

Estudo simulado envolvendo Cartas de Controle Multivariadas

Eduardo de Oliveira Correa¹

Danilo Marcondes Filho²

Resumo: Processos industriais geram dados acerca de inúmeras variáveis de interesse correlacionadas. Buscando um monitoramento mais robusto de tais processos, cartas de controle baseados em técnicas estatísticas multivariadas foram desenvolvidos. Destacam-se as cartas de controle *Qui-Quadrado* (χ^2) e da *Variância Generalizada* (W). Estas estatísticas permitem um monitoramento simultâneo do vetor de médias e da matriz de covariâncias das variáveis, respectivamente, a cada nova amostra do processo. Este trabalho apresenta um estudo por simulação para investigar o poder de detecção das cartas χ^2 e W . A partir de um processo simulado com 4 variáveis e uma estrutura de covariância, descontroles são impostos tanto no vetor de médias quanto na matriz de covariâncias do processo sob controle. Os resultados mostram que a sensibilidade da carta W aumenta para a detecção de modificações maiores na estrutura de covariância original das variáveis. Já em relação à carta χ^2 , podemos notar que alterações no vetor de médias nas direções comuns de variância das variáveis (isto é, na direção das suas covariâncias) são detectadas com menos sensibilidade em relação às alterações que não estão nas suas direções de covariância.

Palavras-chave: *Cartas de Controle Multivariadas, Carta de Controle Qui-Quadrado, Carta de Controle da Variância Generalizada.*

1 Introdução

Com o avanço tecnológico e uma disputa mercadológica extremamente competitiva, tem-se aumentado o interesse das indústrias no estudo dos métodos estatísticos para controle de processos. O *Controle Estatístico do Processo* (CEP) consiste em um grupo de ferramentas desenvolvidas para monitorar o desempenho de um processo, sendo as *cartas de controle* (CCs) possivelmente a ferramenta mais sofisticada; ver (Montgomery, 2007).

As CCs foram introduzidas por Shewhart em 1924, buscando entender as causas que provocam variabilidades no processo. Segundo este autor, a variabilidade pode ocorrer por *causas comuns* (variações

¹UFRGS - Universidade Federal do Rio Grande do Sul. Email: eduardo.correa@ufrgs.br

²UFRGS - Universidade Federal do Rio Grande do Sul. Email: marcondes.danilo@gmail.com

aleatórias inerentes ao processo), e por *causas especiais* (eventos destoantes no processo que prejudicam a qualidade do produto). Através das CCs busca-se monitorar a variabilidade existente nos processos, procurando detectar a possível presença de causas especiais. Eliminando as causas especiais, consegue-se obter a redução sistemática da variabilidade do processo, aprimorando a qualidade, produtividade, confiabilidade e o custo do produto. A CC é uma ferramenta gráfica onde medidas de amostras igualmente espaçadas no tempo são representadas cronologicamente e cujos limites de controle são obtidos a partir de amostras preliminares do processo sob controle estatístico (isto é, apenas causas comuns presentes). Destacam-se as tradicionais cartas de controle univariadas para monitoramento da tendência central dos dados (*Carta de Controle para a Média*) e para o monitoramento da variabilidade (*Carta de Controle para Amplitude*). O trabalho precursor de Shewhart está sumarizado em (Shewhart, 1931).

Processos mais complexos geram uma grande massa de dados acerca de inúmeras de variáveis correlacionadas, tornando inadequado o uso das cartas de controles univariadas tradicionais. Neste caso, versões multivariadas das cartas mencionadas foram desenvolvidas. As CCs *Qui-Quadrado* para minitoramento do vetor de médias e da *Variância Generalizada* para o monitoramento da matriz de covariâncias permitem o monitormanto simultâneo de um conjunto de variáveis e oferecem performance superior as suas versões univariadas.

Este trabalho apresenta um estudo simulado para avaliar o desempenho das cartas de controle multivariadas *Qui-Quadrado* e *Variância Generalizada*. Para tanto, considerando um processo com 4 variáveis sobre interesse, cenários representando descontroles impostos no vetor de médias e na matriz de covariâncias do processo serão investigados.

