

RÔMULO CÉSAR MANULI

**UMA PROPOSTA PARA A CONSTRUÇÃO DE GRÁFICOS DE CONTROLE
POR MEIO DE COMPONENTES PRINCIPAIS**

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Estatística Aplicada e Biometria, para obtenção do título de *Magister Scientiae*.

VIÇOSA
MINAS GERAIS – BRASIL
2015

**Ficha catalográfica preparada pela Biblioteca Central da Universidade
Federal de Viçosa - Câmpus Viçosa**

T

Manuli, Rômulo Cesár, 19-
M294 Uma proposta para a construção de gráficos de controle por
2015 meio de componentes principais / Rômulo Cesár Manuli. –
Viçosa, MG, 2015.
x, 69f. : il. (algumas color.) ; 29 cm.

Inclui anexos.

Orientador: José Ivo Ribeiro Júnior.

Dissertação (mestrado) - Universidade Federal de Viçosa.

Referências bibliográficas: f.63-65.

1. Estatística matemática. 2. Análise multivariada.
3. Estatísticas - Métodos gráficos. 4. Controle de processos .
I. Universidade Federal de Viçosa. Departamento de Estatística.
Programa de Pós-graduação em Estatística Aplicada e Biometria.
II. Título.

CDD 22. ed. 519.535'

RÔMULO CÉSAR MANULI

**UMA PROPOSTA PARA A CONSTRUÇÃO DE GRÁFICOS DE CONTROLE
POR MEIO DE COMPONENTES PRINCIPAIS**

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Estatística Aplicada e Biometria, para obtenção do título de *Magister Scientiae*.

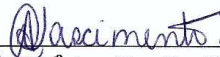
APROVADA: 24 de fevereiro de 2015.



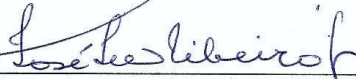
Prof. Carlos Henrique Osório Silva



Prof. Luciano José Minette



Prof.^a Ana Carolina C. Nascimento
(Coorientadora)



Prof. José Ivo Ribeiro Júnior
(Orientador)

*À minha mãe, Eliane Maria Bento.
À minha esposa, Débora Gomes Goularth.*

AGRADECIMENTOS

A Deus, pela oportunidade de estar vivenciando este momento e me conduzir perante todos os obstáculos.

À Universidade Federal de Viçosa e ao Departamento de Estatística, pela oportunidade de realização do curso.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pelo apoio financeiro.

Ao professor, amigo e orientador, José Ivo Ribeiro Júnior, pela dedicação, pela disponibilidade e pelo companheirismo, muito além de uma simples orientação.

Aos professores coorientadores, Ana Carolina Campana Nascimento e Moysés Nascimento, pelo apoio prestado.

Aos professores membros da banca, Carlos Henrique Osório Silva e Luciano José Minette, pelas sugestões e pelos incentivos.

Aos demais docentes e servidores do Departamento de Estatística, pela convivência e enorme contribuição para a minha formação.

À minha esposa, Débora Gomes Goularth, pelo amor, pela amizade, pela presteza, pelo constante incentivo e pela confiança em todos os momentos.

À minha mãe, Eliane Maria Bento, e aos meus irmãos, Adilson Luis Manuli e Leonardo Júnior Manuli, pela família, pelo exemplo de vida e pela amizade.

Aos meus grandes incentivadores da família Fermentech, Nanci H. Ohata Santana, Sebastião Zorzetti e Neveton B. Picciani.

Ao meu amigo, Eduardo Campana Barbosa, pela convivência e pelas experiências vividas diariamente ao longo deste mestrado.

Aos companheiros deste mestrado, pela convivência harmoniosa e valiosa troca de experiências ao longo desta jornada.

À Carla Zinato Campos, pela atenção e prontidão no auxílio das atividades burocráticas.

SUMÁRIO

LISTA DE TABELAS.....	v
LISTA DE FIGURAS.....	vi
RESUMO.....	vii
ABSTRACT.....	ix
1. INTRODUÇÃO.....	1
2. OBJETIVOS.....	3
2.1. Geral.....	3
2.2. Específicos.....	3
3. Revisão de Literatura.....	4
3.1. Controle estatístico de processos.....	4
3.2. Gráficos de controles multivariados.....	10
3.3. Componentes principais.....	13
3.4. Simulação estocástica.....	21
4. MATERIAL E MÉTODOS.....	23
4.1. Simulação dos dados.....	23
4.2. Gráficos de controle.....	26
4.2.1. Gráfico de Controle 1.....	26
4.2.2. Gráfico de Controle 2.....	29
4.2.3. Gráfico de Controle 3.....	32
4.2.4. Avaliação do desempenho.....	35
5. RESULTADOS E DISCUSSÃO.....	37
5.1. Processo sob controle.....	37
5.2. Processos Fora de Controle Estatístico.....	44
5.2.1. Deslocamento da Média.....	44
5.2.2. Aumento da variabilidade.....	51
5.2.3. Deslocamento da média e aumento da variabilidade.....	56
6. CONCLUSÕES.....	62
REFERÊNCIAS BIBLIOGRÁFICAS.....	63
ANEXO A – ROTINA DO R UTILIZADA PARA ANALISAR O CENÁRIO ONDE O PROCESSO ESTAVA SOB CONTROLE ESTATÍSTICO.....	66
ANEXO B – ROTINA DO R UTILIZADA PARA ANALISAR O CENÁRIO ONDE O PROCESSO ESTAVA FORA DE CONTROLE ESTATÍSTICO.....	68

LISTA DE TABELAS

Tabela 1 – Probabilidade do alarme falso (α) em função de k	9
Tabela 2 – Probabilidade do alarme verdadeiro (Pd) em função de k, n e δ	10
Tabela 3 – Probabilidades dos alarmes falso e verdadeiro em função de k, n e γ , nos gráficos de controle R e S	11
Tabela 4 – Autovalores obtidos pelas matrizes Σ , ρ e Σ^* e as respectivas porcentagens de explicação de cada componente principal de acordo com a correlação	37
Tabela 5 – Autovetores normalizados dos componentes principais CP ₁ e CP ₂ , obtidos pelas matrizes Σ , ρ e Σ^*	38
Tabela 6 – Médias e desvios-padrão dos componentes principais obtidos pelas matrizes Σ , ρ e Σ^*	38
Tabela 7 – Médias e desvios-padrão das variáveis plotadas nos gráficos.....	39
Tabela 8 – Linha média e limites de controle obtidos nos três gráficos de controle	39
Tabela 9 – Estimativa da probabilidade do alarme falso em função das correlações existentes entre Y ₁ e Y ₂ , em cada um dos três gráficos de controle propostos.....	44
Tabela 10 – Médias e desvios-padrão dos componentes principais obtidos pelas matrizes Σ , ρ e Σ^*	44
Tabela 11 – Médias e desvios-padrão das variáveis plotadas nos gráficos.....	45
Tabela 12 – Probabilidade do alarme verdadeiro observada em função das correlações existentes entre Y ₁ e Y ₂ em cada um dos gráficos propostos	49
Tabela 13 – Probabilidade do alarme verdadeiro observada em função das correlações existentes entre Y ₁ e Y ₂ plotando a variável V ₁ no gráfico de controle EWMA	50
Tabela 14 – Probabilidade do alarme verdadeiro observada em função das correlações existentes entre Y ₁ e Y ₂ plotando a variável V ₁ no gráfico de controle CUSUM.....	51
Tabela 15 – Médias e desvios-padrão dos componentes principais obtidos pelas matrizes Σ , ρ e Σ^*	51
Tabela 16 – Médias e desvios-padrão das variáveis plotadas nos gráficos.....	52
Tabela 17 – Probabilidade do alarme verdadeiro observada em função das correlações existentes entre Y ₁ e Y ₂ em cada um dos gráficos propostos	56
Tabela 18 – Médias e desvios-padrão dos componentes principais obtidos pelas matrizes Σ , ρ e Σ^*	57
Tabela 19 – Médias e desvios-padrão das variáveis plotadas nos gráficos.....	57
Tabela 20 – Probabilidade do alarme verdadeiro observada em função das correlações existentes entre Y ₁ e Y ₂ em cada um dos gráficos propostos	61

LISTA DE FIGURAS

Figura 1 – Ciclo DMAIC.....	5
Figura 2 – Gráfico de controle que monitora um estimador nos subgrupos racionais t_1, t_2, \dots, t_{14} de um processo sob controle estatístico.....	7
Figura 3 – Probabilidades P_d e β associadas a um processo fora de controle estatístico para a média ($RNRH_0$ = região de não rejeição de H_0 ; RRH_0 = região de rejeição de H_0).	8
Figura 4 – Gráfico de Controle 1 – Processo sob controle estatístico, com correlações de 0,1, 0,3, 0,5, 0,7 e 0,9, da esquerda para direita e de cima para baixo, respectivamente.....	40
Figura 5 – Gráfico de Controle 2 – Processo sob controle estatístico, com correlações de 0,1, 0,3, 0,5, 0,7 e 0,9, da esquerda para direita e de cima para baixo, respectivamente.....	41
Figura 6 – Gráfico de Controle 3 – Processo sob controle estatístico, com correlações de 0,1, 0,3, 0,5, 0,7 e 0,9, da esquerda para direita e de cima para baixo, respectivamente.....	43
Figura 7 – Gráfico de Controle 1 – Processo com deslocamento da média de controle, com correlações de 0,1, 0,3, 0,5, 0,7 e 0,9, da esquerda para direita e de cima para baixo, respectivamente.....	46
Figura 8 – Gráfico de Controle 2 – Processo com deslocamento da média de controle, com correlações de 0,1, 0,3, 0,5, 0,7 e 0,9, da esquerda para direita e de cima para baixo, respectivamente.....	47
Figura 9 – Gráfico de Controle 3 – Processo com deslocamento da média de controle, com correlações de 0,1, 0,3, 0,5, 0,7 e 0,9, da esquerda para direita e de cima para baixo, respectivamente.....	48
Figura 10 – Gráfico de Controle 1 – Processo com aumento da variabilidade, com correlações de 0,1, 0,3, 0,5, 0,7 e 0,9, da esquerda para direita e de cima para baixo, respectivamente.	53
Figura 11 – Gráfico de Controle 2 – Processo com aumento da variabilidade, com correlações de 0,1, 0,3, 0,5, 0,7 e 0,9, da esquerda para direita e de cima para baixo, respectivamente.	54
Figura 12 – Gráfico de Controle 3 – Processo com aumento da variabilidade, com correlações de 0,1, 0,3, 0,5, 0,7 e 0,9, da esquerda para direita e de cima para baixo, respectivamente.	55
Figura 13 – Gráfico de Controle 1 – Processo com deslocamento da média e aumento da variabilidade, com correlações de 0,1, 0,3, 0,5, 0,7 e 0,9, da esquerda para direita e de cima para baixo, respectivamente.	58
Figura 14 – Gráfico de Controle 2 – Processo com deslocamento da média e aumento da variabilidade, com correlações de 0,1, 0,3, 0,5, 0,7 e 0,9, da esquerda para direita e de cima para baixo, respectivamente.	59
Figura 15 – Gráfico de Controle 3 – Processo com deslocamento da média e aumento da variabilidade, com correlações de 0,1, 0,3, 0,5, 0,7 e 0,9, da esquerda para direita e de cima para baixo, respectivamente.	60

