UM ESTUDO COMPARATIVO ENTRE O DESEMPENHO DE GRÁFICOS DE CONTROLE ESTATÍSTICO MULTIVARIADOS COM A APLICAÇÃO DA ANÁLISE DE COMPONENTES PRINCIPAIS

Elisa Henning
UDESC
Campus Universitário Professor Avelino Marcante Joinville - SC Brasil
dma2eh@joinville.udesc.br

Custódio Cunha Alves
UNIVILLE
Campus Universitário Joinville - SC Brasil
custodio@univille.net

Leandro Zvirtes
UDESC
Campus Universitário Professor Avelino Marcante Joinville - SC Brasil
dep2lz@joinville.udesc.br

Nillo Gabriel Alves de Araujo
UDESC
Campus Universitário Professor Avelino Marcante Joinville - SC Brasil
nilloaraujo@yahoo.com.br

RESUMO

Os métodos estatísticos multivariados baseados na projeção de dados tais como Análise de Componentes Principais (ACP), representam ferramentas adequadas para a reduzir a dimensão dos dados e melhorar o aproveitamento das informações disponíveis. Os gráficos de controle multivariados não são suficientemente robustos para o monitoramento de um grande número de variáveis. Este artigo apresenta um estudo comparativo dos gráficos T_2 de Hotelling e MCUSUM para uma série de dados com oito variáveis, considerando a aplicação da análise de componentes principais**.**

PALAVRAS-CHAVE: Gráfico MCUSUM. Gráfico T^2 de Hotelling. Análise de componentes principais. Estatística.

ABSTRACT

The multivariate statistical methods based on projection of data, such as Principal Components Analysis (PCA), represent appropriate tools to reduce the size of the data and better improve the use of available information. Multivariate control charts are not robust enough to monitor a large number of variables. This article presents a comparative study of the Hotelling T2 and Multivariate Cumulative Sum (MCUSUM) charts for a data set with eight variables, considering the application of Principal Components Analysis.

KEYWORDS: MCUSUM chart; Hotelling T^2 chart; Principal Component Analysis. Statistics.

1. Introdução

Em análise estatística de dados multivariados o acompanhamento de um problema pode ter início a partir da redução substancial do número de variáveis. Com isso, é possível descrever com precisão os valores das variáveis de um conjunto de dados a partir de um pequeno subconjunto destas, o que reduz de forma significativa a dimensão do problema à custa de uma pequena perda de informações (PEÑA, 2002).

Segundo Jackson (1991), a análise de componentes principais é utilizada na indústria para o controle estatístico de processos multivariados que envolvam um conjunto de dados com grande número de variáveis correlacionadas, tais como os processos químicos industriais.

A aplicação de métodos multivariados que permitem a redução das dimensões, entre eles a Análise de Componentes Principais (ACP), aparece como ferramenta alternativa para o melhor aproveitamento da informação disponível. Com este método estatístico, é possível considerar todas as variáveis originais no tratamento estatístico e visualizar os dados em duas ou três dimensões. Assim, pode-se representar a maior parte da variância, ou seja, a melhor dispersão dos pontos em relação às componentes principais.

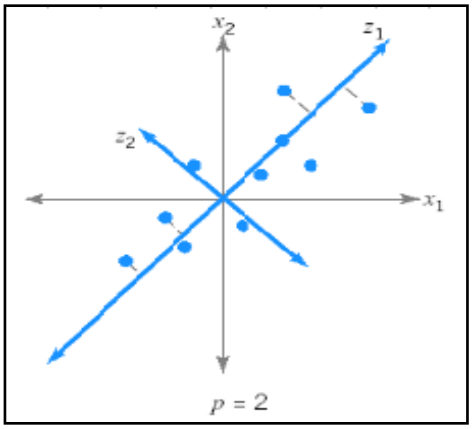
Este trabalho está assim estruturado: a seção 2 descreve a análise de componentes principais, a seção 3 trata do controle estatístico de processos multivariado, abrangendo os gráficos T^2 de Hotelling e MCUSUM, a seção 4 contém os procedimentos metodológicos e na seção 5 os resultados e análise destes. Para finalizar, na seção 6, estão as conclusões e sugestões de trabalhos futuros.

2. Análise de Componentes Principais

A análise de componentes principais é um método estatístico multivariado que visa transformar um grupo de variáveis correlacionadas em variáveis independentes, denominadas componentes principais. O objetivo é facilitar a interpretação dos dados (RENCHER, 2002). Estas componentes são combinações lineares das variáveis originais (1) e representam suas projeções nas direções de máxima variabilidade dos dados.

$$\begin{aligned} z_1 &= c_{11}x_1 + c_{12}x_2 + \ldots + c_{1p}x_p \\ z_2 &= c_{21}x_1 + c_{22}x_2 + \ldots + c_{2p}x_p \\ &\vdots \\ z_p &= c_{p1}x_1 + c_{p2}x_2 + \ldots + c_{pp}x_p \end{aligned} \qquad (1)$$

Geometricamente, as variáveis das componentes principais z_1, $z_{2,\ldots,}$ z_p são os eixos do novo sistema de coordenadas obtido pela rotação do sistema das variáveis originais. Os novos eixos representam as direções de máxima variabilidade. A informação contida no conjunto completo das *p* componentes principais corresponde àquela contida no conjunto completo de todas as variáveis originais do processo (PEÑA, 2002). Segundo o autor, o principal objetivo das componentes principais é determinar o novo conjunto de direções ortogonais que definem a variabilidade máxima dos dados originais. Como ilustração, considere o caso da figura 1: há duas variáveis x_1 e x_2 e duas componentes principais z_1 e z_2.