2 Cartas de Controle Multivariadas

As primeiras publicações na perspectiva multivariada foram feitas por Harold Hotelling [(Hotelling, 1947)], utilizando abordagem multivariada em dados contendo informações sobre bombardeios durante a Segunda Guerra Mundial. Esta seção descreve brevemente a base teórica das tradicionais cartas de controle *Qui-Quadrado* e da *Variância Generalizada*.

2.1 Carta de controle Qui-Quadrado

Considera-se p -características correlacionadas medidas simultaneamente compondo amostras p -variadas de tamanho n . Supõe-se que estas características seguem uma distribuição p -dimensional multivariada normal com vetor de médias $\boldsymbol{\mu} = (\mu_1, \mu_2, \dots, \mu_p)$ e matriz de covariância $\boldsymbol{\Sigma}$, sendo μ_i a média para a i -ésima característica e $\boldsymbol{\Sigma}$ uma matriz consistindo de variâncias e covariâncias das p -características, onde os elementos da diagonal principal são as variâncias de x 's e os elementos fora da diagonal principal representam as covariâncias. O monitoramento futuro de vetores p -variados de tamanho n é dado por:

$$\chi_p^2 = n(\bar{\mathbf{x}} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1}(\bar{\mathbf{x}} - \boldsymbol{\mu}) \sim \chi_p^2 \quad (2.1)$$

onde χ_p^2 segue uma distribuição *Qui-Quadrado* com p graus de liberdade e representa a distância quadrada padronizada, p -dimensional, entre um vetor de observações $\bar{\mathbf{x}}$ e o vetor de médias do processo $\boldsymbol{\mu}$. A raiz quadrada χ_p é conhecida como *distância de Mahalanobis*. Na prática, é necessário estimar $\boldsymbol{\mu}$ e $\boldsymbol{\Sigma}$ a partir de amostras preliminares, tomadas quando se assume que o processo está sob controle. $\bar{\mathbf{x}}$ e \mathbf{S} representam, respectivamente, as estimativas para o vetor das médias e a matriz de covariância do processo. Entretanto, neste trabalho não necessitamos estimar tais parâmetros, visto que será utilizado um estudo de caso simulado, onde conhecemos inteiramente não só a distribuição geradora dos dados, como o vetor de médias e a matriz de covariâncias populacionais. Mais detalhes sobre a construção da carta de controle utilizando as estimativas da média e da covariância podem ser encontrados em (Johnson & Wichern, 2007).

2.2 Carta de controle para variabilidade do processo

Considere novamente um conjunto de observações p -variadas geradas de uma distribuição normal p -variada com vetor de médias $\boldsymbol{\mu}$ e matriz de covariâncias $\boldsymbol{\Sigma}$. A carta de controle da *Variância Generalizada* é empregada para detectar mudanças na estrutura de covariâncias dos dados. Esta carta se constitui numa extensão multivariada da carta de controle univariada S^2 . (Montgomery, 2007) descreve a estatística para monitorar a estrutura de covariâncias de futuras amostras baseada no determinante de matrizes de covariância, como se segue:

$$W_i = -pn + p \ln(n) - n \ln(|\mathbf{A}_i|/|\boldsymbol{\Sigma}|) + \text{tr}(\boldsymbol{\Sigma}^{-1} \mathbf{A}_i) \sim \chi_{p(p+1)/2}^2 \quad (2.2)$$

onde W_i segue uma distribuição *Qui-Quadrado* com $p(p+1)/2$ graus de liberdade, $\boldsymbol{\Sigma}$ é a matriz de covariância populacional, $\mathbf{A}_i = (n-1)\mathbf{S}_i$, \mathbf{S}_i é a matriz de covariância da i -ésima amostra de tamanho n e tr é o operador de traço.

