

Studi Simulasi Tentang Penerapan Grafik Pengendali Berdasarkan Analisis Komponen Utama (Principal Component Analysis)

Wirayanti¹⁾, Adi Setiawan²⁾, Bambang Susanto²⁾

¹⁾ Mahasiswa Program Studi Matematika
Fakultas Sains dan Matematika Universitas Kristen Satya Wacana
Jl. Diponegoro 52-62 Salatiga, email: *wiraH9@yahoo.com*

²⁾ Dosen Program Studi Matematika
Fakultas Sains dan Matematika Universitas Kristen Satya Wacana
Jl. Diponegoro 52-62 Salatiga

Abstrak

Pengendalian kualitas secara statistik dapat dilakukan dengan menerapkan metode *Statistical Process Control (SPC)*, salah satunya dengan grafik pengendali berdasarkan Analisis Komponen Utama (*Principal Component Analysis-PCA*). Metode PCA ini merupakan suatu metode untuk mengurangi atau meringkas jumlah variabel dengan membentuk kombinasi linier yang biasa disebut komponen utama. Komponen utama dapat menjelaskan variabel asli tanpa banyak kehilangan banyak informasi. Data yang digunakan merupakan data simulasi yang dibangkitkan dengan banyaknya variabel dan ukuran sampel tertentu. Berdasarkan data, ditentukan komponen utama selanjutnya digunakan untuk membangun grafik pengendali dalam pendeteksian data yang *out of control*.

Kata kunci : Statistical Process Control, Principal Component Analysis (PCA), grafik pengendali

1. Pendahuluan

1.1 Latar belakang

Mutu suatu produk dapat menentukan lakunya produk dipasaran, sehingga dibutuhkan pengendalian kualitas agar kualitas dari produk tersebut dapat dijaga. Dalam statistik, pengendalian kualitas dapat dilakukan dengan menerapkan metode *Statistical Process Control*. Salah satunya dengan menggunakan grafik pengendali yang berdasarkan *Principal Component Analysis (PCA)*. *Principal Component Analysis (PCA)* adalah suatu analisis yang menjelaskan struktur varian-kovarian dari suatu himpunan variabel yang melalui beberapa kombinasi linear dari variabel – variabel tersebut [2]. Secara sederhana analisis komponen utama ini adalah prosedur pengurangan atau meringkas banyaknya variabel.

1.2 Perumusan masalah

Berdasarkan latar belakang, permasalahan penelitian ini akan membahas antara lain:

1. Bagaimana menerapkan grafik pengendali berdasarkan analisis komponen utama.
2. Bagaimana mengetahui komponen utama yang akan digunakan sebagai komponen atau variabel dalam grafik pengendali.

1.3 Tujuan penelitian

Tujuan dalam penelitian ini antara lain :

1. Menerapkan grafik pengendali berdasarkan analisis komponen utama.
2. Mengetahui komponen utama yang akan digunakan sebagai komponen atau variabel dalam grafik pengendali.

1.4 Manfaat penelitian

Manfaat yang dapat diperoleh dari penyusunan makalah ini adalah untuk dapat membangun grafik pengendali yang berdasarkan analisis komponen utama untuk mengetahui seberapa banyak titik sampel yang tidak terkendali atau di luar kontrol sehingga dapat dilakukan perbaikan secepatnya.

2. Metode penelitian

Data yang digunakan adalah data simulasi yang merupakan data acak berdistribusi normal yang dibangkitkan dengan jumlah variabel dua, tiga dan empat, dengan mean dan ukuran sampel tertentu. Langkah-langkah dalam analisis data dijabarkan sebagai berikut :

