhi atau melebihi harapan.
- Menurut Gaspersz (Erch, 2007) pada dasarnya kualitas mengacu kepada pengertian pokok yaitu kualitas terdiri dari sejumlah keistimewaan produk, baik keistimewaan langsung maupun keistimewaan atraktif yang memenuhi keinginan pelanggan dan dengan demikian memberikan kepuasan atas penggunaan produk serta terdiri dari segala sesuatu yang bebas dari kekurangan atau kerusakan.
- Menurut Montgomery (2001:4) kualitas merupakan kesesuaian untuk penggunaan.

Sehingga dapat ditarik kesimpulan bahwa kualitas adalah segala sesuatu yang dapat memenuhi keinginan semua pihak baik itu produsen, konsumen, dan pihak lain yang berhubungan dengan produk atau jasa tersebut.

Berdasarkan perspektif kualitas, Garvin (1987) yang dikutip dari Montgomery (2001:2) mengembangkan dimensi kualitas ke dalam delapan dimensi yang dapat digunakan sebagai dasar perencanaan strategis terutama bagi perusahaan atau manufaktur yang menghasilkan barang, kedelapan dimensi tersebut adalah sebagai berikut:

1. Kinerja produk *(Performance)*

   Kinerja produk merupakan kemampuan produk melakukan fungsi-fungsi spesifik yang diinginkan dan seberapa baik produk tersebut melakukan fungsinya itu.

2. Keandalan (*Reliability)*

   Keandalan merupakan kemampuan produk untuk tidak mengalami kerusakan sepanjang umur produk tersebut.

3. Durabilitas *(Durability)*

   Durabilitas menunjukkan seberapa lama produk yang dihasilkan dapat bekerja dengan baik dalam jangka waktu tertentu.

4. Kemampuan diperbaiki (*Serviceability)*

   Dimensi ini menunjukkan seberapa mudah produk dapat diperbaiki jika mengalami kerusakan yang meliputi kecepatan, kompetensi, kenyamanan, kemudahan, dalam pemeliharaan dan penanganan keluhan yang memuaskan.

5. Estetika *(Aesthetics)*

   Estetika merupakan penampilan fisik produk yang terlihat oleh konsumen menyangkut corak, rasa dan daya tarik produk tersebut.

6. Fitur *(Features)*

   Fitur merupakan kemampuan produk untuk melakukan fungsi yang lain diluar kemampuan atau fungsi utamanya.

7. Reputasi produk *(Perceived Quality)*

   Dimensi ini menyangkut citra atau reputasi produk serta tanggung jawab perusahaan terhadap produk yang dihasilkannya.

8. Kesesuaian terhadap standar *(Conformance to Standards)*

   Dimensi yang menunjukkan sejauh mana karakteristik desain dan operasi produk memenuhi standar – standar yang telah ditetapkan sebelumnya.

Pada umunya terdapat tiga proses utama yang harus dilakukan dalam memproduksi suatu produk. Juran (Mayaranie, 2009:9) menyebutnya dengan konsep *Trilogi Kualitas* yang isinya mencakup tiga proses utama yaitu:

1) Perencanaan kualitas

Perencanaan kualitas merupakan kegiatan yang dilakukan untuk mengembangkan proses yang dapat menghasilkan produk dengan kualitas yang memenuhi keinginan konsumen. Aktivitas-aktivitas yang dilakukan diantaranya:

- Mengidentifikasi konsumen;
- Menentukan kebutuhan konsumen;
- Menciptakan fitur-fitur produk yang memenuhi kebutuhan konsumen;
- Menentukan tujuan kualitas dengan biaya seminimal mungkin;
- Membangun proses produksi untuk menghasilkan produk yang diinginkan;

- Membuktikan bahwa proses yang direncanakan dapat memenuhi tujuan kualitas.

2) Pengendalian kualitas

Pengendalian kualitas statistik (*statistical quality control*) adalah teknik penyelesaian masalah yang digunakan untuk memonitor, mengendalikan, menganalisis, mengelola, dan memperbaiki produk dan proses menggunakan metode-metode statistik. Aktivitas-aktivitas yang dilakukan diantaranya:

- Memilih karakteristik yang akan dikendalikan;
- Memilih satuan pengukuran;
- Melakukan pengukuran;
- Menetapkan standar performansi;
- Mengukur performansi aktual;
- Menginterpretasikan perbedaan antara keadaan aktual dan keadaan standar;
- Menindaklanjuti perbedaan.

3) Perbaikan kualitas

Perbaikan kualitas merupakan sebuah kegiatan yang dilakukan untuk meningkatkan performansi yang lebih tinggi dibandingkan dengan performansi yang ada. Aktivitas-aktivitas yang dilakukan diantaranya:

- Membuktikan kebutuhan akan perbaikan;
- Mengidentifikasi hal yang akan diperbaiki
- Memverifikasi penyebab masalah;
- Menentukan solusi masalah;
- Membuktikan bahwa solusi yang dipilih dapat berjalan dengan efektif;

- Mengendalikan kegiatan produksi untuk mempertahankan perbaikan yang ada.

Ketiga proses utama tersebut saling bergantian, ketika perencanaan produksi telah dilakukan dan kemudian dilakukan proses produksi jika hasilnya belum mencapai hasil yang diinginkan maka dilakukanlah perbaikan produksi untuk menentukan standar produksi yang baru yang ukurannya lebih tinggi. Setelah diadakan proses perbaikan produksi, kegiatan yang dilakukan adalah mengendalikan kembali kualitas proses dengan standar yang baru, sampai diperoleh hasil yang sesuai dengan spesifikasi yang diinginkan.

