

Um comparativo de desempenho entre gráficos de controle univariados e multivariados

Rodrigo Luiz Pereira Lara^{1 3}

Vladimir Barbosa Carlos de Souza^{2 3}

1. Introdução

O Controle Estatístico do Processo (CEP) é uma metodologia que desempenha papel primordial na indústria moderna por resolver problemas de instabilidade de processos que acarretam em danos e perda na produção. Um poderoso componente do CEP são os gráficos de controle, utilizados no monitoramento da média e variabilidade de diversas características avaliadas.

Como existem diferentes tipos de gráficos de controle com o mesmo objetivo, é importante conhecer o desempenho e rapidez com que cada um deles detecta alterações reais no processo (LARA, 2012). Para processos univariados, existem os gráficos de controle univariados – dentre eles Shewhart, EWMA e CUSUM. Por outro lado, na prática, um mesmo produto possui várias variáveis, as quais precisam ser monitoradas simultaneamente. Assim, os gráficos T^2 de Hotelling constituem-se em uma importante contribuição para verificar a estabilidade de um processo (SOUZA e RIGÃO, 2005).

Uma das dificuldades encontradas com gráficos de controle é visualizar em qual contexto o gráfico de controle multivariado é superior ao univariado. Trazido à tona esse importante assunto, novas discussões e aplicações práticas podem ser estabelecidas, de modo a proporcionar melhorias no âmbito acadêmico e industrial.

Tendo como tais motivações, o presente trabalho objetiva comparar o desempenho do gráfico de controle de Hotelling (multivariado) com o Gráfico de Shewhart (univariado) na detecção da mudança de média de uma variável dentro de um processo, verificar o poder dos gráficos na detecção de causas especiais e verificar o desempenho dos gráficos para processos não correlacionados e altamente correlacionados.

¹ Doutorando em Estatística Aplicada e Biometria pela Universidade Federal de Viçosa. DET / UFV. rodrigo.lara@ufv.br

² Mestrando em Estatística Aplicada e Biometria pela Universidade Federal de Viçosa. DET / UFV. vladimir.souza@ufv.br

³ Agradecimentos ao CNPq, CAPES e FAPEMIG pelo apoio financeiro.

2. Material e Métodos

Quando o monitoramento é feito sobre uma única variável, deseja-se testar se a média de uma amostra é igual à média da característica da qualidade quando o processo está sob controle estatístico. As hipóteses nesse caso são:

$$H_0: \mu = \mu_0$$

$$H_1: \mu \neq \mu_0,$$

e a estatística do teste é

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

Em alguns casos a produção é lenta e, sendo assim, possível analisar baseando-se em observações individuais ao invés de analisar médias de amostras, ou seja, o tamanho da amostra é $n=1$. Para estes casos a estatística de teste é:

$$z = \frac{(X - \mu)}{\sigma} \quad (2)$$

Já os gráficos de controle multivariados baseiam-se na distribuição T^2 de Hotelling que é a generalização da distribuição t de Student, para várias dimensões (SOUZA e RIGÃO, 2005). Generalizando a Equação 1 para o caso em que mais de uma variável é submetida a teste, tem-se a estatística T^2 mostrada na Equação 3.

$$T^2 = (\bar{X} - \mu_0)' \left(\frac{S}{n} \right)^{-1} (\bar{X} - \mu_0) = n(\bar{X} - \mu_0)' S^{-1} (\bar{X} - \mu_0) \quad (3)$$

Consequentemente, para observações individuais, tem-se:

$$T^2 = (\bar{X} - \mu_0)' S^{-1} (\bar{X} - \mu_0) \quad (4)$$

Por fim, as hipóteses no caso multivariado são:

$$H_0: \underline{\mu} = \underline{\mu}_0$$

$$H_1: \underline{\mu} \neq \underline{\mu}_0.$$

A metodologia utilizada no presente trabalho utiliza os seguintes vetores de média e matrizes covariâncias descritos na Tabela 1.