Fonte: Montgomery (2004, página 340)

Figura 1: Componentes principais para p = 2.

De acordo com Montgomery (2004), encontrar os c_{ij} que definem as componentes principais é fácil. Seja X uma matriz de dados preliminares do processo que contém uma observação p-variada em cada linha e seja S a matriz de covariâncias amostrais. No entanto, se utilizarmos no lugar de X a matriz Z de dados padronizados, obtém-se, ao invés de S, a matriz R de correlações amostrais. Como a matriz S é simétrica e não singular, pela decomposição espectral, existe uma matriz ortonormal $\mathbf{U=[u_1,u_2,....,u_p]}$ que transforma a matriz S numa matriz diagonal **L** (FILHO,2001).

$$\boldsymbol{U'SU=L} \tag{2}$$

Os elementos da diagonal **L**, λ_1, λ_2, ...λ_p, são os autovalores obtidos através da solução da equação característica da matriz **S** (Jackson,1980):

$$|\boldsymbol{S} - \lambda \boldsymbol{I}| = \mathbf{0} \tag{3}$$

onde **I** é uma matriz identidade de ordem *p x p*. A partir da expressão (2) pode-se transformar *p* variáveis correlacionadas $x_1, x_2,.....,x_p$ em *p* variáveis independentes z_1, $z_2,.....,z_p$ denominadas componentes principais. A matriz **L** apresenta as covariâncias amostrais dessas novas variáveis independentes. Os autovalores de **L** estão dispostos em ordem decrescente, com λ_p representando a variância de componente z_p. Os autovetores de **S** são obtidos através da solução da equação

$$[\mathbf{S} - \lambda_\iota \mathbf{I}]\mathbf{t_i} = \mathbf{0} \tag{4}$$

tal que

$$\mathbf{u_i} = \frac{\mathbf{t_i}}{\sqrt{\mathbf{t_i't_i}}}, \quad \text{para } i = 1,.....,p \tag{5}$$

representam os autovetores normalizados que descrevem os eixos de coordenadas das novas variáveis (Jackson,1980,1985).

Cada autovetor $\mathbf{u_i}$, i =1,.....,p, de **U**, obtido a partir de λ_i é uma combinação linear das variáveis originais **X,** projetados nas direções de máxima variabilidade em relação à nuvem de pontos p-dimensional, obtida a partir dos dados (Johnson e Wichern,1992). A matriz **U** é chamada matriz das cargas, com cada autovetor $\mathbf{u_i}$ contendo as cargas devidas a componente z_i. Dessa forma, o vetor **Z** de variáveis latentes pode ser escrito como:

$$\mathbf{Z = U'X} . \tag{6}$$

A primeira componente, de acordo com Peña (2002), se define como a combinação linear das variáveis originais que tem variância máxima. Os valores nesta primeira componente $\mathbf{z_1}$, dado por $\mathbf{z_1 = Xu_1}$. Como as variáveis originais têm média zero, pois foram padronizadas, $\mathbf{z_1}$ também terá média nula. Sua variância será:

$$\frac{1}{n}\mathbf{z_1'z_1} = \frac{1}{n}\mathbf{u_1'X'Xu_1} = \mathbf{u_1'Su_1} \tag{7}$$

onde **S** e a matriz de variâncias e covariâncias das observações. Para que (7) seja maximizado e tenha solução, a restrição $\mathbf{u_1'u_1} = 1$ é introduzida mediante o multiplicador de Lagrange $M = \mathbf{u_1'Su_1} - \lambda(\mathbf{u_1'u_1} - 1)$ e derivando respectivamente as componentes principais de $\mathbf{u_1}$ e igualando a zero. Então:

$$\frac{\partial M}{\partial \mathbf{u_1}} = 2\mathbf{Su_1} - 2\lambda\mathbf{u_1} = 0, \tag{8}$$

cuja solução é

$$\mathbf{Su_1} = \lambda_1\mathbf{u_1} \tag{9}$$

que implica que $\mathbf{u_1}$ é um autovetor de **S**, e λ seu correspondente valor próprio Para determinar que autovalor de **S** é a solução de (9), multiplicando esta equação por $\mathbf{u_1'}$ tem-se:

$$\mathbf{u_1'Su_1} = \lambda_1\mathbf{u_1'u_1} = \lambda_1 \tag{10}$$

e concluímos, por (9) que λ_1 é a variância de $\mathbf{z_1}$. Como o autovalor λ_1 é o que desejamos maximizar, λ_1 será o maior autovalor de **S**. Seu vetor associado, autovetor, $\mathbf{u_1}$, define os coeficientes de cada variável na primeira componente principal.