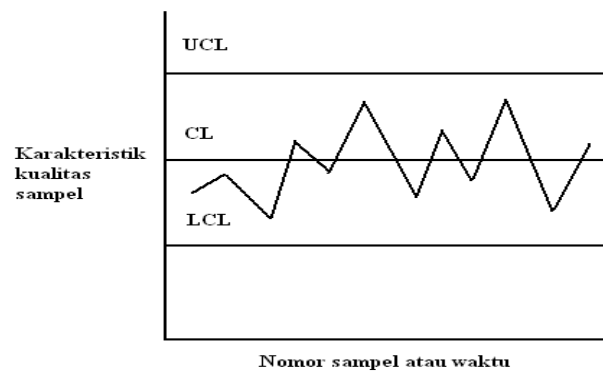
1. Membangkitkan data simulasi dengan banyaknya variabel dua, tiga dan empat, dengan mean dan ukuran sampel tertentu.
2. Mencari matriks kovariansi data simulasi, eigen value dan eigen vektor.
3. Mencari komponen utama dari data simulasi.
4. Menerapkan grafik pengendali yang berdasarkan komponen utama.
5. Mengidentifikasi titik sampel yang di luar kendali.

3. Dasar teori

3.1 Grafik pengendali

Statistical Process Control (pengendalian proses secara statistik) merupakan metode untuk mengendalikan suatu proses untuk menentukan stabilitas dan kemampuannya menghasilkan produk atau jasa bermutu [5]. SPC memiliki kemampuan untuk mendeteksi penyimpangan-penyimpangan yang terjadi dalam suatu proses baik suatu produk, proses maupun sistem, sehingga dapat dilakukan perbaikan agar dihasilkan suatu produk yang berkualitas.

Suatu alat yang digunakan dalam pengendalian kualitas secara statistik pada proses produksi disebut grafik pengendali (*Control Chart*). Dalam grafik pengendali umumnya terdiri dari batas atas (UCL), batas bawah (LCL) dan batas tengah (CL) seperti diperlihatkan seperti Gambar 1. Apabila titik-titik sampel berada di antara UCL dan LCL maka dapat dikatakan bahwa proses dalam keadaan terkendali. Akan tetapi, jika ada titik-titik sampel yang berada di luar UCL atau LCL maka proses dikatakan tidak terkendali.



Gambar 1. Grafik Pengendali

Jika μ dan σ diketahui maka UCL, LCL dan CL dari grafik pengendali adalah

$$\begin{aligned} UCL &= \mu + k\sigma \\ \text{centerline} &= \mu \\ LCL &= \mu - k\sigma \end{aligned} \quad (1)$$

dengan

μ = rata-rata (*mean*),

σ = deviasi standar,

k = kelipatan deviasi standar.

Biasanya kelipatan deviasi standar dalam teknik statistik digunakan $k = 3$, dan berkaitan dengan tingkat signifikansi (tingkat kesalahan tipe I) $\alpha=0.0027$ [3].

3.2 Principal Componen Analysis (PCA)

Analisis komponen utama merupakan suatu teknik statistik untuk mengubah dari sebagian besar variabel asli yang digunakan dan saling berkorelasi satu dengan yang lainnya menjadi satu set variabel baru yang lebih kecil dan tidak berkorelasi [4]. Setiap

pengukuran multivariat (atau observasi), komponen utama merupakan kombinasi linier dari variabel p awal. Tujuan utama analisis komponen utama ialah untuk mengurangi dimensi peubah-peubah yang saling berhubungan dan cukup banyak variabelnya sehingga lebih mudah untuk menginterpretasikan data-data tersebut [2]. Metode yang digunakan yaitu menentukan komponen utama dengan melakukan alih ragam orthogonal atau membentuk kombinasi linier $Y = A'X$ [6]. Dari sini akan dipilih beberapa komponen utama yang dapat memberikan sebagian besar keragaman total data semula.

3.3 Menentukan Komponen Utama

Komponen utama merupakan suatu kombinasi linear vektor p variabel acak X_1, \dots, X_p . Misalkan matriks $X = [X_1, \dots, X_p]$ mempunyai matriks kovariansi Σ . Dalam hal ini, Σ adalah matriks simetris dan positif tegas (*positive definite*) dengan nilai eigen $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p > 0$ dan sebutlah vektor eigen yang bersesuaian untuk setiap $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p > 0$ adalah $\vec{e}_1, \dots, \vec{e}_p$ yang saling orthogonal, dengan mencari kombinasi linier yaitu

$$Y_i = \vec{e}_i^T X = e_{i1}X_1 + e_{i2}X_2 + \dots + e_{ip}X_p, \quad i=1, 2, \dots, p. \quad (2)$$