Simulou-se 50 dados de uma variável aleatória $X \sim N\left(\underline{\mu}_1, \Sigma_1\right)$, ou seja, X segue distribuição normal bivariada de média $\underline{\mu}_1$ e matriz de covariância Σ_1 . Note que X_1 e X_2 são variáveis altamente correlacionadas uma vez que $\rho_{X_1, X_2} = -0,9$. Note ainda

Tabela 1: Parâmetros utilizados na simulação de dados

Vetores de média		Matrizes de covariâncias	
$\mu_1 = \begin{bmatrix} 10 \\ 200 \end{bmatrix}$	$\mu_2 = \begin{bmatrix} 10 \\ 200 + L\sqrt{3600} \end{bmatrix}$	$\Sigma_1 = \begin{bmatrix} 9 & -162 \\ -162 & 3600 \end{bmatrix}$	$\Sigma_2 = \begin{bmatrix} 9 & 0 \\ 0 & 3600 \end{bmatrix}$

L: valor do deslocamento da média em número de desvios padrão.

que os parâmetros escolhidos para compor o vetor de médias e matriz de covariâncias foram tais que mantiveram o mesmo coeficiente de variação para cada variável.

$$CV_{X_1} = \frac{\sigma_1}{\mu_1} = \frac{3}{10} = \frac{60}{200} = \frac{\sigma_2}{\mu_2} = CV_{X_2}$$

Essas 50 realizações obtidas da variável aleatória X compuseram os 50 primeiros subgrupos racionais de um processo considerado sob controle estatístico.

Em seguida, simulou-se mais 50 dados de uma distribuição $N(\mu_2, \Sigma_1)$. Esses dados compuseram os subgrupos racionais de ordem 51 a 100 de modo a representar um processo que sofreu um aumento de L desvios-padrão na média da variável X_2 .

Foi feita também outra simulação, semelhante à descrita nos parágrafos anteriores, dessa vez para dados não correlacionados. Para representar os 50 primeiros subgrupos racionais de um processo sob controle estatístico, simulou-se dados de uma distribuição $N(\mu_1, \Sigma_2)$. E, para representar os subgrupos de ordem 51 a 100, fora de controle estatístico, simulou-se dados de uma distribuição $N(\mu_2, \Sigma_2)$.

O valor L indicado na Tabela 1 representa o deslocamento da média em número de desvios padrão. Para ambos os processos descritos acima, utilizou-se valores de L iguais a 0,5; 1; 1,5; 2; 2,5 e 3.

Para monitorar os processos, foram utilizados o gráfico de controle de Shewhart (univariado) e T^2 de Hotelling (multivariado), comparando o desempenho desses em detectar mudança na média de um processo.

Em seguida, cada um dos dois processo supracitados foram repetidos 5 mil vezes de modo a proporcionar melhor precisão dos resultados.

Os gráficos de controle foram planejados utilizando $\alpha = 0,05$, de modo que os limites superiores de controle dos gráficos de Shewhart e Hotelling foram iguais a $\mu_0 + 1,96\sigma$ e $\chi^2(0,05; 2)$ respectivamente.

As simulações foram realizadas com o uso do *software* livre R.

3. Resultados e Discussões

Na Figura 1 é visualizada a probabilidade do alarme falso obtida dos 50 primeiros subgrupos racionais para variáveis de correlação alta (a) e nula (b).

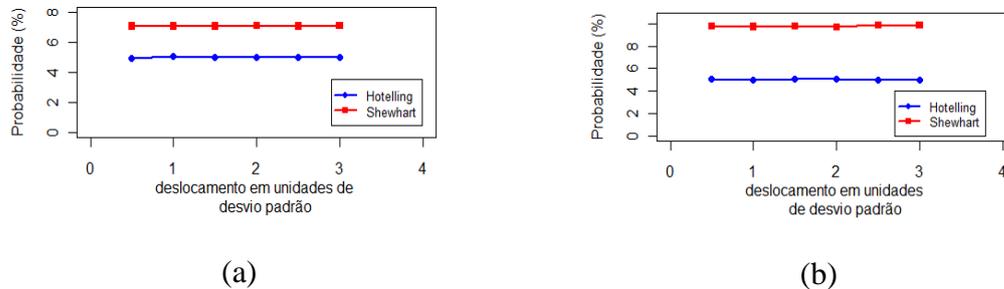


Figura 1: Probabilidade de alarme falso em processos altamente correlacionados (a) e sem correlação (b), utilizando $\alpha = 0,05$.