A segunda componente, segundo Peña (2002), representa a projeção das variáveis originais na segunda maior direção de variabilidade. Para obter o melhor plano de projeção das variáveis **X,** estabelecemos como função objetivo que a soma das variâncias $\mathbf{z_1 = Xu_1}$ e $\mathbf{z_2 = Xu_2}$ seja máxima, onde $\mathbf{u_1}$ e $\mathbf{u_2}$ são vetores que definem o plano. Esta função objetivo será:

$$\phi = \mathbf{u_1'Su_1} + \mathbf{u_2'Su_2} - \lambda_1(\mathbf{u_1'u_1} - 1) - \lambda_2(\mathbf{u_2'u_2} - 1) \tag{11}$$

que incorpora as restrições de que as direções devem ter módulo unitário $u_i'u_i = 1$, com i =1,2. Derivando e igualando a zero:

$$\begin{cases} \dfrac{\partial\phi}{\partial\mathbf{u_1}} = 2\mathbf{Su_1} - 2\lambda_1\mathbf{u_1} = 0 \\ \dfrac{\partial\phi}{\partial\mathbf{u_2}} = 2\mathbf{Su_2} - 2\lambda_2\mathbf{u_2} = 0 \end{cases}$$

a solução deste sistema é:

$$\mathbf{S}\mathbf{u}_1 = \lambda_1 \mathbf{u}_1 \quad (12)$$

$$\mathbf{S}\mathbf{u}_2 = \lambda_2 \mathbf{u}_2 \quad (13)$$

indicando que $\mathbf{u_1}$ e $\mathbf{u_2}$ são vetores próprios de **S,** e λ_1 e λ_2 seus valores próprios correspondentes. Esses valores próprios são as quantidades que queremos maximizar, λ será o maior valor próprio da matriz **S.** Seu vetor associado $\mathbf{u_1}$ e $\mathbf{u_2}$, define os coeficientes de cada variável na primeira e segunda componente principal, respectivamente.

3. Controle Estatístico Multivariado Baseado em Métodos de Projeção de Dados

O controle estatístico multivariado de processos está fundamentado nos métodos de projeção de dados via Análise de Componentes Principais (ACP) e via Projeção de Estruturas Latentes (PEL) (FILHO, 2001). O objetivo destes é transformar variáveis de processo (características da qualidade) em um grupo reduzido de variáveis latentes (combinação linear das originais), preservando informações relevantes contidas nestas, e eliminando redundâncias no sistema inicial, representadas por variáveis colineares. Dessa forma, quando os dados originais são projetados num espaço de dimensões menores, trazem consigo a estrutura de covariância das variáveis originais. Normalmente, duas ou três variáveis latentes são suficientes para representar variáveis originais (FILHO, 2001).

Os tradicionais métodos estatísticos de controle multivariado tais como os gráficos de controle T^2 de Hotelling, MCUSUM e MEWMA não são suficientemente robustos para tratar com um grande número de variáveis correlacionadas, pois foram desenvolvidos para monitorar um pequeno número dessas variáveis (PHAM, 2006). A estrutura colinear das variáveis do processo faz com que as estatísticas de controle utilizadas nestes gráficos forneçam sinalizações distorcidas acerca do estado do processo. Em algumas situações tais estatísticas não poderão ser calculadas, devido à dificuldade de inversão da matriz de covariâncias das variáveis de processo. Surge então a necessidade de utilizar gráfico de controle multivariado baseados em métodos de projeção de dados (PHAM, 2006).

O desenvolvimento de gráfico de controle multivariado é separado em duas fases. A primeira fase (fase I) consiste em obter-se uma amostra representativa dos dados com o objetivo de determinar os limites de controle, sendo em geral um estudo retrospectivo dos dados. A segunda fase (fase II) é direcionada ao monitoramento do processo (MASON & YOUNG, 2002). O desempenho de um gráfico de controle é comumente avaliado através de parâmetros relacionados com a distribuição do tempo necessário para o gráfico emitir um sinal. O número médio de amostras coletadas até emissão de um sinal (ARL-Average Run Length), é um desses parâmetros. Num gráfico de controle o ARL_0 indica o número médio de amostras coletadas até a emissão de um sinal durante o período sob controle enquanto o ARL_1 representa o número médio de amostras coletadas até a emissão de um sinal que indique uma situação de fora de controle (ALVES e SAMOHYL, 2004).

3.1 Gráficos de Controle T^2 de Hotelling

Hotelling (1947) foi pioneiro na pesquisa sobre controle multivariado de qualidade. Ele utilizou uma abordagem multivariada de controle em dados contendo informações sobre localizações de bombardeios, durante a Segunda Guerra Mundial. Os desenvolvimentos teóricos propostos pelo autor estão descritos nesta seção.

Entre os gráficos multivariados existentes, o gráfico de controle multivariado T^2 *de* Hotelling é o mais conhecido na literatura. Inicialmente, Hotelling (1947) supôs que as variáveis de interesse seguiam uma distribuição normal multivariada, com vetor de médias $\bar{X}$ e matriz de

covariâncias *S*. Assim, tomam-se amostras de tamanho n para cada uma das *p* variáveis de interesse (a serem monitoradas) e considerando as estimativas dos parâmetros, a equação para a obtenção das estimativas da estatística T^2 é dada por:

$$T^2 = (\boldsymbol{X} - \bar{\boldsymbol{X}})' \boldsymbol{S^{-1}} (\boldsymbol{X} - \bar{\boldsymbol{X}}) \quad (14)$$

onde $\bar{\boldsymbol{X}}$ e **S** representam, respectivamente, as estimativas para o vetor de médias e matriz de covariâncias do processo. A expressão (14) é utilizada como base para a construção do gráfico de controle T^2 de Hotelling (Lowry & Montgomery, 1995). Na primeira fase, os limites de controle, para Tracy et al. (1992) são baseados na distribuição beta e, são expressos por (15):