3 Metodologia

Consideramos um processo simulado sob controle com quatro variáveis. Suponha que os dados seguem uma distribuição normal 4-variada com vetor de médias $\boldsymbol{\mu}$ e a matriz das covariâncias $\boldsymbol{\Sigma}$, dados por:

$$\boldsymbol{\mu} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad \boldsymbol{\Sigma} = \begin{bmatrix} 1 & 0.9 & 0.07 & 0.05 \\ 0.9 & 1 & 0.04 & 0.03 \\ 0.07 & 0.04 & 1 & 0.9 \\ 0.05 & 0.03 & 0.9 & 1 \end{bmatrix} \quad (3.1)$$

Consideraremos no monitoramento vetores de médias amostrais 4-variados de tamanho $n = 50$ observações. Para avaliação do desempenho da carta de controle χ^2 , em cada cenário simulado de descontrole são gerados 100 vetores de médias e replicados 500 vezes. Novas amostras são descontroladas no vetor de médias. Para cada amostra os escores referentes à estatística χ^2 são comparados ao limite de controle correspondente [equação (2.1)].

Considere descontroles realizados para três casos distintos: **(I)** monitoramento do processo simulando o descontrole no vetor de médias para duas variáveis correlacionadas, mantendo as outras duas variáveis com média fixa em 0. Escolhemos deslocar para esse caso a média da primeira (μ_1) e segunda (μ_2) variável, dado a alta correlação entre elas ($\rho_{12} = 0.9$); **(II)** monitoramento do processo simulando o descontrole para duas variáveis fracamente correlacionadas, com médias fixas em 0 para as outras duas variáveis. Escolhemos a média da primeira (μ_1) e quarta (μ_4) variável, com correlação $\rho_{14} = 0.05$; **(III)** monitoramento simulando o descontrole para duas variáveis correlacionadas conjuntamente com outra variável que apresenta fraca correlação com as duas, sendo mantida a outra variável com média 0. Escolhemos a primeira e segunda, dado a alta correlação entre elas, juntamente com a quarta variável, que apresenta fraca correlação com as duas primeiras ($\rho_{14} = 0.05$ e $\rho_{24} = 0.03$, respectivamente). Temos assim:

$$\begin{array}{ccc} \text{Caso (I)} & \text{Caso (II)} & \text{Caso (III)} \\ \boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ 0 \\ 0 \end{bmatrix} & \boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ 0 \\ 0 \\ \mu_4 \end{bmatrix} & \boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ 0 \\ \mu_4 \end{bmatrix} \end{array} \quad (3.2)$$

onde $\mu_1 = \mu_2 = \mu_3 = 0, 0.1, 0.2, 0.3, 0.4, \dots, 2.9, 3$. Em cada cenário variamos na razão de 0.1 na média para cada variável descontrolada.

Para o monitoramento da variabilidade consideramos a matriz de covariância do processo sob con-

trole dado na equação (3). Para avaliar o desempenho da carta de controle W , em cada cenário simulado de descontrolo, novamente foram gerados 100 vetores de médias de 50 amostras e computada a matriz de covariâncias associada a cada vetor. Os cenários foram replicados 500 vezes. Essa estrutura de covariância é descontrolada para novas amostras. Para cada amostra os escores referentes à estatística W_i são comparados ao limite de controle correspondente [equação (2.2)]. Serão escolhidos dois casos a serem simulados: **(I)** monitoramento do processo simulando o descontrolo na estrutura de covariância para duas variáveis altamente correlacionadas. Escolhemos para esse caso a primeira e segunda variável, cuja correlação é de $\rho_{12} = 0.9$; **(II)** monitoramento do processo simulando o descontrolo na primeira e na quarta variável, cuja a correlação é fraca ($\rho_{14} = 0.05$). Temos assim:

$$\Sigma = \begin{matrix} & \text{Caso (I)} & & \text{Caso (II)} \\ \Sigma = & \begin{bmatrix} 1 & \rho_{12} & 0.07 & 0.05 \\ \rho_{21} & 1 & 0.04 & 0.03 \\ 0.07 & 0.04 & 1 & 0.9 \\ 0.05 & 0.03 & 0.9 & 1 \end{bmatrix} & \Sigma = & \begin{bmatrix} 1 & 0.9 & 0.07 & \rho_{14} \\ 0.9 & 1 & 0.04 & 0.03 \\ 0.07 & 0.04 & 1 & 0.9 \\ \rho_{41} & 0.03 & 0.9 & 1 \end{bmatrix} \end{matrix} \quad (3.3)$$

onde $\rho_{12} = \rho_{21} = 0.6, 0.61, 0.63, 0.63, 0.64, \dots, 0.98, 0.99$ e $\rho_{14} = \rho_{41} = -0.14, -0.13, -0.12, -0.11, -0.10, \dots, 0.24, 0.25$. Em cada cenário variamos na razão de 0.01 a correlação nas variáveis descontroladas.

