

Diagram Kontrol *Data Depth* untuk Memonitor Proses Multivariat

¹Mutia Laksmi, ²Suwanda, ³Lisnur Wachidah

^{1,2,3}Prodi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Islam Bandung,
Jl. Tamansari No. 1 Bandung 40116

e-mail : ¹mutia.laksmiutami@gmail.com, ²wanda_100358@yahoo.co.id, ³lisnur_w@yahoo.co.id

Abstrak. Di era globalisasi yang semakin kompetitif ini, setiap pelaku bisnis yang ingin memenangkan persaingan akan memberikan perhatian penuh pada mutu atau kualitas. Memahami variasi dan upaya mereduksinya adalah kunci dalam peningkatan kualitas. Pengendalian kualitas multivariat dalam proses produksi tidak hanya difokuskan untuk pengontrolan rata-rata, akan tetapi pengontrolan terhadap variabilitas proses juga sama pentingnya dengan pengontrolan rata-rata proses. Metode yang bisa mendeteksi kasus multivariat, seperti T^2 Hotelling, MEWMA dan lain-lain, namun metode tersebut juga memiliki kelemahan karena bergantung pada distribusi normal, sedangkan kenyataannya tidak semua proses memenuhi distribusi normal. Dalam skripsi ini akan dibahas diagram kontrol *Data Depth* dengan menggunakan kedalaman mahalnobis sebagai alternatif karena tidak memerlukan asumsi apapun tentang distribusi yang mendasari proses. Penerapannya dilakukan dengan menganalisis data kepuasan pelanggan sebuah hotel di Cihampelas pada tipe kamar standar. Hasil analisis menunjukkan data tidak berdistribusi normal pada taraf nyata 10%. Pengontrolan kepuasan pelanggan dengan menggunakan diagram kontrol *data depth* menunjukkan bahwa proses dalam keadaan *out of control*. Perkiraan pergeseran terjadi pada periode ke 21. Jika menggunakan diagram kontrol T^2 Hotelling, proses dalam keadaan *in control*. Dengan demikian, jika asumsi normalitas dilanggar dan terjadi pergeseran parameter lokasi, diagram kontrol T^2 Hotelling tidak segera memberikan *signal out of control*.

Kata Kunci: Statistik Pengendalian Proses, Diagram Kontrol, *Data Depth*.

A. Pendahuluan

Kebutuhan konsumen terhadap barang dan jasa bukan hanya dari segi kuantitas tetapi juga kualitas. Kualitas hasil produksi menjadi salah satu dasar keputusan konsumen akan kepuasan produk yang mereka beli sesuai dengan keinginan dan harapannya. Pengendalian kualitas sangat diperlukan dalam memproduksi suatu barang untuk menjaga kestabilan mutu (Leavenworth, 1998).

Statistik Pengendalian Proses (*Statistical Process Control*) merupakan sekumpulan metode yang sering digunakan dalam memonitor suatu proses. Perbedaan jumlah variabel tentu memiliki pengaruh terhadap perhitungan statistik yang harus dijalankan, dimana Statistik Pengendalian Proses (SPP) univariat lebih sederhana karena hanya melibatkan satu variabel. Tetapi pada kenyataannya, pada dunia industri banyaknya variabel yang berpengaruh terhadap kualitas suatu produk dimungkinkan lebih dari satu variabel. Karena itulah SPP multivariat perlu dilakukan. Pengontrolan proses dimana beberapa variabel berhubungan secara kolektif dikenal dengan *Multivariate Statistical Process Control* (MSPC) (Bersimis dkk, 2009).

Salah satu alat yang digunakan dalam pengendalian kualitas secara statistik pada proses produksi adalah diagram kontrol (*control chart*). Diagram kontrol dapat memberikan informasi tentang stabilitas proses. Diagram kontrol multivariat shewhart pertama kali diperkenalkan pada tahun 1947 dan dikenal sebagai diagram T^2 Hotelling. Kemudian, sejumlah diagram kontrol multivariat dirancang agar sesuai dengan situasi yang berbeda seperti diagram *Exponentially Weighted Moving Average* (EWMA) multivariat dan diagram *Cumulative Sum* (CUSUM) multivariat. Diagram-diagram kontrol klasik ini telah dikembangkan untuk memonitor vektor rata-rata. Sedangkan diagram kontrol yang lazim digunakan untuk memonitor dispersi di antaranya adalah

diagram kontrol Uji Rasio Kemungkinan (URK), *Generalized Variance* (GV) dan Rasio *Generalized Variance*. Diagram-diagram kontrol tersebut digunakan untuk memonitor karakteristik kualitas di bawah sejumlah asumsi (Alloway, 1995).