$$LSC = \frac{(m-1)^2}{m} \beta_{\alpha, p/2, (m-p-1)/2} \quad (15)$$

onde $\beta_{\alpha, p/2, (m-p-1)/2}$ é o ponto percentual α superior de uma distribuição β com parâmetros $p/2$ e $(m-p-1)/2$. Na segunda fase, os novos limites são estabelecidos apenas para monitorar as observações futuras, utilizando os limites de controle mostrados na equação (16):

$$LSC = \frac{p(m+1)(m-1)}{m^2 - mp} F_{\alpha, p, m-p} \quad (16)$$

onde *p* é o número de variáveis (característica de qualidade) e m o número de amostras. Se o valor de T^2 excede o limite superior de controle (LSC), diz-se que o processo está fora de controle estatístico. O limite inferior de controle (LIC), para as duas fases, é considerado igual a zero $LIC = 0$.

Um problema significativo, no caso das observações individuais é a estimação da matriz de covariâncias do processo. Sullivan e Woodall (1996) *apud* Montgomery (2004, p.332) apresentam critérios para estimar a matriz de covariâncias de processos. Esses autores propõem alguns procedimentos para obtenção de S (estimadores) que tornam o deslocamento abrupto (mudança súbita) no processo e deslocamento gradativo (tendência ou mudança gradativa) no vetor de médias do processo. Nesse artigo utiliza-se a matriz de covariâncias amostral definida por:

$$S_{p \times p} = \frac{1}{2(m-1)} \sum_{i=1}^{m-1} (x_{i+1} - x_i)' (x_{i+1} - x_i). \quad (17)$$

com S representando um estimador de variabilidade de processo para a matriz de covariância do processo. Este estimador usa a diferença entre os sucessivos pares de observações.

Além do gráfico de controle T^2 de Hotelling, outros tipos de gráficos de controle para processos multivariados, são indispensáveis para atender a especificidade de um determinado processo tais como os gráficos de controle memória MCUSUM (multivariado de soma acumulada) e o MEWMA (multivariado de móvel exponencialmente ponderada). Estes gráficos multivariados de controle estatístico possuem a característica de detectar pequenas mudanças no processo (abaixo de 2σ), ao contrário do gráfico de controle multivariado T^2 Hotelling que detecta grandes alterações.

3.2 Gráfico Multivariado de Soma Acumulada

O modelo de gráfico de controle univariado CUSUM (Soma Acumulada) foi desenvolvido, de acordo com Alves & Samohyl (2004), para oferecer maior sensibilidade a pequenos e moderados desvios na média de um processo que passam despercebidos pelo gráfico

de Shewhart. Os procedimentos de controle estatístico multivariado baseados na filosofia CUSUM são discriminados em duas principais categorias: (i) procedimentos de controle que utilizam múltiplos gráficos de controle CUSUM univariados (abreviados por MCU), desconsiderando assim a correlação entre as variáveis e (ii) procedimentos de controle que utilizam um gráfico de controle CUSUM multivariado (abreviado por MCUSUM), isto é, utilizam a matriz de covariâncias das variáveis para obter uma aproximação do gráfico CUSUM em processos multivariados. Portanto, a primeira delas consiste em reduzir as observações multivariadas a um escalar enquanto a outra consiste em elaborar um gráfico MCUSUM para analisar diretamente estas observações multivariadas (CROSIER, 1988).

O gráfico de controle CUSUM multivariado foi proposto por Crosier (1988) a partir de dois procedimentos de controle. O primeiro procedimento baseado na raiz quadrada da estatística T^2 de Hotelling denominado de gráfico de controle CUSUM COT (CUSUM of T) consiste em reduzir as observações multivariadas a escalares. O segundo procedimento denominado MCUSUM (CUSUM de vetores) se constitui numa extensão multivariada do gráfico de controle CUSUM univariado onde as quantidades escalares são substituídas por vetores. Define-se C_i como:

$$C_i = \sqrt{(\boldsymbol{S}_{i-1} + \boldsymbol{X}_i - \boldsymbol{\mu}_0)' \Sigma^{-1} (\boldsymbol{S}_{i-1} + \boldsymbol{X}_i - \boldsymbol{\mu}_0)} \tag{18}$$

onde S_i são as somas acumuladas expressadas como:

$$\boldsymbol{S}_i = \begin{cases} 0 & \text{se} \quad C_i \leq k \\ (\boldsymbol{S}_{i-1} + \boldsymbol{X}_i)\left(1 - {}^{k}/_{C_i}\right) & \text{se} \quad C_i > k \end{cases} \tag{19}$$

com $\boldsymbol{S_0} = \boldsymbol{0}$ e valor de referência $k > 0$, relacionado à magnitude de mudança. A estatística de controle a ser plotada no gráfico de controle MCUSUM é dada por:

$$Y_i = \sqrt{\boldsymbol{S}_i' \Sigma^{-1} \mathbf{S}_\mathrm{i}} \tag{20}$$

O método sinaliza uma situação fora de controle se $Y_i > H$ onde H é o intervalo de decisão (limite de controle). Crosier (1988) demonstra que, de uma maneira geral, este procedimento tem desempenho de ARL melhor do que o procedimento escalar. Além disso, ainda de acordo com o autor, este tipo de gráfico de controle apresenta um desempenho ARL superior em relação ao gráfico T^2 de Hotelling na detecção de deslocamentos no vetor de médias do processo.