2.1.2 Konsep Pengendalian kualitas

Secara umum pengendalian kualitas didefinisikan sebagai suatu usaha yang dilakukan untuk mencapai, memperbaiki, dan mempertahankan kualitas produk atau jasa. Menurut Montgomery (2001:7), pengendalian kualitas adalah aktivitas keteknikan dan manajemen yang dengan aktifitas itu dapat diukur ciri-ciri kualitas produk, membandingkannya dengan spesifikasi atau persyaratan dan mengambil tindakan perbaikan yang sesuai apabila ada penyimpangan dari keadaan yang standar. Sedangkan Juran (1974) (dalam Mayaranie, 2009:10) mendefinisikan pengendalian kualitas sebagai aktivitas yang dilakukan untuk mengukur performansi aktual dari proses yang sedang dijalankan, membandingkan performansi aktual tersebut dengan standar yang telah ditetapkan, dan melakukan tindakan perbaikan apabila terdapat perbedaan. Dari beberapa definisi di atas maka dapat disimpulkan bahwa pengendalian kualitas merupakan suatu kegiatan yang dilakukan untuk memenuhi tujuan kualitas selama proses produksi.

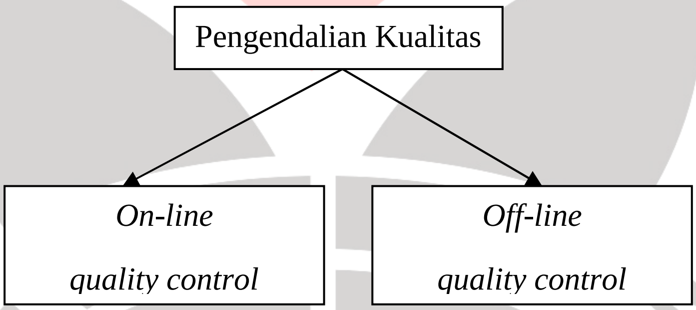
Pengendalian kualitas dibagi menjadi dua bagian, yaitu :

1) *On-line quality control*

Pengendalian kualitas ini merupakan pengendalian kualitas yang dilakukan pada saat produksi berlangsung. Kegiatan yang dilakukannya misalnya penyesuaian proses, pengendalian proses, dan inpeksi hasil proses.

2) *Off-line quality control*

Pengendalian kualitas ini merupakan pengendalian kualitas yang dilakukan di luar proses produksi, baik sebelum proses produksi atau setelah proses produksi dilakukan. Kegiatan yang dilakukan misalnya perbaikan rancangan proses dan produk.



Gambar 2.1 Pengendalian kualitas

Kestabilan proses dan kapabilitas proses merupakan dua hal yang diukur dalam proses pengendalian kualitas. Kestabilan proses merupakan alat ukur untuk mengetahui apakah keadaan proses produksi yang dilakukan sudah terkontrol. Sedangkan kapabilitas proses merupakan alat ukur untuk mengetahui seberapa baik kemampuan proses produksi yang dilakukan dalam memenuhi spesifikasi yang ditentukan.

UNIVERSITAS PENDIDIKAN INDONESIA PERPUSTAKAAN

Menurut Montgomery (2001:336), terdapat beberapa kondisi yang terkait dengan kestabilan proses dan kapabilitas proses, yang ditunjukkan pada gambar di bawah ini.

		IS THE PROCESS CAPABLE?	
		Yes	***No***
IS THE PROCESS IN CONTROL?	***Yes***	SPC	SPC *Experimental Design* *Investigate specifications* *Change process*
	No	SPC	SPC *Experimental Design* *Investigate specifications* *Change process*

Gambar 2.2 Klasifikasi Kondisi Proses

Dari gambar di atas dapat terlihat bahwa ada empat kondisi proses yang terkait dua hal tersebut. Keempat kondisi tersebut adalah sebagai berikut.

1. Kondisi untuk kotak kiri atas

Kondisi ini merupakan kondisi yang baik dan dapat dikatakan kondisi yang ideal, karena pada kondisi ini proses sudah terkontrol dan kemampuan prosesnya juga sudah cukup baik. Meskipun kondisi ini sudah stabil, namun sebaiknya tetap dilakukan pengendalian proses.

2. Kondisi untuk kotak kanan atas

Kondisi ini merupakan kondisi di mana proses sudah terkontrol tetapi kemampuan proses tidak baik.

3. Kondisi untuk kotak kiri bawah

Kondisi ini merupakan kondisi di mana proses tidak terkendali, namun memiliki kemampuan proses yang baik.

4. Kondisi untuk kotak kanan bawah

Kondisi ini merupakan kondisi yang buruk karena proses tidak stabil atau tidak terkendali dan memiliki kemampuan proses yang tidak baik pula.

2.2 Aspek Analisis Multivariat

Analisis multivariat merupakan metode analisis data statistik yang memperlakukan sekelompok variabel kriteria yang saling berkorelasi sebagai satu sistem, dengan memperhitungkan korelasi atau hubungan antara variabel-variabel itu (Iriawan dan Astuti, 2006:367).

Dalam analisis multivariat, *p* pengukuran (karakteristik) pada sebuah pengamatan yang dilakukan pada setiap individu atau objek sebanyak *n* sampel dengan x_{ij} menunjukkan nilai tertentu pada variabel ke-*i*, pengamatan ke-*j* dapat ditulis sebagai berikut:

	Objek ke-1	…	Objek ke-j	…	Objek ke-n
Variabel 1	x_{11}	…	x_{1j}	…	x_{1n}
Variabel 2	x_{21}	…	x_{2j}	…	x_{2n}
.	.		.		.
.	.		.		.
.	.		.		.
Variabel i	x_{i1}	…	x_{ij}	…	x_{in}
.	.		.		.
.	.		.		.
.	.		.		.
Variabel p	x_{p1}	…	x_{pj}	…	x_{pn}

UNIVERSITAS PENDIDIKAN INDONESIA PERPUSTAKAAN

Data tersebut dapat ditulis sebagai matriks **X** dengan *p* baris dan *n* kolom, yaitu:

$$\mathbf{X}=\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1j} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2j} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{i1} & x_{i2} & \cdots & x_{ij} & \cdots & x_{in} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{p1} & x_{p2} & \cdots & x_{pj} & \cdots & x_{pn} \end{bmatrix}$$

2.3 Vektor Acak dan Matriks Acak

Vektor acak adalah vektor yang elemen-elemennya variabel acak. Matriks acak adalah matriks yang elemen-elemennya variabel acak. Nilai harapan matriks acak adalah nilai harapan dari setiap elemen-elemennya.

