

# Identificação de variáveis fora de controle em processos produtivos multivariados

**ADRIANO MENDONÇA SOUZA, DR**

Universidade Federal de Santa Maria – UFSM

**MARIA HELENA RIGÃO**

Universidade Federal de Santa Maria – UFSM

## Resumo

A maioria dos processos produtivos, envolvendo múltiplas variáveis, geram grande quantidade de dados, nos quais o monitoramento simultâneo se faz necessário. Assim, o objetivo principal deste trabalho é mostrar, por meio de um exemplo, um procedimento de identificação de variáveis responsáveis pela instabilidade do processo. Ele considera variáveis fraca e fortemente correlacionadas. O comportamento destas variáveis é analisado, inicialmente, por meio do gráfico de controle  $T^2$  de Hotelling, para a verificação da estabilidade do processo. Quando o processo é considerado fora de controle, uma investigação adicional é realizada. Para as variáveis fracamente correlacionadas o gráfico X-barra, com limites de Bonferroni, é utilizado. Para as variáveis com forte correlação, os dados são decompostos em componentes principais e estas são analisadas pelo gráfico de controle X-barra. Em ambos os casos, as técnicas utilizadas identificaram corretamente a variável responsável pela instabilidade do processo, bem como o período do distúrbio.

## Palavras-chave

Controle estatístico de processos, Gráfico de controle  $T^2$  de Hotelling, Gráfico X-barra com limites de Bonferroni, Análise de componentes principais.

## Identification of out-of-control variables in a multivariate productive process

### Abstract

*Most productive processes are affected by multiple variables and generate a great amount of data and simultaneous monitoring becomes necessary. Thus, the main purpose of this work is to show, by means of an example, procedures of identification of variables responsible for process instability. It considers weakly and strongly correlated variables. The behavior of these variables is analyzed, initially by means of a Hotelling's  $T^2$  chart for the verification of the process stability. When the process is considered to be out-of-control, another investigation is carried out. For variables of weak correlation an X-bar chart with Bonferroni limits is used. For variables of strong correlation, the data are decomposed in principal components and analyzed by means of a X-bar control chart. In both cases, the techniques used can correctly identify the variable responsible for the process instability, as well as the period of disturbance.*

### Key words

*Statistical process control, Hotelling's  $T^2$  chart, X-bar chart with Bonferroni limits, Principal components analysis.*

## INTRODUÇÃO

O Controle Estatístico do Processo (CEP) é uma metodologia que atua preventivamente sobre o processo produtivo, utilizando a estatística como instrumento básico para avaliar suas alterações, em busca do aperfeiçoamento contínuo da qualidade. Dentro deste contexto, os gráficos de controle são considerados uma das suas ferramentas mais importantes. Esses gráficos podem ser por atributos, ou por variáveis. Os gráficos de controle por atributos se referem às características de qualidade que classificam itens em conformes e não-conformes, enquanto que os gráficos de controle por variáveis baseiam-se na medida das características de qualidade do produto em uma escala contínua. Os gráficos de controle por variáveis ainda podem ser ditos univariados ou multivariados, de acordo com a quantidade de variáveis envolvidas na avaliação.

Em processos univariados, os gráficos de monitoramento da média desempenham um excelente papel, pois conseguem identificar quando uma fonte de variação está afetando o processo. Porém, é comum que um produto ou processo possuam várias características de qualidade que precisem ser monitoradas conjuntamente. Os gráficos univariados, então, não mais devem ser utilizados, pois as várias características de qualidade, podem apresentar correlação entre si, o que prejudica o desempenho destes para sinalizar uma falta de controle. Uma alternativa, para monitorar essas características, é ignorar essa correlação e usar gráficos univariados para cada uma das características. Infelizmente essa metodologia pode levar a erros na tomada de decisões. Nesses casos, um gráfico multivariado pode fornecer melhores resultados no monitoramento das variáveis.

De acordo com Montgomery (1991), a análise multivariada torna-se útil quando existem múltiplos parâmetros e seus efeitos não são independentes, ou quando alguns parâmetros são parciais ou medidas de outros parâmetros, o que configuraria uma correlação. Ao utilizar-se gráficos de controle multivariados, em lugar de vários gráficos de controle univariados, os ganhos são evidentes, tanto em relação ao tempo quanto aos recursos humanos utilizados para gerir esse monitoramento e, sobretudo, à acurácia que esses gráficos fornecem. Assim, os gráficos de controle multivariados, conhecidos como gráficos  $T^2$  de Hotelling, constituem-se em uma importante contribuição para verificar a estabilidade de um processo.

Uma vez que o gráfico multivariado tenha registrado um ponto fora dos limites, um diagnóstico realizado por meio de gráficos univariado pode ajudar a decidir quais das variáveis têm seus efeitos causando essa instabilidade (KOURTI e MACGREGOR, 1996).

Uma das dificuldades encontradas com gráficos de controle multivariados é a interpretação do sinal de que o processo está fora de controle, ou seja, identificar qual das múltiplas variáveis é responsável pelo sinal de alarme. Existem, pelo menos, três procedimentos recomendados para a identificação das variáveis responsáveis pela ocorrência do alarme: gráficos de controle X-barras com limites de Bonferroni; gráficos de controle baseados nos componentes principais das variáveis originais; decomposição da estatística  $T^2$  em componentes que refletem a contribuição individual de cada variável (MASON, TRACY e YOUNG, 1995).

Este trabalho tem como objetivo mostrar, mediante um exemplo, o funcionamento de duas técnicas de identificação de variáveis fora de controle, em processos multivariados: uma delas para variáveis fracamente correlacionadas, e a outra para variáveis fortemente correlacionadas. Para isso, foram gerados dois conjuntos de variáveis por simulação. O conjunto fracamente correlacionado foi analisado pelo gráfico de controle X-barras, com limites de Bonferroni. O conjunto de variáveis fortemente correlacionadas foi avaliado pela análise de componentes principais, seguida do gráfico de controle X-barras. Essas duas técnicas foram precedidas de uma avaliação conjunta das variáveis, com base no gráfico de controle multivariado  $T^2$  de Hotelling. Este artigo justifica sua importância pelo fato de divulgar um exemplo detalhado do uso bem-sucedido de técnicas de controle estatístico de processos multivariados.

Trazido à tona, esse importante assunto pode gerar novas discussões no meio acadêmico e industrial e levar ao aprimoramento das técnicas para novas aplicações práticas.

## GRÁFICO $T^2$ DE HOTELLING

Os gráficos de controle multivariados se baseiam na distribuição  $T^2$  de Hotelling que é a generalização da distribuição “ $t$ ” de Student, para várias dimensões.

Hotelling (1947) foi o primeiro a perceber a fragilidade do uso de múltiplos gráficos de controle univariados, quando havia correlação entre as variáveis. Nas décadas subseqüentes, muitos autores se ocuparam do assunto e uma vasta bibliografia pode ser encontrada, como, por exemplo: Aparise (1997), Jackson (1956, 1959), Jackson e Morris (1957), Ryan (1989), Montgomery (1991), Tracy *et al.* (1992).

No caso univariado, para determinar se um valor específico  $\mu_0$  pertence a uma população que possui média  $\mu$ , formulam-se as hipóteses:

$$H_0: \mu = \mu_0$$

$$H_1: \mu \neq \mu_0.$$

Se  $X_i$ , com  $i = 1, 2, \dots, n$ , é uma amostra aleatória de uma população normal, estatística do teste é:

$$t = \frac{(\bar{X} - \mu_0)}{\frac{s}{\sqrt{n}}} \quad (1)$$

Considere-se o problema de determinar se um dado vetor  $\mu_{0(p \times 1)}$  é um valor plausível para a média de uma distribuição normal multivariada, uma vez que  $\bar{X}$  e  $S^2$  são valores observados. Deve-se, então, proceder, analogamente, como se fez no caso univariado citado anteriormente (JOHNSON e WICHERN, 1992).