Proporsi total variansi komponen prinsip ke- i didefinisikan sebagai

$$\frac{\lambda_k}{\lambda_1 + \lambda_2 + \dots + \lambda_p}, \quad k=1, 2, \dots, p. \quad (3)$$

Nilai e_{ki} menyatakan ukuran pentingnya variabel ke- k terhadap komponen prinsip ke- i . Secara khusus, e_{ki} menyatakan korelasi antara komponen-komponen Y_i dan variabel-variabel X_k . Hal ini dijelaskan dengan menggunakan koefisien korelasi antara komponen-komponen Y_i dan variabel-variabel X_k adalah

$$\rho_{Y_i, X_k} = \frac{e_{ki} \sqrt{\lambda_i}}{\sqrt{\sigma_{kk}}}, \quad i, k=1, 2, \dots, p. \quad (5)$$

dengan σ_{kk} adalah simpangan baku variabel ke- k [2].

4. Analisis dan Pembahasan

Dalam bab ini akan dilakukan analisis berdasarkan data simulasi.

4.1 Studi Simulasi untuk 2 variabel

Pada simulasi ini akan dibangkitkan data acak berdistribusi normal dengan ukuran sampel (*sample size*) n yang berbeda-beda yaitu $n=100$, $n=500$, $n=1000$ dan $n=5000$. Dengan menggunakan matriks kovariansi $\Sigma = \begin{bmatrix} 5 & 2 \\ 2 & 2 \end{bmatrix}$ [2] dan diperoleh eigen value yaitu $\lambda_1 = 6$ dan $\lambda_2 = 1$, sedangkan vektor eigen yaitu $e_1' = (-0.8945, -0.44721)$ dan $e_2' = (0.44721, -0.8945)$, kombinasi liniernya adalah sebagai berikut:

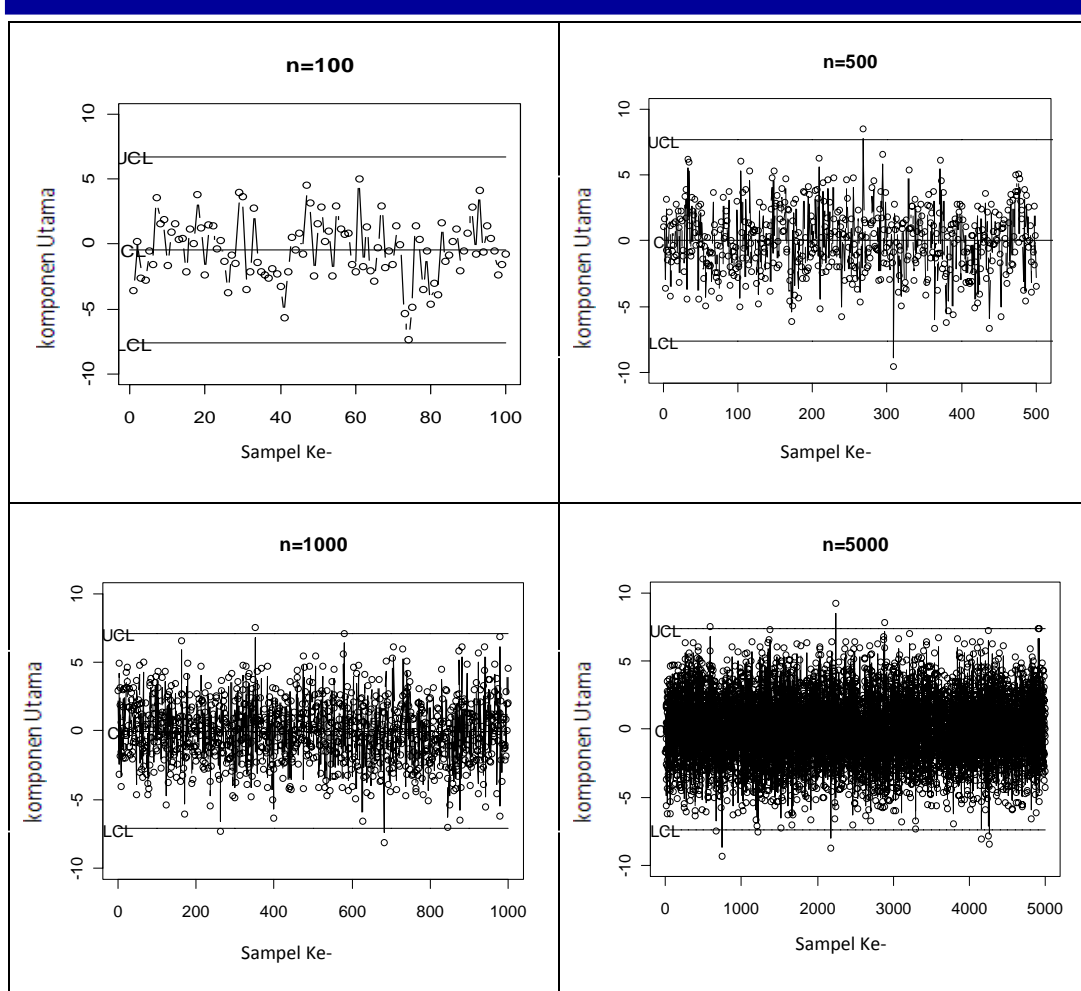
$$Y_1 = e_1' X = -0.8945X_1 - 0.44721X_2,$$