4. Procedimentos Metodológicos

Os procedimentos adotados neste artigo envolvem a utilização da análise de componentes principais a um conjunto de dados da literatura e posterior aplicação dos gráficos de controle T^2 de Hotelling e MCUSUM às componentes. Dois procedimentos foram utilizados para escolha do número de componentes. O primeiro, conhecido como critério da raiz latente, implica que apenas autovalores maiores que um (1) são considerados significantes. O segundo é determinado por meio de um gráfico ("screeplot") onde relaciona-se os autovalores à ordem de extração. A forma da curva resultante é usada para avaliar o número de componentes. O ponto no qual o gráfico começa a ficar horizontal é um indicativo do número de componentes (HAIR et al., 2005).

O conjunto de dados (Tabela 1) é composto de oito variáveis que representam a leitura de temperaturas nos queimadores de uma caldeira (Mason e Young, 2002). As temperaturas são medidas seqüencialmente ao longo do tempo. Este exemplo contempla apenas a fase I do processo. A taxa de alarmes adotada será de 0,2%, ou seja ARL_0 =500. A observação 9 foi retirada pelos autores, portanto os limites são calculados para as 24 observações restantes.

	t1	t2	t3	t4	t5	t6	t7	t8
1	507	516	527	516	499	512	472	477
2	512	513	533	518	502	510	476	475
3	520	512	537	518	503	512	480	477
4	520	514	538	516	504	517	480	479
5	530	515	542	525	504	512	481	477
6	528	516	541	524	505	514	482	480
7	522	513	537	518	503	512	479	477
8	527	509	537	521	504	508	478	472
9	533	514	528	529	508	512	482	477
10	530	512	538	524	507	512	482	477
11	530	512	541	525	507	511	482	476
12	527	513	541	523	506	512	481	476
13	529	514	542	525	506	512	481	477
14	522	509	539	518	501	510	476	475
15	532	515	545	528	507	511	481	478
16	531	514	543	525	507	511	482	477
17	535	514	542	530	509	511	483	477
18	516	515	537	515	501	516	476	481
19	514	510	532	512	497	512	471	476
20	536	512	540	526	509	512	482	477
21	522	514	540	518	497	514	475	478
22	520	514	540	518	501	514	475	478
23	526	517	546	522	502	516	477	480
24	527	514	543	523	502	512	475	476
25	529	518	544	525	504	516	479	481

Tabela 1: Temperatura da caldeira
Fonte: Mason e Young (2002) página 86

Os gráficos de controle das componentes são comparados com o gráfico T_2 de Hotelling aplicado às variáveis originais. Todavia o MCUSUM aplicado às componentes irá trazer informações importantes para o estudo.

A análise estatística dos dados será feita com o GNU R (R Core Development Team, 2010), com auxílio dos pacotes qcc (Scrucca, 2004) para o gráfico T^2 de Hotelling e acesso ao conjunto de dados. Para o gráfico MCUSUM será aplicada a rotina desenvolvida por Alves et al. (2008), com o limite superior de controle h determinado pelo Método da Equação Integral (Alves, 2009). Não é possível comparar com o gráfico MCUSUM às variáveis originais, pois a rotina desenvolvida não contempla oito variáveis.

5. Resultados

A análise de componentes principais e os gráficos de controle pressupõem que os dados tenham distribuição normal e sejam independentes. Para a verificação da normalidade foi utilizado o teste de Mardia. Os dados apresentam uma distribuição normal multivariada (p-valor assimetria = 0,30 e p-valor curtose = 0,15). Verificou-se a ausência de autocorrelação mediante a análise dos correlogramas. É importante salientar que a independência, uma condição para o uso dos gráficos T_2 de Hotelling e MCUSUM, deve ser garantida preferencialmente pelo processo de amostragem, que deve ser aleatório.

Aplicou-se o gráfico de controle T^2 de Hotelling às oito variáveis do conjunto de dados, conforme figura 2. Pode-se concluir que o processo está sob controle. O limite para a fase I é LSC=16,55 (figura 2). O limite calculado para a fase II, de monitoramento é LSC=65,10, conforme Mason & Young (2002).