RESUMO

MANULI, Rômulo César, M.Sc., Universidade Federal de Viçosa, fevereiro de 2015. **Uma proposta para a construção de gráficos de controle por meio de componentes principais.** Orientador: José Ivo Ribeiro Júnior. Coorientadores: Ana Carolina Campana Nascimento e Moysés Nascimento.

Idealizados inicialmente por Shewhart, os gráficos de controle são ferramentas utilizadas para o monitoramento de processos. Quando o objetivo é controlar várias características de um mesmo processo, é recomendável a aplicação dos gráficos de controle multivariados, ou seja, gráficos que monitorem todas essas características simultaneamente. O objetivo deste trabalho foi elaborar uma proposta de um gráfico de controle multivariado construído por meio de componentes principais, sendo esta uma técnica de análise multivariada que reduz a dimensionalidade dos dados. Os componentes principais são indicados para o controle estatístico de processos multivariados que envolvam um conjunto de dados com grande número de variáveis correlacionadas. Para tanto, foram analisados 20 cenários compostos, cada um, por mil pares de valores simulados para as duas variáveis-resposta correlacionadas com as seguintes correlações: 0,1; 0,3; 0,5; 0,7 e 0,9. Em cinco dos vinte cenários avaliados, o processo encontrava-se sob controle estatístico. Nos demais quinze cenários, o processo se encontrava fora de controle estatístico, dos quais, cinco apresentavam deslocamento na média de controle, em outros cinco os processos apresentaram aumento no desvio-padrão de controle e, por fim, os cinco restantes cenários apresentaram alterações na média de controle e no desvio-padrão de controle, simultaneamente. Para o total dos vinte cenários, foram construídos gráficos de controle empregando a matriz de variâncias e covariâncias ($s\Sigma$), onde os elementos são as variâncias de controle (σ_{01}^2 e σ_{02}^2) e as covariâncias estipuladas para o estudo, a matriz de correlações (ρ), onde os elementos são obtidos a partir da matriz de variâncias e covariâncias (Σ) e, por último, a matriz de coeficientes de variação (Σ^*), realizando a transformação dos dados. Nesta transformação, os dados, após serem padronizados com média zero e variância um, têm suas variâncias ponderadas pelos respectivos coeficientes de variação (CVs). Em todas as três propostas, as probabilidades dos alarmes falsos obtidas foram constantes e iguais ao valor teórico esperado de 0,0027. As probabilidades dos alarmes verdadeiros foram

baixas em todos os casos. A matriz Σ mostrou-se superior às demais nos cenários onde ocorreu o deslocamento da média e no cenário onde ocorreu o deslocamento da média e o aumento da variabilidade, simultaneamente. Já no cenário onde ocorreu somente o aumento da variabilidade, as matrizes Σ e Σ^* mostraram-se exatamente iguais. As correlações existentes entre as variáveis-resposta não influenciaram nas probabilidades dos alarmes falsos e verdadeiros. Assim, quando o objetivo do controle for apontar pequenos desvios na média de controle ou pequenos aumentos na variabilidade, nenhuma das três propostas são indicadas, pois não são poderosas para apontar pequenas oscilações do processo. Uma alternativa para aumentar o poder no apontamento destas pequenas oscilações é plotar esta nova variável criada em outros gráficos de controle, como EWMA e CUSUM, por exemplo. Esta nova variável é uma combinação das variáveis-resposta e pode ser plotada em outros gráficos univariados.

ABSTRACT

MANULI, Rômulo César, M.Sc., Universidade Federal de Viçosa, February of 2015. **Using principal component for the construction of control charts.** Adviser: José Ivo Ribeiro Júnior. Co-adviser: Ana Carolina Mota Campana and Moysés Nascimento.

When tools used for the monitoring of processes were idealized initially by Shewhart, the printers of healthy control. When the objective is to control several characteristics of the same process, there is recommendable the application of the multivaried printers of control, in other words, graphic what monitor all these characteristics simultaneously. The objective of this work prepared a proposal of a printer of multivaried control built through principal components, when there is this a technique of multivaried analysis that reduces the dimensionalidade of the data. The principal components are indicated for the statistical control of multivaried processes that wrap a set of data with great number of correlated variables. For so much, 20 serious sceneries were analysed, each one, by thousand couples of values when two simulated for variable-answer correlated with the next correlations: 0,1; 0,3; 0,5; 0,7 and 0,9. In five of twenty evaluated sceneries, the process was under statistical control. In too many fifteen sceneries, the process was out of statistical control, of which, five were presenting dislocation on the average of control, in other five the processes presented increase in the diversion-standard of control and, finally, the remaining five sceneries presented alterations on the average of control and in the diversion-standard of control, simultaneously. For the total of twenty sceneries printers of control were built employing to womb of variances and covariances (Σ), where the elements are the variances of control (σ_{01}^2 e σ_{02}^2) and the covariances stipulated for the study, the womb of correlations (ρ), where the elements are obtained from the womb of variances and covariances (Σ) and, for last, the womb of coefficients of variation (Σ^*), carrying out the transformation of the data. In this transformation, after they were standardized by average I reduce the data to zero and variance one, they have his variance considered by the respective coefficients of variation (CVs). In all three proposals false alarms probabilities obtained were constant and equal to the expected theoretical value of 0,0027. The probabilities of true alarms were low in all cases. The matrix Σ was superior to the other in

scenarios where there was the average displacement and displacement occurred in the scenario where the average and increased variability, simultaneously. In the scenario where only occurred the increased variability, matrices Σ and Σ^* proved to be exactly alike. The existent correlations between the variable-answer did not influence the probabilities of the false and true alarms. So, when the objective of the control will be going to point to small diversions on the average of control or small increases in the variability, none of three proposals are indicated, since are not powerful to point to small oscillations of the process. An alternative to increase the power in the note of these small oscillations is to represent this new variable created in other printers of control, like EWMA and CUSUM, for example. This new variable is a combination of the variable-answer and it can be represented in other printers univariateds.

1. INTRODUÇÃO

Em decorrência da competitividade do mercado e das exigências dos clientes, os esforços para melhoria dos padrões de qualidade de processos industriais tornam-se cada vez mais necessários. Desta forma, instrumentos/metodologias que permitam um maior controle do processo de produção são requeridos. Dentre as metodologias existentes para esse fim, destacam-se as ferramentas do Controle Estatístico de Processos (CEP).

Os gráficos de controle são as ferramentas do CEP mais utilizadas para o monitoramento dos processos. Quando se deseja monitorar uma característica de um produto, utilizam-se os gráficos de controle univariados. Para os casos onde é preciso controlar várias características de um mesmo processo, é recomendável a aplicação dos gráficos de controle multivariados, ou seja, gráficos que monitoram todas essas características simultaneamente.

As técnicas de análise multivariada de dados possibilitam avaliar um conjunto de características, levando em consideração as correlações existentes, que permitem que as inferências sobre o conjunto de variáveis sejam feitas em um nível global de significância conhecido.

Dentre os métodos multivariados que permitem a redução da dimensionalidade dos dados, os componentes principais destacam-se como uma ferramenta para um bom aproveitamento da informação disponível. Os componentes principais foram introduzidos por Pearson em 1901 e desenvolvidos de forma independente por Hotteling em 1933 e caracterizam-se como variáveis latentes que representam combinações lineares de um grupo de variáveis relacionadas sob estudo.

Segundo Jackson (1991), a análise de componentes principais é utilizada na indústria moderna para o controle estatístico de processos multivariados que envolvam um conjunto de dados com grande número de variáveis correlacionadas. Com a aplicação desse método estatístico de projeção de dados, 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.

O controle estatístico multivariado de processos está fundamentado nos métodos de projeção de dados via componentes principais e via projeção de estruturas latentes. O objetivo destes métodos é transformar variáveis de processo em

um grupo reduzido de variáveis latentes, normalmente, duas ou três, que são suficientes para representar as variáveis originais (FILHO, 2001). Filho (2001) trabalhou com gráficos de controle de Shewhart, gráficos de controle Qui-Quadrado, gráficos de Hotelling e $|S|$ controlando os escores padronizados. Neste trabalho, foi criada uma nova variável para ser plotada nos gráficos de controle, empregando os escores e os autovalores.

