

Evaluasi Performa dari Diagram Kontrol Multivariat berbasis Independen Principal Component Analysis (PCA)

Muhammad Ahsan dan Hidayatul Khusna

Statistics Department, Faculty of Mathematics, Computing, and Data Science
Institut Teknologi Sepuluh Nopember Surabaya

Email: ahsan4th@gmail.com

Abstract— *Diagram kontrol multivarian akan efektif ketika jumlah karakteristik kualitas yang terlibat tidak terlalu besar. Sejumlah besar karakteristik kualitas dapat mengurangi kemampuan untuk mendeteksi setiap perubahan dalam suatu proses dan juga menyebabkan masalah multikolinieritas. Untuk mengatasi masalah ini, integrasi Principal component analysis (PCA) dan diagram kontrol digunakan. PCA adalah metode yang dapat mengubah sejumlah besar variabel berkorelasi menjadi beberapa komponen utama yang tidak berkorelasi tanpa kehilangan informasi. Paper ini akan fokus untuk mengevaluasi kinerja diagram kontrol multivariat berdasarkan Independen PCA menggunakan Average Run Length (ARL) melalui studi simulasi. Dari proses simulasi dapat dilihat bahwa Independen PCA memiliki probabilitas kinerja yang mirip untuk mendeteksi false alarm untuk semua jenis korelasi dan jumlah karakteristik. Namun, kemampuan untuk mendeteksi pergeseran menurun ketika terjadi peningkatan korelasi dan jumlah karakteristik kualitas.*

Keywords: *Multivariate Control Chart, Principal Component Analysis, Average Run Length*

I. PENDAHULUAN

Statistical Process Control (SPC) adalah teknik yang banyak digunakan untuk kontrol kualitas. Teknik yang paling populer dalam SPC adalah diagram kontrol. Diagram Kontrol dapat digunakan untuk memantau suatu proses dan mendeteksi penyimpangan dari operasi. Awalnya, diagram kontrol hanya fokus pada diagram kontrol univariat, diagram kontrol ini dapat memonitor satu variabel untuk menentukan kondisi proses. Namun, dengan kemajuan teknologi dan kompleksitas proses industri, sebagian besar proses modern membutuhkan pemantauan tidak hanya satu variabel tetapi beberapa variabel. Diagram kontrol multivariat telah banyak digunakan di banyak industri untuk memantau dan mengidentifikasi proses yang ditandai oleh sejumlah besar karakteristik kualitas. Karakteristik kualitas ini sangat berkorelasi satu sama lain [1].

Keefektifan diagram kontrol multivariat konvensional akan maksimal ketika jumlah karakteristik kualitas kecil. Sejumlah besar karakteristik kualitas dapat mengurangi kemampuan untuk mendeteksi setiap perubahan dalam suatu proses dan juga menyebabkan masalah multikolinieritas [2]. Sejumlah besar karakteristik kualitas yang sangat berkorelasi satu sama lain sering terjadi dalam proses industri modern. Realitas ini dapat mempengaruhi perhitungan statistik T^2 karena matriks kovarians dapat tunggal dan dengan demikian tidak dapat diinverskan [3].

Proses perhitungan statistik T^2 melibatkan kebalikan dari matriks kovarians. Ketika sejumlah besar karakteristik proses sangat berkorelasi, matriks kovarians menjadi matriks yang hampir singular, yang sulit untuk diinverskan [4]. Principal Component Analysis (PCA) adalah teknik analisis multivariat yang mengekstraksi serangkaian variabel baru dengan memproyeksikan variabel asli ke ruang komponen utama. Variabel yang diekstraksi, yang disebut Principal Component (PC), adalah kombinasi linear dari variabel asli di mana koefisien kombinasi linear dapat diperoleh dari eigen-vektor dari matriks kovarians (atau korelasi) dari data asli [5]. PCA mengambil data dimensi tinggi, dan menggunakan dependensi antara variabel-variabel untuk mewakilinya dalam bentuk dimensi-lebih rendah yang dapat ditelusuri, tanpa kehilangan informasi. Sementara itu, karakteristik kualitas dapat berkorelasi satu sama lain, PC didefinisikan sedemikian rupa sehingga mereka ortogonal, atau independen satu sama lain, yang diperlukan untuk analisis. Integrasi teknik PCA dan diagram kontrol dapat meningkatkan kemampuan untuk mendeteksi kesalahan lebih awal dan mendeteksi perubahan dalam struktur kovarian dari variabel proses [3].