Misal $\boldsymbol{X} = \{x_{ij}\}$ adalah matriks acak ukuran $p \times n$. Nilai harapan dari $\boldsymbol{X}$ ditulis $E(\boldsymbol{X})$ adalah matriks $p \times n$, yaitu:

$$E(\boldsymbol{X}) = \begin{bmatrix} E(X_{11}) & E(X_{12}) & \cdots & E(X_{1n}) \\ E(X_{21}) & E(X_{22}) & \cdots & E(X_{2n}) \\ \vdots & \vdots & \ddots & \vdots \\ E(X_{p1}) & E(X_{p2}) & \cdots & E(X_{pn}) \end{bmatrix}$$

dengan

$$E(X) = \begin{cases} \int_{-\infty}^{\infty} x_{ij} f_{ij}(x_{ij}) dx_{ij} & \text{Jika } X_{ij} \text{ variabel acak kontinu dengan fungsi densitas } f_{ij}(x_{ij}) \\ \sum_{x_{ij}} x_{ij} p_{ij}(x_{ij}) & \text{Jika } X_{ij} \text{ variabel acak diskrit dengan fungsi peluang } p_{ij}(x_{ij}) \end{cases}$$

2.4 Vektor Mean dan Matriks Kovarian

Statistik adalah ukuran-ukuran yang merupakan wakil kumpulan suatu data. Pada dasarnya, statistik deskriptif memaparkan secara numerik ukuran *tendency central*, dispersi (penyebaran), distribusi suatu data, dan sebagainya. Pada metode statistika univariat, salah satu ukuran *tendency central* yaitu *mean* (nilai harapan) dan ukuran dispersi, $Var(X) = \sigma^2$, sedangkan pada metode multivariat yang menjadi ukuran *tendency central* adalah vektor *mean* yang merupakan vektor rata-rata semua variabelnya. Vektor *mean* didefinisikan sebagai berikut:

$$E(\boldsymbol{X}) = \begin{bmatrix} E(X_1) \\ \vdots \\ E(X_i) \\ \vdots \\ E(X_p) \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_i \\ \vdots \\ \mu_p \end{bmatrix} = \boldsymbol{\mu},$$

sedangkan yang menjadi ukuran dispersi adalah matriks kovarian $\sum$. Matriks kovarian didefinisian sebagai berikut :

$$\sum = E(\boldsymbol{X} - \boldsymbol{\mu})(\boldsymbol{X} - \boldsymbol{\mu})'$$

$$= E\begin{bmatrix} (X_1 - \mu_1)^2 & (X_1 - \mu_1)(X_2 - \mu_2) & \cdots & (X_1 - \mu_1)(X_p - \mu_p) \\ (X_2 - \mu_2)(X_1 - \mu_1) & (X_2 - \mu_2)^2 & \vdots & (X_2 - \mu_2)(X_p - \mu_p) \\ \vdots & \vdots & \ddots & \vdots \\ (X_p - \mu_p)(X_1 - \mu_1) & (X_p - \mu_p)(X_2 - \mu_2) & \cdots & (X_p - \mu_p)^2 \end{bmatrix}$$

$$= \begin{bmatrix} E(X_1 - \mu_1)^2 & E(X_1 - \mu_1)(X_2 - \mu_2) & \cdots & E(X_1 - \mu_1)(X_p - \mu_p) \\ E(X_2 - \mu_2)(X_1 - \mu_1) & E(X_2 - \mu_2)^2 & \vdots & E(X_2 - \mu_2)(X_p - \mu_p) \\ \vdots & \vdots & \ddots & \vdots \\ E(X_p - \mu_p)(X_1 - \mu_1) & E(X_p - \mu_p)(X_2 - \mu_2) & \cdots & E(X_p - \mu_p)^2 \end{bmatrix}$$

$$= \begin{bmatrix} E(X_1)^2 - \mu_1{}^2 & E(X_1X_2) - \mu_1\mu_2 & \cdots & E(X_1X_P) - \mu_1\mu_P \\ E(X_2X_1) - \mu_2\mu_1 & E(X_2)^2 - \mu_2{}^2 & \vdots & E(X_2X_P) - \mu_1\mu_P \\ \vdots & \vdots & \ddots & \vdots \\ E(X_PX_1) - \mu_P\mu_1 & E(X_1X_2) - \mu_2\mu_P & \cdots & E(X_P)^2 - \mu_P{}^2 \end{bmatrix}$$

$$\sum = Cov(\boldsymbol{X}) = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \vdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{bmatrix}$$

UNIVERSITAS PENDIDIKAN INDONESIA PERPUSTAKAAN

Karena $\sigma_{12} = \sigma_{21}$, maka : $\sum = Cov(\boldsymbol{X}) = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{12} & \sigma_{22} & \vdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1p} & \sigma_{2p} & \cdots & \sigma_{pp} \end{bmatrix}$