Performance dari diagram kontrol multivariat ini sangat bergantung pada hipotesis dari distribusi yang mendasari proses yaitu berdistribusi normal multivariat yang dalam prakteknya tidak semua proses berdistribusi normal. Penanganan ketika proses tidak berdistribusi normal adalah dengan transformasi *Box Cox* (Winchern dan Johnson, 2007), tetapi teknik ini tidak menjamin memperoleh hasil yang optimal. Dalam situasi ini, dikembangkan diagram kontrol dengan pendekatan nonparametrik salah satunya dengan menggunakan *Data Depth* (Liu, 1995). Diagram kontrol *Data Depth* diimplementasikan oleh Hajlaoui (2011) sebagai prosedur visual yang dirancang sebagai alternatif untuk memonitor pergeseran di lokasi dan di dispersi.

Pada penelitian ini akan dibahas mengenai diagram kontrol *Data Depth* yang dibatasi hanya untuk diagram kontrol individu. Untuk aplikasinya menggunakan data sekunder. Karakteristik mutu yang dikontrol adalah kepuasan pelanggan dari sebuah hotel pada tipe kamar standar di daerah Cihampelas Bandung.

B. Landasan Teori

Diagram Kontrol *Data Depth*

Misalkan F merupakan distribusi probabilitas dalam \mathfrak{R}^p , $p \geq 1$, diasumsikan F adalah kontinu absolut (skala pengukuran minimal interval) dan (X_1, X_2, \dots, X_n) sampel referensi yang berasal dari F untuk proses dalam keadaan *in control*, oleh karena itu jika kualitas pengamatan ke i dilambangkan dengan ω , maka

$$X'_i(\omega) = (X_{i1}, X_{i2}, \dots, X_{ip})$$

Data depth adalah cara untuk mengukur seberapa dalam atau memusatnya titik tertentu $X \in \mathfrak{R}^p$ terhadap F . *Data Depth* yang sebanding dengan diagram kontrol T^2 adalah *Data Depth* Mahalanobis yang didefinisikan sebagai:

$$D(F, x) = \frac{1}{1+(x(\omega)-\mu)' \Sigma^{-1}(x(\omega)-\mu)} \quad \dots(2.1)$$

Versi sampel dari Persamaan (2.24) diperoleh dengan mengganti μ dan Σ dengan sampel masing-masing penaksir, maka,

$$D_n(F, x) = \frac{1}{1+(x(\omega)-\bar{X})' S^{-1}(x(\omega)-\bar{X})} \quad \dots(2.2)$$

Selanjutnya D atau D_n akan digunakan untuk menunjukkan *datum depth*, dan nilai $D(F, x)$ yang lebih besar selalu menyiratkan lebih dalam atau lebih sentral dari $x(\omega)$ terhadap F .

Merujuk definisi (2.1) atau (2.2), sampel menjadi $X_{[1]}, X_{[2]}, \dots$ dari sampel yang diurut ini dapat dicari pusat untuk distribusi pengamatan. Andaikan, Pusat v adalah titik yang paling pusat, yaitu:

$$v = X_{[1]} \quad \dots(2.3)$$

Ketika Persamaan (2.3) dipunyai lebih dari satu titik, yaitu $X_{[11]}, X_{[12]}, \dots, X_{[1k]}$, menurut Liu *et al* (1999) median atau dari data adalah rata-rata dari titik terdalam, sehingga dalam kasus ini:

$$v = \frac{1}{k} \sum_{k=1}^k X_{[1k]} \quad \dots(2.4)$$

Untuk mengontrol proses selanjutnya, ambil sampel (Y_1, Y_2, \dots, Y_m) yang selanjutnya disebut sebagai sampel empiris berdistribusi G yang mencirikan kualitas dari proses pengamatan. Dalam hal $m = n$ plot DD diperoleh ketika memplotkan nilai *data depth* sampel referensi *versus data depth* sampel empiris. Tepatnya, diagram plot

