

BAB II

TINJAUAN PUSTAKA

2.1 Pendahuluan

Dalam bab ini akan diuraikan mengenai teori-teori yang mendukung serta berkaitan dengan metode bootstrap untuk pembentukan diagram kendali *minimax*. Uraian dimulai dengan penjelasan mengenai diagram kendali Shewhart, diagram kendali multivariat T^2 Hotelling, diagram kendali *minimax*, uji normalitas multivariat serta metode t-bootstrap untuk pembentukan diagram kendali *minimax* multivariat.

2.2 Diagram Kendali Shewhart

Diagram Kendali Shewhart dilengkapi dengan tiga buah garis mendatar yang berfungsi sebagai batas kendali. Batas maksimum dari variasi mutu yang diijinkan dinamakan *Upper Control Limit (UCL)*, sedangkan batas minimum dari variasi mutu yang diijinkan dinamakan *Lower Control Limit (LCL)*. Diantara kedua garis UCL dan LCL terdapat *Centre Line (CL)* yang merupakan pangkal dari pengkajian variasi.

Karakteristik yang diamati



Gambar 2.1

Diagram Kendali Shewhart

Batas-batas pengendalian ini dipilih sedemikian rupa sehingga apabila proses terkendali, maka hampir semua titik sampel akan jatuh diantara kedua garis. Selama titik-titik berada dalam garis batas pengendalian, maka dapat dikatakan bahwa proses

dalam keadaan terkendali. Namun apabila terdapat satu titik yang berada di luar batas kendali maka dapat dikatakan bahwa proses tidak terkendali dan diperlukan tindakan penyelidikan dan perbaikan untuk menghilangkan hal-hal yang menyebabkan proses tidak terkendali. Proses dikatakan tidak terkendali apabila adanya pencaran titik sebagai berikut :

1. Ada satu atau lebih titik berada di luar batas kendali.
2. Ada 7 titik yang berurutan terletak pada belahan sisi yang sama.
3. Dari 11 titik yang digambarkan secara berurutan, paling sedikit ada 10 diantaranya terletak pada belahan sisi yang sama.
4. Dari 14 titik yang digambarkan secara berurutan, paling sedikit ada 12 diantaranya terletak pada belahan sisi yang sama.
5. Dari 17 titik yang digambarkan secara berurutan, paling sedikit ada 14 diantaranya terletak pada belahan sisi yang sama.
6. Dari 20 titik yang digambarkan secara berurutan, paling sedikit ada 16 diantaranya terletak pada belahan sisi yang sama.

2.3 Diagram Kendali Multivariat T^2 Hotelling

Statistik T^2 Hotelling pada digram kendali multivariat digunakan untuk dapat mendeteksi tanda-tanda yang berada di luar kendali, apakah proses terkendali secara statistik atau tidak. Dalam diagram T^2 Hotelling titik pencaran yang masih berada di bawah batas kendali atas dapat kita asumsikan bahwa proses produksi terkendali, dan fungsi dari statistik T^2 Hotelling ini adalah dapat menunjukkan uji statistik yang optimal untuk mendeteksi pergeseran secara umum vektor rata-rata. Namun kelemahan dari diagram T^2 Hotelling ini yaitu tidak bisa mendeteksi dan menjelaskan karakteristik atau variabel mana yang menyebabkan proses tidak terkendali.

Jika kualitas produk terdiri dari beberapa karakteristik, variable acak ($p > 2$) misal $X = (X_1, X_2, \dots, X_p)$, maka pengendalian kualitas melalui pembentukan diagram kendali multivariat. Struktur data selengkapnya dapat kita lihat pada tabel berikut ini

Tabel 2.1
Struktur Data Pengamatan untuk Data Multivariat

No Pengamatan	Variabel				
	X ₁	X ₂	X ₃	X _p
1	X ₁₁	X ₁₂	X ₁₃	X _{1p}
2	X ₂₁	X ₂₂	X ₂₃	X _{2p}
3	X ₃₁	X ₃₂	X ₃₃	X _{3p}
⋮	⋮	⋮	⋮	⋮
m	X _{m1}	X _{m2}	X _{m3}	X _{mp}

Dan vektor rata-ratanya adalah :

$$\bar{X} = (\bar{X}_1 \quad \bar{X}_2 \quad \dots \quad \bar{X}_p) \quad \dots(2.1)$$

Dimana : $\bar{X}_j = \frac{1}{m} \sum_{i=1}^m X_{ij}$, $i = 1, 2, \dots, m$, $j = 1, 2, \dots, p$