4 Estudo de casos simulados

Nessa seção apresentamos os resultados das simulações dos cenários descritos da seção anterior. Em cada tabela apresentamos os resultados de ambas as cartas em função da média e do desvio-padrão entre replicações. O limite de controle das cartas χ^2 e W foram obtidos considerando probabilidade de alarme falso de $\alpha = 0.05$.

A tabela 1 apresenta os resultados das simulações para os três casos descritos na seção anterior para avaliação da carta χ^2 . Observamos no caso **(I)** que a detecção de descontroles cresce conforme o tamanho do descontrolo imposto. Entretanto, a carta apresenta baixa sensibilidade de detecção, pelo fato do descontrolo representar uma alteração na direção comum de variabilidade destas variáveis. No caso **(II)** observamos acentuada sensibilidade na detecção de descontroles comparado ao caso **(I)**. Isto se deve a baixa correlação nas variáveis descontroladas, dado que esses descontroles estão em direções opostas as direções comuns de variabilidade. Dessa forma, mesmo pequenas alterações são detectadas com

frequência relativa alta. Observamos no caso (III) a boa sensibilidade da carta na detecção de descontroles . Notamos que a sensibilidade na detecção de pequenos descontroles é superior ao caso mostrado no caso (I) e inferior descrito no caso (II). Isto se justifica plenamente pelo fato de que neste cenário simulamos descontroles simultâneos em duas variáveis fortemente correlacionadas entre si e uma terceira fracamente correlacionada com as demais. Dessa forma, este descontrole está numa direção de variabilidade próxima (não oposta) a direção comum entre as duas variáveis correlacionadas.

Tabela 1: Quantidade média (e desvio padrão) de amostras perturbadas identificadas pela carta χ^2 para cada caso.

Descontroles	caso (I)		caso (II)		caso (III)	
	média	desvio padrão	média	desvio padrão	média	desvio padrão
0.0	0.054	0.022	0.052	0.024	0.048	0.022
0.1	0.049	0.027	0.057	0.021	0.054	0.024
0.2	0.055	0.022	0.071	0.023	0.063	0.023
0.3	0.052	0.020	0.106	0.033	0.082	0.027
0.4	0.057	0.024	0.155	0.035	0.110	0.032
0.5	0.059	0.022	0.231	0.046	0.143	0.035
0.6	0.071	0.024	0.322	0.050	0.195	0.040
0.7	0.078	0.027	0.431	0.049	0.250	0.042
0.8	0.087	0.025	0.542	0.048	0.327	0.046
0.9	0.096	0.030	0.661	0.042	0.411	0.048
1.0	0.110	0.032	0.760	0.045	0.502	0.049
1.1	0.122	0.036	0.852	0.036	0.592	0.051
1.2	0.139	0.033	0.909	0.028	0.680	0.047
1.3	0.154	0.037	0.952	0.021	0.755	0.044
1.4	0.176	0.036	0.977	0.016	0.825	0.039
1.5	0.201	0.039	0.988	0.011	0.880	0.033
1.6	0.222	0.039	0.995	0.007	0.920	0.028
1.7	0.252	0.047	0.998	0.004	0.949	0.023
1.8	0.275	0.043	0.999	0.003	0.970	0.017
1.9	0.302	0.044	1	0	0.984	0.013
2.0	0.336	0.047	1	0	0.991	0.010
2.1	0.376	0.046	1	0	0.995	0.007
2.2	0.405	0.047	1	0	0.998	0.004
2.3	0.444	0.048	1	0	0.999	0.003
2.4	0.474	0.052	1	0	1	0
2.5	0.512	0.052	1	0	1	0
2.6	0.553	0.048	1	0	1	0
2.7	0.587	0.056	1	0	1	0
2.8	0.615	0.048	1	0	1	0
2.9	0.662	0.051	1	0	1	0
3.0	0.688	0.042	1	0	1	0

Tabela 2: quantidade média (e desvio padrão) de amostras perturbadas identificadas pela carta W para cada caso.