$$Y_2 = e_2' X = 0.44721X_1 - 0.8945X_2.$$

Proporsi dari Y_1 telah menjelaskan 86% dari data, dan proporsi untuk Y_2 hanya menjelaskan 14% dari seluruh data. Apabila dilihat dari korelasi antara Y_1 dan X_1 lebih mendekati -1 sebesar -0.9798 yang artinya korelasi cukup besar, sedangkan untuk Y_1 dan X_2 adalah sebesar -0.7745 yang juga relatif dekat ke -1. Oleh karena itu dapat dibangun grafik pengendali berdasarkan kombinasi liniernya Y_1 sebagai komponen utama, yang ditunjukkan pada Gambar 2. Sedangkan untuk titik yang di luar batas pengendali untuk masing-masing simulasi dengan ukuran sampel n yang berbeda yaitu $n=100$, $n=500$, $n=1000$ dan $n=5000$ diperoleh batas UCL, LCL dan CL pada Tabel 1. Selain itu, dapat dilihat rata-rata banyaknya titik yang di luar batas pengendali untuk 1000 kali pengulangan, hal ini dilakukan agar diperoleh hasil yang lebih akurat dan memperlihatkan bahwa proporsi atau prosentase lebih mendekati tingkat signifikansi $\alpha = 0.0027$ yang ditunjukkan pada Tabel 2.

Tabel 1. Titik di luar batas pengendali untuk 4 ukuran sampel yang berbeda dengan 2 variabel

No	Banyaknya n	UCL	LCL	CL	Titik di luar batas pengendali
1	100	7.69	-7.52	0.09	0
2	500	7.66	-7.64	0.01	2
3	1000	7.54	-7.49	0.02	3
4	5000	7.38	-7.39	0.01	10



Gambar 2. Grafik pengendali dua variabel dengan sampel size n berturut-turut $n=100$, $n=500$, $n=1000$ dan $n=5000$

Tabel 2. Rata-rata banyaknya titik yang berada di luar batas pengendali untuk 1000 kali pengulangan.

No	Banyaknya n	Rata-rata banyaknya titik yang berada di luar batas pengendali	Proporsi banyaknya titik yang berada di luar batas pengendali
1	100	0.228	$0.228/100 = 0.00228$
2	500	1.286	$1.286/500=0.002572$
3	1000	2.67	$2.67/1000 = 0.00267$
4	5000	13.41	$13.41/5000 = 0.002682$

4.2 Studi simulasi untuk 3 variabel

Dalam simulasi untuk 3 variabel ini akan dibangkitkan data acak berdistribusi normal dengan ukuran sampel (*sample size*) n yang berbeda-beda yaitu $n=100$, $n=500$,