Assim, objetivou-se propor uma metodologia para a construção de gráficos de controle univariados, por meio de componentes principais, para o monitoramento de alterações das médias e também das variabilidades, simultaneamente, e que seja melhor do que os gráficos de controle tradicionais existentes para indicar alarmes falsos e alarmes verdadeiros.

2. OBJETIVOS

2.1. Geral

Elaborar uma proposta para a construção de gráficos de controle, por meio da utilização dos escores dos componentes principais e dos autovalores obtidos a partir das matrizes contendo os elementos de controle do processo, para realizar o monitoramento da média e da variabilidade das variáveis-resposta importantes para a avaliação da qualidade do processo.

2.2. Específicos

- Verificar os efeitos das correlações entre as variáveis-resposta sobre o desempenho da metodologia proposta.
- Verificar os efeitos das alterações das médias e das variabilidades das variáveis-resposta sobre o desempenho dos gráficos de controle propostos.
- Verificar os efeitos das matrizes empregadas sobre o desempenho dos gráficos de controle propostos.

3. REVISÃO DE LITERATURA

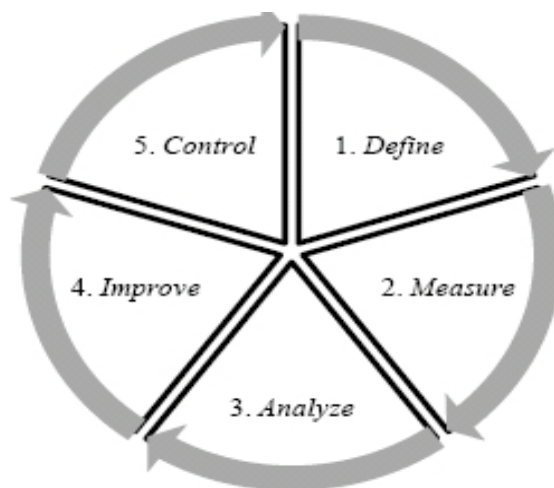
3.1. Controle estatístico de processos

Se um produto deve corresponder às exigências do cliente, ele deve, em geral, ser produzido por um processo que seja estável ou replicável. Mais precisamente, o processo deve ser capaz de operar com pequena variabilidade em torno das dimensões-alvo ou nominais das características de qualidade do produto. O controle estatístico de processo (CEP) é uma poderosa coleção de ferramentas de resolução de problemas útil na obtenção da estabilidade do processo e na melhoria da capacidade através da redução da variabilidade (MONTGOMERY, 2004).

A metodologia Seis Sigma é a aplicação de ferramentas, inclusive do CEP, para o projeto e operação dos sistemas de gerenciamento e processos de negócios, os quais dão condições aos funcionários de entregarem mais valor aos clientes e empregadores. Trata-se de uma implantação rigorosa, focada e altamente eficiente das técnicas e princípios e qualidade comprovadas. As ferramentas são aplicadas com um modelo simples para melhoria de desempenho (PYZDEK; KELLER, 2011). Para atingir um alto nível de qualidade, próximo ao “zero defeito”, faz-se uso do ciclo de melhoria DMAIC (*Define, Measure, Analyze, Improve, Control*). Esta metodologia de solução de problemas é utilizada pela maioria das companhias americanas que empregam a metodologia Seis Sigma com o objetivo de realizar as melhorias nos produtos, serviços e processos e para projetá-los e, ou, reprojeta-los (AGUIAR, 2002). As cinco etapas do DMAIC (Figura 1) podem ser descritas como: definir (*define*), medir (*measure*), analisar (*analyze*), melhorar (*improve*) e controlar (*control*). A etapa de controle, última fase do ciclo, visa o permanente monitoramento do processo. Desse modo, busca-se aprimorar os parâmetros que definem quando um processo está ou não sob controle. Jalali et al. (2008) ressaltam ainda que os fatores mais importantes dessa fase são a documentação, monitoramento visando melhoria, preservação dos resultados obtidos na fase *improve*, além de um bom preparo da equipe técnica.

De acordo com Ribeiro Júnior (2013), um processo de produção pode ser representado por um sistema formado por variáveis de entrada e de saída. As de entrada são constituídas pelos fatores controláveis ou não, e as de saída, pelas

características de qualidade, denominadas de variáveis-resposta (dependentes), medidas nos itens de um mesmo produto ou serviço. No CEP, os valores observados de uma dada variável-resposta, são resumidos a partir de medidas de posição e de dispersão e gráficos, de forma a serem mais facilmente interpretáveis. Dentre as medidas, as mais utilizadas são a média e o desvio-padrão.



Fonte: Carvalho (2010).

Figura 1 – Ciclo DMAIC.

Qualidade é sinônimo de atender à especificação, sendo esta constituída por um valor-alvo e limitada pelos limites inferior e superior ou por um deles. Todos os itens de um mesmo produto ou serviço, produzidos por um mesmo processo, terão que apresentar valores da variável-resposta localizados dentro desta especificação. Quanto menor o deslocamento da média e a variabilidade dos valores observados em relação ao valor-alvo, melhor será a qualidade do processo.

Na melhoria da qualidade, quando se busca centralizar a média da variável-resposta sobre o valor-alvo, o objetivo também consistirá em colocar toda a distribuição dos seus valores entre os limites inferior (LIE) e superior (LSE) de especificação. No controle da qualidade, o objetivo consiste em não deixar nenhum dos valores sair, como também em diminuir, ainda mais, a variabilidade entre eles e em torno do valor-alvo (RIBEIRO JÚNIOR, 2013). Os gráficos de controle são os principais métodos estatísticos do CEP utilizados para o monitoramento da média e da variabilidade de uma ou mais características avaliadas nos produtos ou serviços que respondem sobre a qualidade do processo.

Segundo Montgomery (2004), em qualquer processo de produção sempre existirá uma quantidade de variabilidade inerente ou natural. No CEP estas causas inerentes são denominadas como causas aleatórias ou também como causas comuns. Diz-se que um processo que opera apenas com as causas aleatórias está sob controle estatístico. Ocasionalmente, outros tipos de variabilidade podem estar presentes. Tal variabilidade é, geralmente, muito grande quando comparada com a variabilidade natural ou inerente. Essas fontes de variabilidade são conhecidas como causas atribuíveis ou também como causas especiais. Um processo que opera na presença de causas especiais está fora de controle.

Ribeiro Júnior (2013) explica que não é possível, tecnicamente, produzir todos os produtos ou serviços com o mesmo valor, tornando-se importante estimar o limite de variação natural imposto pelas causas aleatórias. No CEP, a média e o desvio-padrão de controle são parâmetros de referência definidos por critérios técnicos. Assim, quanto mais próximas forem as estimativas destes parâmetros, melhor será o controle do processo.

Para que o gráfico de controle traga interpretações apropriadas sobre o processo, a linha média (LM) deverá representar a média de controle e os limites inferior de controle (LIC) e superior (LSC), o intervalo da variabilidade aleatória permitida às suas estimativas obtidas desse processo. As estimativas mais apropriadas do LIC, LM e LSC são para aqueles processos isentos de causas especiais. Os valores observados em um processo fora de controle estatístico não são apropriados para a construção dos gráficos de controle, pois poderão acarretar estimativa da LM distante do valor-alvo e, ou, estimativas dos limites de controle associados a um maior intervalo, em razão de a estimativa do desvio-padrão ser superior ao valor de referência (COSTA *et al.*, 2004). Portanto, torna-se necessário implantar os procedimentos operacionais de padronização do processo e eliminar as influências das causas especiais, objetivando a redução da variabilidade, antes do controle pelos respectivos gráficos (RIBEIRO JÚNIOR, 2013).

Os gráficos de controle são gráficos temporais, onde os estimadores estudados de uma característica de qualidade são plotados no eixo vertical, sendo o eixo horizontal composto pela evolução temporal, dividida em subgrupos racionais. Considera-se o processo como fora de controle estatístico quando algum ponto do gráfico estiver acima do limite superior de controle (LSC) ou abaixo do limite

inferior de controle (LIC), ou ainda quando os pontos apresentarem algum comportamento sistemático ou não aleatório (Figura 2).

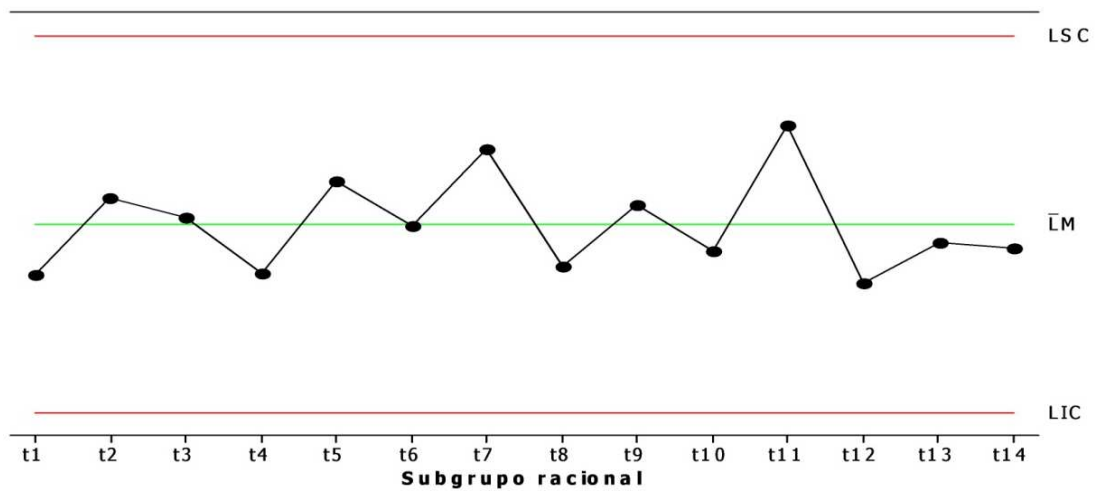


Figura 2 – Gráfico de controle que monitora um estimador nos subgrupos racionais t_1, t_2, \dots, t_{14} de um processo sob controle estatístico.