DD didefinisikan oleh:

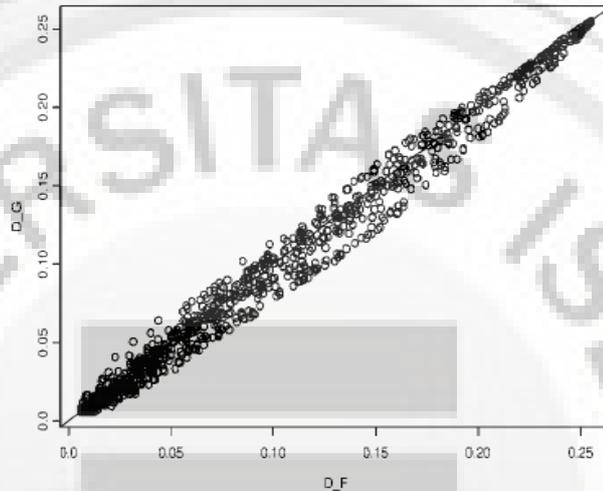
$$DD(F, G) = \{(D(F, x), D(F, y)), \text{ untuk semua } x, y \in \mathfrak{R}^p\} \quad \dots(2.5)$$

Secara umum, distribusi F jarang diketahui jadi sebagai gantinya kita menggunakan versi empiris plot DD. Jika F dan G adalah distribusi yang tidak diketahui untuk sampel \mathbf{X} dan \mathbf{Y} , maka plot DD untuk $m = n$ diperoleh ketika memplotkan:

$$DD(F_n, G_m) = \{(D_n(F, x), D_m(F, y)), \text{ untuk semua } x, y \in \mathfrak{R}^p\} \quad \dots(2.6)$$

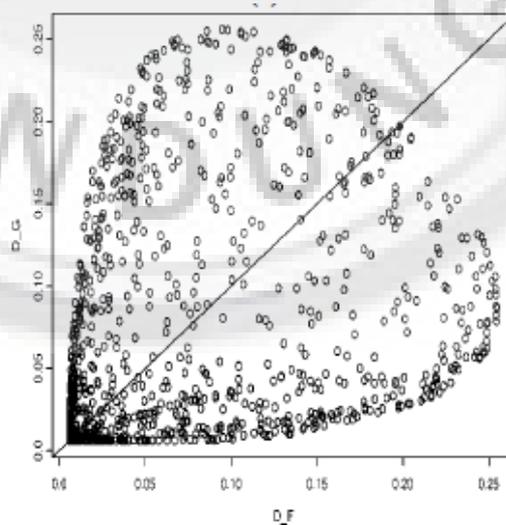
Dalam hal $m \neq n$, plot DD diperoleh dengan memplotkan:

$$DD(F_n, G_m) = \{(D_{F_n}(x), D_{G_m}(x)), x \in \{X \cup Y\}\} \quad \dots(2.7)$$



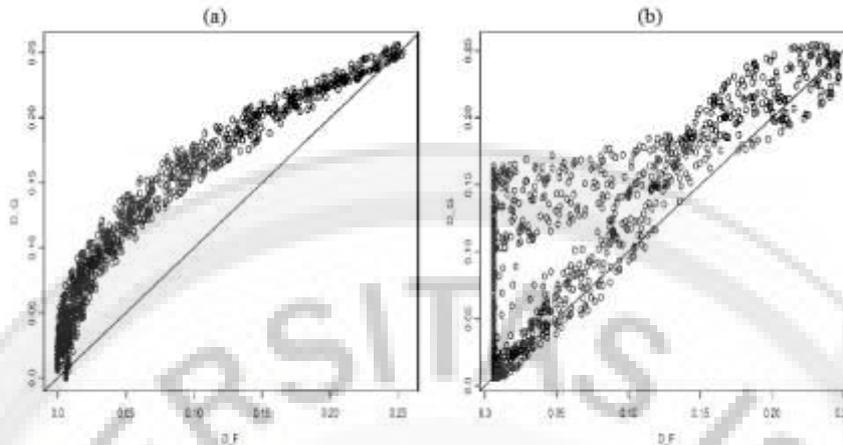
Gambar 2.1 Plot-DD dari Distribusi Identik