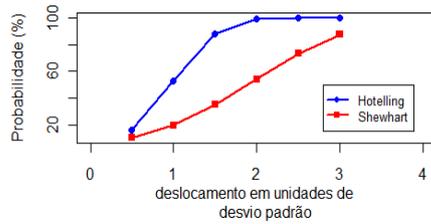
Tanto para processos de correlação alta ou nula, o gráfico de controle de Hotelling manteve a probabilidade de alarme falso igual a 0,05, conforme admitido durante o planejamento do gráfico.

Para processos de correlação nula, a probabilidade de alarme falso do gráfico de Shewhart subiu para o valor médio de 9,760%. Esse resultado condiz com Montgomery (2004) ao afirmar que se há p características da qualidade independentes para um produto particular e se o gráfico de Shewhart é mantido com $P(\text{erro tipo I}) = \alpha$ para cada uma, então a verdadeira probabilidade de um erro tipo I para o procedimento do controle conjunto é:

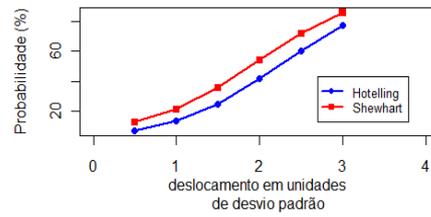
$$\alpha' = 1 - (1 - \alpha)^p \quad (5)$$

Já para processos de correlação alta, a probabilidade de alarme falso do gráfico de Shewhart subiu para o valor médio de 7,067%. Nesse caso o aumento foi menos acentuado em relação ao caso não correlacionado, pois, quanto maior a correlação entre duas variáveis, mais é possível que apenas um único gráfico de controle univariado monitore esse processo. Daí, para duas variáveis de correlação 1 (ou -1), espera-se que a probabilidade de alarme falso do Gráfico de Shewhart permaneça em 5%.

A Figura 2 mostra a probabilidade em detectar descontrole, ou seja, alarme verdadeiro ou poder. Ao analisá-la observa-se que o desempenho do Gráfico de Controle de Hotelling é visivelmente superior ao de Shewhart para processos altamente correlacionados (Figura 2a).



(a)



(b)

Figura 2: Probabilidade de alarme verdadeiro em processos altamente correlacionados (a) e sem correlação (b), utilizando $\alpha = 0,05$.

Para dados não correlacionados (Figura 2b), o gráfico de controle de Shewhart se mostrou ligeiramente superior ao de Hotelling. Tal fato ocorre, pois, quando há variação de somente um componente, o peso da variável fora de controle é apenas 50%, o que justifica o menor poder em relação ao gráfico univariado.

4. Conclusões

Os gráficos de controle multivariados apresentam maior desempenho que os univariados no monitoramento de um conjunto de médias, principalmente quando as variáveis em questão são correlacionadas.

O poder do gráfico de controle univariado é superior ao multivariado em processos envolvendo variáveis não correlacionadas.

5. Bibliografia

- [1] LARA, R. L. P. *Desempenho do Gráfico de Controle CUSUM Tabular para o Monitoramento da Média*. Viçosa, 2012. 85 p. Dissertação (Mestrado em Estatística Aplicada e Biometria) – Programa de Pós-Graduação em Estatística Aplicada e Biometria – PPESTBIO/UFV, Universidade Federal de Viçosa, 2012.
- [2] MONTGOMERY, D. C. *Introdução ao Controle Estatístico da Qualidade*. 4 ed. Tradução Ana Maria Lima de Farias e Vera Regina Lima de Farias e Flores. Rio de Janeiro: LTC, 2004.
- [3] SOUZA, A. M.; RIGÃO, M. H. Identificação de Variáveis Fora de Controle em Processos Produtivos Multivariados. *Revista Produção*, vol. 15, n. 1, p. 74-86, Jan./Abr. 2005.