A largura dos limites de controle é uma das decisões críticas no planejamento de um gráfico de controle. Pelo deslocamento dos limites de controle para mais longe da LM, diminui-se a probabilidade do alarme falso (α), ou seja, a probabilidade de um ponto cair além dos limites de controle, indicando a condição de falta de controle quando nenhuma causa atribuída estiver presente. No entanto, o alargamento dos limites de controle aumentará a probabilidade da falta de detecção (β), isto é, a probabilidade de um ponto cair entre os limites de controle, quando o processo estiver realmente fora de controle. Se os limites de controle forem movidos para mais perto da LM, o efeito oposto será obtido: o α será aumentado, enquanto o β será diminuído (MONTGOMERY e RUNGER, 2009).

Segundo Pyzdek e Keller (2011), um processo funcionando com seis sigma produzirá 3,4 partes por milhão (ppm) de não conformidades. Isto porque ocorre um deslocamento, em longo prazo, da média da variável-resposta em relação ao valor-alvo de, no máximo, 1,5 desvios-padrão. Assim, a área de uma distribuição normal que esteja 4,5 desvios-padrão acima da média é, de fato, 3,4 ppm. Contrastando com a qualidade seis sigma, o padrão tradicional de qualidade três sigma significa uma produção de 66.811 ppm de não conformidades, pois a variabilidade aleatória em torno da linha média é de 1,5 desvios-padrão, quando a média está deslocada do alvo.

A consequência de ordem prática associada ao alarme falso é intervir no processo na hora errada, quando ele está isento de causas especiais, o que em si já acarreta um custo de interrupção do processo, de mão de obra, além de um risco de desajustar o processo que estava ajustado. E, a consequência de ordem prática associada à falta de detecção é não intervir no processo na hora certa, quando ele está sob a influência de causas especiais (COSTA *et al.*, 2004).

Ademais, o poder (Pd) é a probabilidade do gráfico de controle acusar, corretamente, que o processo está fora de controle, isto é, a probabilidade do alarme verdadeiro.

Como se pode observar, as probabilidades β e Pd estão associadas à distribuição alternativa de $H_0: \mu = \mu_0$, ou seja, à $H_1: \mu \neq \mu_0$. No caso mais específico, à $H_1: \mu > \mu_0$ (Figura 3).

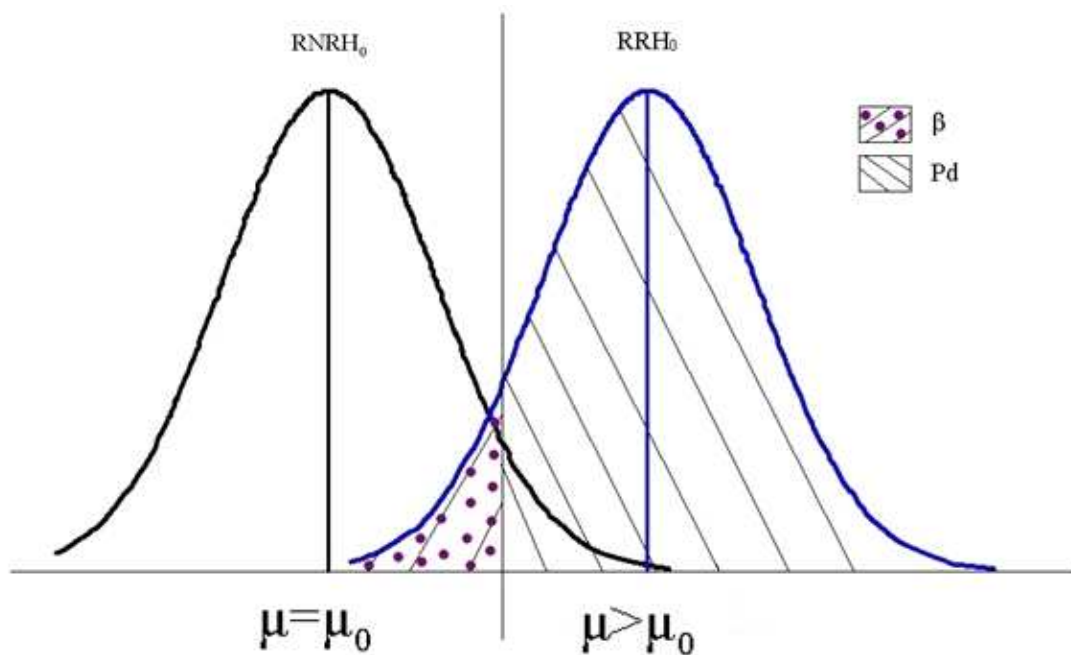


Figura 3 – Probabilidades Pd e β associadas a um processo fora de controle estatístico para a média (RNRH₀ = região de não rejeição de H₀; RRH₀ = região de rejeição de H₀).

De acordo com Ribeiro Júnior (2013), o desempenho de um determinado tipo de gráfico de controle pode ser medido pelas probabilidades dos alarmes falso e verdadeiro. Um gráfico de controle bem planejado é aquele que apresenta baixa

probabilidade do alarme falso e alta do verdadeiro, considerando preestabelecidos os valores do deslocamento da média de controle que se deseja detectar como variação especial em número de desvios-padrão (δ) e também o fator de aumento do desvio-padrão de controle (γ), que é a razão entre os desvios-padrão considerados como fora e de controle.

Nos gráficos de controle de Shewhart para o monitoramento da média, a probabilidade do alarme falso não depende do tamanho do subgrupo racional (n) e diminui somente em função do aumento do número de desvios-padrão da variável-resposta utilizado para incluir a quantidade de variabilidade aleatória permitida (k) (COSTA *et al.*, 2004; SOUZA *et al.*, 2008) (Tabela 1).

Tabela 1 – Probabilidade do alarme falso (α) em função de k

K	α	K	α	k	α
0,1	0,9203	1,1	0,2713	2,1	0,0357
0,2	0,8415	1,2	0,2301	2,2	0,0278
0,3	0,7642	1,3	0,1936	2,3	0,0214
0,4	0,6892	1,4	0,1615	2,4	0,0164
0,5	0,6171	1,5	0,1336	2,5	0,0124
0,6	0,5485	1,6	0,1096	2,6	0,0093
0,7	0,4839	1,7	0,0891	2,7	0,0069
0,8	0,4237	1,8	0,0719	2,8	0,0051
0,9	0,3681	1,9	0,0574	2,9	0,0037
1,0	0,3173	2,0	0,0455	3,0	0,0027

Já a probabilidade do alarme verdadeiro aumenta em função dos aumentos de δ , n e α (diminuição de k). Entretanto, o efeito positivo do n sobre o Pd é menor à medida que diminui o valor de δ e aumenta o valor de k (diminui o α). Isto significa que, para aumentar o Pd dos gráficos de controle de Shewhart para detectar pequenas diferenças na média de controle, em número de desvios-padrão, é necessário aumentar consideravelmente o número de n . Dependendo da redução do valor de δ , o aumento do Pd em função do aumento de n poderá se tornar economicamente inviável. O aumento do Pd em função da redução de k poderá ser considerado somente em casos de menores custos, maiores facilidades e rapidez de intervenção. Ribeiro Júnior (2013) recomenda a redução do valor de k de 3, este que é o valor mais usual, até no máximo 2,5, que significa aumentar a probabilidade α de 0,0027 para aproximadamente, 0,01 (Tabela 2).

Tabela 2 – Probabilidade do alarme verdadeiro (Pd) em função de k, n e δ

K	n	δ			
		0,5	1	2	3
3	1	0,0064	0,0228	0,1587	0,5000
	2	0,0110	0,0564	0,4319	0,8930
	3	0,0165	0,1024	0,6787	0,9860
	4	0,0228	0,1587	0,8413	0,9987
	5	0,0299	0,2225	0,9295	0,9999
	6	0,0379	0,2910	0,9712	1,0000
	7	0,0468	0,3616	0,9890	1,0000
	8	0,0564	0,4319	0,9961	1,0000
	9	0,0668	0,5000	0,9987	1,0000
	10	0,0780	0,5645	0,9996	1,0000

Para o monitoramento do aumento da variabilidade de processos sob controle estatístico para a média, ou seja, quando todos os subgrupos racionais apresentarem a mesma média de controle μ_0 , a probabilidade do alarme falso (α) nos gráficos de controle de Shewhart da amplitude (R) e do desvio-padrão (S) diminui em função do aumento de k. No gráfico de controle R, ela sofre ainda pequenos efeitos de n. A probabilidade do alarme verdadeiro (Pd) aumenta em função dos aumentos de n e γ e da diminuição de k (COSTA et al., 2004). As probabilidades do alarme falso são baixas nos gráficos de controle R e S para $k \geq 2,5$. Mas as probabilidades dos alarmes verdadeiros também são baixas, principalmente para $n \leq 5$ e $1 < \gamma \leq 2$, significando que eles são eficientes somente para detectar causas especiais associadas ao aumento da variabilidade nas razões de $\gamma > 2$ e, mesmo assim, utilizando altos valores de n (Tabela 3).