Partindo do teste de hipótese, com base na distribuição “ $t$ ” de Student, faz-se uma generalização de (1) para o caso em que mais de uma variável é submetida a teste, a fim de verificar se o seu vetor média pertence ou não à amostra em estudo. Essa generalização dá origem à estatística  $T^2$  mostrada em (2).

$$T^2 = (\bar{X} - \mu_0)' \left( \frac{S}{n} \right)^{-1} (\bar{X} - \mu_0) = n(\bar{X} - \mu_0)' S^{-1} (\bar{X} - \mu_0) \quad (2)$$

em que:

$$\bar{X}_{(p \times 1)} = \frac{1}{n} \sum_{j=1}^n X_j ;$$

$$S_{(p \times p)} = \frac{1}{n-1} \sum_{j=1}^n (X_j - \bar{X})(X_j - \bar{X})'$$

$$\mu_{0(p \times 1)} = \begin{bmatrix} \mu_{10} \\ \mu_{20} \\ \vdots \\ \mu_{p0} \end{bmatrix}$$

$T^2$  é distribuída como  $\frac{(n-1)p}{(n-p)} F_{p, n-p}$

onde

$F_{p, n-p}$  é uma variável aleatória com uma distribuição  $F$ ;  $p$  e  $n-p$  são os números dos graus de liberdade;  $p$  é o número de variáveis;  $n$  é o número total de observações.

Para a utilização do procedimento de Hotelling, é necessário que as  $p$ -variáveis correlacionadas sejam oriundas de medidas realizadas simultaneamente e sigam uma distribuição normal multivariada, com vetor média  $\mu$  e matriz de variância-covariância  $\Sigma$ . Quando uma  $i$ -ésima amostra é retirada, têm-se  $n$  valores de cada característica e, então, é possível calcular os vetores  $\mu_0$  e  $\Sigma_0$ ; que representam a média e a matriz de variância-covariância das  $p$ -características, e comparar com a situ-

ação na qual o processo está sob controle (MASON e YOUNG, 2002).

Se os valores populacionais são desconhecidos, pode-se estimá-los por meio da informação amostral. A média populacional será estimada por  $\bar{\bar{X}}$ , e a matriz de variância-covariância populacional por meio de  $S$ . Então, a estatística (2) se torna:

$$T^2 = n(\bar{X} - \bar{\bar{X}})' S^{-1} (\bar{X} - \bar{\bar{X}}) \quad (3)$$

Segundo Lowry e Montgomery (1995) e Tracy *et al.* (1995), o gráfico de Hotelling apresenta duas fases distintas de avaliação do processo. A primeira fase é utilizada para estabelecer o controle, testando se o processo estava sob controle quando a amostra foi retirada e as estatísticas  $\bar{\bar{X}}$  e  $S$  foram calculadas. Esses limites são definidos em (4).

$$\begin{aligned} \text{LSC} &= \frac{p(m-1)(n-1)}{mn-m-p+1} F_{\alpha, p, mn-m-p+1} \\ \text{LIC} &= 0 \end{aligned} \quad (4)$$

em que:

$p$  - é o número de variáveis que estão sendo analisadas;  
 $m$  - é o número de subgrupos;  
 $n$  - é o tamanho das amostras;  
 $F$  - é o valor da estatística tabelada;  
 $\alpha$  - é o nível de significância especificado.

Após ter verificado a estabilidade do processo, é possível estabelecer novos limites de controle que terão a função de monitorá-lo, caracterizando, assim, uma segunda fase de controle, e mostrados em (5).

$$\begin{aligned} \text{LSC} &= \frac{p(m+1)(n-1)}{mn-m-p+1} F_{\alpha, p, mn-m-p+1} \\ \text{LIC} &= 0 \end{aligned} \quad (5)$$

Muitas vezes, faz-se necessário estabelecer o controle estatístico quando o processo é considerado com o subgrupo de amostra de tamanho igual a um ( $n = 1$ ). Nesse caso, a estatística  $T^2$ , em (3), torna-se:

$$T^2 = (X - \bar{\bar{X}})' S^{-1} (X - \bar{\bar{X}}) \quad (6)$$

Assim, os novos limites de controle a serem utilizados estão em (7).

$$\begin{aligned} \text{LSC} &= \frac{p(m+1)(m-1)}{m^2-mp} F_{\alpha, p, m-p} \\ \text{LIC} &= 0 \end{aligned} \quad (7)$$

## GRÁFICOS DE CONTROLE X-BARRA COM LIMITES DE BONFERRONI

É um método alternativo para comparações múltiplas, desenvolvido com base na desigualdade de probabilidades de Bonferroni em que se faz um ajuste estatístico para o nível de significância ( $\alpha$ ), em relação ao número de testes realizados, simultaneamente, para um conjunto de dados.

Dewey (2002) reinterpreta a desigualdade geral de Bonferroni, em (8), como sendo a probabilidade de não-ocorrência do evento A que denota uma afirmação de confiança sobre o valor da combinação linear com  $P(A, \text{ser verdadeiro}) = 1 - \alpha_i$ .

$$P\left(\bigcap_{i=1}^m \bar{A}_i\right) \geq 1 - \sum_{i=1}^m P[A_i] \quad (8)$$

em que  $A_i$  e seu complemento são eventos quaisquer, e  $m$  é o número de comparações ou combinações.

Quando o gráfico de controle multivariado indicar que o processo está fora de controle, a desigualdade de Bonferroni pode ser usada para investigar qual das  $p$  características de qualidade é a responsável por tal alteração. Para tanto, seja  $A_i$  o evento em que a média da  $i$ -ésima característica de qualidade  $i = 1, 2, \dots, p$  está dentro dos limites de controle especificados, com  $\alpha$  substituído por  $\frac{\alpha}{p}$ . A desigualdade de Bonferroni, nesse caso, afirma que:

$$P(A_1 \cap A_2 \cap \dots \cap A_p) \geq 1 - \alpha \quad (9)$$

Então, se são desejadas estimativas simultâneas, de múltiplos intervalos, com um intervalo de confiança geral  $(1-\alpha)$ , pode-se construir cada intervalo com confiança de  $(1-\frac{\alpha}{m})$ , e a desigualdade de Bonferroni em (9) assegura que o coeficiente de confiança geral é de, pelo menos,  $(1-\alpha)$ .

Esse ajuste possui alguns inconvenientes e a desigualdade de Bonferroni, em gráficos de controle, deve ser utilizada com cautela. Diferentes situações para a decisão sobre o uso da correção de Bonferroni devem ser analisadas (PERNEGER, 1998; BLAND e ALTMAN, 1995).

## ANÁLISE DE COMPONENTES PRINCIPAIS

A Análise de Componentes Principais (ACP) é um método estatístico multivariado, que pode ser utilizado para compressão de dados, redução de dimensionalidade de conjuntos de dados multivariados, extração de caracté-

terísticas, e projeção de dados multivariados. Mediante essa técnica, é possível obter uma combinação linear de um conjunto de variáveis, que retenha o máximo possível da informação contida nas variáveis originais. Esse novo conjunto é chamado de componentes principais (CPs) (KHATTREE e NAIK, 2000).

As CPs são independentes, favorecendo a análise, especialmente quando vários fatores estão envolvidos, como no controle estatístico de processos. Elas são extraídas na ordem da mais explicativa para a menos explicativa. Teoricamente, o número de CPs é sempre igual ao número de variáveis, entretanto, algumas poucas componentes são responsáveis por grande parte da explicação total.

### Determinação das componentes principais de uma população

Seja o vetor aleatório  $X' = [X_1, X_2, \dots, X_p]$ , com matriz de covariância  $\Sigma$  e com autovalores  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$ . Considerando-se a combinação linear, baseada nos autovalores ( $\lambda$ ):

$$\begin{aligned} Y_1 &= \ell'_1 X = \ell_{11} X_1 + \ell_{21} X_2 + \dots + \ell_{p1} X_p \\ Y_2 &= \ell'_2 X = \ell_{12} X_1 + \ell_{22} X_2 + \dots + \ell_{p2} X_p \\ &\dots \\ Y_p &= \ell'_p X = \ell_{1p} X_1 + \ell_{2p} X_2 + \dots + \ell_{pp} X_p \end{aligned} \quad (10)$$

Sabe-se que:

$$\text{Var}(Y_i) = \ell'_i \Sigma \ell_i \quad i = 1, 2, \dots, p \quad (11)$$

$$\text{Cov}(Y_i, Y_k) = \ell'_i \Sigma \ell_k \quad i, k = 1, 2, \dots, p \quad (12)$$

As CPs são essas combinações lineares não correlacionadas de  $Y_1, Y_2, \dots, Y_p$ , cuja variância é a maior possível.

A primeira componente é a combinação linear com máxima variância, ou seja, é a combinação que maximiza a  $\text{Var}(Y_i) = \ell'_i \Sigma \ell_i$ . Por conveniência, restringe-se o vetor de coeficientes à unidade. Então, define-se que a  $i$ -ésima componente principal ( $CP_{i\text{-ésima}}$ ) é a combinação linear  $\ell'_i X$  que maximiza  $\text{Var}(\ell'_i X)$  sujeita às restrições de  $\ell'_i \ell_i = 1$ , e  $\text{Cov}(\ell'_i X, \ell'_k X) = 0$  para  $k < i$ .