Serta memiliki matriks kovarians sebagai berikut :

$$S = \begin{pmatrix} S_{11} & S_{12} & S_{13} & \dots & S_{1p} \\ S_{21} & S_{22} & S_{23} & \dots & S_{2p} \\ S_{31} & S_{32} & S_{33} & \dots & S_{3p} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ S_{p1} & S_{p2} & S_{p3} & \dots & S_{pp} \end{pmatrix} \quad \dots(2.2)$$

Dimana : $S = \frac{1}{m-1} \sum_{i=1}^m (X_{ij} - \bar{X}_j)(X_{ij} - \bar{X}_j)'$, $i = 1, 2, \dots, m$, $j = 1, 2, \dots, p$

Dalam mengamati proses multivariat dimana data memiliki observasi individu, bentuk statistik T^2 Hotelling untuk membentuk diagram kendali multivariat persamaan yang digunakan adalah :

$$T^2 = (X_{ij} - \bar{X}_j)' S^{-1} (X_{ij} - \bar{X}_j), \quad i = 1, 2, \dots, m \quad , \quad j = 1, 2, \dots, p \quad \dots(2.3)$$

Dimana : T_i^2 merupakan sebaran data dari karakteristik yang diamati.

Untuk diagram kendali multivariat T^2 Hotelling hanya ada satu batas kendali yaitu *Upper Control Limit* sedangkan *Lower Control Limit* nilainya pasti 0 karena T^2 Hotelling merupakan bentuk kuadrat sehingga nilainya tidak mungkin negatif.

Nilai *Upper Control Limit* yang diberikan adalah :

$$T^2 \sim \frac{p(m-1)(m+1)}{m(m-p)} F_{1-\alpha, p, m-p} \quad \dots(2.4)$$

Dimana : p = banyaknya karakteristik yang diteliti

m = banyak sampel.

Karakteristik yang diamati



Gambar 2.2

Diagram Kendali Multivariat T^2 Hotelling

2.4 Diagram Kendali *Minimax*

Diagram kendali *minimax* memonitor p karakteristik kualitas dalam bersamaan. Pada interval waktu yang tetap, sebuah sampel berukuran n diambil dan nilai dari masing-masing karakteristik kualitas ditentukan. Misalkan X_{ij} merupakan nilai dari karakteristik kualitas ke- i pada item ke- j untuk $i=\{1,2,\dots,p\}$ dan $j=\{1,2,\dots,n\}$ dan misalkan $\bar{X} = \{\bar{X}_1 \bar{X}_2 \dots \bar{X}_p\}^T$ nilai rata-rata sampel untuk setiap karakteristik kualitas. Meskipun pengamatan mungkin berkorelasi dalam setiap sampel, sampel diasumsikan saling bebas satu sama lain. Selain itu, diasumsikan \bar{X} berdistribusi normal multivariat dengan proses rata-rata vektor $\mu = \{\mu_1 \mu_2 \dots \mu_p\}^T$ dan matriks kovarians $\Sigma = \sigma_{ij}^2$, dimana σ_{ii}^2 merupakan varians dari i karakteristik kualitas dan σ_{ij}^2 merupakan kovarians antara karakteristik kualitas ke- i dan karakteristik kualitas ke- j . Struktur kovarians diasumsikan tetap stabil dari waktu ke waktu hanya menyisakan pergeseran vektor rata-rata untuk situasi yang berada di luar kendali.

Diagram kendali *minimax* membutuhkan nilai rata-rata minimum dan rata-rata maksimum dari sampel yang telah dibakukan untuk diplotkan ke dalam diagram kendali. Kemudian lakukan standarisasi dengan menggunakan persamaan sebagai berikut (Sapulveda, 1996) :

$$Z_i = \frac{\sqrt{n}(\bar{X}_i - \mu_i)}{\sigma_{ii}}, i = 1, 2, \dots, p \quad \dots(2.5)$$

dimana : Z_i = sampel terstandar rata-rata vektor.