Saat ini, ada tiga jenis peta kendali berdasarkan PCA yang dikembangkan. Pertama, teknik PCA untuk konstruksi diagram T^2 disajikan menggunakan Principal Component (PC) yang pertama [6]. Selanjutnya, diagram Q adalah versi lain dari bagan kendali yang didasarkan pada PCA yang dapat dibangun dengan menggunakan residu yang diperoleh dari set p -k PC yang tersisa [5], dan diagram kontrol untuk setiap skor PC yang disebut Diagram Independen PCA [7].

Paper ini akan fokus untuk mengevaluasi kinerja diagram kontrol multivariat berbasis PCA menggunakan Average Run Length (ARL) terutama untuk diagram PCA berbasis Independen PC. Pada proses simulasi, data dibangkitkan dengan berbagai karakteristik kualitas p dan koefisien korelasi.

II. DIAGRAM KONTROL INDEPENDEN PCA

Analisis komponen utama (PCA) adalah metode untuk mengubah pengamatan dalam dataset menjadi pengamatan baru yang tidak berkorelasi satu sama lain dan memperhitungkan penurunan proporsi total varian dari variabel asli. Setiap pengamatan baru merupakan kombinasi linear dari pengamatan asli.

Untuk melakukan PCA langkah pertama yang dilakukan adalah melakukan estimasi kovarians matriks C yang berukuran $p \times p$ dari data:

$$\mathbf{x}_j, j = 1, \dots, n, \mathbf{x}_j \in \mathbb{R}^p, \sum_{j=1}^n \mathbf{x}_j = \mathbf{0}$$

yang didefinisikan sebagai berikut

$$\mathbf{C} = \frac{1}{n} \sum_{j=1}^n \mathbf{x}_j \mathbf{x}_j^T.$$

Dalam notasi matriks persamaan diatas dapat dituliskan sebagai berikut:

$$\mathbf{C} = \frac{1}{n} \mathbf{X}^T \mathbf{X}.$$

Standarisasi data sering lebih ketika variabel dalam unit yang berbeda atau ketika varians dari kolom yang berbeda dari data yang substansial. PCA dilakukan dengan mengaplikasikan metode eigen-dekomposisi dari matrix \mathbf{C} , sehingga

$$\mathbf{C} = \mathbf{V} \mathbf{\Lambda} \mathbf{V}^T,$$

dimana

$\mathbf{V} = (\mathbf{v}_1, \dots, \mathbf{v}_p)$ adalah $p \times p$ matriks *eigenvector*

$\mathbf{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_p)$ adalah diagonal matriks dari *eigenvalue*.

Nilai-nilai dari *eigenvalue* tersebut merupakan varians yang dapat dijelaskan oleh masing-masing *principal component scores* (PCs). Untuk menghitung matriks *principal component* \mathbf{Y} digunakan persamaan:

$$\mathbf{Y} = \mathbf{X} \mathbf{V} = (y_1, \dots, y_p)^T,$$

atau jika menggunakan matiks korelasi

Koordinat baru dimana data direpresentasikan dinamakan *principal component* dimana variansi data mencapai maksimum.