$n=1000$ dan $n=5000$. Dengan menggunakan matriks kovariansi $\Sigma = \begin{bmatrix} 1 & -2 & 0 \\ -2 & 5 & 0 \\ 0 & 0 & 2 \end{bmatrix}$ [2]

dan diperoleh eigen value yaitu $\lambda_1 = 5.83$, $\lambda_2 = 2.00$ dan $\lambda_3 = 0.17$ sedangkan vektor eigen yaitu $e'_1 = [0.383, -0.924, 0]$, $e'_2 = [0, 0, 1]$ dan $e'_3 = [0.924, 0.383, 0]$, sehingga kombinasi liniernya adalah sebagai berikut:

$$Y_1 = e'_1 X = 0.383X_1 - 0.924X_2,$$

$$Y_2 = e'_2 X = X_3,$$

$$Y_3 = e'_3 X = 0.924X_1 + 0.383X_2.$$

Proporsi dari total variansi untuk komponen utama pertama telah menjelaskan 73%. Selanjutnya proporsi untuk pertama dan kedua adalah 98,3% dari total variansi populasi, dalam hal ini komponen Y_1 dan Y_2 akan bisa menggantikan ketiga variabel asli tanpa kehilangan banyak informasi. Pemilihan komponen utama sangat relatif, dapat disesuaikan dengan tingkat kepuasan yang diinginkan, apabila cukup menjelaskan seluruh total variansi dengan proporsi sebesar 73% maka digunakan komponen utama Y_1 , namun jika belum cukup dengan pemilihan tersebut dapat ditambahkan dengan komponen Y_2 sehingga proporsi untuk kedua komponen utama Y_1 dan Y_2 menjadi 98,3% .

Dilihat dari korelasi antara Y_1 dan X_1 relatif dekat 1 sebesar 0.925 yang artinya korelasi cukup besar, begitu pula untuk Y_1 dan X_2 adalah sebesar -0.998 yang juga relatif dekat -1. Dapat disimpulkan bahwa X_1 dan X_2 sama pentingnya dengan komponen utama pertama.

Dalam pembuatan grafik pengendali dapat digunakan dua cara yaitu grafik pengendali yang berdasarkan komponen utama Y_1 dan grafik pengendali yang berdasarkan dua komponen utama Y_1 dan Y_2 . Hal tersebut dikarenakan adanya tingkat kepuasan yang digunakan. Salah satu grafik pengendali yang berdasarkan komponen utama Y_1 dapat dilihat pada Gambar 3, sedangkan rata-rata banyaknya titik sampel yang

di luar batas pengendali dapat diperoleh untuk 1000 kali pengulangan seperti pada Tabel 3. Pada Tabel 3 menunjukkan bahwa proporsi dari banyaknya titik yang berada di luar batas pengendali mendekati nilai $\alpha=0.0027$. Namun, dalam pembuatan grafik pengendali dengan dua komponen utama yang dipilih, untuk menggambarkan grafik pengendali dua variabel tersebut dapat menggunakan metode yang dijelaskan pada Darmawan (2010) yaitu dengan grafik pengendali *Hotteling T²* [1].

Tabel 3. Rata-rata banyaknya titik yang berada di luar batas pengendali untuk 1000 kali pengulangan.

No.	Banyaknya n	Rata-rata banyaknya titik yang berada di luar batas pengendali	Proporsi banyaknya titik yang berada di luar batas pengendali
1	100	0.244	$0.244/100 = 0.00244$
2	500	1.33	$1.33/500=0.00266$
3	1000	2.625	$2.625/1000 = 0.002625$
4	5000	13.529	$13.529/5000 = 0.0027$

4.3 Studi simulasi dengan mean masing-masing variabel pada data kandungan Kapsul Herbal Glucoser

Pada simulasi ini akan dibangkitkan data acak berdistribusi normal dengan $n=1000$ dan mean (197.97, 148.49, 98.99, 49.51) yang diperoleh dari data kandungan Kapsul Herbal Glucoser [7] dan menggunakan matriks kovariansi

$$\Sigma = \begin{bmatrix} 8.172314 & 6.126621 & 4.091408 & 2.0513710 \\ 6.126621 & 4.593824 & 3.067336 & 1.5382090 \\ 4.091408 & 3.067336 & 2.048997 & 1.0269717 \\ 2.051371 & 1.538209 & 1.026972 & 0.5154635 \end{bmatrix}.$$