3.2. Gráficos de controles multivariados

Há muitas situações nas quais é necessário o monitoramento simultâneo ou controle de duas ou mais características da qualidade relacionadas. De acordo com Montgomery (2004), em determinadas situações, monitorar duas características separadamente pode ser muito enganoso. Em um processo sob controle, a probabilidade de duas variáveis aleatórias com distribuições normais independentes, excederem os limites de controle três sigma é de 0,0027. No entanto, a probabilidade conjunta de que pelo menos uma de duas variáveis-resposta exceda os seus

Tabela 3 – Probabilidades dos alarmes falso e verdadeiro em função de k, n e γ , nos gráficos de controle R e S

K	n	α		Pd			
				$\gamma = 2$		$\gamma = 3$	
		R	S	R	S	R	S
3	2	0,0092	0,0027	0,1924	0,2523	0,3849	0,4453
	3	0,0058	0,0027	0,2721	0,3679	0,5598	0,6412
	4	0,0049	0,0027	0,3443	0,4598	0,6848	0,7651
	5	0,0046	0,0027	0,41	0,5377	0,7746	0,8464
	6	0,0045	0,0027	0,4687	0,6051	0,8385	0,9001
	7	0,0044	0,0027	0,5211	0,6634	0,884	0,9354
	8	0,0043	0,0027	0,5673	0,7139	0,9163	0,9584
	9	0,0044	0,0027	0,6088	0,7576	0,9397	0,9734
	10	0,0044	0,0027	0,6459	0,7951	0,9562	0,9831

respectivos limites de controle, quando estão sob controle, para $k = 3$, é igual a: $(0,0027) (0,0027) + (0,0027) (0,9973) + (0,9973) (0,0127) = 0,0054$, o que é consideravelmente maior que 0,0027. Já a probabilidade de que ambas as variáveis sejam plotadas simultaneamente dentro dos limites de controle quando estão realmente sob controle é $(0,9973) (0,9973) = 0,99460729$. Esta distorção no procedimento de monitoramento do processo aumenta à medida que aumenta o número de características de qualidade estudadas. Em geral, se há p características de qualidade estatisticamente independentes para um produto particular, realizando o monitoramento independentemente, a verdadeira probabilidade de um erro tipo I para o procedimento de controle conjunto é:

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

Além disso, se as p características de qualidade não forem independentes, o que usualmente ocorre quando elas se relacionam no mesmo produto, não há uma maneira fácil de medir a distorção no procedimento de controle conjunto.

Hotelling (1947), pioneiro na pesquisa sobre controle multivariado de qualidade, utilizou uma abordagem multivariada de controle em dados contendo informações sobre localização de bombardeios, durante a Segunda Guerra Mundial. Entre os gráficos de controle multivariados existentes, o gráfico T^2 de Hotelling é o mais conhecido na literatura, sendo um análogo direto do gráfico de controle da média de Shewhart univariado. O gráfico desenvolvido por Hotelling (1947) supõe

que as variáveis de interesse tenham uma distribuição normal multivariada, com vetor de médias μ e matriz de covariâncias Σ . A equação para a obtenção da estatística T^2 é dada por:

$$T^2 = (Y - \mu)^T \Sigma^{-1} (Y - \mu),$$

em que Y = vetor das variáveis originais; μ é o vetor de médias; e Σ é a matriz de variâncias e covariâncias de Y .

Esta equação é utilizada como base para a construção do gráfico T^2 de Hotelling (LOWRY e MONTGOMERY, 1995). O maior cuidado deve ser na seleção dos limites de controle para a estatística T^2 de Hotelling. Há duas fases distintas no uso desse gráfico de controle. A fase 1 é o uso dos gráficos para o estabelecimento do controle, isto é, testar se o processo estava, ou não, sob controle quando os m subgrupos racionais preliminares foram extraídos e as estatísticas amostrais \bar{Y} e S calculadas. O objetivo desta primeira fase é a obtenção de um conjunto de observações sob controle, de modo que os limites de controle possam ser estabelecidos para a fase 2, que é o monitoramento das observações futuras. Este procedimento é chamado, às vezes, de análise retrospectiva.

Além do gráfico de controle T^2 de Hotelling, outros tipos de gráficos de controle para processos multivariados são indispensáveis para atender à especificidade de um determinado processo, tais como os gráficos de controle MCUSUM (Multivariado de Soma Acumulada) e o MEWMA (Multivariado de Média Móvel Exponencialmente Ponderada). Estes dois gráficos de controle possuem a característica de detectar pequenas mudanças no processo, abaixo de 2σ (dois desvios-padrão), ao contrário do gráfico T^2 de Hotelling, indicado para detectar grandes alterações (ALVES, 2009; HENNING et al., 2011).

Um procedimento potencialmente útil, desenvolvido por Hawkins em 1991, é denominado ajustamento de regressão. Consiste em plotar gráficos de controle univariados para os resíduos de cada variável, obtidos quando aquela variável é regredida sobre todas as outras. Uma aplicação muito importante do ajustamento de regressão ocorre quando o processo tem uma hierarquia distinta de variáveis, tal como um conjunto de variáveis de entrada do processo e um conjunto de variáveis de saída. Essa situação pode ser chamada de processo em cascata. Uma boa propriedade

deste modelo é o fato dos resíduos serem não-correlacionados, mesmo quando a variável de interesse original tenha exibido correlação. Por causa dessa propriedade, este procedimento tem muitas aplicações possíveis nas indústrias químicas e de processamento, onde há, frequentemente, processos em cascata com várias entradas e apenas poucas saídas, e onde muitas das variáveis são altamente correlacionadas (MONTGOMERY, 2004).

3.3. Componentes principais

De acordo com Ferreira (2011), a análise de componentes principais é uma técnica multivariada de modelagem da estrutura de covariância. Foi introduzida por Pearson em 1901 e desenvolvida de forma independente por Hotelling (1933). Apesar de Hotelling ter denominado seu método de análise de fatores, que é uma outra técnica multivariada, tratava-se da derivação teórica dos componentes principais. O termo fator utilizado decorreu da aplicação ter sido proveniente da área social. A técnica consiste em transformar um conjunto de variáveis originais em outro conjunto de variáveis de mesma dimensão denominadas de componentes principais.

Os componentes principais apresentam propriedades importantes: cada componente principal é uma combinação linear de todas as variáveis originais, são independentes entre si e estimados com o propósito de reter, em ordem de estimação, o máximo de informação, em termos da variação total contida nos dados. A análise de componentes principais é associada à ideia de redução de massa de dados, com menor perda possível da informação. Procura-se redistribuir a variação observada nos eixos originais de forma a se obter um conjunto de eixos ortogonais não correlacionados. Esta técnica pode ser utilizada para geração de índices e agrupamento de indivíduos. A análise agrupa os indivíduos de acordo com sua variação, isto é, os indivíduos são agrupados segundo suas variâncias, ou seja, segundo seu comportamento dentro da população, representado pela variação do conjunto de características que define o indivíduo, ou seja, a técnica agrupa os indivíduos de uma população segundo a variação de suas características.

Apesar das técnicas de análise multivariada terem sido desenvolvidas para resolver problemas específicos, principalmente de Biologia e Psicologia, podem ser também utilizadas para resolver outros tipos de problemas em diversas áreas do

conhecimento. A análise de componentes principais é a técnica multivariada mais conhecida. O objetivo dos componentes principais (CPs) é substituir a informação contida nas p variáveis originais pela informação contida em k ($k \leq 3$) CPs não correlacionados e com a menor perda de informação possível (BARBOSA et al., 2005). É importante salientar que a análise de componentes principais é considerada uma técnica de análise intermediária utilizada para auxiliar em outras análises.

A partir do vetor das p variáveis originais $Y=[Y_1, Y_2, \dots, Y_p]^T$ com vetor de médias $\mu=[\mu_1, \mu_2, \dots, \mu_p]^T$ e matriz de variâncias e covariâncias Σ , $p \times p$, o i -ésimo componente principal é definido por:

$$CP_i = a_{i1}^* Y_1 + a_{i2}^* Y_2 + \dots + a_{ip}^* Y_p,$$

em que a_i^* = autovetor normalizado do CP_i .

Os p componentes CP_1, CP_2, \dots, CP_p também podem ser baseados nas variáveis padronizadas Z 's, como segue, para o i -ésimo componente principal:

$$CP_i = a_{i1}^* Z_1 + a_{i2}^* Z_2 + \dots + a_{ip}^* Z_p,$$

em que $Z_i = \frac{Y_i - \mu_i}{\sigma_i}$.

Ferreira (2011) mostra que a definição dos componentes principais é, na verdade, a obtenção dos autovalores e autovetores da matriz Σ . Para as p variáveis-resposta Y s, serão obtidos p autovalores (λ_i) por meio do determinante da expressão, baseada na matriz amostral ($p \times p$) das variâncias e covariâncias (S), dada por:

$$S - \lambda_i I = 0,$$

em que

$$S = \begin{bmatrix} s_1^2 & s_{12} & \dots & s_{1p} \\ s_{12} & s_2^2 & \dots & s_{2p} \\ \dots & \dots & \dots & \dots \\ s_{1p} & s_{2p} & \dots & s_p^2 \end{bmatrix};$$

I = matriz identidade de ordem p ;

s_i^2 = estimativa da variância da variável-resposta Y_i ;

s_{ij} = estimativa da covariância entre as variáveis-resposta Y_i e Y_j .

Os p autovalores também podem ser obtidos por meio do determinante da expressão baseada na matriz amostral ($p \times p$) das correlações (R) entre as variáveis-resposta Y_s , que é a mesma matriz ($p \times p$) das variâncias e covariâncias entre as variáveis padronizadas Z_s :

$$R - \lambda_i I = 0,$$

em que

$$R = \begin{bmatrix} 1 & r_{12} & \dots & r_{1p} \\ r_{12} & 1 & \dots & r_{2p} \\ \dots & \dots & \dots & \dots \\ r_{1p} & r_{2p} & \dots & 1 \end{bmatrix}; e$$

r_{ij} = estimativa da correlação entre as variáveis-resposta Y_i e Y_j .

Apesar das estimativas dos autovalores serem diferentes quando baseadas em S e R , por estarem captando informações diferentes para ambos os casos, têm-se, de forma paramétrica (JONHSON e WICHERN, 2002; FERREIRA, 2011):

$$V(CP_i) = \lambda_i,$$

em que

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p; V(CP_1) \geq V(CP_2) \geq \dots \geq V(CP_p); e$$

$$\text{Cov}(CP_1, CP_2) = \text{Cov}(CP_1, CP_p) = \dots = \text{Cov}(CP_{p-1}, CP_p) = 0.$$

Portanto, o CP_1 tem maior poder de explicação da variabilidade total do que o CP_2 , que tem maior poder de explicação do que o CP_3 , e assim sucessivamente. O CP_1 contém mais informações sobre os dados do que o CP_2 , que não contém informações do CP_1 , e assim sucessivamente.