Ukuran keeratan variabel acak X_i dan X_j disebut koefisien korelasi populasi yang dinotasikan $\boldsymbol{\rho}_{ij}$. Koefisien korelasi ini didefinisikan sebagai berikut:

$$Cor(X_i, X_j) = \rho_{ij} = \frac{Cov(X_i, X_j)}{\sqrt{Var(X_i)Var(X_j)}} = \frac{\sigma_{ij}}{\sqrt{\sigma_i{}^2}\sqrt{\sigma_j{}^2}}$$

dengan σ_{ij} dinamakan kovariansi, sedangkan σ_{jj} dan σ_{i} dinamakan variansi. Sehingga matriks koefisien korelasi populasinya adalah matriks simetri berukuran $p \times p$, yang didefinisikan sebagai berikut :

$$\boldsymbol{\rho} = \begin{bmatrix} \frac{\sigma_{11}}{\sqrt{\sigma_{11}}\sqrt{\sigma_{11}}} & \frac{\sigma_{12}}{\sqrt{\sigma_{11}}\sqrt{\sigma_{22}}} & \cdots & \frac{\sigma_{1p}}{\sqrt{\sigma_{11}}\sqrt{\sigma_{pp}}} \\ \frac{\sigma_{12}}{\sqrt{\sigma_{11}}\sqrt{\sigma_{22}}} & \frac{\sigma_{22}}{\sqrt{\sigma_{22}}\sqrt{\sigma_{22}}} & \cdots & \frac{\sigma_{2p}}{\sqrt{\sigma_{22}}\sqrt{pp}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\sigma_{1p}}{\sqrt{\sigma_{11}}\sqrt{\sigma_{pp}}} & \frac{\sigma_{2p}}{\sqrt{\sigma_{22}}\sqrt{\sigma_{pp}}} & \cdots & \frac{\sigma_{pp}}{\sqrt{\sigma_{pp}}\sqrt{\sigma_{pp}}} \end{bmatrix} = \begin{bmatrix} 1 & \rho_{12} & \cdots & \rho_{1p} \\ \rho_{12} & 1 & \cdots & \rho_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{1p} & \rho_{2p} & \cdots & 1 \end{bmatrix}$$

Jadi, parameter untuk data yang memiliki karakteristik multivariat adalah $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$, dan $\boldsymbol{\rho}$.

2.5 Vektor Mean Sampel dan Matriks Variansi – Kovariansi Sampel

Ukuran yang dihitung dari kumpulan data dalam sampel dinamakan *statistik*, sedangkan ukuran yang dihitung dari kumpulan data dalam populasi dinamakan *parameter*. Menurut Herrhyanto (2003:81) penaksir titik dari sebuah *parameter* populasi adalah sebuah nilai yang diperoleh dari sampel dan digunakan sebagai penaksir dari *parameter* yang nilainya tidak diketahui yang disebut *statistik* sampel. Misalnya X adalah variabel acak dengan fungsi kepadatan peluang $f(x)$ dengan parameter populasi θ. Jika $X_1, X_2, \dots, X_n$ adalah sampel acak berukuran n dari X, maka statistik $\hat{\theta} = h(x_1, x_2, \dots, x_n)$ yang berkaitan dengan θ dinamakan *penaksir (estimator)* dari θ.

Berikut ini adalah sifat-sifat penaksir sebuah parameter populasi, diantaranya:

1) Tak bias

Penaksir $\hat{\theta}$ dikatakan *penaksir tak bias* bagi parameter θ jika:

$$E(\hat{\theta}) = \theta,$$

dan sebaliknya $\hat{\theta}$ dikatakan *penaksir bias* bagi parameter θ jika:

$$E(\hat{\theta}) \neq \theta.$$

Penaksir bias dapat diubah menjadi penaksir tak bias dengan cara mengalikan atau menambahkan konstanta tertentu di ruas kanan.

2) Variansi minimum

Dalam pembahasan penaksir sebuah parameter yang mempunyai varians minimum, kita harus membandingkan dua buah penaksir dalam hal variansinya. Kedua penaksir tersebut harus merupakan penaksir yang tak bias.

Misalkan ada dua buah penaksir tak bias $\hat{\theta}_1$ dan $\hat{\theta}_2$ untuk θ. Jika $\hat{\theta}_1$mempunyai variansi yang lebih kecil dibandingkan dengan $\hat{\theta}_2$, maka $\hat{\theta}_1$ dikatakan *penaksir tak bias bervariansi minimum*.

Sebuah penaksir tak bias akan mencapai variansi minimum di antara semua penaksir tak bias lainnya, apabila variansi dari penaksir itu minimal sama dengan batas bawah *Cramer-Rao*. Perumusan batas bawah Cramer-Rao untuk variansi dari $\hat{\theta}$ adalah

$$\frac{1}{n.E\left(\frac{d}{d\theta}lnf(x;\theta)\right)^2} = \frac{1}{-n.E\left(\frac{d^2}{d\theta^2}\ln f(x;\theta)\right)}$$

3) Konsisten

Jika $\hat{\theta}_n$ adalah penaksir untuk θ yang didasarkan pada sampel acak berukuran n, maka kita katakan $\hat{\theta}_n$adalah konsisten untuk θ, jika

$$\lim_{n\to\infty} P(|\theta_n - \theta| < \varepsilon) = 1$$

Untuk membuktikan bahwa sebuah penaksir adalah konsisten bagi parameternya bisa menggunakan *pertidaksamaan Chebyshev's* yang dirumuskan sebagai berikut:

$$P(|X - \mu_x| < k\sigma) \geq 1 - \frac{1}{k^2}$$

Jika

$$\lim_{n\to\infty} \frac{1}{k^2} = 0$$

maka dikatakan penaksir adalah konsisten bagi parameternya.

4) Statistik cukup

Statistik $T = T(X_1, X_2, \ldots, X_n)$ dikatakan statistik cukup untuk parameter $\theta \in \Theta$, jika fungsi kepadatan peluang bersyarat

$$P[X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n | T(X_1, X_2, \ldots, X_n) = t]$$

tidak bergantung pada θ.

Jika sifat 1, 2, dan 3 telah dipenuhi maka $\hat{\theta}$ sudah dapat dikatakan sebagai penaksir terbaik untuk θ.