Skor dari principal components y_i dapat dibuat menjadi diagam untuk memonitor proses dalam bentuk diagram kontrol univariat untuk setiap i . Dengan asumsi *confidence interval* yang diinginkan adalah 99.7% maka *upper control limit* (UCL), the *center line* (CL), dan the *lower control limit* (LCL) diperoleh dengan cara sebagai berikut:

$$\begin{aligned} UCL &= 3\sqrt{\lambda_i}, \\ CL &= 0, \\ LCL &= -3\sqrt{\lambda_i}, \end{aligned}$$

dengan λ_i merupakan *eigenvalue* yang berkorespondensi dengan PC ke i .

III. AVERAGE RUN LENGTH

Average run length (ARL) adalah alat populer yang banyak digunakan oleh para peneliti untuk mengukur kinerja dari diagram kontrol. Performa dari diagram kontrol dinilai oleh dua tipe ARL yaitu, ARL_0 dan ARL_1 . ARL_0 adalah jumlah yang diharapkan dari sampel sebelum titik yang pertama kali keluar dari batas kontrol terdeteksi saat proses sebenarnya dalam kendali. ARL_1 jumlah yang diharapkan dari sampel sebelum sinyal keluar dari batas kendali diterima saat proses sebenarnya dalam keadaan diluar kendali. Untuk nilai ARL_0 yang tetap, diagram kontrol dianggap lebih efektif jika memiliki ARL_1 yang lebih kecil [8].

Saat proses terkendali:

$$ARL_0 = \frac{1}{\alpha}.$$

Jika proses diluar kendali:

$$ARL_1 = \frac{1}{1-\beta}.$$

dimana α adalah peluang suatu diagram kontrol memberikan sinyal *out of control* padahal sebenarnya berada dalam keadaan *in control*. Sedangkan β adalah peluang suatu diagram kontrol memberikan sinyal *in control* padahal sebenarnya berada dalam keadaan *out of control*. Dengan demikian suatu diagram kontrol diharapkan memiliki ARL_0 sebesar mungkin dan ARL_1 yang sangat kecil.

IV. HASIL DAN PEMBAHASAN

A. Setup Studi Simulasi

Proses simulasi akan digunakan dalam makalah ini untuk mengevaluasi kinerja diagram kontrol Independen PCA untuk sejumlah karakteristik kualitas yang berbeda. Evaluasi untuk diagram kontrol Independen PCA hanya melibatkan komponen pertama. Data akan dibangkitkan mengikuti distribusi Normal Multivariat dengan korelasi yang berbeda. Jumlah karakteristik kualitas akan dibagi menjadi lima kategori, yaitu 3, 5, 10, 20 dan 50. Akhirnya, kinerja dari masing-masing diagram kontrol akan dievaluasi menggunakan Average Run Length untuk semua kategori dengan pergeseran rata-rata δ adalah 0,1 dengan $\alpha = 0,0027$ mengacu pada tiga sigma. Jumlah *run* untuk simulasi ini adalah 30 dengan 1000 kali menghasilkan data untuk setiap run.

B. Hasil Studi Simulasi

Hasil simulasi untuk diagram kontrol Independen PCA disajikan oleh tabel dan gambar di bawah ini. Simulasi ini menggunakan komponen pertama karena paling variabilitas pada data yang ditangkap oleh komponen pertama. Berdasarkan hasil simulasi, nilai ARL^0 untuk semua jenis korelasi dan jumlah karakteristik sekitar 370 yang sesuai dengan ARL^0 untuk tiga sigma. Dengan demikian, diagram kontrol kepala sekolah independen memiliki kinerja yang sama ketika proses ini mengendalikan semua jenis korelasi dan jumlah karakteristik.