As restrições garantem que o sistema tem solução única, que as CPs são não-correlacionadas e apresentam variância máxima (JOHNSON & WICHERN, 1992).

### Seleção das componentes principais

Seja o vetor aleatório  $X' = [X_1, X_2, \dots, X_p]$ , com matriz de covariância  $\Sigma$ , com  $(\lambda_1, \ell_1), (\lambda_2, \ell_2) \dots (\lambda_p, \ell_p)$  pares de autovalores e autovetores associados, em que  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$ , e a  $p$ -ésima componente principal é dada por

$$Y_p = \ell'_p X = \ell_{1p} X_1 + \ell_{2p} X_2 + \dots + \ell_{pp} X_p$$

Para se obter a proporção da variância total devida, ou explicada por uma componente principal qualquer, utiliza-se o seguinte resultado obtido por Johnson e Wichern (1992).

A variância total da população é

$$\sigma_{11} + \sigma_{22} + \dots + \sigma_{pp} = \sum_{i=1}^p \text{Var}(X_i) = \lambda_1 + \lambda_2 + \dots + \lambda_p = \sum_{i=1}^p \text{Var}(Y_i) = \text{tr}(\Sigma)$$

e a proporção da variância total explicada pela i-ésima CP é dada por

$$\frac{\lambda_i}{\lambda_1 + \lambda_2 + \dots + \lambda_p} \quad i = 1, 2, \dots, p$$

De acordo com Johnson & Wichern (1992), se grande parte da variabilidade total, mais de 70%, pode ser explicada por uma, duas ou três componentes principais, então essas componentes poderão substituir as variáveis originais, sem grande perda de informações.

Existem outros critérios para a seleção de componentes principais, entre eles se pode destacar o critério sugerido por Kaiser *apud* Mardia (1979) e o método gráfico de Cattell (1966).

**Componentes principais no controle estatístico de processos**

A busca da variável, ou variáveis, causadora(s) de uma instabilidade em um processo, quando se utiliza CPs, realiza-se pelo gráfico X-barra das CPs. A componente que sinalizar uma instabilidade será, então, investigada com base no estudo dos coeficientes de correlação entre ela e as variáveis originais. Observando-se esse cálculo, identifica-se qual é a variável que está mais fortemente correlacionada com a CP e que, possivelmente, está afetando a componente.

Hawkins e Fatti (1984) dizem que as primeiras CPs não são sensíveis para detectar perturbações ocorridas nas variáveis, assim a análise das componentes de menor importância, isto é, as combinações lineares das variáveis X, que possuem variância menor, podem ser usadas para detectar as possíveis causas de uma falta de controle no sistema. Para isso, segundo Gnanadesikan e Ketterink (1972), analisam-se os diagramas de dispersão dos pares das últimas componentes principais para buscar por pontos considerados suspeitos, ou seja, pontos que não seguem o padrão usual dos demais.

A determinação da variável, que possui maior influência na combinação linear, será encontrada pelas correlações:

$$r_{\hat{Y}_i, X_k} = \frac{\hat{e}_{ki} \sqrt{\hat{\Lambda}_i}}{\sqrt{S_{kk}}} \quad ; i, k = 1, 2, \dots, p \quad (13)$$

$$r_{\hat{Y}_i, Z_k} = \hat{e}_{ki} \sqrt{\hat{\Lambda}_i} \quad ; i, k = 1, 2, \dots, p \quad (14)$$

que apontam, mediante seus valores absolutos, as variáveis que exercem maior influência sobre a componente principal. O sinal da correlação indica o modo dessa influência (Johnson & Wichern, 1992), sendo, assim, identificada aquela que deve ser monitorada para manter o sistema estável.

A equação (13) deve ser utilizada quando os autovetores são obtidos com base na matriz de variância S, e a equação (14) quando obtidos da matriz de correlação R.

Em controle estatístico de qualidade, quando duas ou mais componentes apresentam valores observados fora dos limites de controle, deve-se estabelecer uma ordem hierárquica entre elas, para auxiliar na solução do conflito quanto à variável de maior influência sobre a perda de controle. Pode-se, nesse caso, ficar em dúvida quanto a dar mais atenção a uma componente em detrimento de outra; então, deve-se levar em consideração o maior autovalor que originou essa componente, optando-se por esta.

**APLICAÇÃO DA METODOLOGIA DE GRÁFICOS X-BARRA COM LIMITES DE BONFERRONI**

Para essa aplicação foi gerado um conjunto de quatro variáveis, cada uma com 100 observações, seguindo uma distribuição normal, com fraca correlação, e sob controle estatístico. Essas variáveis foram denominadas V1, V2, V3 e V4. Depois de assegurado o controle estatístico das quatro variáveis, a um nível de significância de 5%, as primeiras 10 observações da variável V1 foram propositalmente geradas com uma distribuição normal, e com média diferente da original, com uma diferença de magnitude de 3,5 desvios-padrão. O mesmo foi feito, em seguida, com as observações de 11 a 20, e assim, sucessivamente, até as observações 91 a 100. Procedimento similar foi adotado para as variáveis V2, V3 e V4.

Aqui serão apresentados apenas os resultados do procedimento envolvendo a Variável 1, quando foram retiradas de controle as observações de 1 a 10 (indicada por V1,1-10); a Variável 2 (V2), a Variável 3 (V3) e a Variável 4 (V4), estas três últimas, sem observações alteradas. Uma vez que o mesmo procedimento é repetido para todas as demais variáveis, os outros resultados serão apenas comentados posteriormente.

Nos gráficos X-barra, usados para testar se as quatro variáveis estavam sob controle estatístico, foi utilizado o valor de  $\alpha = 5\%$ , para manter o mesmo nível de significância utilizado no gráfico de controle  $T^2$  de Hotelling. De acordo com Kourti e MacGregor (1996), para se utilizarem os limites de Bonferroni, em controle de qualidade, é necessário, primeiro, que as variáveis sejam fracamente correlacionadas e, a seguir, seja realizada uma avaliação conjunta destas, pelo gráfico de

controle  $T^2$  de Hotelling, a fim de assegurar a significância do teste. Assim, em primeiro lugar será realizada a verificação da estabilidade do sistema, pelo gráfico de controle  $T^2$  de Hotelling.

Estando o processo sob controle estatístico, segue-se o monitoramento dos dados, caso contrário, havendo um sinal de instabilidade, imediatamente segue-se a investigação da causa geradora deste, para que a devida correção seja realizada e, assim, o processo retorne a sua estabilidade.

A Figura 2 mostra o gráfico de controle  $T^2$  de Hotelling, para as variáveis em estudo, com  $\alpha = 5\%$ .

Observa-se que o gráfico de controle  $T^2$  de Hotelling assinala uma instabilidade no processo, porém não identifica a variável, ou as variáveis, causadora(s) dessa instabilidade. A identificação será realizada a seguir, após verificar-se a fraca correlação entre as variáveis,

indicada na Tabela 1.

Tendo em vista esse resultado, a identificação da variável causadora da instabilidade será realizada utilizando-se o gráfico de controle X-barra com limites de Bonferroni, uma vez que já foram verificados, anteriormente, os requisitos para o uso dessa metodologia, ou seja, as variáveis são fracamente correlacionadas e o conjunto de observações, no gráfico de controle  $T^2$  de Hotelling, foi significativo a um nível de 5%.

Os gráficos de controle X-barra, para as quatro variáveis, utilizando-se os limites de Bonferroni, isto é, com o valor de  $\alpha$  dividido pelo número de variáveis em estudo, ou seja:

$$\alpha' = \frac{\alpha}{n} = \frac{\alpha}{4} = \frac{5\%}{4} = 1,25\%$$

são mostrados nas Figuras 3, 4, 5 e 6.

Figura 2: Gráfico  $T^2$  de Hotelling para as variáveis V1, 1-10, V2, V3 e V4.