Cada autovalor λ_i ($i = 1, 2, \dots, p$) possui um autovetor correspondente, normalizado ou não. Assim, para o autovalor λ_i , obtém-se o autovetor não normalizado a_i (FERREIRA, 2011).

De acordo com Ferreira (2011), utilizando a decomposição espectral da matriz Σ , dada por $\Sigma = P\Lambda P^T$, em que P é a matriz composta pelos autovetores de

Σ em suas colunas e Λ é a matriz diagonal de autovalores de Σ , pode-se demonstrar que:

$$\text{tr}(\Sigma) = \text{tr}(P\Lambda P^T) = \text{tr}(\Lambda P^T P) = \text{tr}(\Lambda I) \text{ e}$$

$$\text{tr}(\Sigma) = \text{tr}(\Lambda) = \sum_{i=1}^p \lambda_i .$$

Mas, $\text{tr}(\Sigma)$ é obviamente dada pela soma dos elementos da diagonal, ou seja:

$$\text{tr}(\Sigma) = \sum_{i=1}^p \sigma_i^2 .$$

Portanto, pode-se concluir que:

$$\sum_{i=1}^p \sigma_i^2 = \sum_{i=1}^p \lambda_i .$$

Isto significa que a variabilidade total contida nas variáveis originais é igual à variabilidade total contida nos componentes principais.

A explicação do modelo reduzido em relação ao modelo completo pode ser obtida pela relação das variâncias dos dois modelos, ou seja, o quanto da variação total das variáveis originais é explicada pelo modelo de k componentes principais. Essa proporção acumulada de explicação da variação total, expressa em porcentagem, é formalmente dada por (FERREIRA, 2011):

$$p^k = \frac{\sum_{i=1}^k \lambda_i}{\sum_{i=1}^p \lambda_i} \times 100 .$$

A explicação individual de cada componente pode também ser considerada. Pode-se obter a proporção de explicação do k -ésimo componente principal por (FERREIRA, 2011):

$$p_k = \frac{\lambda_k}{\sum_{i=1}^p \lambda_i} \times 100 .$$

Com base na proporção de explicação da variabilidade total que o modelo de k componentes principais é responsável, pode-se determinar o número de componentes que deve ser retido. Em geral, para interpretar os dados com sucesso,

basta escolher os primeiros k componentes que envolvam pelo menos 70% da variância total (FERREIRA, 2011). Outro critério para a escolha do número de componentes, conhecido como critério da raiz latente, implica que apenas autovalores maiores que 1 (um), quando estimados a partir da matriz ρ , sejam considerados significantes. Além destes critérios para a definição do número de componentes, pode-se empregar também o gráfico “screeplot”, em que se relacionam 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). Se o número mínimo de componentes para explicar a proporção mínima da variação total for muito grande, em muitas aplicações práticas não haverá vantagem em utilizar o método dos componentes principais. Neste presente trabalho foram utilizados os dois componentes principais, considerando que foram simuladas apenas duas variáveis-resposta. Não faz sentido trabalhar com todos os componentes na prática.

Uma outra propriedade importante é a ortogonalidade dos componentes principais. Pode-se mostrar que são não-correlacionados mutuamente. Considerando o vetor dos componentes principais $CP = [CP_1, CP_2, \dots, CP_p]^T$ tem-se que $Cov(CP) = \Lambda$, que é uma matriz diagonal, como segue:

$$Cov(CP) = Cov(P^T Y) = P^T Cov(Y) P;$$

$$Cov(CP) = P^T \Sigma P = P^T P \Lambda P^T P; e$$

$$Cov(CP) = \Lambda .$$

Uma outra informação interessante é a importância ou a influência que cada variável-resposta Y exerce sobre os CPs, ou seja, o grau de associação existente entre os componentes e as variáveis originais. A medida naturalmente utilizada para este fim é o coeficiente de correlação definido por:

$$\rho_{CP_i, Y_i} = \frac{a_{ii} \sqrt{\lambda_i}}{\sigma_i} .$$

A matriz de correlação entre o vetor dos CPs e o vetor das variáveis originais é dada por:

$$\rho_{CP, Y} = \Lambda^{-1/2} \Lambda P^T V^{-1/2}$$

$$\rho_{CP,Y} = \Lambda^{1/2} P^T V^{-1/2},$$

em que

$$V = \text{diag}(\sigma_{ii}).$$

Assim, a *i*-ésima linha e a *k*-ésima coluna dessa matriz de correlações contêm o coeficiente de correlação entre o CP e a variável original. Conforme já citado anteriormente, a importância relativa dos CPs decresce do primeiro para o último. Desta forma, conclui-se que os últimos componentes serão responsáveis pela explicação de uma pequena parte da variação total dos dados. Assim, a variável que apresentar maior correlação, em valor absoluto, com o componente de menor autovalor, terá menor importância em explicar a variabilidade em relação aos elementos amostrais (MARDIA *et al.*, 1997; JONHSON; WICHERN, 2002).

Os valores de cada componente, os quais são obtidos por meio da combinação linear das variáveis originais, são denotados escores. As estimativas dos escores obtidas em relação aos *p* componentes CP₁, CP₂, ..., CP_p baseadas na matriz Σ são dadas por:

$$CP_1 = a_{11}^* Y_1 + a_{12}^* Y_2 + \dots + a_{1p}^* Y_p;$$

$$CP_2 = a_{21}^* Y_1 + a_{22}^* Y_2 + \dots + a_{2p}^* Y_p;$$

...

$$CP_p = a_{p1}^* Y_1 + a_{p2}^* Y_2 + \dots + a_{pp}^* Y_p.$$

As estimativas dos escores obtidas em relação aos *p* componentes CP₁, CP₂, ..., CP_p baseadas na matriz ρ são dadas por:

$$CP_1 = a_{11}^* Z_1 + a_{12}^* Z_2 + \dots + a_{1p}^* Z_p;$$

$$CP_2 = a_{21}^* Z_1 + a_{22}^* Z_2 + \dots + a_{2p}^* Z_p;$$

...

$$CP_p = a_{p1}^* Z_1 + a_{p2}^* Z_2 + \dots + a_{pp}^* Z_p.$$

Segundo Ferreira (2011), a extração de componentes principais da matriz de variâncias e covariâncias (matriz Σ) representa a essência e o objetivo natural do método dos componentes principais. No entanto, em algumas situações o efeito da escala das variáveis pode fazer com que haja predominância na representação dos

componentes principais por parte de algumas variáveis com grandes variâncias. Neste caso, as demais variáveis contribuirão muito pouco. Nessas situações, os CPs podem ser definidos a partir da matriz das correlações (matriz ρ) ao invés da matriz de variâncias e covariâncias, com o objetivo de contornar o efeito das escalas. A decomposição da matriz Σ das variáveis padronizadas é equivalente a decompor a matriz das correlações (ρ) das variáveis originais. As variáveis originais padronizadas têm média zero e variância um, conforme:

$$Z_i = \frac{Y_i - \mu_i}{\sigma_i},$$

em que

Z_i = variável original Y_i padronizada com média zero e variância um;

μ_i = média da variável original Y_i ; e

σ_i = desvio-padrão da variável original Y_i .

Campana *et al.* (2010) apresentaram uma alternativa para controlar o problema causado pela grande diferença entre as escalas, característico da matriz Σ , e ao mesmo tempo, levar em consideração a variação relativa das variáveis, o que não acontece com a utilização da matriz ρ . Trata-se de uma transformação, onde as variáveis, após serem padronizadas com média zero e variância um, têm suas variâncias ponderadas pelos respectivos coeficientes de variação (CVs). Após esta transformação, obtém-se uma matriz Σ^* , de variâncias e covariâncias das variáveis transformadas.

Como mencionado anteriormente, os dados originais são padronizados com média zero e variância um e, em seguida, transformados considerando o coeficiente de variação (CV) dos mesmos, como segue:

$$Z_i^* = Z_i \times CV_i,$$

em que

Z_i^* = variável original Y_i transformada, com média zero e variância ponderada pelo respectivo CV_i ;

Z_i = variável original Y_i padronizada, com média zero e variância um; e

CV_i = coeficiente de variação da variável original $Y_i = \frac{\sigma_i}{\mu_i}$.

Segundo Jackson (1991), a análise de componentes principais é utilizada na indústria moderna para o controle estatístico de processos multivariados que envolvam um conjunto de dados com grande número de variáveis correlacionadas. Com a aplicação desse método estatístico de projeção de dados, é possível considerar todas as variáveis originais no tratamento estatístico e visualizar os dados em duas ou três dimensões. Com isso, é possível descrever com 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).

De acordo com Montgomery (2004), os procedimentos de gráfico de controle multivariado convencionais são razoavelmente eficazes desde que p (número de variáveis do processo a serem monitoradas) não seja muito grande. Afirmam que funcionam bem quando o número de variáveis do processo não é muito grande, isto é, menor ou igual a dez. No entanto, à medida em que o número de variáveis cresce, os gráficos de controle multivariados tradicionais perdem eficiência em detectar mudanças. Uma abordagem muito utilizada nessas situações é diminuir a dimensão do problema, como a aplicação da análise de componentes principais.