n = banyaknya sampel

\bar{X}_i = rata-rata dari sampel untuk variabel ke-i, $i=1,2,\dots,p$

μ_i = rata-rata dari populasi untuk variabel ke-i, $i=1,2,\dots,p$

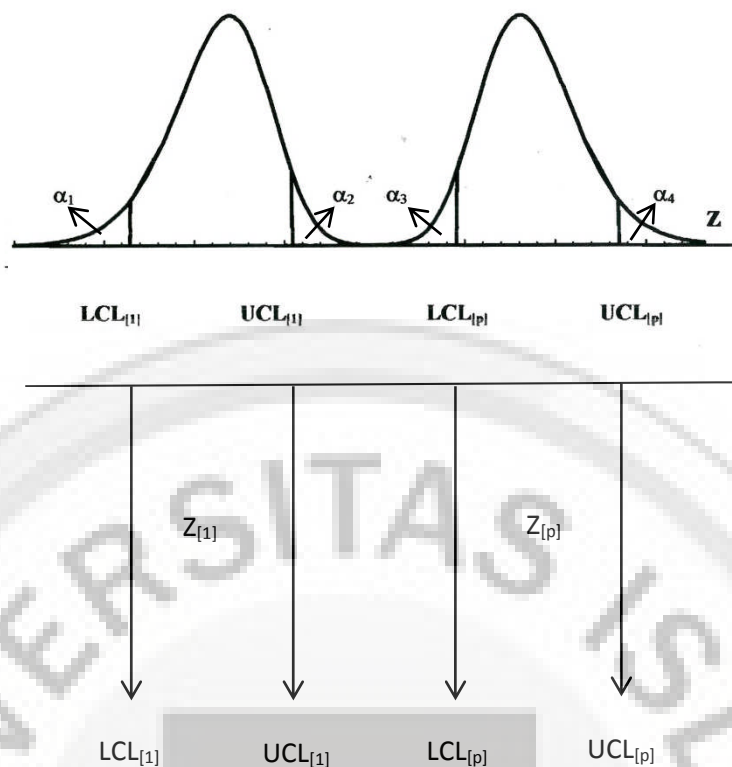
σ_{ii} = simpangan baku dari populasi untuk variabel ke-i, $i=1,2,\dots,p$

Perhatikan bahwa Z_i merupakan sampel rata-rata terstandar untuk mengukur kualitas ke-i, sehingga akan diperoleh $Z_i = \{Z_1, Z_2, \dots, Z_p\}$. Selanjutnya nilai-nilai Z_i disusun dari nilai terkecil sampai nilai terbesar, disimbolkan dengan $Z_{[i]} = \{Z_{[1]}, Z_{[2]}, \dots, Z_{[p]}\}$. Kemudian ambil nilai terkecil $Z_{[1]}$ dan nilai terbesar $Z_{[p]}$ untuk diplotkan ke dalam diagram kendali *minimax*. Ada empat batas kontrol untuk diagram kendali *minimax* diantaranya $UCL_{[1]}$ dan $LCL_{[1]}$ yaitu batas kontrol atas dan bawah masing-masing untuk $Z_{[1]}$ dan $UCL_{[p]}$ dan $LCL_{[p]}$ yaitu batas kontrol atas dan bawah masing-masing untuk $Z_{[p]}$. Perlu dicatat bahwa grafik *minimax* pada dasarnya adalah dua grafik yang saling berhubungan. Satu grafik untuk memantau dan mengendalikan $Z_{[1]}$ dan grafik yang lainnya untuk memantau dan mengendalikan $Z_{[p]}$. Grafik tersebut berhubungan karena distribusinya gabungan dari $Z_{[1]}$ dan $Z_{[p]}$ yang digunakan untuk menentukan batas kontrol. $Z_{[1]}$ dalam diagram kendali menunjukkan nilai $\alpha_{[1]}$ dan $Z_{[p]}$ menunjukkan nilai $\alpha_{[p]}$.

Batas kontrol diagram kendali *minimax* adalah sebagai berikut :

$$\begin{aligned} UCL_{[p]} &= \{\tau : \int_{-\infty}^{\tau} \dots \int_{-\infty}^{\tau} f(Z) dz = 1 - \alpha\} \\ LCL_{[p]} &= \{\tau : \int_{-\infty}^{\tau} \dots \int_{-\infty}^{\tau} f(Z) dz = \alpha\} \\ UCL_{[1]} &= \{\tau : 1 - \int_{\tau}^{\infty} \dots \int_{\tau}^{\infty} f(Z) dz = 1 - \alpha\} \\ LCL_{[1]} &= \{\tau : 1 - \int_{\tau}^{\infty} \dots \int_{\tau}^{\infty} f(Z) dz = \alpha\} \end{aligned} \quad \dots(2.6)$$

Batas kontrol di atas adalah fungsi probabilitas error tipe I (α), τ dan distribusi dari variabel yang sudah dibakukan. Batas kontrol untuk diagram kendali *minimax* sangat tergantung pada nilai α yang dipilih oleh peneliti.