Tabel 1 Performa diagram untuk $\rho=0$

δ	Number of Quality Characteristics p						
	2	3	4	5	10	20	50
0	341.8	332.82	349.95	344.67	352.68	352.89	344.85
0.1	234.37	226.09	226.98	232.96	230.52	230.07	227.73
0.2	161.89	161.69	161.86	159.28	158.51	155.72	156.05
0.3	114.75	116.23	120.77	115.99	115.8	113.99	114.56
0.4	86.17	87.57	89.06	89.29	86.89	87.55	90.66
0.5	69.56	68.85	70.57	69.54	72.41	68.14	68.98
0.6	54.85	55.66	57.92	55.21	56.79	55.68	57.91
0.7	44.62	46.48	47.05	46.62	46.23	46.53	46.04
0.8	39.47	38.38	39.61	40.83	39.48	38.67	39.66
0.9	32.57	33.92	33.24	33.24	32.89	33.73	34.11

Tabel 2 Performa diagram untuk $\rho=0.3$

δ	Number of Quality Characteristics p						
	2	3	4	5	10	20	50
0	342.8	342.5	341.5	352.9	350.8	341.9	345.2
0.1	258.9	271.8	278.2	294.8	313.5	322.1	331.4
0.2	190.4	206.9	231.7	241.4	283.3	309.7	325.5
0.3	147.9	167.2	183.6	203.9	247.8	292.4	324.9
0.4	117.7	137.4	152.5	174.1	216.8	271.1	311.9
0.5	93.39	116.4	130.7	148.1	205.7	258	308.4
0.6	75.37	94.31	110.9	128.7	183.6	241.3	298.2
0.7	63.14	81.02	97.22	110.5	168.5	224.3	277.6
0.8	55.17	72.6	83.6	96.55	148.7	209.8	292.6
0.9	45.89	63.42	72.48	86.28	143.9	201.4	277.6

Tabel 3 Performa diagram untuk $\rho=0.5$

δ	Number of Quality Characteristics p						
	2	3	4	5	10	20	50
0	349.8	344.62	342.25	342.97	344.37	353.95	345.7
0.1	269.31	285.1	305.89	303.04	324.62	327.8	336.97
0.2	205.83	236.31	252.78	268.53	294.9	320.19	343.61
0.3	162.51	196.98	216.88	230.76	275.7	304.85	324.92
0.4	136.64	155.08	186.21	209.44	257.3	292.83	334.69
0.5	106.33	132	160.61	178.93	240.06	285.38	327.32
0.6	86.72	120.42	146.82	159.7	227.12	269.29	323.45
0.7	73.3	102.54	122.84	146.59	206.46	263.02	297.49
0.8	64.18	88.7	110.89	131.25	196.79	255.33	302.99
0.9	57.84	78.26	98.86	118.09	184.2	237.2	296.31

Tabel 4 Performa diagram untuk $\rho=0.7$

δ	Number of Quality Characteristics p						
	2	3	4	5	10	20	50
0	347.7	343.1	346.2	348.3	350.2	348.6	353.8
0.1	278.4	283	301.4	312.3	330.4	332.2	347.2
0.2	214.8	246.1	270.5	277.6	306.1	324.9	344.2
0.3	180.4	215.3	232.4	245.8	297.5	316	339.9
0.4	139.2	180.2	211.1	233	276.5	308.6	339.8
0.5	118.9	157.3	181.1	206.3	268.5	302.1	333.8
0.6	100.7	136.7	167	182.1	248.1	298.8	320.2
0.7	84.83	120.6	151.5	165.3	242.2	288.5	317.6
0.8	75.45	104.6	132.3	152.2	218.4	267.6	313.1
0.9	67.07	93.51	120.4	145.3	212.8	264.1	312.4