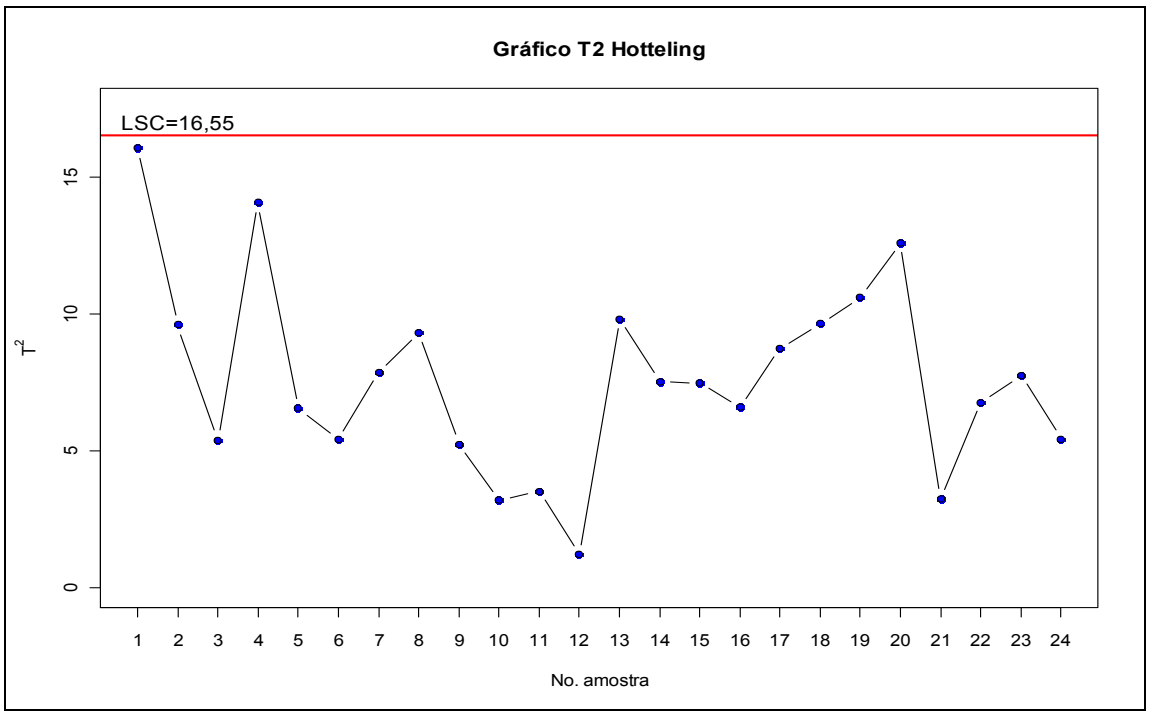


Figura 2: Gráfico T2 de Hotelling para as oito variáveis originais

Em seguida, foi feita a análise de componentes principais, pela matriz de correlações. A análise revelou que apenas duas componentes são responsáveis por 85% da variabilidade dos dados (Tabela 2). Na Tabela 2, e no gráfico da figura 3, pode-se ver que apenas duas componentes tem autovalor maior que 1 (um) e, a partir da terceira componente (figura 3) a curva começa a ficar horizontal. Portanto, os gráficos de controle podem ser aplicados a apenas duas componentes, ao invés das oito variáveis originais.

Componentes	Autovalores	Proporção da variância	Proporção da variância acumulada
1	2,03	0,5171	0,5171
2	1,63	0,3341	0,8512
3	0,73	0,066	0,9172
4	0,66	0,0539	0,9711
5	0,31	0,0119	0,9831
6	0,26	0,0088	0,9919
7	0,23	0,0065	0,9985
8	0,11	0,0002	1,0000

Tabela 2: Autovalores, variância explicada por cada componente e proporção da variância acumulada para os dados

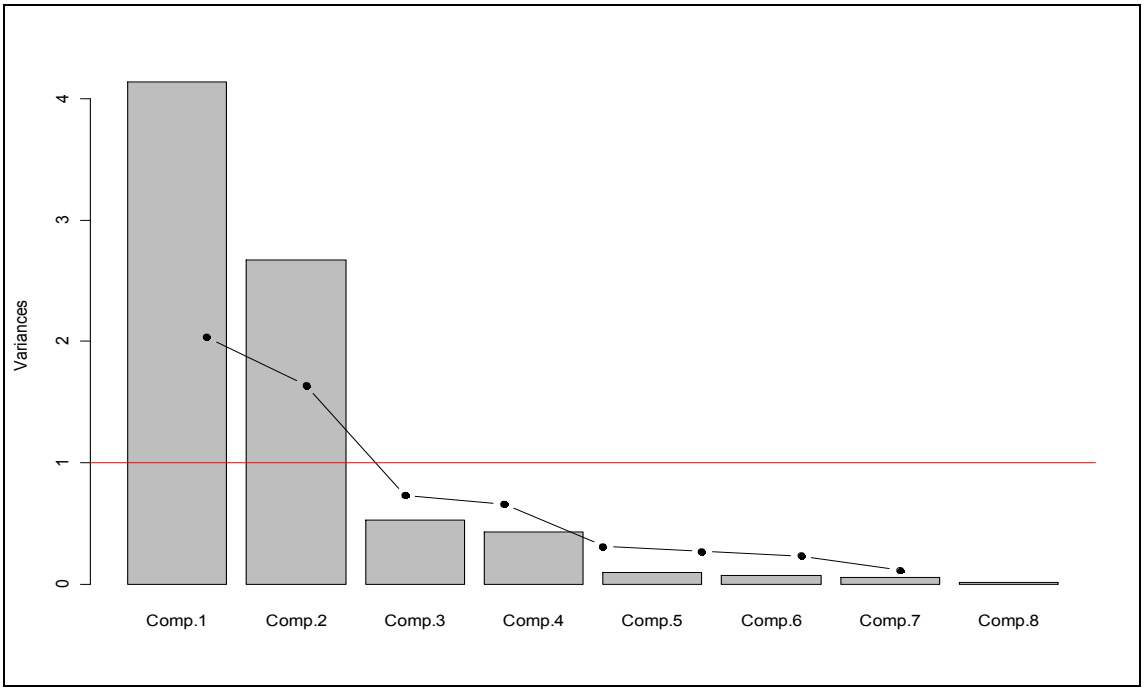


Figura 3: Gráfico screeplot

Uma análise feita às cargas (não presentes neste documento) de cada componente às temperaturas, revela que t_1, t_4, t_5 e t_7 têm moderada correlação com a primeira componente e, t_2, t_6 e t_8 têm moderada correlação a segunda componente.