Dalam Statistika multivariat, parameter populasi untuk ukuran *tendency central*, dan ukuran dispersi adalah vektor *mean* $\boldsymbol{\mu}$, dan matriks kovarian $\boldsymbol{\Sigma}$. Pembahasan selanjutnya akan menjelaskan mengenai penaksir populasi bagi $\boldsymbol{\mu}$ dan $\boldsymbol{\Sigma}$.

Parameter populasi untuk ukuran *tendency central* dan ukuran dispersi dalam statistika univariat adalah mean μ dan variansi σ. Sedangkan penaksir yang digunakan untuk menaksir kedua parameter tersebut adalah $\bar{x}$ dan s. Sedangkan untuk statistika multivariat, vektor *mean* sampel $\bar{\boldsymbol{X}}$ digunakan untuk menaksir vektor *mean* populasi yaitu $\boldsymbol{\mu}$, yang didefinisikan sebagai berikut :

$$\bar{\boldsymbol{X}} = \begin{bmatrix} \bar{X}_1 \\ \bar{X}_2 \\ \vdots \\ \bar{X}_n \end{bmatrix}, \text{ dengan}$$

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Sedangkan untuk menaksir matriks kovarian populasi $\boldsymbol{\Sigma}$ adalah matriks variansi-kovariansi sampel, $\boldsymbol{S}$ yang didefinisikan sebagai berkut :

$$\boldsymbol{S} = \frac{1}{n-1}\sum_{i=1}^{n} (\boldsymbol{X}_i - \bar{\boldsymbol{X}})\,(\boldsymbol{X}_i - \bar{\boldsymbol{X}})'$$

$$\boldsymbol{S} = \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1p} \\ s_{12} & s_{22} & \vdots & s_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ s_{1p} & s_{2p} & \cdots & s_{pp} \end{bmatrix}$$

Misalnya $X_1, X_2, \ldots, X_n$ adalah sampel acak yang merupakan barisan peubah acak yang memiliki distribusi bersama dengan vektor rata-rata $\boldsymbol{\mu}$ dan matriks kovarians populasi $\boldsymbol{\Sigma}$. Maka $\bar{\boldsymbol{X}}$ adalah penaksir tak bias dari $\boldsymbol{\mu}$ atau $E(\bar{\boldsymbol{X}}) = \boldsymbol{\mu}$, dan matriks varians kovarians dari $\bar{\boldsymbol{X}}$ adalah $\frac{1}{n}\boldsymbol{\Sigma}$ atau $Cov\ (\bar{\boldsymbol{X}}) = \frac{1}{n}\boldsymbol{\Sigma}$.

Bukti.

Jika rata-rata sampel $\bar{\boldsymbol{X}}$ merupakan penaksir tak bias untuk vektor rata-rata $\boldsymbol{\mu}$, maka harus dibuktikan bahwa nilai harapan dari $\bar{\boldsymbol{X}}$ sama dengan $\boldsymbol{\mu}$ atau $E(\bar{\boldsymbol{X}}) = \boldsymbol{\mu}$, yaitu :

Misalkan $\bar{\boldsymbol{X}} = (\boldsymbol{X_1}, \boldsymbol{X_2}, \ldots, \boldsymbol{X_n})/n$, maka:

$$E(\bar{\boldsymbol{X}}) = E\left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{X_i}\right)$$

$$= \frac{1}{n}E\left(\sum_{i=1}^{n}\boldsymbol{X_i}\right)$$

$$= \frac{1}{n}\left(\sum_{i=1}^{n}E(\boldsymbol{X_i})\right)$$

$$= \frac{1}{n}\left(\sum_{i=1}^{n}\boldsymbol{\mu}\right)$$

$$= \frac{1}{n}(n.\boldsymbol{\mu})$$

$$E(\bar{\boldsymbol{X}}) = \boldsymbol{\mu}$$

Jadi terbukti bahwa $E(\bar{\boldsymbol{X}}) = \boldsymbol{\mu}$. Berdasarkan sifat penaksir, $\bar{\boldsymbol{X}}$ adalah penaksir tak bias dari $\boldsymbol{\mu}$.

Matriks varians kovarians dari $\bar{\boldsymbol{X}}$ adalah $\frac{1}{n}\boldsymbol{\Sigma}$ atau $Cov\ (\bar{\boldsymbol{X}}) = \frac{1}{n}\boldsymbol{\Sigma}$.

Bukti.

$$Cov(\overline{\boldsymbol{X}}) = E(\overline{\boldsymbol{X}} - \boldsymbol{\mu})(\overline{\boldsymbol{X}} - \boldsymbol{\mu})'$$

$$Cov(\bar{X}) = E\left(\left(\frac{1}{n}\sum_{j=1}^{n}(\boldsymbol{X}_j - \boldsymbol{\mu})\right)\left(\frac{1}{n}\sum_{l=1}^{n}(\boldsymbol{X}_l - \boldsymbol{\mu})\right)'\right)$$

$$= E\left(\frac{1}{n^2}\sum_{j=1}^{n}\sum_{l=1}^{n}(\boldsymbol{X}_j - \boldsymbol{\mu})(\boldsymbol{X}_l - \boldsymbol{\mu})'\right)$$

$$Cov(\overline{\boldsymbol{X}}) = \frac{1}{n^2}\left(\sum_{j=1}^{n}\sum_{l=1}^{n} E(\boldsymbol{X}_j - \boldsymbol{\mu})(\boldsymbol{X}_l - \boldsymbol{\mu})'\right) \tag{2.1}$$

untuk $j \neq l$, setiap elemen $E(\overline{\boldsymbol{X}} - \boldsymbol{\mu})(\overline{\boldsymbol{X}} - \boldsymbol{\mu})'$ bernilai nol karena elemen-elemennya adalah kovarians antara komponen dari $\boldsymbol{X}_j$ dan komponen $\boldsymbol{X}_l$, dan independen. Oleh karena persamaan (2.2) menjadi:

$$Cov(\overline{\boldsymbol{X}}) = \frac{1}{n^2}\left(\sum_{j=1}^{n} E\,(\boldsymbol{X}_j - \boldsymbol{\mu})(\boldsymbol{X}_l - \boldsymbol{\mu})'\right). \tag{2.2}$$