Henning *et al.* (2011) compararam os gráficos de controle multivariados T^2 de Hotelling, McUsum (Soma Acumulada) e Mewma (Média Móvel Exponencialmente Ponderada), utilizando a aplicação da análise de componentes principais em um conjunto de dados reais com oito variáveis originais. No estudo, as oito variáveis iniciais foram reduzidas a dois CPs. Ao final, eles concluíram que a utilização dos CPs é adequada para o monitoramento simultâneo de um grande número de variáveis, sem perda de informação. De forma semelhante, Henning *et al.* (2010) estudaram os gráficos de controle multivariados T^2 de Hotelling e MCUSUM, utilizando também a aplicação da análise de componentes principais em um outro conjunto de dados reais com oito variáveis originais. A aplicação do gráfico T^2 de Hotelling aos componentes principais conduziu à mesma conclusão de quando aplicado às oito variáveis originais. O gráfico MCUSUM dos CPs levou ao mesmo resultado, concluindo que o componente principal é uma alternativa para o controle estatístico de processos multivariados, sendo possível reduzir o número de variáveis analisadas, com pouca perda de informação.

Machado (2009) monitorou o vetor de médias através do gráfico de controle T^2 baseado em componentes principais. Para o monitoramento da matriz de covariâncias o autor propôs uma nova estatística baseada nas variâncias amostrais denominada de VMAX. O gráfico de VMAX mostrou-se mais eficiente do que o gráfico da variância amostral generalizada $|S|$, que é o gráfico usual para o monitoramento da matriz de covariâncias. Uma vantagem adicional dessa nova estatística é que o usuário já está bem familiarizado com o cálculo de variâncias amostrais; o mesmo não pode ser dito em relação à variância amostral generalizada $|S|$.

Além desses, Henning *et al.* (2011) utilizaram os gráficos T^2 de Hotelling, McUsum e Mewma aplicados aos componentes principais para analisar uma série de dados reais de um processo de montagem de cabines de caminhões. Os dados foram compostos de oito variáveis que representavam medidas relativas ao assento do capuz do motor de um caminhão. No gráfico T^2 de Hotelling o processo encontrou-se sob controle. No entanto, os gráficos McUsum e Mewma revelaram uma alteração no vetor de médias do processo, sendo sensíveis para indicar que o processo saiu de controle. Realizou também um estudo comparativo aplicando os gráficos T^2 de Hotelling e McUsum aos componentes principais e também às oito variáveis originais. Os gráficos aplicados aos componentes e aos dados originais levaram aos mesmos resultados. Sendo assim, concluiu que os componentes principais é uma alternativa para o controle estatístico de processos multivariados, sendo possível reduzir o número de variáveis analisadas, sem perda de informação. Os componentes principais podem ser utilizados quando tem-se um grande número de variáveis correlacionadas. Quando as variáveis forem independentes não existe vantagem em utilizar esta técnica multivariada, pois teria de utilizar todos os componentes, não havendo a possibilidade de redução da dimensionalidade.

3.4. Simulação estocástica

Segundo Bustos e Orgarnbide (1992), pode-se definir simulação estocástica ou probabilística como a arte de gerar amostras de variáveis aleatórias em um ambiente computacional e de usar tais amostras para a obtenção de um determinado resultado. É um ensaio que consiste de uma série de cálculos numéricos com

decisões de escolha limitada, realizada segundo um conjunto de regras predeterminadas e apropriadas ao emprego de computadores digitais. Deve ser realizada em um ambiente computacional, pois supõe-se que a solução do problema seria praticamente impossível de ser feita manualmente.

Na engenharia, o termo simulação designa situações nas quais se tenta compreender as características de um sistema pelo conhecimento de outro que lhe é similar. É a técnica de solução de um problema pela análise de um modelo que descreve o comportamento de um sistema usando um computador digital (PRADO, 1999). É uma amostragem estatística controlada, para estimar o desempenho de sistemas estocásticos complexos quando os modelos analíticos são inadequados. Consiste, também, em imitar, por meio de recursos computacionais, o comportamento de um sistema real, para estudar o seu funcionamento em condições alternativas.

A simulação estocástica apresenta também várias outras aplicações, tais como: avaliação de violações nas pressuposições de certos procedimentos, checagem da melhoria de certas análises e auxílio na determinação do tamanho de parcelas em experimentos de campo (PRADO, 1999).

Magalhães *et al.* (2007) aplicaram a simulação de dados para comparar o desempenho dos gráficos CUSUM e EWMA em relação aos gráficos de Shewhart e Shewhart modificado. Foram utilizadas simulações de dados nos quais houve uma modificação na média para a realização de uma análise crítica de quando cada um deles deverá ser utilizado.

4. MATERIAL E MÉTODOS

4.1. Simulação dos dados

Todos os procedimentos estatísticos, simulação de dados e gráficos de controle foram implementados com o auxílio do software R, utilizando os pacotes MASS e qcc.

Foram simulados valores aleatórios da distribuição normal para as duas variáveis-resposta Y_1 e Y_2 de acordo com os parâmetros de média e variância apresentados por Campana *et al.* (2010) para Y_1 e Y_2 , respectivamente:

$$Y_1 \sim \text{Normal}(\mu_1=300; \sigma_1^2=900); \text{ e}$$

$$Y_2 \sim \text{Normal}(\mu_2=100; \sigma_2^2=100).$$

No total, foram realizadas cinco simulações para Y_1 e Y_2 , simultaneamente, de acordo com os seguintes valores para os coeficientes de correlação (ρ_{12}): 0,1; 0,3; 0,5; 0,7 e 0,9.

Em cada simulação, foram gerados aleatoriamente, 1.000 pares de valores para Y_1 e Y_2 .

Apesar das variáveis-resposta Y_1 e Y_2 apresentarem diferentes médias e variâncias paramétricas, elas apresentaram os mesmos coeficientes de variação, dados, respectivamente, por:

$$CV_1 = \frac{\sigma_1}{\mu_1} = \frac{30}{300} = 0,10; \text{ e}$$

$$CV_2 = \frac{\sigma_2}{\mu_2} = \frac{10}{100} = 0,10.$$

Foram obtidos cinco arquivos de dados referentes aos cinco coeficientes de correlação entre as variáveis-resposta Y_1 e Y_2 , fornecidos por um processo sob controle estatístico. Portanto, definiram-se as médias $\mu_1=300$ e $\mu_2=100$ como as respectivas médias de controle μ_{01} e μ_{02} . Do mesmo modo, os desvios-padrão $\sigma_1=30$ e $\sigma_2=10$ foram definidos como os desvios-padrão de controle σ_{01} e σ_{02} , para Y_1 e Y_2 , respectivamente.

Por outro lado, os 1.000 pares dos valores das variáveis-resposta Y_1 e Y_2 , fornecidos por um processo fora de controle estatístico, com correlações de 0,1; 0,3; 0,5; 0,7 e 0,9, foram simulados de acordo com três diferentes situações: deslocamento da média de controle, aumento do desvio-padrão de controle e ambas as alterações.

O deslocamento da média de controle foi da ordem de um desvio-padrão de controle ($\delta=1$). Portanto, para esta primeira situação, foram realizadas as seguintes simulações:

$$Y_1 \sim \text{Normal}(\mu_{11}=330; \sigma_{01}^2=900); \text{ e}$$

$$Y_2 \sim \text{Normal}(\mu_{12}=110; \sigma_{02}^2=100) .$$

Para o aumento da variabilidade, foi proposto um desvio-padrão equivalente ao dobro do desvio-padrão de controle ($\gamma=2$). Desse modo, as simulações foram obtidas a partir de:

$$Y_1 \sim \text{Normal}(\mu_{01}=300; \sigma_{11}^2=3.600); \text{ e}$$

$$Y_2 \sim \text{Normal}(\mu_{02}=100; \sigma_{12}^2=400) .$$

E para a terceira situação fornecida por um processo fora de controle estatístico, foram realizadas as seguintes simulações:

$$Y_1 \sim \text{Normal}(\mu_{11}=330; \sigma_{11}^2=3.600); \text{ e}$$

$$Y_2 \sim \text{Normal}(\mu_{12}=110; \sigma_{12}^2=400) .$$

Neste presente trabalho foram investigados somente estes quatro cenários. Na prática, outros cenários podem acontecer como, por exemplo, somente uma variável-resposta ter a média deslocada ou a variância aumentada, dentre outros.

Os parâmetros das médias e variâncias descritos anteriormente são definidos por:

$$\mu_{01} = \text{média de controle para } Y_1;$$

$$\mu_{11} = \text{média fora de controle para } Y_1;$$

$$\mu_{02} = \text{média de controle para } Y_2;$$

$$\mu_{12} = \text{média fora de controle para } Y_2;$$

σ_{01}^2 = variância de controle para Y_1 ;

σ_{11}^2 = variância fora de controle para Y_1 ;

σ_{02}^2 = variância de controle para Y_2 ; e

σ_{12}^2 = variância fora de controle para Y_2 .

Além das variáveis-resposta Y_1 e Y_2 , foram utilizadas no estudo as variáveis transformadas Z_1^* e Z_2^* , conforme apresentado por Campana et al. (2010):

$$Z_1^* = Z_1 \times CV_1 = \frac{Y_1 - \mu_1}{\mu_1};$$
$$Z_2^* = Z_2 \times CV_2 = \frac{Y_2 - \mu_2}{\mu_2};$$

em que μ_1 = media de Y_1 e μ_2 = média de Y_2 .

Desse modo, as matrizes paramétricas de variâncias e covariâncias das variáveis originais Y_1 e Y_2 (Σ), de correlações das variáveis originais Y_1 e Y_2 (ρ) e de variâncias e covariâncias das variáveis transformadas Z_1^* e Z_2^* (Σ^*), do processo sob controle estatístico, foram definidas por:

$$\Sigma = \begin{bmatrix} \sigma_{01}^2 & \sigma_{12} \\ \sigma_{21} & \sigma_{02}^2 \end{bmatrix};$$
$$\rho = \begin{bmatrix} 1 & \rho_{12} \\ \rho_{21} & 1 \end{bmatrix};$$
$$\Sigma^* = \begin{bmatrix} CV_1^2 & \rho_{12} CV_1 CV_2 \\ \rho_{21} CV_1 CV_2 & CV_2^2 \end{bmatrix},$$

em que $\rho_{12} = \rho_{21}$ e $\sigma_{12} = \sigma_{21} = \rho_{12} \sigma_{01} \sigma_{02}$.