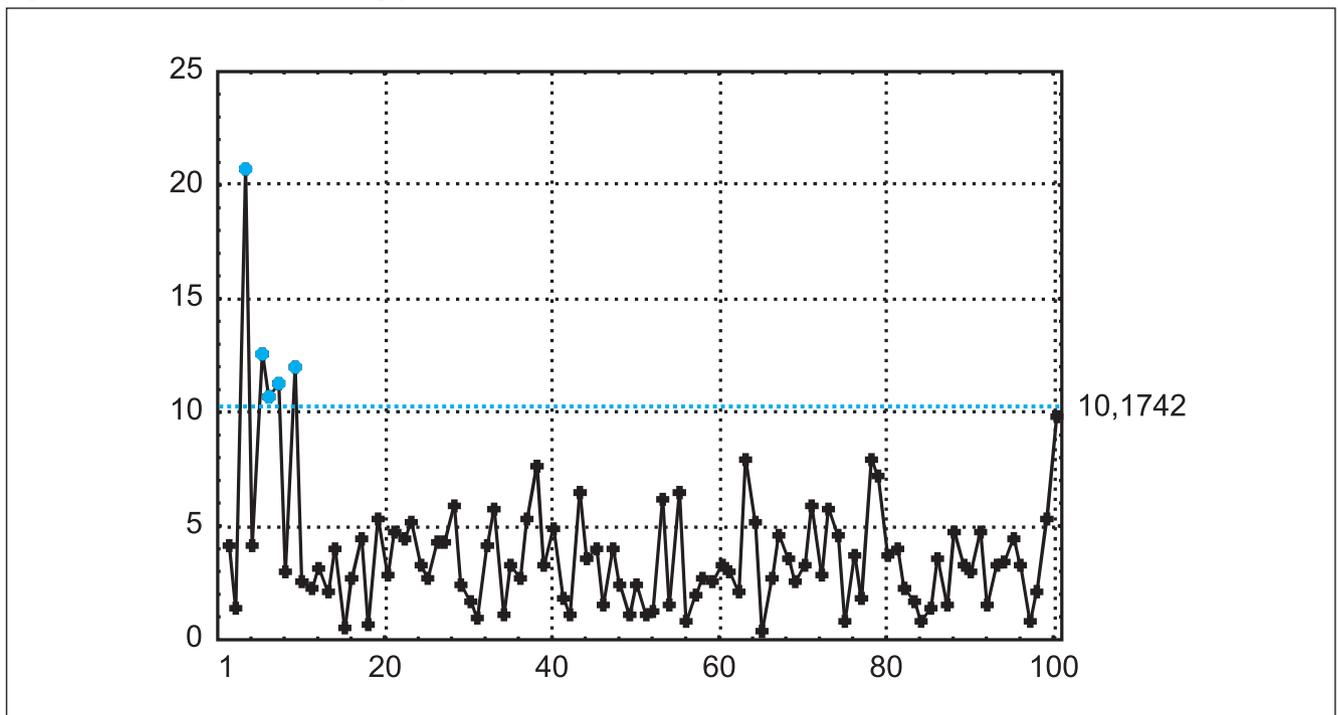


Tabela 1: Matriz de correlação entre as variáveis V1, 1-10, V2, V3 e V4.

	VARIÁVEL 1,1-10	VARIÁVEL 2	VARIÁVEL 3	VARIÁVEL 4
Variável 1,1-10	1,00000			
Variável 2	0,02422	1,00000		
Variável 3	-0,01205	0,10575	1,00000	
Variável 4	0,06644	-0,02603	0,05082	1,00000

A Figura 3 mostra que essa metodologia identificou corretamente que a variável V1 estava fora de controle estatístico, da observação 1 a 10, sendo, portanto, a causadora da instabilidade do processo, uma vez que as outras variáveis estavam sob controle estatístico.

Uma situação que merece bastante atenção é o caso em

que o gráfico de controle multivariado com nível de significância  $\alpha$  sinaliza uma situação fora de controle, e os gráficos individuais com limites de Bonferroni ao nível de significância  $\alpha/n$  não o captam. Nesse caso, uma das opções sugeridas é a de estreitar os limites de controle dos gráficos individuais.

Figura 3: Gráfico X-barra com limites de Bonferroni para a variável V1, 1-10.

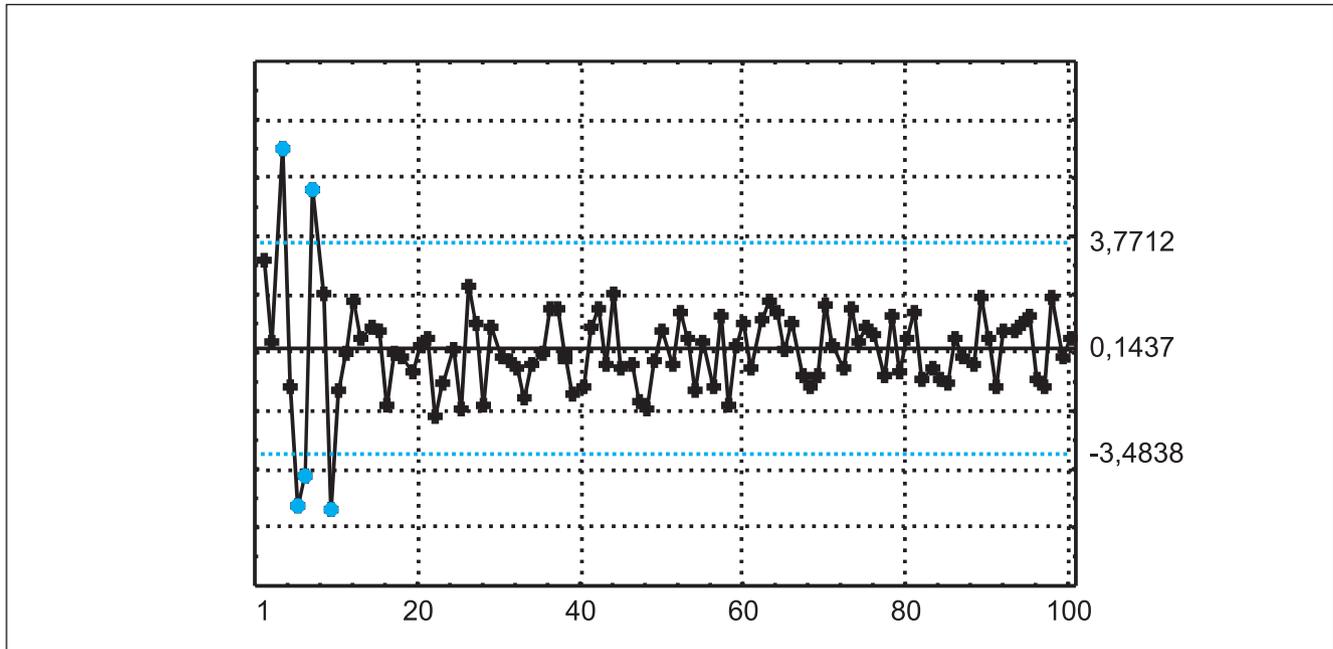
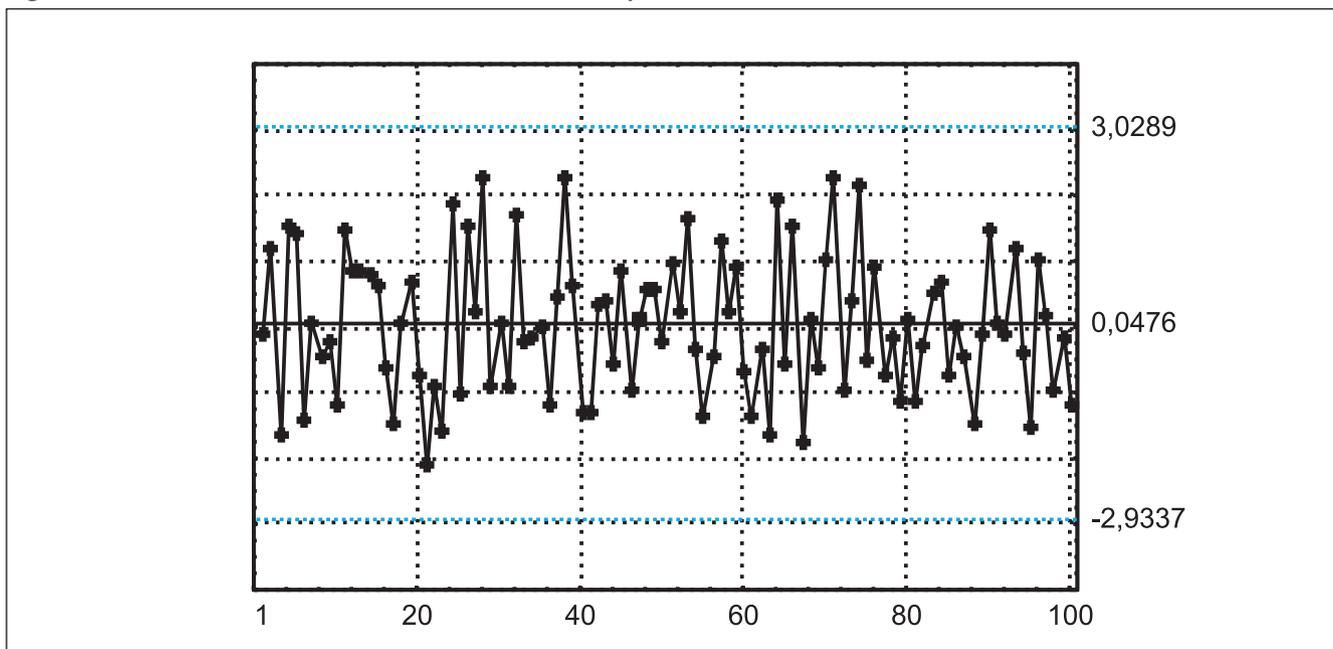


Figura 4: Gráfico X-barra com limites de Bonferroni para a variável V2.