Karena $\boldsymbol{\Sigma} = E(\boldsymbol{X}_j - \boldsymbol{\mu})(\boldsymbol{X}_l - \boldsymbol{\mu})'$ adalah matriks varians kovarians populasi untuk setiap $\boldsymbol{X}_j$, maka persamaan (2.3) menjadi:

$$Cov(\overline{\boldsymbol{X}}) = \frac{1}{n^2}\underbrace{(\boldsymbol{\Sigma} + \boldsymbol{\Sigma} + \cdots + \boldsymbol{\Sigma})}_{\text{sebanyak n}}$$

$$= \frac{1}{n^2}(n\boldsymbol{\Sigma})$$

$$Cov(\overline{\boldsymbol{X}}) = \frac{1}{n}\boldsymbol{\Sigma}.$$

Jadi diperoleh $Cov(\overline{\boldsymbol{X}}) = \frac{1}{n}\boldsymbol{\Sigma}$.

2.6 Distribusi Normal Multivariat

Dalam menyelesaikan beberapa masalah statistik, kebanyakan teknik-teknik yang dibicarakannya berdasarkan pada anggapan bahwa data yang dipakai adalah data yang dihasilkan dari distribusi normal multivariat. Distribusi normal multivariat merupakan perluasan dari distribusi normal univariat. Distribusi normal multivariat ini memegang peranan penting dalam analisis multivariat.

Misal X suatu peubah acak yang berdistribusi normal dengan rata-rata (μ) dan varian(σ^2), dinotasikan $X \sim N(\mu, \sigma^2)$, maka fungsi peluang dari X adalah:

$$f(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}}, \qquad -\infty < x < \infty \qquad (2.3)$$

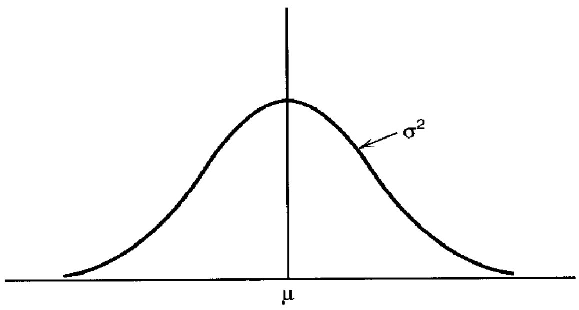
dimana: $e = 2{,}718$

$\pi = 3{,}141$

$\mu =$ rata-rata populasi

$\sigma =$ deviasi standar populasi

Berikut ini adalah gambar kurva normal univariat.



Gambar 2.3 Kurva Normal Univariat

Standarisasi dari X adalah $Z = (X - \mu)/\sigma$, maka fungsi kepadatan dari X pada persamaan (2.4) menjadi:

$$f(x;\mu,\sigma^2)=\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{1}{2}(x-\mu)\sigma^{-2}(x-\mu)}$$

$$f(x;\mu,\sigma^2)=\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{1}{2}z^2} \tag{2.4}$$

Pada kasus multivariat, misalkan $\boldsymbol{X}$ adalah vektor berukuran $(1\times p)$ dengan elemen $x_1,x_2,\dots,x_p$ yang memiliki vektor rata-rata $\boldsymbol{\mu}$ ($\boldsymbol{\mu}$ adalah vektor berukuran $(p\times 1)$ dengan elemen $\mu_1,\mu_2,\dots,\mu_p$) dan matriks varians kovarians $\boldsymbol{\Sigma}$ ($\boldsymbol{\Sigma}$ adalah matriks simetri definit positif yang berukuran $(p\times p)$), maka fungsi kepadatan dari X adalah

$$f(\boldsymbol{x};\boldsymbol{\mu},\boldsymbol{\Sigma})=f\left(x_1,x_2,\dots,x_p;\mu_1,\mu_2,\dots,\mu_p,\sigma_1^2,\sigma_2^2,\dots,\sigma_p^2\right)$$

$$=\prod_{i=1}^{p}\frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{\sigma_i^2}}e^{-\frac{1}{2}(x_i-\mu_i)\sigma_i^{-2}(x_i-\mu_i)}$$

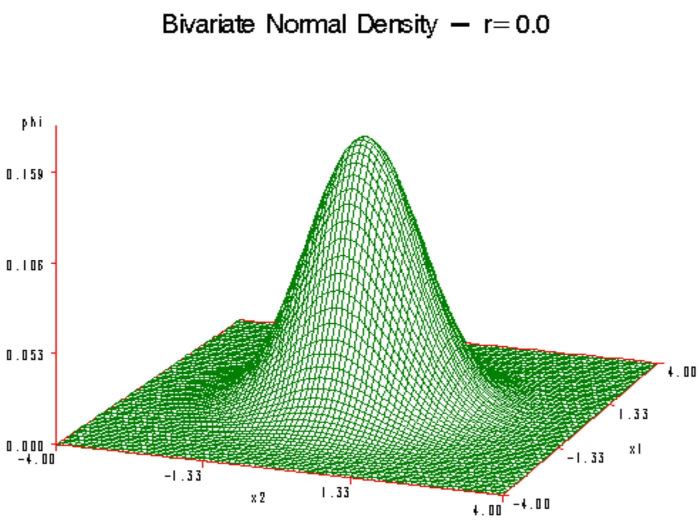
$$=\frac{1}{(2\pi)^{p/2}}\frac{1}{\sqrt{\prod_{i=1}^{p}\sigma_i^2}}e^{-\frac{1}{2}\sum_{i=1}^{p}(x_i-\mu_i)\sigma_i^{-2}(x_i-\mu_i)}$$

$$=\frac{1}{(2\pi)^{p/2}}\frac{1}{|\boldsymbol{\Sigma}|^{1/2}}e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}$$

$$f(x;\mu,\Sigma)=\frac{1}{(2\pi)^{p/2}|\boldsymbol{\Sigma}|^{1/2}}e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})},\quad -\infty<x<\infty,\quad i=1,2,\dots,p.$$

Fungsi kepadatan normal multivariat dari $\boldsymbol{X}$ dinotasikan oleh $\boldsymbol{X}\sim\mathrm{N}_p(\boldsymbol{\mu},\boldsymbol{\Sigma})$ dengan p menunjukkan banyaknya variabel bebas $\boldsymbol{X}$.