Gambar 2.3

Diagram Kendali *Minimax* Multivariat

Penentuan batas kendali bergantung kepada probabilitas error tipe I atau taraf nyata (α). Diagram kendali *minimax* memiliki 4 buah nilai α . Asumsikan nilai $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = \alpha^*$, sehingga $\alpha = \alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 = 4\alpha^*$. Misalkan untuk taraf nyata 0,01, maka dalam menentukan LCL dan UCL nilai taraf nyata menjadi 0,04.

2.5 Uji Normalitas Multivariat

Seringkali sebelum melakukan pemodelan statistik, penting untuk mengetahui apakah data yang sedang diamati memenuhi asumsi di bawah distribusi tertentu (Hajarisman, 2008). Hal ini perlu dilakukan untuk analisis residu setelah mencocokkan berbagai model. Karena dalam diagram kendali *minimax* diasumsikan \bar{X} berdistribusi normal multivariat, maka harus dilakukan pengujian terlebih dahulu terhadap data yang akan dianalisis. Ada beberapa pengujian yang dapat dilakukan untuk melihat apakah data tersebut berdistribusi normal multivariat atau bukan, diantaranya adalah uji Mardia serta uji Lin dan Mudholkar.

Dalam kasus univariat, asumsi kenormalan secara individu dapat diperiksa melalui plot Q-Q atau plot P-P. Metode grafik yang analog dengan plot Q-Q atau plot P-P untuk menentukan normalitas multivariat dari segugus data yang didasarkan pada ukuran jarak kuadrat (*squared generalized distances*), yang didefinisikan oleh :

$$d^2_i = (x_i - \bar{x})'S^{-1}(x_i - \bar{x}) \quad \dots(2.7)$$

Dimana x_1, x_2, \dots, x_n adalah pengamatan sampel, sedangkan jarak kuadrat $d^2_1, d^2_2, \dots, d^2_n$ merupakan variabel acak yang berdistribusi chi-kuadrat. Walaupun jarak ini tidak saling bebas atau tidak berdistribusi chi-kuadrat yang eksak, tetapi hal ini akan sangat bermanfaat untuk memplotkannya. Hasil plot ini disebut juga sebagai plot chi-kuadrat. Adapun algoritma untuk pembentukan plot chi-kuadrat ini adalah sebagai berikut :

1. Urutkan jarak kuadrat dalam persamaan (2.7) dari yang terkecil sampai dengan yang terbesar yaitu $d^2_{(1)} \leq d^2_{(2)} \leq \dots \leq d^2_{(n)}$.
2. Plotkan pasangan $(d^2_{(j)}, \chi^2_p((j - \frac{1}{2})/n))$, dimana $\chi^2_p((j - \frac{1}{2})/n)$ adalah $100(j - \frac{1}{2})/n$ persentil dari distribusi chi-kuadrat dengan derajat bebas p

Untuk segugus data tertentu pengujian normalitas multivariat dapat dilakukan melalui SAS/STAT dan SAS/IML. Perhatikan bahwa perhitungan jarak akan dilakukan dalam SAS/IML, sedangkan proses pengurutan jarak kuadrat serta pembentukan plotnya dilakukan melalui prosedur PROC SORT dan PROC PLOT dalam SAS/STAT. Prosedur untuk pemeriksaan normalitas multivariat dilihat dari pola yang nonlinear didalam plot memberikan indikasi bahwa normalitas tidak terpenuhi.

2.6 Metode T-Bootstrap untuk Pembentukan Diagram Kendali *Minimax* Multivariat

Dalam ilmu statistika, metode bootstrap adalah suatu metode simulasi berdasarkan pada data untuk menarik kesimpulan secara statistika atau statistika inferens (Mutaqin, 2005). Metode bootstrap digunakan untuk menurunkan sifat-sifat dari penaksir (misalnya rata-rata). Metode bootstrap termasuk ke dalam kelas metode resampling.