Nas figuras 4 e 5 estão os gráfico de controle T^2 de Hotelling e MCUSUM aplicados às duas componentes resultantes. Pode-se perceber que aplicando o gráfico de controle T^2 de Hotelling às componentes, o resultado é similar ao encontrado com as oito variáveis originais. O limite para a fase II é LSC=18,19.

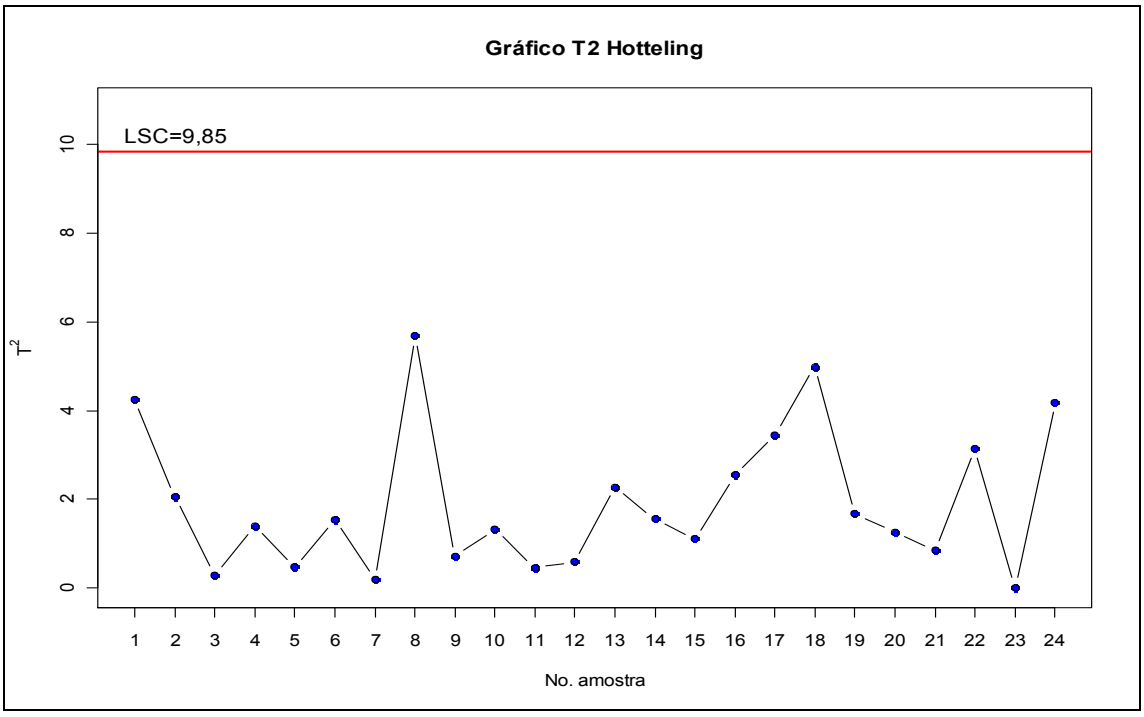


Figura 4: Gráfico T^2 de Hotelling para as componentes

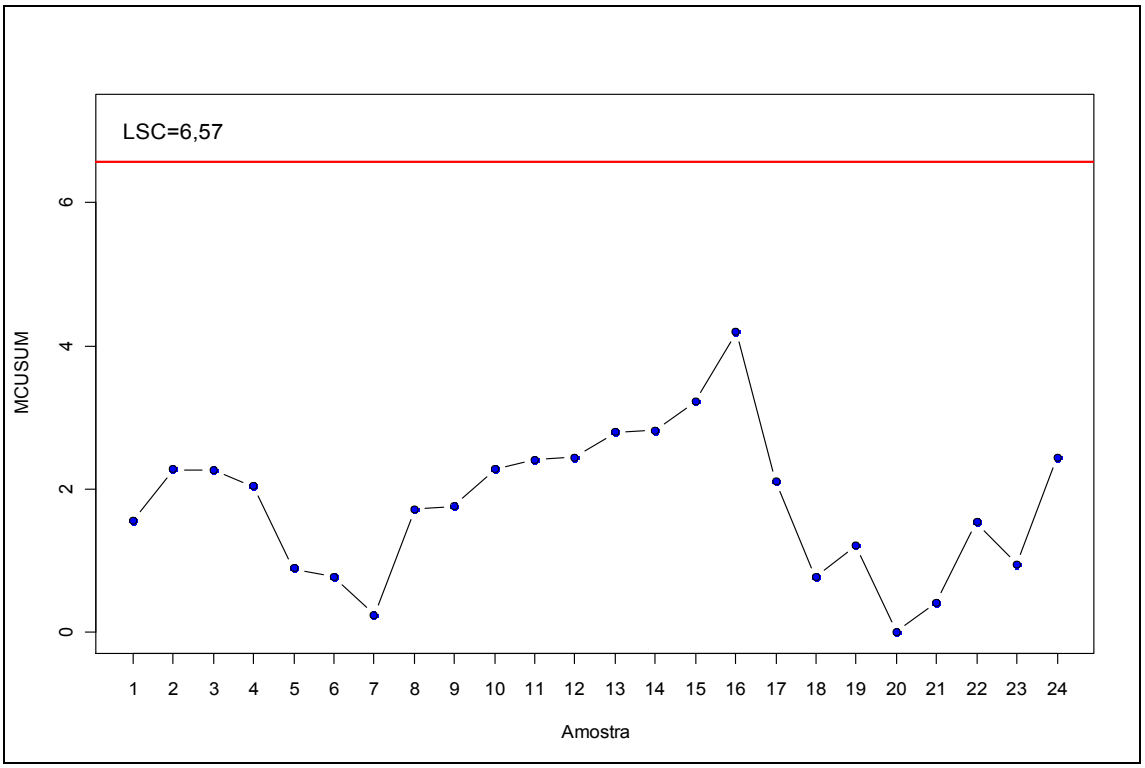


Figura 5: Gráfico MCUSUM para as componentes

O gráfico MCUSUM, neste caso, também indica um processo sob controle. Lembra-se que este último gráfico, quando aplicado, é mais eficaz na detecção de pequenas e persistentes alterações no processo.