Tabel 1 menunjukkan ARL_1 yang dihasilkan dari proses simulasi ketika $\rho = 0$. Untuk setiap jumlah karakteristik kualitas, hasil simulasi menghasilkan nilai ARL_1 hampir sama yang menunjukkan kesamaan kinerja diagram kontrol ketika tidak ada korelasi terjadi. Nilai ARL_1 mulai meningkat ketika jumlah karakteristik kualitas meningkat ditunjukkan oleh Tabel 2. Nilai ARL_1 yang lebih besar ditunjukkan pada Tabel 3 dengan $\rho = 0,3$. Dalam hal ini, ARL_1 memiliki pola yang sama dengan kasus sebelumnya, namun nilainya sedikit lebih besar. Tabel 4 dan Tabel 5 mengilustrasikan ARL_1 tetap menghasilkan kenaikan nilai seiring dengan meningkatnya jumlah karakteristik kualitas. Dari hasil ini dapat disimpulkan, nilai ARL_1 untuk diagram kontrol independen PCA akan meningkat ketika sejumlah besar karakteristik kualitas dan korelasi tinggi terjadi. Ini terjadi karena batas kontrol akan menjadi lebih luas ketika jumlah karakteristik kualitas meningkat dan korelasi menjadi lebih tinggi. Batas kontrol dari diagram kontrol independen PCA tergantung pada nilai eigen masing-masing komponen. Namun, karena jumlah karakteristik kualitas dan korelasi meningkatkan nilai-nilai

eigen dari komponen pertama yang digunakan untuk simulasi ini menjadi lebih besar. Akibatnya, batas kontrol yang lebih besar dapat mengurangi kemampuan kontrol mendeteksi hingga bergeser dalam rata-rata.

Tabel 5 Performa diagram untuk $\rho=0.9$

δ	Number of Quality Characteristics p						
	2	3	4	5	10	20	50
0	350.08	351.66	349.35	350.8	342.93	352.22	342.79
0.1	288.94	297.86	307.05	313.34	328.58	336.9	343.76
0.2	232.91	252.42	279.28	287.67	315.01	328.89	342.62
0.3	189.93	226.06	253.6	272.36	303.55	323.53	332
0.4	154.39	201.99	227.36	243.26	291.46	320.24	330.59
0.5	132.01	178.24	203.67	222.76	281.24	307.98	336.94
0.6	110.53	153.23	183.26	214.24	253.84	302.39	330.92
0.7	97.08	141.5	171.91	185.86	248.01	293.36	323.83
0.8	88.01	123.96	151.38	180.38	241.47	286.46	326.11
0.9	76.55	111.71	142.19	164.1	237.09	278.59	324.14

Gambar 1.a menunjukkan grafik ARL untuk $\rho = 0$, ARL dari semua jumlah karakteristik kualitas yang ditumpuk menjadi satu garis. Hasil ini menunjukkan kinerja identik dari diagram kontrol komponen utama independen ketika korelasi antara kualitas karakteristik adalah 0. Hasil ini mengungkapkan bahwa kemampuan grafik kontrol komponen utama independen adalah sama dalam kasus tidak ada korelasi. Di sisi lain, kemampuan diagram kontrol komponen utama independen untuk mendeteksi pergeseran dalam proses menjadi lebih lambat ketika ada korelasi antara variabel yang dikonfirmasi pada Gambar 1.b. Kemampuan untuk mendeteksi shift akan berkurang dengan meningkatnya jumlah karakteristik kualitas. Gambar 1.c menggambarkan pola yang sama dengan nilai ARL yang lebih besar yang membuat plot menjadi lebih rata. Plot yang lebih datar menunjukkan kemampuan untuk mendeteksi pergeseran akan berkurang ketika jumlah karakteristik kualitas yang terlibat dan meningkat. Kemampuan untuk mendeteksi shift akan berkurang selama peningkatan korelasi yang dikonfirmasi pada Gambar 1.d dan 1.e. Itu bisa dilihat dari plot yang menjadi lebih rata.

V. KESIMPULAN

Dalam paper ini, studi Simulasi digunakan untuk mengevaluasi kinerja diagram kontrol Multivariat berdasarkan PCA. Simulasi dilakukan dengan menggunakan berbagai skenario dan pendekatan. Hasil simulasi menunjukkan bahwa grafik kontrol Independen PCA memiliki probabilitas kinerja yang sama untuk mendeteksi *false alarm* untuk semua jenis korelasi dan jumlah karakteristik. ARL_1 dari diagram kontrol Independen PCA akan terus meningkat seiring meningkatnya jumlah karakteristik kualitas. Kemampuan dari diagram kontrol Independen PCA dalam mendeteksi pergeseran akan berkurang selama peningkatan korelasi dan jumlah karakteristik kualitas.