Dalam definisi lain metode bootstrap adalah alat untuk memperkirakan distribusi sampling dari statistik. Perkiraan bootstrap dari distribusi sampling

umumnya lebih baik dari pendekatan normal berdasarkan dalil limit pusat, bahkan jika statistik tidak terstandar. Andaikan $\{x_1, \dots, x_n\}$ adalah sampel yang mengikuti distribusi normal dengan rata-rata μ dan varians σ^2 . Kemudian dari sampel tersebut dilakukan resampling dengan metode bootstrap, sehingga akan diperoleh rata-rata sebanyak B kali $\{\hat{\theta}^*_b, b = 1, 2, \dots, B\}$. Dari rata-rata setiap bootstrap akan diperoleh rata-rata dari semua penaksir bootstrap yaitu sebagai berikut :

$$\bar{\hat{\theta}}^* = \frac{\sum_{b=1}^B \hat{\theta}^*_b}{B} \quad \dots(2.8)$$

Dimana : $\hat{\theta}^*_b$ = rata-rata sampel bootstrap ke- b , $b=1, 2, \dots, B$

B = banyaknya bootstrap.

Metode bootstrap menangani kasus apabila distribusi tidak diketahui. Melalui resampling dapat memperkirakan distribusi dari variabel yang diamati. Dalam banyak kasus statistik, mengetahui distribusi dari sampel itu sangat penting. Apabila hanya terdapat sedikit informasi tentang distribusi dari sampel, maka metode bootstrap menjadi solusinya. Melalui metode bootstrap, sampel yang ada akan menjadi petunjuk untuk mengetahui distribusi sampling.

Selanjutnya akan diperoleh sampel bootstrap $X_1^*, X_2^*, \dots, X_B^*$ dan untuk masing-masing dihitung :

$$Z^*(b) = \frac{\hat{\theta}^*_b - \bar{X}}{Se \hat{\theta}^*_b}, \text{ dimana } Se \hat{\theta}^*_b = \frac{SD}{\sqrt{n}} \quad \dots(2.9)$$

dimana : $Z^*(b)$ = nilai Z bootstrap dari setiap sampel bootstrap ke- b , $b=1, 2, \dots, B$

$\hat{\theta}^*_b$ = rata-rata sampel bootstrap ke- b , $b=1, 2, \dots, B$

\bar{X} = rata-rata sampel asli untuk setiap variabel

$Se \hat{\theta}^*_b$ = standar error dari $\hat{\theta}^*_b$ untuk masing-masing sampel bootstrap

SD = standar deviasi dari $\hat{\theta}^*_b$ untuk masing-masing sampel bootstrap

n = banyaknya sampel

Jika dilakukan resampling sebanyak B kali, maka akan diperoleh rata-rata dari $Z^*(b)$ sebagai berikut :

$$\bar{Z}^*_i = \frac{\sum_{b=1}^B Z^*(b)}{B}, i = 1, 2, \dots, p \quad \dots(2.10)$$

dimana : \bar{Z}^*_i = rata-rata dari keseluruhan nilai bootstrap variabel ke-i, $i=1, 2, \dots, p$

$Z^*(b)$ = nilai bootstrap dari resampling ke-b, $b=1, 2, \dots, B$

B = banyaknya bootsrap

Selanjutnya setiap persentil untuk $Z^*(b)$ ditaksir oleh nilai $\hat{t}(\alpha)$ sehingga :

$$\#\{Z^*(b) \leq \hat{t}(\alpha)\}/B = \alpha \quad \dots(2.11)$$

Dimana $\#\{.\}$ adalah B dikali nilai α yang di order mulai dari nilai terkecil $Z^*(b)$ dalam suatu interval $\{.\}$. Misalnya terdapat 5 variabel yang dilakukan bootstrap sebanyak $B=10000$, sehingga diperoleh $Z_1^*, Z_2^*, \dots, Z_B^*$ untuk setiap variabel. Selanjutnya tentukan nilai Z^*_{Min} dan Z^*_{Max} dari nilai $Z^*(b)$ untuk setiap observasi dari kelima variabel. Kemudian urutkan nilai Z^*_{Min} dan Z^*_{Max} dari nilai terkecil sampai terbesar, nilai-nilai tersebut digunakan untuk menentukan batas kendali. Misalnya untuk taraf nyata atau $\alpha = 0,01$ LCL diperoleh pada urutan ke-200 (0,04 dibagi 2 dikali 10000) dan UCL diperoleh pada urutan ke-9800 ($[1 - (0,04 \text{ dibagi } 2)]$ dikali 10000). Maka dari nilai Z^*_{Min} dan Z^*_{Max} yang sudah diorder akan diperoleh LCL min dan LCL max pada urutan ke-200, selain itu akan diperoleh UCL min dan UCL max pada urutan ke-9800.