Data yang dibangkitkan mempunyai matriks kovariansi sampel yang baru yaitu

$$\Sigma = \begin{bmatrix} 8.623499 & 6.466907 & 4.319825 & 2.1653782 \\ 6.466907 & 4.850434 & 3.239552 & 1.6242007 \\ 4.319825 & 3.239552 & 2.164657 & 1.0846241 \\ 2.165378 & 1.624201 & 1.084624 & 0.5443009 \end{bmatrix},$$

eigen value dan eigen vektor

$$\lambda_1 = 16.18125, \quad e_1 = [-0.7300120, -0.5474799, -0.3657058, -0.1833236],$$

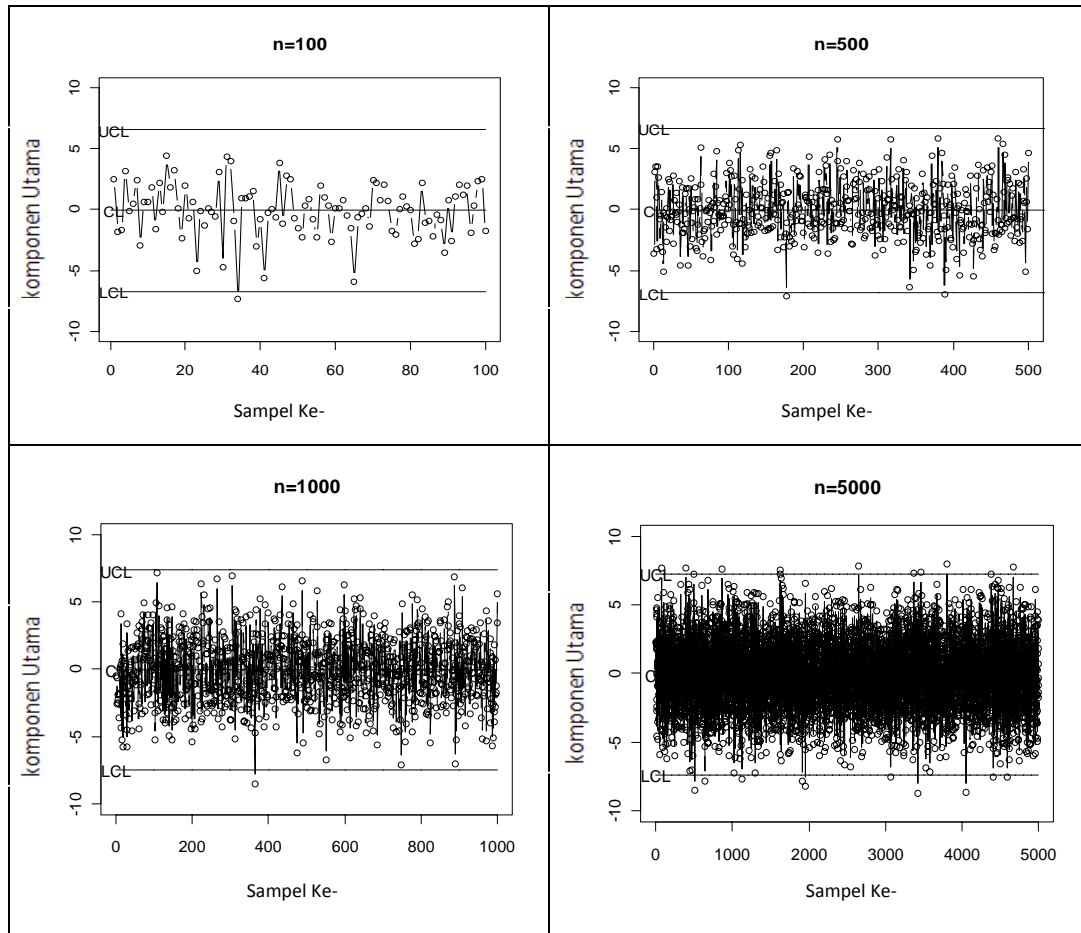
$$\lambda_2 = 0.00088, \quad e_2 = [-0.1918813, 0.4830829, -0.6296164, 0.5774043],$$

$$\lambda_3 = 0.00046 , \quad e_3 = [0.6298910, -0.3104681, -0.6645000, -0.2555124],$$

$$\lambda_4 = 0.00028 , \quad e_4 = [0.1830338, -0.6086922, 0.1681732, 0.7534654].$$