Observa-se que, em todos os outros conjuntos de variáveis analisados nesse trabalho, a metodologia utilizada identificou corretamente a variável causadora da instabilidade do processo, bem como o período em que ela se encontrava fora de controle, ficando, assim, demonstrando que é perfeitamente possível sua utilização em contro-

le de qualidade. Isto é verdadeiro, desde que sejam assegurados os requisitos de que as variáveis a serem monitoradas tenham uma correlação fraca, e de que seja garantido o nível de significância global, utilizando inicialmente no gráfico de controle  $T^2$  de Hotelling, quando da avaliação conjunta das variáveis.

Figura 5: Gráfico X-barra com limites de Bonferroni para a variável V3.

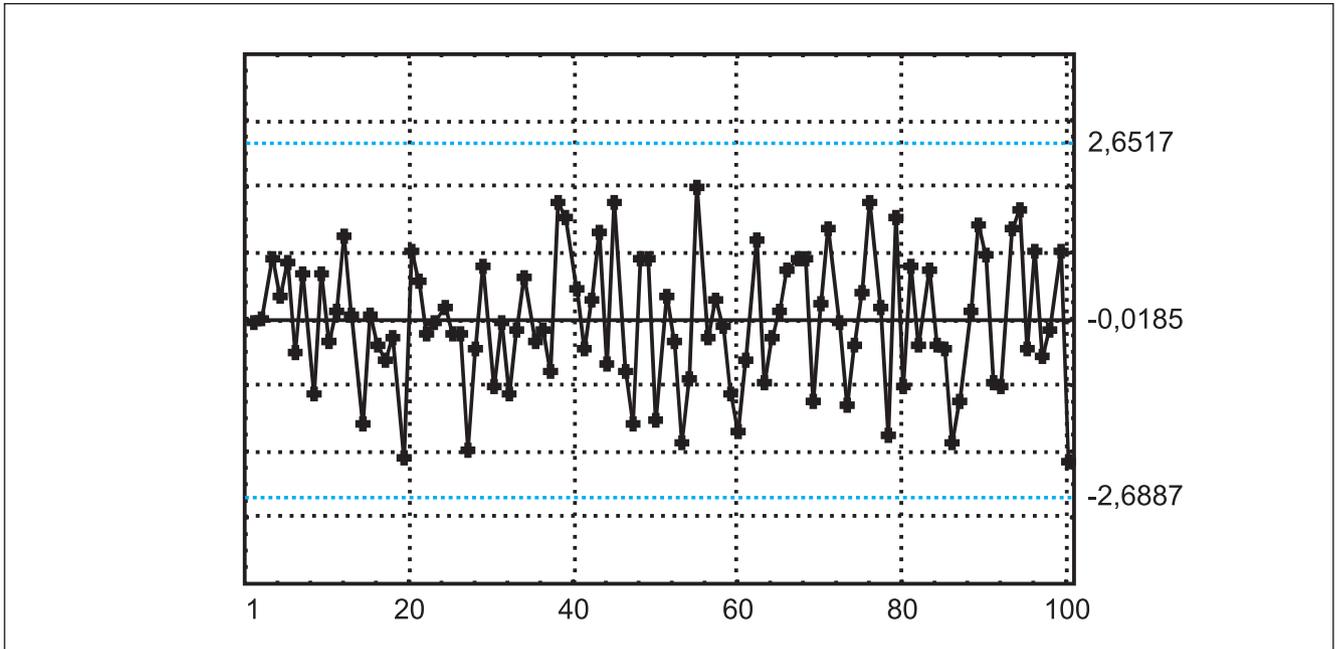
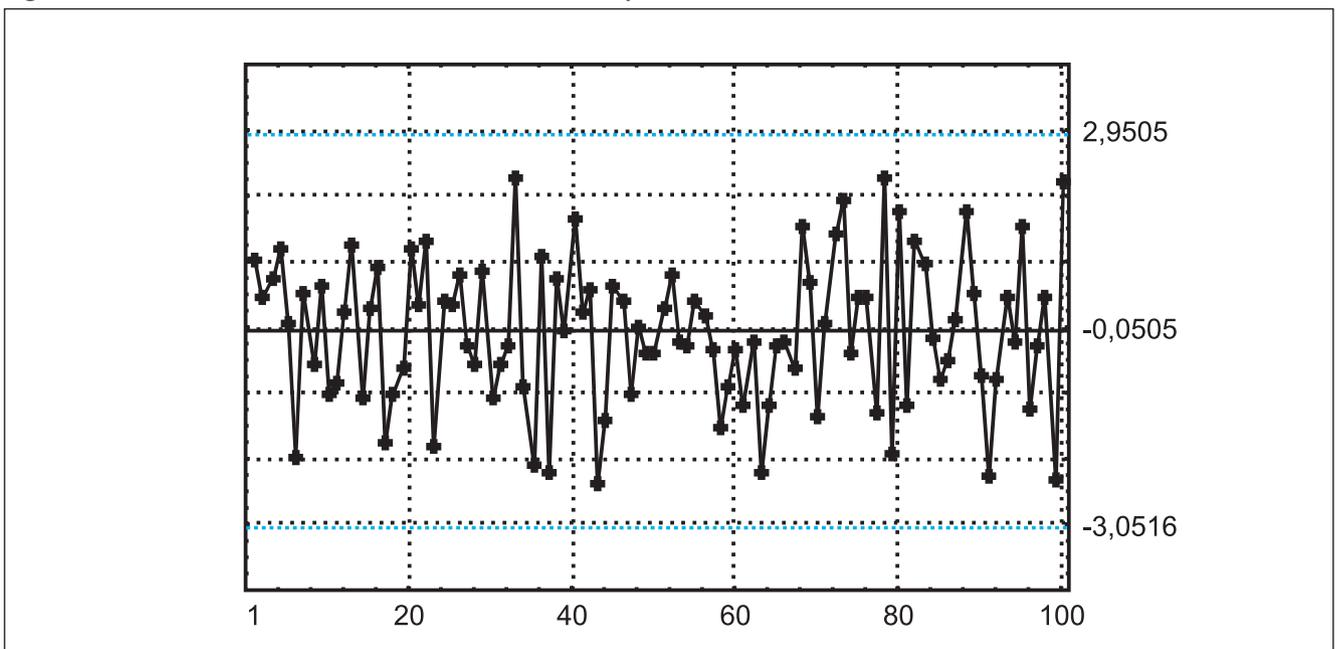


Figura 6: Gráfico X-barra com limites de Bonferroni para a variável V4.



## APLICAÇÃO DA METODOLOGIA DE ANÁLISE DE COMPONENTES PRINCIPAIS

Para a aplicação da análise de componentes principais também foram geradas 4 variáveis, porém fortemente correlacionadas. O procedimento para a geração das variáveis foi o mesmo utilizado na aplicação da metodologia de limites de Bonferroni.

A verificação da estabilidade do sistema se realiza partindo do gráfico de controle  $T^2$  de Hotelling. Se o processo estiver sob controle estatístico, não há problemas, e segue-se o monitoramento dos dados, até que uma causa especial afete o sistema, seguida de sua identificação e posterior intervenção para manutenção da estabilidade deste.

A Figura 7 mostra o gráfico de controle  $T^2$  de Hotelling para as variáveis em estudo.

O gráfico  $T^2$  de Hotelling identifica, claramente, a instabilidade do processo nas primeiras observações,

porém não é capaz de identificar qual ou quais são as variáveis causadoras desta instabilidade, assim, uma investigação mais minuciosa será realizada.

Para prosseguir com a investigação de qual ou quais variáveis estão fora de controle, faz-se necessário confirmar a correlação existente entre elas. A Tabela 2 mostra essas correlações.