Jika dua distribusi adalah identik, maka plot-DD adalah segmen dari garis diagonal dari (0,0) ke (1,1) seperti Gambar (2.1). Plot yang menyimpang dari garis lurus ini menunjukkan perbedaan di antara kedua distribusi, pola penyimpangan yang berbeda pada plot berhubungan dengan perbedaan jenis variasi antara distribusi, misalnya pergeseran lokasi atau skala



Gambar 2.2 Plot-DD Pergeseran Lokasi

Gambar (2.2) di atas memperlihatkan plot-DD jika terjadi pergeseran lokasi pada kedua distribusi, plot-DD akan membentuk seperti daun dan titik puncak berada pada garis diagonal. Variasi dari bentuk daun mencerminkan besarnya lokasi bergeser seperti halnya kesimetrisan dari distribusi yang mendasari.



Gambar 2.3 Plot-DD (a) Perbedaan Skala dan (b) Perbedaan Skewness

Gambar (2.3) di atas memperlihatkan pola plot-DD (pola *half-moon* dan pola *wedge-like*) yang mengindikasikan adanya perubahan skala (Gambar 2.3a) dan perbedaan skewness (Gambar 2.3b).

Untuk mengetahui pada periode mana perkiraan terjadinya pergeseran, diagram kontrol bisa dibangun dengan menggunakan batas kontrol yang dikenal dengan Diagram kontrol *Data Depth* Shewhart. Liu (1995) tidak menggunakan batas kontrol atas karena perubahan *data depth* yang cenderung meningkat mengindikasikan adanya peningkatan kualitas produk, oleh karena itu tidak harus menjadikannya sebagai alarm. Asumsinya bahwa distribusi $D_m(F, y)$ adalah kontinu dan menuju distribusi uniform $\sim U[0,1]$, ketika $m \rightarrow \infty$ (berdasarkan hasil simulasi, konvergen pada $m = 3000$ dengan $D \approx 0,889$ untuk distribusi normal multivariat dan $m = 300$ dengan $D \approx 0,958$ untuk distribusi student multivariat) dengan nilai ekspektasi 0,5 yang akan berfungsi sebagai garis pusat (GP). Proses dikatakan *out of control* ketika nilai $D_m(F, y) < \alpha$, jadi

$$BKB = \alpha \quad \dots(2.8)$$

dengan α merupakan kekeliruan Tipe I atau probabilitas yang menyatakan proses *out of control* padahal *in control*, biasanya ditentukan dengan menggunakan diagram kontrol Shewhart dengan 3σ yang identik dengan $\alpha = 0,0027$.

Dalam hal m kecil, Hajloui (2011) telah mengusulkan BKB sebagai nilai kritis sebagai berikut:

$$BKB = \frac{1}{(p-1)[D.(F,v)+Log(m+p-1)-1]} \quad \dots(2.9)$$

di mana $D.(F, v)$ merupakan titik terpusat (*deepest point*) dari sampel referensi, dan $Log(m + p - 1)$ merupakan fungsi Log modifikasi dari derajat kebebasan sampel empiris, di mana p adalah jumlah variabel, dan m adalah ukuran sampel dari sampel empiris, kemudian dikurangi kedalaman titik terdekat ke pusat yaitu 1. Oleh karena itu, jika titik yang dihitung dengan menggunakan Persamaan (2.1) atau (2.2) lebih rendah dari nilai BKB pada Persamaan (2.8) atau Persamaan (2.9), proses yang diamati adalah di luar kendali dan penyelidikan harus dilakukan untuk mencari faktor penyebab.

C. Hasil dan Pembahasan

Data yang digunakan merupakan data sekunder yang diperoleh dari hasil penelitian Dinihayati (2012). Penelitiannya mengenai pengontrolan kualitas proses pelayanan di sebuah hotel di Cihampelas. Karakteristik mutu yang dikontrol adalah kepuasan pelanggan pada tipe kamar standar. Data diambil terdiri dari dimensi *Tangibles* (X1), *Reliability* (X2), *Responsiveness* (X3), *Assurance* (X4), *Empathy* (X5), sehingga di dalam penelitian ini banyaknya variabel $p = 5$. Data terdiri dari data referensi (*in control*) sebanyak 42 responden dan data empiris sebanyak 21 responden.