DAFTAR PUSTAKA

[1] P. Phaladiganon, S. B. Kim, V. C. P. Chen, and W. Jiang, "Principal component analysis-based control charts for multivariate nonnormal distributions," *Expert Systems with Applications*, vol. 40, no. 8, pp. 3044–3054, 2013.

[2] D. Montgomery, *Introduction to statistical quality control*. New York, 2009.

[3] T. Kourti, "Application of latent variable methods to process control and multivariate statistical process control in

industry,” *International Journal of Adaptive Control and Signal Processing*, vol. 19, no. 4, pp. 213–246, 2005.

[4] R. L. Mason and J. C. Young, *Multivariate Statistical Process Control with Industrial Application*, 1st ed. Philadelphia: American Statistical Association and Society for Industrial and Applied Mathematics, 2002.

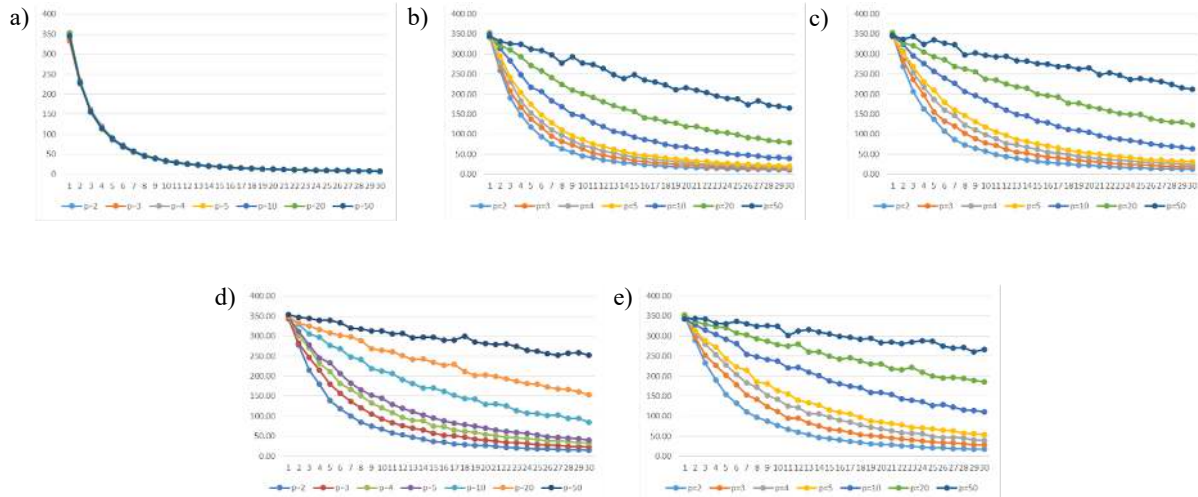
[5] J. E. Jackson and G. S. Mudholkar, “Control Procedures for Residuals Associated with Principal Component Analysis,” *Technometrics*, vol. 21, no. 3, pp. 341–349, 1979.

[6] J. E. Jackson, “Quality Control Methods for Several Related

Variables,” *Technometrics*, vol. 1, no. 4, pp. 359–377, 1959.

[7] J. E. Jackson, *A user’s guide to principal components*, vol. 587. John Wiley & Sons, 2005.

[8] Z. Wu, J. Jiao, M. Yang, Y. Liu, and Z. Wang, “An enhanced adaptive CUSUM control chart,” *IIE Transactions (Institute of Industrial Engineers)*, vol. 41, no. 7, pp. 642–653, 2009.



Gambar 1 Grafik ARL untuk: a) $\rho=0$, b) $\rho=0.3$, c) $\rho=0.5$, d) $\rho=0.7$, e) $\rho=0.9$