Como se pode observar na Tabela 2, a correlação entre as variáveis é forte, o que permite a investigação da instabilidade do processo com base na análise de componentes principais (ACP).

Todas as CPs serão observadas pelo gráfico de controle X-barra. Quando um gráfico assinalar a falta de controle da CP, a matriz de correlação desta CP com as variáveis originais será observada, para verificar qual a variável que mais contribui com esta CP e se, realmente, corresponde à variável cujos valores observados foram alterados.

Figura 7: Gráfico  $T^2$  de Hotelling para as variáveis V1, 1-10, V2, V3 e V4.

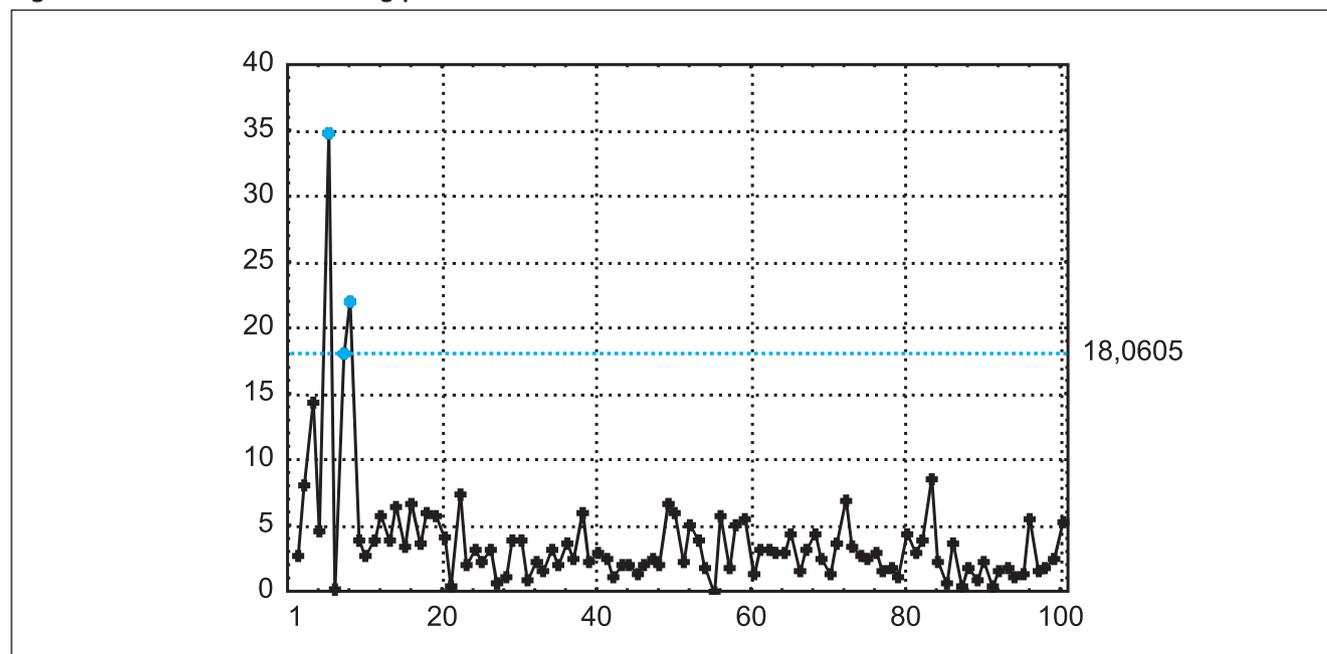


Tabela 2: Matriz de correlação entre as variáveis V1,1-10, V2, V3 e V4.

	VARIÁVEL 1,1-10	VARIÁVEL 2	VARIÁVEL 3	VARIÁVEL 4
Variável 1,1-10	1,00000			
Variável 2	0,80302	1,00000		
Variável 3	-0,82804	0,92347	1,00000	
Variável 4	0,82158	-0,90797	0,92456	1,00000

Para proceder à análise de componentes principais, calculam-se os autovalores, e seus correspondentes autovetores, que são mostrados nas Tabelas 3 e 4. Neste estudo, utilizou-se a matriz das correlações para o cálculo das componentes principais (CP), pois, dessa forma, se trabalha com as variáveis padronizadas, evitando-se, assim, a ocorrência de problemas em razão da influência das magnitudes das variáveis, ou a utilização de unidades de medidas diferentes destas.

Desse modo, tem-se as seguintes CPs

$$Y1 = 0,25136V1_{1-10} + 0,26565V2 + 0,26869V3 + 0,26706V4$$

$$Y2 = -1,8158V1_{1-10} + 0,72894V2 + 0,48914V3 + 0,49191V4$$

$$Y3 = 0,23818V1_{1-10} + 2,26349V2 - 0,06850V3 - 2,40680V4$$

$$Y4 = -0,16436V1_{1-10} - 1,48796V2 + 3,12022V3 - 1,50442V4$$

As Figuras 8, 9, 10 e 11 mostram os gráficos X-barra para as quatro componentes.

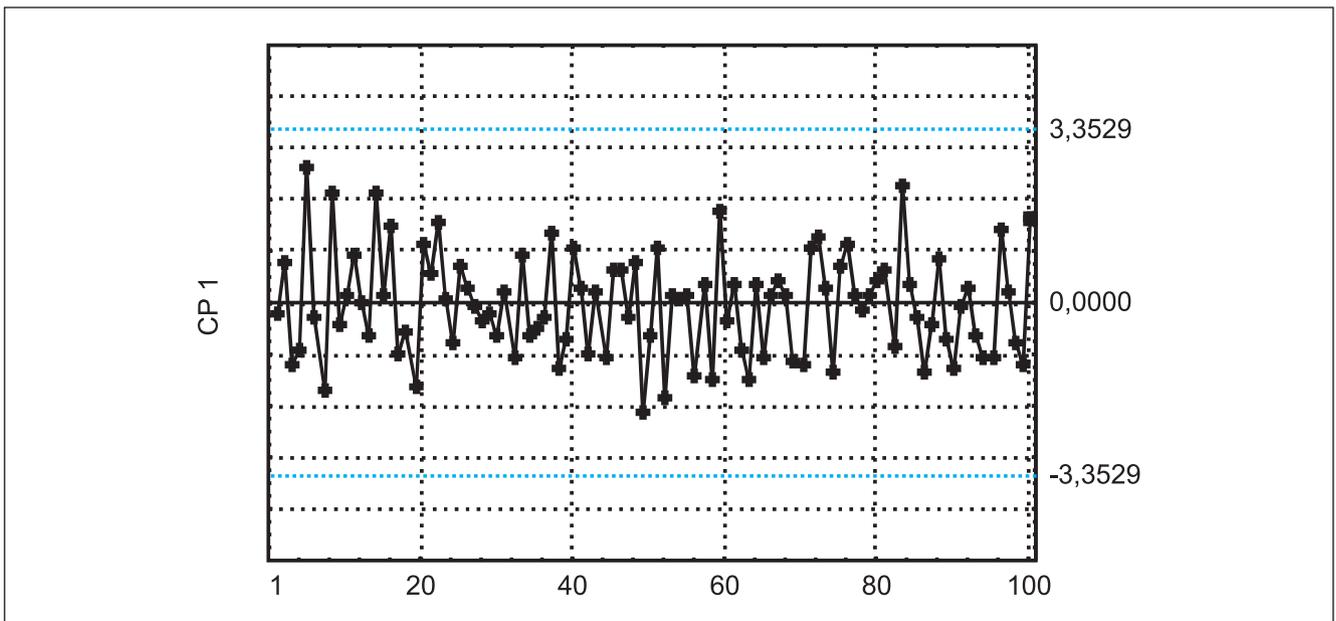
Tabela 3: Autovalores e porcentagem da variância explicada.

	AUTOVALOR	% DA VARIÂNCIA EXPLICADA	AUTOVALOR ACUMULADO	% DE VARIÂNCIA ACUMULADA
1	3,60666	90,16640	3,60666	90,16640
2	0,23202	5,80062	3,83868	95,96701
3	0,09110	2,27738	3,92978	98,24440
4	0,07022	1,75560	4,00000	100,00000

Tabela 4: Autovetores, coeficientes das componentes principais.

	$l_1$	$l_2$	$l_3$	$l_4$
Variável 1,1-10	0,25136	-1,81585	0,23818	-0,16436
Variável 2	0,26565	0,72894	2,26349	-1,48796
Variável 3	0,26869	0,48914	-0,06850	3,12022
Variável 4	0,26706	0,49191	-2,40680	-1,50442

Figura 8: Gráfico X-barra para a primeira componente CP1.