Berdasarkan hasil uji kecocokan distribusi normal multivariat dengan menggunakan nilai *skewness* dan *kurtosis*, nilai *P-value kurtosis* adalah sebesar 0,0620 yang lebih kecil dari α (10%) dapat disimpulkan jika dilihat dari nilai *kurtosis*nya bahwa data referensi distribusinya tidak mengikuti pola distribusi normal multivariat pada taraf nyata 10%.

Pengolahan data yang diamati dimulai dengan menggabungkan pengamatan \mathbf{X} dan \mathbf{Y} karena ukuran sampel tidak sama. Pada pada kasus ini $\boldsymbol{\mu}$ dan Σ tidak diketahui maka harus diperkirakan dari pengamatan periode dasar. Dari sampel referensi yang sudah terkontrol, dengan mengganti $\boldsymbol{\mu}$ dan Σ dengan penaksirnya. Kemudian hitung nilai $\bar{\mathbf{X}}$ dan \mathbf{S} dari sampel, maka didapatkan vektor rata-rata sampel referensi:

$$\bar{X}_{5 \times 1} = \frac{1}{42} \sum_{j=1}^{42} X_j = \begin{pmatrix} 71,190 \\ 71,429 \\ 68,762 \\ 70,095 \\ 72,119 \end{pmatrix}$$

dan matriks kovarians sampel referensi:

$$S_{5 \times 5} = \frac{1}{41} \sum_{j=1}^{42} (X_j - \bar{X})(X_j - \bar{X})'$$

$$= \begin{pmatrix} 43,1865 & 9,5505 & -2,7340 & -3,4820 & 6,8060 \\ 9,5505 & 35,1289 & -1,4077 & -1,6028 & 9,5575 \\ -2,7340 & -1,4077 & 35,6005 & -3,9524 & -4,1417 \\ -3,4820 & -1,6028 & 3,9524 & 39,2590 & 6,1591 \\ 6,8060 & 9,5575 & -4,1417 & 6,1591 & 35,6835 \end{pmatrix}$$

Kemudian menghitung nilai *data depth* sampel gabungan \mathbf{X} dan \mathbf{Y} terhadap distribusi F dengan menggunakan persamaan (2.2).

Selanjutnya mengitung nilai *data depth* gabungan dari sampel \mathbf{X} dan \mathbf{Y} terhadap distribusi G dengan Persamaan (2.2) dengan mengganti F oleh G , parameter $\bar{\mathbf{X}}$ dan \mathbf{S} yang digunakan yaitu dari sampel empiris. Vektor rata-rata sampel empiris distribusi G yaitu:

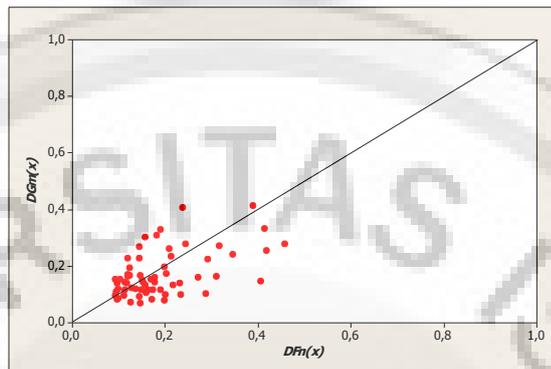
$$\bar{X}_{5 \times 1} = \frac{1}{21} \sum_{j=1}^{21} X_j = \begin{pmatrix} 69,1905 \\ 70,2381 \\ 72,4286 \\ 70,1908 \\ 69,0952 \end{pmatrix}$$

dan matriks kovarians sampel empiris distribusi G :

$$S_{5 \times 5} = \frac{1}{20} \sum_{j=1}^{21} (X_j - \bar{X})(X_j - \bar{X})'$$

$$= \begin{pmatrix} 48,2619 & 5,1524 & -0,8357 & -3,2381 & 22,3810 \\ 5,1524 & 35,7905 & 1,9429 & -15,0976 & 4,8762 \\ -0,8357 & 1,9429 & 33,7571 & 3,1643 & 14,8571 \\ -3,2381 & -15,0976 & 3,1643 & 43,9616 & 13,3310 \\ 22,3810 & 4,8762 & 14,8571 & 13,3310 & 47,6905 \end{pmatrix}$$

Oleh karena dalam kasus ini ukuran sampel dari distribusi F dan G tidak sama, maka plot DD yang dibuat adalah memplotkan distribusi F dan distribusi G dengan menggunakan Persamaan (2.7).