Berikut ini adalah contoh kurva distribusi normal multivariat dengan $p=2$ (distribusi normal bivariat).



Gambar 2.4 Kurva Normal Bivariat

Salah satu cara untuk menaksir kenormalan multivariat adalah dengan menggunakan plot *Chi-square*. Distribusi normal multivariat standar memiliki $\boldsymbol{\mu} = \mathbf{0}$ dan $\boldsymbol{\Sigma} = \mathrm{I}_{\boldsymbol{p}}$. $(\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})$ mendekati distribusi $\mathcal{X}_p^2$ (*Chi-square* dengan derajat bebas *p*) dengan nilai harapan *p* dan varians 2*p*.

Teorema 2.1

Misalkan $\mathbf{X}$ berdistribusi $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ dengan $|\boldsymbol{\Sigma}| > 0$ maka $(\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})$ berdistribusi $\mathcal{X}_p^2$ dengan derajat bebas *p*.

Bukti.

$\mathcal{X}_p^2$ didefinisikan sebagai distribusi dari jumlah $z_1^2 + z_2^2 + \cdots + z_p^2$ dimana $z_1 + z_2 + \cdots + z_p$ adalah variabel acak independen yang masing-masing berdistribusi $N(0,1)$. Dengan dekomposisi spektral $\boldsymbol{\Sigma}^{-1} = \sum_{i=1}^{p} \frac{1}{\lambda_i} \boldsymbol{e}_i \boldsymbol{e}_i{}'$ dimana $\sum e_i = \lambda_i e_i$ sehingga $\boldsymbol{\Sigma}^{-1}\boldsymbol{e}_i = \frac{1}{\lambda_i}\boldsymbol{e}_i$, maka diperoleh bahwa:

$$(\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) = \sum_{\boldsymbol{i=1}}^{\boldsymbol{p}} \frac{1}{\lambda_i}(\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{e}_i\boldsymbol{e}_i{}'(\boldsymbol{x}-\boldsymbol{\mu})$$

$$(\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) = \sum_{i=1}^{p} \frac{1}{\lambda_i}\left(\boldsymbol{e}_i'(\boldsymbol{x}-\boldsymbol{\mu})\right)^2$$

$$= \sum_{i=1}^{p} \left(\frac{1}{\sqrt{\lambda_i}}\boldsymbol{e}_i'(\boldsymbol{x}-\boldsymbol{\mu})\right)^2$$

$$(\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) = \sum_{i=1}^{p} \boldsymbol{z}_i^2$$

dimana $z_i = \frac{1}{\sqrt{\lambda_i}}\boldsymbol{e}_i'(\boldsymbol{x}-\boldsymbol{\mu})$. $\boldsymbol{Z} = \boldsymbol{A}(\boldsymbol{X}-\boldsymbol{\mu})$ dimana

$$\underset{p\times 1}{\boldsymbol{Z}} = \begin{bmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_p \end{bmatrix}, \qquad \underset{p\times p}{\boldsymbol{A}} = \begin{bmatrix} \frac{1}{\sqrt{\lambda_1}}\boldsymbol{e}_1' \\ \frac{1}{\sqrt{\lambda_2}}\boldsymbol{e}_2' \\ \vdots \\ \frac{1}{\sqrt{\lambda_p}}\boldsymbol{e}_p' \end{bmatrix}$$

$(\boldsymbol{X}-\boldsymbol{\mu})$ berdistribusi $N_p(\mathbf{0}, \boldsymbol{\Sigma})$.

Dengan bantuan teorema yaitu jika $\mathbf{X}$ berdistribusi $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ maka $\mathbf{AX}$ berdistribusi $N_p\left(\boldsymbol{A\mu}, \boldsymbol{A}\boldsymbol{\Sigma}\mathbf{A}'\right)$. Sehingga untuk $\boldsymbol{Z} = \boldsymbol{A}(\boldsymbol{X}-\boldsymbol{\mu})$ berdistribusi $N_p\left(\mathbf{0}, \boldsymbol{A}\boldsymbol{\Sigma}\mathbf{A}'\right)$ dimana

$$\boldsymbol{A}\boldsymbol{\Sigma}\mathbf{A}' = \begin{bmatrix} \frac{1}{\sqrt{\lambda_1}}\boldsymbol{e}_1' \\ \frac{1}{\sqrt{\lambda_2}}\boldsymbol{e}_2' \\ \vdots \\ \frac{1}{\sqrt{\lambda_p}}\boldsymbol{e}_p' \end{bmatrix} \left[\sum_{i=1}^{p} \frac{1}{\lambda_i}\boldsymbol{e}_i\,\boldsymbol{e}_i'\right] \begin{bmatrix} \frac{1}{\sqrt{\lambda_1}}\boldsymbol{e}_1' & \frac{1}{\sqrt{\lambda_2}}\boldsymbol{e}_2' & \cdots & \frac{1}{\sqrt{\lambda_p}}\boldsymbol{e}_p' \end{bmatrix}$$

$$\boldsymbol{A\Sigma}\mathbf{A}' = \begin{bmatrix} \sqrt{\lambda_1}\boldsymbol{e}_1{}' \\ \sqrt{\lambda_2}\boldsymbol{e}_2{}' \\ \vdots \\ \sqrt{\lambda_p}\boldsymbol{e}_p{}' \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{\lambda_1}}\boldsymbol{e}_1{}' & \frac{1}{\sqrt{\lambda_2}}\boldsymbol{e}_2{}' & \cdots & \frac{1}{\sqrt{\lambda_p}}\boldsymbol{e}_p{}' \end{bmatrix} = \mathbf{I}$$