Descontroles	caso (I)		Descontroles	caso (II)	
	média	desvio padrão		média	desvio padrão
0.60	1.000	0.002	-0.14	1	0
0.61	0.999	0.003	-0.13	1	0
0.62	0.999	0.003	-0.12	1	0
0.63	0.999	0.003	-0.11	1	0
0.64	0.998	0.004	-0.10	1	0
0.65	0.998	0.004	-0.09	1	0
0.66	0.997	0.005	-0.08	0.999	0.003
0.67	0.996	0.007	-0.07	0.990	0.010
0.68	0.994	0.008	-0.06	0.955	0.022
0.69	0.991	0.010	-0.05	0.873	0.034
0.70	0.989	0.011	-0.04	0.739	0.043
0.71	0.982	0.013	-0.03	0.593	0.050
0.72	0.974	0.016	-0.02	0.456	0.053
0.73	0.965	0.018	-0.01	0.323	0.049
0.74	0.952	0.021	0.00	0.227	0.037
0.75	0.933	0.025	0.01	0.161	0.036
0.76	0.909	0.030	0.02	0.108	0.031
0.77	0.875	0.032	0.03	0.089	0.027
0.78	0.830	0.039	0.04	0.070	0.028
0.79	0.777	0.041	0.05	0.064	0.025
0.80	0.710	0.047	0.06	0.074	0.026
0.81	0.632	0.046	0.07	0.091	0.028
0.82	0.546	0.048	0.08	0.112	0.030
0.83	0.451	0.047	0.09	0.164	0.037
0.84	0.357	0.049	0.10	0.214	0.040
0.85	0.267	0.044	0.11	0.309	0.047
0.86	0.189	0.039	0.12	0.403	0.048
0.87	0.131	0.034	0.13	0.541	0.048
0.88	0.095	0.029	0.14	0.689	0.048
0.89	0.070	0.024	0.15	0.811	0.037
0.90	0.067	0.025	0.16	0.914	0.026
0.91	0.082	0.026	0.17	0.970	0.015
0.92	0.114	0.031	0.18	0.994	0.007
0.93	0.193	0.039	0.19	1	0
0.94	0.354	0.048	0.20	1	0
0.95	0.613	0.050	0.21	1	0
0.96	0.884	0.030	0.22	1	0
0.97	0.994	0.007	0.23	1	0
0.98	1	0	0.24	1	0
0.99	1	0	0.25	1	0

A tabela 2 apresenta os resultados das simulações para os dois casos descritos na seção anterior para avaliação da carta W . Observamos no caso (I) uma alta sensibilidade na detecção dos descontroles, visto que pequenas alterações em relação a correlação de referência já são detectadas em 100% das amostras. O caso (II) apresenta resultados semelhantes aos do caso (I), isto é, alta sensibilidade nas detecções dos descontroles impostos. Dessa forma, verificamos o bom desempenho da carta W independente do grau da correlação das variáveis no processo sob controle estatístico.

5 Considerações Finais

Este trabalho apresentou um estudo do desempenho das cartas de controle χ^2 e W . As cartas χ^2 e W são abordagens multivariadas clássicas para o monitoramento de médias e covariâncias, respectivamente.

Através de um estudo simulado utilizando quatro variáveis apresentando uma estrutura de covariância, exibindo correlações fortes e fracas, diferentes cenários foram investigados incluindo diversos descontroles impostos no vetor de médias e na matriz de covariâncias.

Em relação a carta χ^2 verificamos que descontroles impostos nas direções comuns de variabilidade são detectados com menos sensibilidade quando comparados aos descontroles impostos fora das direções comuns. Já em relação a carta W verificamos a boa sensibilidade na detecção de descontroles independente do tamanho da correlação entre as variáveis no processo sob controle.

5.1 Bibliografia

Referências

- Hotelling, H. (1947). Multivariate quality control. *Techniques of statistical analysis*.
- Johnson, R., & Wichern, D. (2007). Applied multivariate statistical analysis. *INC., New Jersey*.
- Montgomery, D. C. (2007). *Introduction to statistical quality control*. John Wiley & Sons.
- Shewhart, W. A. (1931). *Economic control of quality of manufactured product*. ASQ Quality Press.