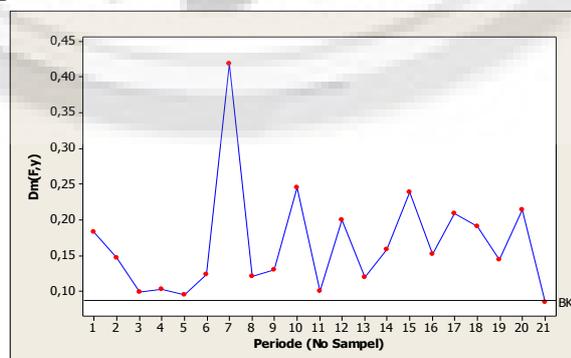
Gambar 3.1 Plot *Depth to Depth*

Dari plot DD (Gambar 3.1) di atas, terlihat bahwa kedua distribusi F dan G tidak identik, ini terlihat dari plot yang tidak membentuk garis lurus atau tidak terkonsentrasi di sepanjang garis diagonal, ini berarti proses tidak berada dalam kontrol. Plot pencar terlihat membentuk seperti daun, ini berarti telah terjadi pergeseran di lokasi seperti yang telah dijelaskan pada Gambar (2.2).

Untuk lebih mengetahui titik mana yang menyebabkan pergeseran, maka dibangun diagram kontrol *data depth shewhart*. Nilai terbesar dari *data depth* sampel referensi adalah $D_{18}(F, x(\omega)) = 0,4575$ untuk ukuran sampel sebanyak 42, ini merupakan pengamatan $X_{18} = (74 \ 74 \ 70 \ 65 \ 75)$ yang merupakan titik terpusat pada sampel F , Sehingga batas kontrol dapat dihitung sebagai berikut:

$$BKB = \frac{1}{(5 - 1)(0,4575 + \text{Log}(21 + 5 - 1) - 1)} = 0,0934$$

Kemudian menghitung *data depth* $D_m(F, y)$ yaitu *data depth* sampel empiris Y dengan menggunakan parameter dari distribusi F . Diagram kontrol *data depth Shewhart* dibangun dengan memplotkan nilai *data depth* $D_m(F, y)$ dengan nomor sampel, hasilnya disajikan pada Gambar 3.2.



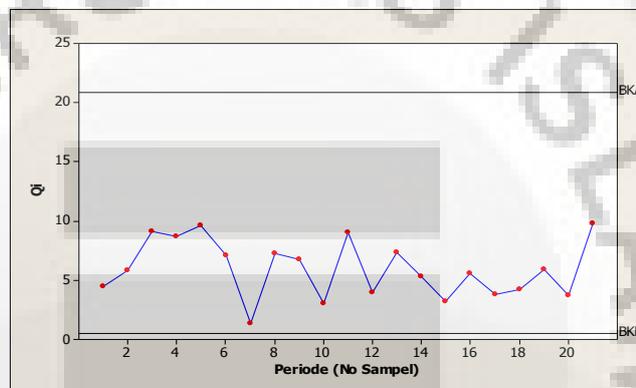
Gambar 3.2 Diagram Kontrol *data depth Shewhart*

Dari diagram kontrol di atas, terlihat bahwa dari periode ke 1 sampai dengan periode ke 20 proses dalam keadaan *in control*, tetapi pada periode ke 21 terdapat sinyal *out of control*. Dari hasil tersebut dapat dikatakan bahwa perkiraan proses tidak terkontrol terjadi pada periode ke 21. Karena dari hasil pengontrolan dengan diagram kontrol *data depth* menunjukkan telah terjadi pergeseran di lokasi maka untuk mengetahui sensitivitas hasil proses dalam hal *in* dan *out of control* perlu dibandingkan dengan diagram kontrol T^2 Hotelling. Hitung statistik Q_i data empiris dengan menggunakan Persamaan $\widehat{Q}_i = (\mathbf{X}_i - \bar{\mathbf{X}})' \mathbf{S}^{-1} (\mathbf{X}_i - \bar{\mathbf{X}})$. Adapun batas kontrol menggunakan batas kontrol untuk fase II yang dihitung dengan menggunakan $\alpha = 5\%$ perhitungannya adalah sebagai berikut:

$$BKA = \frac{p(n-1)(n+1)}{n(n-p)} F_{(1-\alpha/2; p, n-p)} = \frac{5(21-1)(21+1)}{21(21-5)} F_{(1-\frac{0,05}{2}; 5, 21-5)} = 22,9305$$

$$BKB = \frac{p(n-1)(n+1)}{n(n-p)} F_{(\alpha/2; p, n-p)} = \frac{5(21-1)(21+1)}{21(21-5)} F_{(\frac{0,05}{2}; 5, 21-5)} = 1,0023$$

diagram kontrol T^2 Hotelling disajikan pada Gambar 4.5.



Gambar 3.3 Diagram Kontrol T^2 Hotelling

Dari diagram kontrol di atas terlihat bahwa diagram kontrol T^2 Hotelling dalam kasus ini tidak bisa mendeteksi sinyal *out of control*, ini terbukti dari titik yang dihitung dengan statistik Q_i semuanya berada di daerah *in control*, dengan demikian jika asumsi normalitas dilanggar dan terjadi pergeseran parameter lokasi, diagram kontrol T^2 Hotelling tidak segera memberikan sinyal *out of control*.

D. Kesimpulan

1. Plot *Depth to Depth* (DD) adalah perbandingan grafis untuk mengetahui pergeseran lokasi dan/atau dispersi, sedangkan diagram kontrol *data depth* shewhart adalah diagram kontrol untuk mengetahui perkiraan pada periode ke berapa telah terjadi pergeseran.
2. Diagram kontrol *data depth* dapat dijadikan sebagai alternatif dari diagram kontrol T^2 Hotelling dan diagram kontrol rasio *generalized value* ketika data tidak berdistribusi normal dan terdapat *outlier*.
3. Implementasi diagram kontrol *data depth* pada kasus kepuasan pelanggan di sebuah hotel di Cihampelas pada tipe kamar standar menunjukkan proses tidak terkontrol pada periode ke 21, sehingga perlu dicari faktor penyebab. Sedangkan implementasi pada diagram kontrol T^2 Hotelling menunjukkan bahwa tidak terdeteksi adanya *out of control*.
4. Implementasi diagram kontrol *data depth* pada pembahasan sebelumnya memungkinkan untuk mengatakan bahwa diagram kontrol *data depth* lebih baik

dari diagram kontrol T^2 Hotelling karena sinyal *out of control* diagram kontrol *data depth* tidak bergantung pada tingkat kesalahan α seperti halnya pada diagram kontrol T^2 Hotelling.

Daftar Pustaka

- Alloway Jr., J.A. 1995. *Visual Evaluation of Multivariate Control Chart Assumptions*. Proceedings of the American Statistical Association, the Section on Quality and Productivity, 49-54.
- Bersimis, S., Panaretos, J., dan Psarakis, S., 2009, *Multivariate Statistical Process Control Charts and The Problem of Interpretation: A Short Overview and Some Applications in Industry*. eprint arXiv:0901.2880.
- Dinihayati, E. 2012. *Diagram Kontrol Dispersi Multivariat Rasio Generalized Variance untuk Ukuran Sampel Lebih Kecil dari Banyaknya Variabel*. Universitas Padjadjaran. Bandung
- Hajlaoui, M. 2010. *A Graphical Quality Control Procedure Using Data Depth*. *Advances & Applications in Statistics*, **19**, 97-111.
- Hajlaoui, M. 2011. *On the Charting Procedures: T^2 and DD-Diagram*. *International Journal of Quality, Statistics, and Reliability*, Article ID: 830764. <http://dx.doi.org/10.1155/2011/830764>
- Leavenworth, R.S dan Grant, L.E. 1998. *Pengendalian Mutu Statistik*. Jilid 1. Alih bahasa: H. Kandahjaya. Jakarta: Erlangga.
- Liu, R. Y. 1995. *Control charts for multivariate processes*. *Journal of the American Statistical Association*, 90(432), 1380-1387.
- Liu, R.Y., Parelius, J.M. and Singh, K. 1999. *Multivariate Analysis by Data Depth: Descriptive Statistics, Graphics and Inference*. *The Annals of Statistics*, **27**, 783-858. <http://dx.doi.org/10.1214/aos/1018031259>