Dengan menggunakan komponen utama

$$Y_1 = -0.7300120X_1 - 0.5474799X_2 - 0.3657058X_3 - 0.1833236X_4,$$



Gambar 3. Grafik pengendali tiga variabel dengan sampel size n berturut-turut $n=100$, $n=500$, $n=1000$ dan $n=5000$

maka diperoleh rata-rata banyaknya titik di luar batas pengendali dengan proporsi untuk masing-masing ukuran sampel dengan 1000 kali pengulangan seperti pada Tabel 4. Proporsi dari banyaknya titik yang berada di luar batas pengendali menunjukkan bahwa nilai proporsinya mendekati nilai $\alpha=0.0027$.

Tabel 4. Rata-rata banyaknya titik yang berada di luar batas pengendali untuk 1000 kali pengulangan.

No	Banyaknya n	Rata-rata banyaknya titik yang berada di luar batas pengendali	Proporsi banyaknya titik yang berada di luar batas pengendali
1	100	0.231	$0.231/100 = 0.0023$
2	500	1.312	$1.312/500 = 0.002624$
3	1000	2.61	$2.61/1000 = 0.00261$
4	5000	13.418	$13.418/5000 = 0.0026836$

5. Kesimpulan dan Saran

5.1 Kesimpulan

Dari pembahasan diatas dapat dibuat kesimpulan sebagai berikut:

1. Grafik pengendali yang dibuat berdasarkan komponen utama diperoleh dari pemilihan dua variabel, tiga variabel dan empat variabel dengan ukuran sampel n yang berbeda-beda yaitu $n = 100$, $n=500$, $n=1000$ dan $n=5000$ diperoleh rata-rata titik sampel yang di luar batas pengendali dan proporsi banyaknya titik yang berada di luar batas pengendali yang mendekati tingkat signifikansi $\alpha=0.0027$.
2. Titik sampel yang berada di luar batas pengendali (di luar kontrol) memberikan arti bahwa sampel terjadi penyimpangan atau terjadi suatu kesalahan (cacat) yang mungkin diakibatkan kesalahan dalam proses produksi.

5.2 Saran

Data yang digunakan dapat berupa data karakteristik produksi dari suatu produk yang akan dilakukan pengendalian kualitasnya.

6. Daftar Pustaka

- [1] Darmawan. 2010. *Pengendalian Kualitas Frestea Green Menggunakan Grafik Pengendali Hotelling T^2 Univariat Dan Multivariat*. Salatiga: Fakultas Sains dan Matematika Universitas Kristen Satya Wacana.
- [2] Johnson, Richard. Dean Wichern. 2007. *Applied Multivariate Statistical Analysis*, 6th ed. New Jersey : Prentice Hall.

-
- [3] Montgomery, Douglas C. 1990. *Pengantar Pengendalian Kualitas Statistik*. Yogyakarta: Gadjah Mada University Press.
- [4] *Principal Component Control Chart*
<http://www.itl.nist.gov/div898/handbook/pmc/section3/pmc342.htm> (Diunduh pada 2 Oktober 2011)
- [5] Sugian O, Syahu. 2006. *Kamus Manajemen (Mutu)*. Jakarta : PT Gramedia Pustaka Utama.
- [6] Sumarga, H.1996. *Eksplorasi Data Peubah Ganda*. Salatiga: Fakultas Sains dan Matematika Universitas Kristen Satya Wacana.
- [7] Wirayanti. Setiawan, A., & Susanto, B. 2011. *Pembuatan Grafik Pengendali Berdasarkan Analisis Komponen Utama (Principal Component Analysis)*. Seminar Nasional Matematika dan Pendidikan Matematika FKIP UNS tanggal 26 November 2011.