Menurut teorema bahwa jika $\boldsymbol{X_1}, \boldsymbol{X_2}$ saling independen maka masing-masing akan berdistribusi $N_{p1}(\boldsymbol{\mu_1}, \boldsymbol{\Sigma_1})$ dan $N_{p2}(\boldsymbol{\mu_2}, \boldsymbol{\Sigma_2})$ dan hanya jika $Cov(\boldsymbol{X_1}, \boldsymbol{X_2}) = \mathbf{0}$. Maka menurut teorema tersebut $z_1 + z_2 + \cdots + z_p$ merupakan variabel normal yang saling independen. Sehingga $(\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-\mathbf{1}}(\boldsymbol{x}-\boldsymbol{\mu})$ berdistribusi $\mathcal{X}_p^2$ dengan derajat bebas *p*.

$(\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-\mathbf{1}}(\boldsymbol{x}-\boldsymbol{\mu})$ pada fungsi kepadatan normal multivariat merupakan jarak kuadrat dari x ke μ, atau dikenal dengan jarak *Mahalanobis* yang dinotasikan dengan d_i^2, yaitu:

$$d_i^2 = (\boldsymbol{x_i} - \overline{\boldsymbol{x}})\mathbf{S}^{-\mathbf{1}}(\boldsymbol{x_i} - \overline{\boldsymbol{x}})', i = 1,2,\ldots,n \tag{2.5}$$

Untuk mendeteksi asumsi normal multivariat tersebut dapat menggunakan plot antara jarak *Mahalanobis* (d_i^2) dan *Chi-square* ($\mathcal{X}_p^2$). Jika plot mendekati garis lurus dan lebih dari 50% nilai $d_i^2 \leq \mathcal{X}_p^2$, maka dapat dikatakan bahwa data yang dideteksi berdistribusi normal multivariat (Johnson, 1956:162).

Langkah-langkah untuk membuat plot antara jarak *Mahalanobis* dan *Chi-square* adalah sebagai berikut :

1. Tentukan nilai vektor rata-rata: $\overline{\boldsymbol{x}}$.
2. Tentukan nilai matriks varians kovarians: *S*.
3. Tentukan nilai jarak *Mahalanobis* pada setiap titik pengamatan dengan vektor rata-ratanya. Jarak *Mahalanobis* dirumuskan sebagai berikut:

$$d_i^2 = (\boldsymbol{x_i} - \overline{\boldsymbol{x}})\mathbf{S}^{-1}(\boldsymbol{x_i} - \overline{\boldsymbol{x}})', i = 1,2, \ldots, n$$

dengan $\boldsymbol{S}$ merupakan nilai matriks varians kovarians sampel untuk semua variabel bebas.

4. Urutkan nilai d_i^2 dari yang terkecil ke besar, sehingga $d_{(1)}^2 \leq d_{(2)}^2 \leq \cdots \leq d_{(n)}^2$.
5. Tentukan nilai persentil dari *Chi-square* untuk setiap d_i^2, yaitu:

$$p_i = \frac{i - 0,5}{n}, \quad i = 1,2, \ldots, n.$$

6. Nilai $\mathcal{X}_p^2$ untuk persentil diperoleh dari distribusi *Chi-square* dengan derajat bebas *p*, di mana *p* dalah banyaknya variabel bebas.
7. Buat *scatter-plot* untuk d_i^2 dan $\mathcal{X}_p^2$.
8. Jika *scatter-plot* yang dihasilkan cenderung membentuk garis lurus dan lebih dari 50% nilai $d_i^2 \leq \mathcal{X}_p^2$, maka data dapat dikatakan data berdistribusi normal multivariat (Johnson, 1956).

Plot yang dihasilkan diharapkan berbentuk menyerupai garis lurus. Jika tidak maka hal ini mengindikasikan bahwa data yang ada tidak berdistribusi normal. Jika ada satu atau dua titik yang berada jauh dari garis lurus tersebut, maka hal ini mengindikasikan bahwa adanya *outliers* atau pencilan.

Menurut Mardia (1974) dalam Rencher (2002) cara lain yang dapat dilakukan untuk memeriksa kenormalan multivariat data adalah dengan mengkaji nilai *multivariate skewnewss* ($b_{1,p}$) dan *kurtosis*nya ($b_{2,p}$). Nilai *multivariate skewnewss* ($b_{1,p}$) dan *kurtosis*nya ($b_{2,p}$) dapat dicari dengan menggunakan rumus berikut :

$$b_{1,p} = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} g_{ij}^3$$

$$b_{2,p} = \frac{1}{n} \sum_{i=1}^{n} g_{ij}^2$$

dengan $$g_{ij} = (X_i - \bar{X})' S^{-1} (X_j - \bar{X}).$$

Jika $X_1, X_2, \ldots, X_n$ dikatakan berdistribusi normal multivriat, maka :

$$z_1 = \frac{(p+1)(n+1)(n+3)}{6[(n+1)(p+1)-6]} b_{1,p}$$

berdistribusi $\mathcal{X}^2_{p(p+1)(p+2),6}$ dan

$$z_2 = \frac{b_{2,p} - p(p+2)}{\sqrt{\frac{8p(p+)}{n}}}$$

berdistribusi normal baku.

Kedua cara di atas dapat digunakan untuk memeriksa kenormalan multivariat suatu data multivariat, namun pada tugas akhir ini penulis menggunakan cara yang pertama yaitu dengan menggunakan plot antara jarak *Mahalanobis* (d_i^2) dan *Chi-square* ($\mathcal{X}_p^2$).