A observação dos gráficos de controle X-barra mostra que somente o gráfico da CP 2 está fora de controle. Realizando-se uma investigação sobre a contribuição de cada variável original, com as CP, temos, na Tabela 5, as suas correlações.

A Tabela 3 mostra que a primeira CP explica 90,16% da variância. Se o estudo fosse para redução de dimensionalidade, seria suficiente a utilização da pri-

meira componente para a análise.

Interessa neste estudo tão somente identificar as componentes mais sensíveis a causas especiais que são geradas por autovalores menores. Segue-se, então, para a observação do comportamento da segunda CP, que se mostrou fora de controle estatístico, quando avaliada pelo gráfico de controle X-barra. A segunda CP possui maior correlação (em valor absoluto) com a Var1,1-10,

Figura 9: Gráfico X-barra para a segunda componente CP2.

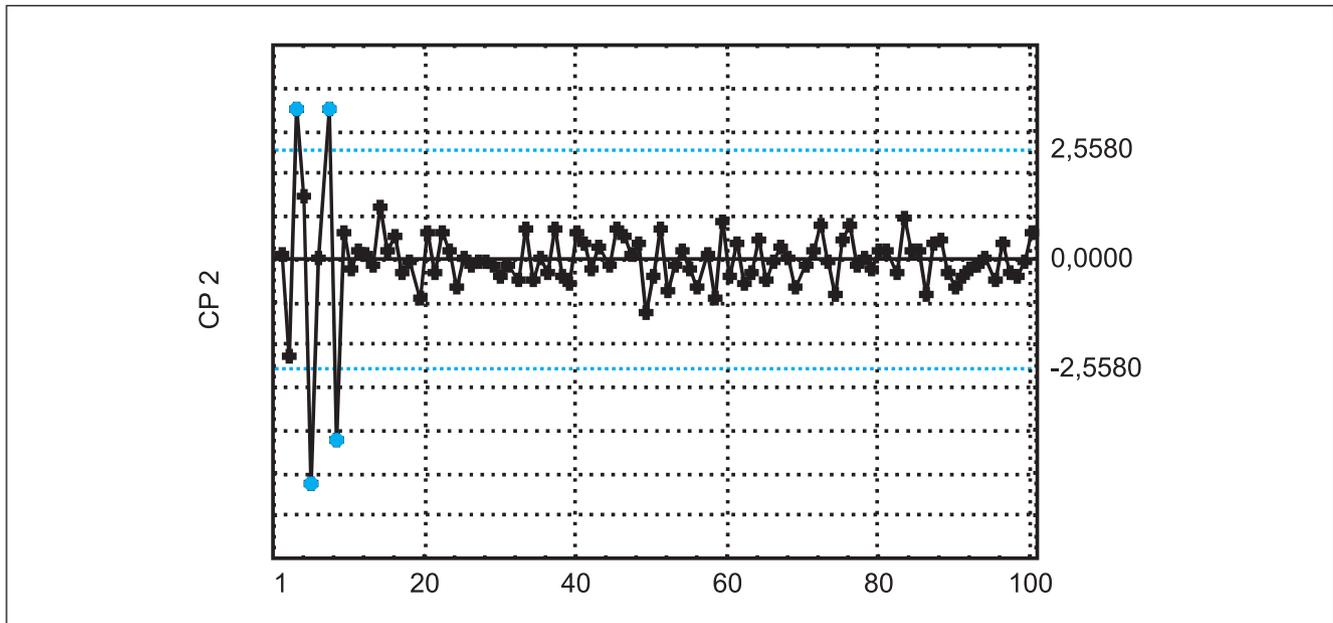
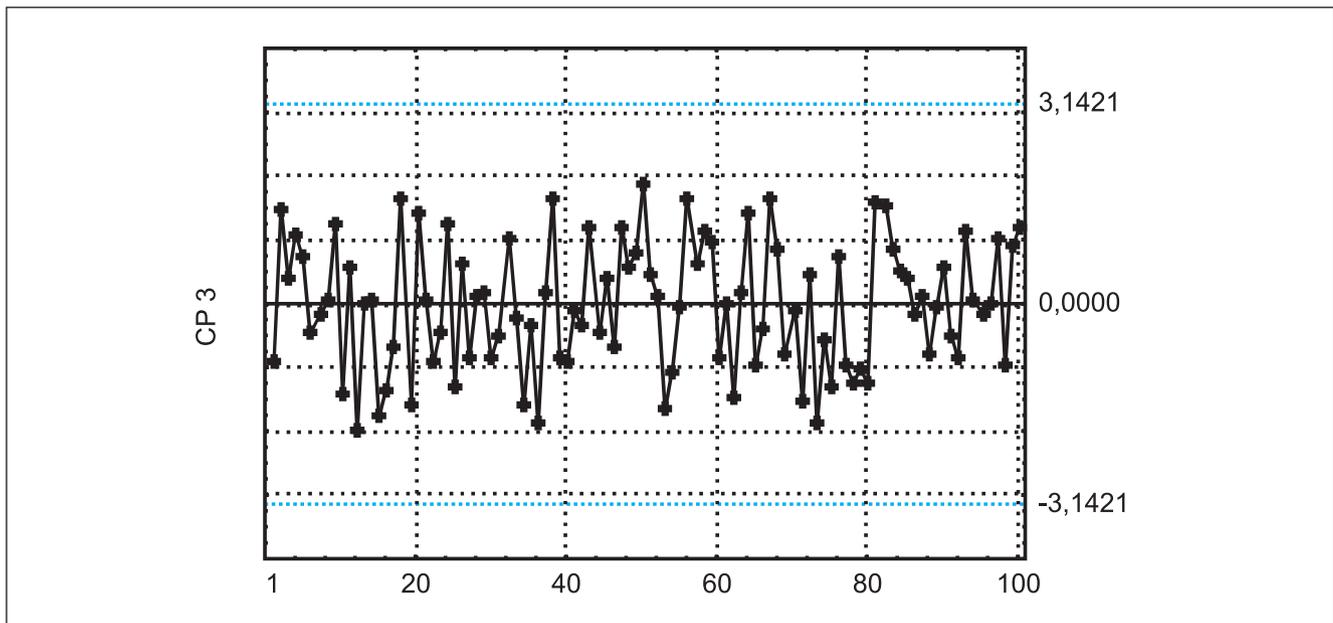


Figura 10: Gráfico X-barra para a terceira componente CP3.



indicando, assim, que essa variável é a que possui maior contribuição para a componente de interesse, justificando, assim, o fato de ter sido a única, sinalizada fora de controle, pelo gráfico X-barra.

Em todos os outros conjuntos de dados avaliados, os resultados foram equivalentes.

## CONCLUSÃO

O desenvolvimento desta pesquisa permitiu que fossem mostradas duas técnicas estatísticas multivariadas, aplicadas em controle de qualidade. Uma delas considerando variáveis fracamente correlacionadas, e a outra considerando variáveis fortemente correlacionadas.

No caso das variáveis fracamente correlacionadas, após serem diagnosticadas fora de controle, pela utilização do gráfico de controle  $T^2$  de Hotelling, a técnica utilizada foi o gráfico de controle X-barra com limites de Bonferroni. Essa técnica identificou, corretamente, a variável causadora da instabilidade do sistema e o período em que esta esteve fora de controle.

No caso das variáveis fortemente correlacionadas, a avaliação conjunta foi realizada pelo gráfico de controle  $T^2$  de Hotelling. Quando este sinalizou que o processo estava fora de controle, utilizou-se a técnica de análise de componentes principais, seguida da utilização do gráfico X-barra. Essa técnica identificou, em todos os casos, qual foi a variável colocada fora de controle, bem como assinalou o período em que esteve fora de controle.

Uma importante contribuição deste trabalho foi mostrar que, quando as variáveis possuem forte correlação, é possível a utilização das menores componentes principais para a identificação das variáveis fora de controle. Em todos os casos aqui analisados, a segunda componente principal sinalizou a variável que havia sido colocada, propositalmente, fora de controle, e não a primeira componente, ou seja, aquela que seria a selecionada pelos métodos de Johnson e Wichern (1992), Kaiser *apud* Mardia (1979) ou Cattel (1966).

A utilização, apenas, das primeiras componentes principais, como geralmente ocorre nos estudos de compressão e análise de dados, nem sempre revela pontos fora de

Figura 11: Gráfico X-barra para a quarta componente CP4.