6. Conclusões e Considerações Finais

Neste artigo foi realizado um estudo comparativo entre gráficos de controle multivariados com a aplicação da análise de componentes principais em um conjunto de dados da literatura. As oito variáveis iniciais deste conjunto são reduzidas a duas componentes. A aplicação dos gráficos T_2 de Hotelling às componentes conduz à mesma conclusão de quando aplicado às oito variáveis originais. O gráfico MCUSUM das componentes leva ao mesmo resultado. Conclui-se que a análise de componentes é uma alternativa para o controle estatístico de processos multivariados, sendo possível reduzir o número de variáveis analisadas, sem perda de informação. Com esse estudo foi possível verificar que trabalhando com apenas duas componentes os resultados de ambos os gráficos apresentam o mesmo resultado do que com as oito variáveis originais. Este tipo de análise pode ser importante para casos com muitas variáveis, como por exemplo, processos químicos industriais.

Como sugestão para trabalhos futuros está a investigação da sensibilidade dos gráficos, aplicando a análise de componentes principais, em situações sob e fora de controle estatístico.

Referências

Alves, C.C, *O método de Equação Integral com Quadratura Gaussiana para otimizar os parâmetros do gráfico de controle multivariado de Somas Acumuladas*. Tese de Doutorado.

Programa de Pos-Graduação em Engenharia de Produção e Sistemas. Universidade Federal de Santa Catarina.2009

Alves,C.C., Henning,E. Samohyl,R.W.*O desenvolvimento de gráficos de controle MCUSUM e MEWMA em ambiente R como um procedimento alternativo para análise estatística de processos multivariado***.** XXIVIII Encontro Nacional de Engenharia de Produção – Rio de Janeiro, RJ, Brasil, 2008a.

Alves, C.C. & Samohyl, R.W. *A utilização dos gráficos de controle CUSUM para o monitoramento de processos industriais*. XXIV Encontro Nacional de Engenharia de Produção - Florianópolis, SC, Brasil, 2004.

Crosier, R.B. Multivariate Generalizations of Cumulative Sum Quality Control Schemes. *Technometrics*, 30(3), 291-303, 1988.

Filho, D. M. *Monitoramento de Processos em Batelada através de Cartas de Controle Multivariadas Baseada na Análise de Componentes Principais Multidirecionais (ACPM).* Dissertação de Mestrado em Engenharia de Produção, UFRGS, Porto Alegre, 2001.

Hair, J. F., Tatham, R.L., Anderson R. E., Black, W. *Análise Multivariada de dados*. 5ª Ed. Porto Alegre: Bookman, 2005.

Hotelling, H. *Multivariate Quality Control-illustrated by the air testing of sample bombsights, in Techniques of Statistical Analysis*. p. 111-184, New York : McGraw Hill, 1947.

Jackson, J.E. *Multivariate Quality Control*. Commun.Statist.Theor.Meth., 14 (11), 2657-2658, 1985

Jackson, J.E.*Principal Components and factor Analysis: Part I*, Journal of Quality Technology, 12(4), 201-213,1980.

Jackson, J.E. *A user guide to Principal Components*. Wiley. New York: 1991.

Johnson, R.A.; *WICHERN, D.W. Applied Multivariate Statistical Analysis*, 4a ed., Prentice Hall, 1998.

Lowry, A.C.; Woodall, W.H.; Champ, C.W. & Rigdon, C.C. *A Multivariate Exponentially Weighted Moving Average Control Chart*. Technometrics, 34(1), 46-53, 1992.

Lowry, A.C. & Montgomery, D.C. *A Review of Multivariate Control Charts*. IIE Transactions, v. 27, p. 800-810, 1995.

Montgomery, D. C. *Introdução ao Controle Estatístico da Qualidade*. LTC Editora, 4ª edição, 2004.

Mason, R.L, & Young, J.C. *Multivariate Statistical Process Control with Industrial Applications*. ASA/SIAM: Philadelphia, PA, 2002.

Peña, D. *Análisis de Datos Multivariantes.* 1a ed., McGraw-Hill Interamericana de España, 2002.

Pham, H. *Springer Handbook of Engineering Statistics*. Rutgers the State University of New Jersey, 2006

R Development Core Team. ***R****: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL: http://www.r-project.org, 2009

Rencher, A.C. *Methods of Multivariante Analysis*, .2a ed., Wiley-Interscience, 2002.

Scrucca, L.. *qcc: an R package for quality control charting and statistical process control*. R News 4/1, 11-17. 2004.

Tracy, N.D.; Young, J. C. & Mason, R.L. Multivariate Control Charts for Individual Observations. *Journal of Quality Technology*. Vol. 24, No. 2, 1992.