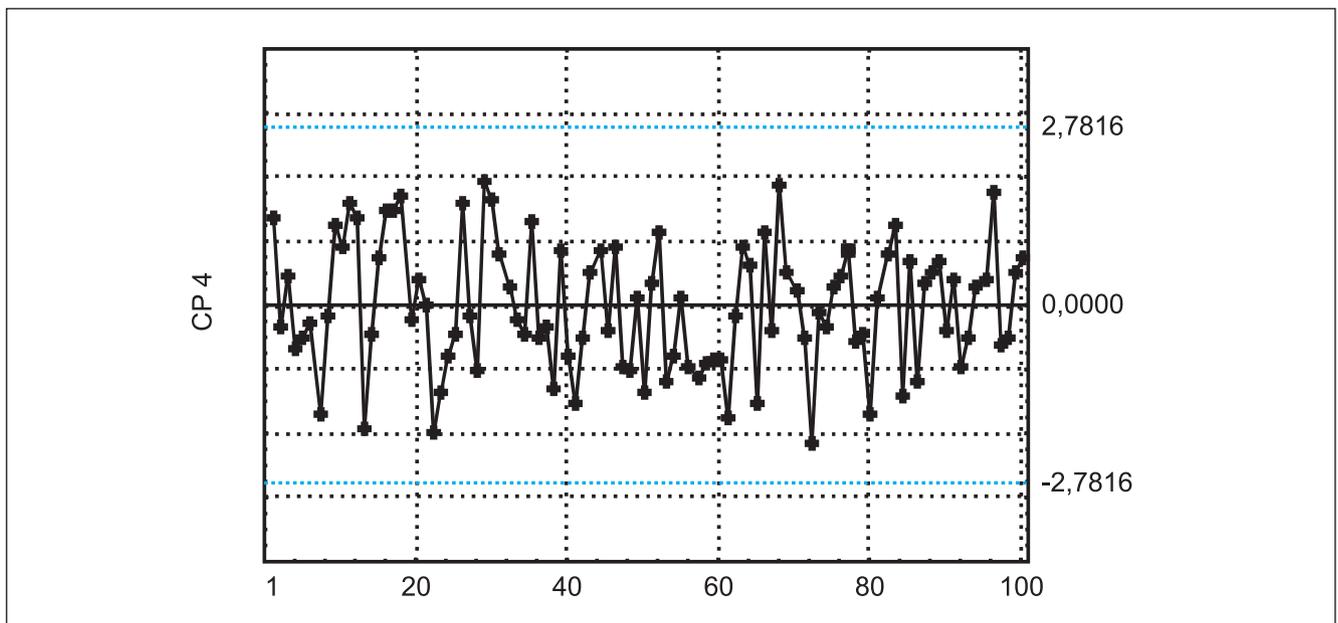


Tabela 5: Correlação entre as variáveis originais e as CPs, com destaque para as que são significativas a um nível de 5%.

	CP 1	CP 2	CP 3	CP 4
Variável 1,1-10	<b>0,906578</b>	<b>-0,421321</b>	0,021697	-0,011542
Variável 2	<b>0,958103</b>	0,169131	<b>0,206194</b>	-0,104491
Variável 3	<b>0,969056</b>	0,113493	-0,00624	<b>0,219115</b>
Variável 4	<b>0,963194</b>	0,114136	<b>-0,219248</b>	-0,105647

controle. Pequenas variações, esperadas em processos produtivos, são melhor captadas pelas menores componentes, e não pelas primeiras componentes, pois estas são as responsáveis por grandes variabilidades em um conjunto de dados. Deve-se levar em consideração que, nesta pesquisa, as variáveis originais foram geradas com uma correlação superior a 0,90.

Sugere-se, para trabalhos futuros, uma análise de sen-

sibilidade das metodologias apresentadas, para delimitar situações em que essas técnicas são eficazes e o emprego de outras magnitudes de média e desvio-padrão. E, no caso da metodologia de componentes principais, sugere-se a investigação de processos produtivos em que as características de qualidade possuam correlação menos forte do que a investigada.

**Artigo recebido em 27/08/2003**

**Aprovado para publicação em 30/09/2004**

### ■ Referências Bibliográficas

APARISE, F. Sampling plans for the multivariate  $T^2$  control chart. *Quality Engineering*, v. 10, n. 1, p.141-147, 1997.

BLAND, J.M. & ALTMAN, D.G. Multiple significances test: the Bonferroni method. *British Medical Journal*, n. 310:170, 21, jan. 1995.

CATTEL, R.B. The screen test for the number of factors. *Multivariate Behavior Research*, v.1, p. 245-276, 1966.

DEWEY, M.E. *Bonferroni e le diduguaglianze*. Disponível em: <<http://www.nottingham.ac.uk/~mh2md/bonf.html>>. Acesso em: 02 abr.2002.

GNANADESIKAN, R & KETTEREING, J. R. Robust estimates, residual and outlier detection with multiresponse data. *Biometrics*, v. 28, p. 81-124, 1972.

HAWKINS, D.M & FATTI, L.P. Exploring multivariate data using minor principal components. *The Statistician*, v. 33, p. 325-338, 1984.

HOTELLING, H. *Multivariate quality control*. Techniques of statistical analysis. New York: Mc Graw Hill, p. 111-184, 1947.

JACKSON, J. EDWARD. Quality control methods. *Industrial Quality Control*, p. 4-8, Jan. 1956.

JACKSON, J. EDWARD. Quality control methods for several related variables. *Technometrics*, p. 359-377, nov. 1959.

JACKSON, J.E. & MORRIS, R.H. An application of multivariate quality control to photographic processing. *Journal of the American Statistical Association*, v. 52, p. 186-199, 1957.

JOHNSON, R.A. & WICHERN, D.W. *Applied multivariate statistical analysis*. 3. ed. New Jersey : Prentice-Hall, 1992.

KHATTREE, R. & NAIK, D.N. *Multivariate data reduction and discrimination*. Cary, USA: SAS Institute Inc., 558 p., 2000.

KOURTI, T. & MACGREGOR, J.F. Multivariate SPC methods for process and product monitoring. *Journal of Quality Technology*, v. 28, n. 4, p. 409-428, Oct. 1996.

LOWRY, C.A. & MONTGOMERY, D.C.A review of multivariate control chart. *IIE Transaction*, v. 27, p. 800-810, 1995.

MARDIA, K.V. *et al. Multivariate analysis*. London : Academic, 1979.

MASON, Robert L. & YOUNG, John C. *Multivariate statistical process control with industrial applications*. Philadelphia: Society for Industrial and Applied Mathematics, 263 p., 2002.

MEYER, P. L. *Probabilidade: aplicações à estatística*. 2. ed. Rio de Janeiro: Livros Técnicos e Científicos, 426 p., 1983

MONTGOMERY, D.C. *Introduction to statistical quality control*. 2. ed. New York: John Wiley & Sons, 674 p., 1991.

PERNEGER, T. V. What's wrong with Bonferroni adjustment. *British Medical Journal*, n. 316, p. 1236-1238, 18 Apr., 1998.

PFEIFFER, P.E. *Concepts of probability theory*. 2. ed. New York: Dover Publications, 403 p., 1978.

RYAN, T.P. *Statistical methods for quality improvement*. New York: John Wiley D& Sons, Inc, 1989.

TRACY, N.D. *et al. Multivariate Control Charts for Individual Observations*. *Journal of Quality Technology*, v. 24, p. 88-95, 1992.

TRACY, N.D. *et al. A bivariate control chart for paired measurements*. *Journal of Quality Technology*, v. 27, p. 370-376, 1995.

### ■ Agradecimentos

Agradecemos à Fapergs (Fundo de Amparo à Pesquisa do Rio Grande do Sul) pelo auxílio financeiro parcial a esta pesquisa. Agradecemos também aos revisores pelas contribuições e sugestões.

### ■ Sobre os autores

#### **Adriano Mendonça Souza, Dr.**

Professor do Departamento de Estatística

Prédio 13 - Sala 1205 C

Universidade Federal de Santa Maria - UFSM

Endereço: Faixa de Camobi, Km 9 - Santa Maria - RS - 97105-900

Tel.: (0xx55) 220 8612

E-mail: amsouza@ccne.ufsm.br

#### **Maria Helena Rigão**

Mestranda do Programa de Pós-Graduação em Engenharia de Produção - PPGEP

Universidade Federal de Santa Maria - UFSM

Endereço: Rua Visconde de Pelotas, 1722/44 - Santa Maria - RS - 97015-140

Tel.: (0xx55) 217 8986

E-mail: hrigao@terra.com.br