Formaram-se, portanto, quatro arquivos com 1.000 pares de dados (Y_1 , Y_2), referentes a um processo sob controle e a três processos fora de controle estatístico. Cada arquivo foi constituído de cinco sub-arquivos com 1.000 pares de dados (Y_1 , Y_2), referentes às correlações de 0,1; 0,3; 0,5; 0,7 e 0,9. No total, foram gerados 20 conjuntos de 1.000 pares de dados.

Posteriormente, em todos os 20 conjuntos, foram aplicadas três metodologias propostas para a construção dos gráficos de controle, totalizando as construções de 60 gráficos de controle.

4.2. Gráficos de controle

Foram apresentadas três metodologias distintas e baseadas na análise de componentes principais (CPs) para as construções dos gráficos de gráficos de controle. Na primeira metodologia, os dois CPs foram obtidos a partir da matriz de variâncias e covariâncias (Σ), considerando as variâncias de controle das variáveis-resposta Y_1 e Y_2 e os cinco respectivos coeficientes de correlação estudados entre as duas variáveis. Na segunda metodologia, os dois CPs foram obtidos a partir da matriz de correlações (ρ) obtida a partir da matriz Σ citada anteriormente. E na terceira metodologia, os dois CPs foram obtidos a partir da matriz de variâncias e covariâncias das variáveis transformadas Z_1^* e Z_2^* (Σ^*), cujas variâncias foram ponderadas pelos respectivos coeficientes de variação de controle (CVs).

4.2.1. Gráfico de Controle 1

Dada a simulação dos 1.000 pares de valores para as duas variáveis-resposta (Y_1 e Y_2), em cada um dos 20 conjuntos de dados, o objetivo foi obter as novas variáveis (CP_1 e CP_2), tal que CP_j ($j = 1, 2$) seja a combinação linear das duas variáveis Y_S . Os dois CPs obtidos a partir da matriz Σ foram dados por:

$$CP_1 = a_{11}^* Y_1 + a_{12}^* Y_2 ; e$$

$$CP_2 = a_{21}^* Y_1 + a_{22}^* Y_2 ,$$

em que $a_j^* = [a_{j1}^* \ a_{j2}^*]$ = autovetor normalizado do CP_j , para $j = 1, 2$, obtido a partir da matriz Σ .

Para os dois CPs, foram obtidos os dois autovalores (λ_j) por meio do determinante da expressão baseada na matriz de variâncias e covariâncias Σ , que contém as variâncias de controle e as correlações preestabelecidas, como segue:

$$\Sigma - \lambda_j I = 0,$$

em que

$$\Sigma = \begin{bmatrix} \sigma_{01}^2 & \sigma_{12} \\ \sigma_{21} & \sigma_{02}^2 \end{bmatrix}, \text{ I = matriz identidade de ordem 2;}$$

σ_{01}^2 = variância de controle para Y_1 ;

σ_{02}^2 = variância de controle para Y_2 ; e

$\sigma_{12} = \sigma_{21}$ = covariância entre as variáveis Y_1 e $Y_2 = \rho_{12} \sigma_{01} \sigma_{02}$.

Para cada autovalor obtido, λ_1 e λ_2 , corresponde um autovetor normalizado ou não. O autovetor não normalizado a_j , foi obtido a partir de uma das soluções do sistema de equações, que é dado por:

$$[\Sigma - \lambda_j I] a_j = \phi,$$

em que $a_j = [a_{j1} \ a_{j2}]$ = autovetor a_j não normalizado do CP_j e ϕ = vetor nulo de dimensão 2×1 .

O autovetor a_j^* normalizado foi obtido por:

$$a_j^* = \begin{bmatrix} a_{j1}^* \\ a_{j2}^* \end{bmatrix} = \frac{1}{\sqrt{a_{j1}^2 + a_{j2}^2}} \begin{bmatrix} a_{j1} \\ a_{j2} \end{bmatrix},$$

em que $a_{j1}^{*2} + a_{j2}^{*2} = 1$.

Em seguida, foram obtidos os escores, ou seja, os valores de cada componente, originados das seguintes combinações lineares:

$$cp_{1s} = a_{11}^* y_{1s} + a_{12}^* y_{2s}; \text{ e}$$

$$cp_{2s} = a_{21}^* y_{1s} + a_{22}^* y_{2s},$$

em que y_{1s} = valor simulado de Y_1 na ordem s , y_{2s} = valor simulado de Y_2 na ordem s e $s = 1, 2, \dots, 1.000$.

A partir dos escores, foram obtidos os valores da variável a ser plotada no gráfico de controle 1 (V_1) proposto, a partir da seguinte expressão:

$$V_{1s} = \frac{\sum_{i=1}^k \lambda_i CP_i}{\sum_{i=1}^k \lambda_i}, \text{ para } s = 1, 2, \dots, 1.000,$$

em que k = número de componentes principais.

A partir dos 1.000 valores de V_1 e para cada conjunto de dados, foi construído o gráfico de controle 1 composto de 1.000 subgrupos racionais com observações individuais ($n = 1$).

Os limites de controle, inferior e superior, foram obtidos de acordo com o processo sob controle estatístico, a partir da seguinte expressão:

$$LC = LM \pm k \sqrt{V(V_1)},$$

em que

LM = linha média; e

$V(V_1)$ = variância da variável V_1 a ser plotada no gráfico de controle 1 (GC1).

A variância da variável V_1 a ser plotada foi definida como:

$$\begin{aligned} V(V_1) &= V\left(\frac{\lambda_1 CP_1 + \lambda_2 CP_2}{\lambda_1 + \lambda_2}\right) \\ &= \frac{1}{(\lambda_1 + \lambda_2)^2} [\lambda_1^2 V(CP_1) + \lambda_2^2 V(CP_2) + 2\lambda_1 \lambda_2 \text{Cov}(CP_1, CP_2)] \\ &= \frac{1}{(\lambda_1 + \lambda_2)^2} (\lambda_1^3 + \lambda_2^3) \\ &= \frac{\lambda_1^3 + \lambda_2^3}{\lambda_1^2 + 2\lambda_1 \lambda_2 + \lambda_2^2}. \end{aligned}$$

Segundo Ribeiro Júnior (2013), o valor de k é definido como o deslocamento em torno da LM, em número de desvios-padrão da variável-resposta, utilizado para incluir a quantidade de variabilidade aleatória permitida. O que se procura é

estabelecer limites de controle cuja distância da LM equivale a k desvios-padrão da variável que será utilizada para o monitoramento. Isso significa que existirá uma probabilidade menor que um de as estimativas da variável-resposta devido à causas aleatórias, estarem situadas entre os limites de controle (LIC e LSC) e uma probabilidade maior que zero de estarem situadas além desses limites. Ou seja, a probabilidade do alarme verdadeiro não será igual a um e existirá um valor maior que zero para a probabilidade do alarme falso.

Nesta presente proposta foi escolhido o valor $k = 3$, estabelecendo um intervalo de controle contendo 99,73% de todos os valores dos CPs associados a um processo sob controle estatístico.

Para as construções dos 20 gráficos de controle, que incluíram os processos sob e fora de controle estatístico, os elementos da matriz Σ foram definidos por:

$$\sigma_{01}^2 = 900;$$

$$\sigma_{02}^2 = 100; \text{ e}$$

$$\sigma_{12} = \rho_{12} \sigma_{01} \sigma_{02}, \text{ para:}$$

$$\rho_{12} = 0,1; 0,3; 0,5; 0,7 \text{ e } 0,9.$$

Por outro lado, a LM foi definida por:

$$LM = E(V_1) = E\left(\frac{\lambda_1 CP_1 + \lambda_2 CP_2}{\lambda_1 + \lambda_2}\right) = \frac{1}{\lambda_1 + \lambda_2} (\lambda_1 \mu_{CP_1} + \lambda_2 \mu_{CP_2}).$$

4.2.2. Gráfico de Controle 2

Nesta abordagem, os CPs foram obtidos a partir da matriz de correlações ρ , ao invés da matriz de variâncias e covariâncias, com o objetivo de contornar o efeito das escalas das variáveis-resposta Y_1 e Y_2 .

Dada a simulação dos 1.000 pares de valores para as duas variáveis-resposta (Y_1 e Y_2), em cada um dos 20 conjuntos de dados, o objetivo foi obter as novas variáveis (CP_1 e CP_2), tal que CP_j ($j=1,2$) seja a combinação linear das duas variáveis Y_s . Os dois CPs obtidos a partir da matriz ρ foram dados por:

$$CP_1 = a_{11}^* Z_1 + a_{12}^* Z_2; \text{ e}$$

$$CP_2 = a_{21}^* Z_1 + a_{22}^* Z_2,$$

em que $a_j^* = [a_{j1}^* \ a_{j2}^*]$ = autovetor normalizado do CP_j , para $j = 1, 2$, obtido a partir da matriz ρ ;

$$Z_1 = \frac{Y_1 - \mu_1}{\sigma_1}; \text{ e}$$

$$Z_2 = \frac{Y_2 - \mu_2}{\sigma_2}.$$

Para os dois CPs, foram obtidos os dois autovalores (λ_j) por meio do determinante da expressão baseada na matriz de correlações (ρ), como segue:

$$\rho - \lambda_j I = 0,$$

em que $\rho = \begin{bmatrix} 1 & \rho_{12} \\ \rho_{21} & 1 \end{bmatrix}$, I = matriz identidade de ordem 2 e $\rho_{12} = \rho_{21}$ = correlação preestabelecida entre as variáveis-resposta Y_1 e Y_2 .

Para cada autovalor obtido, λ_1 e λ_2 , corresponde um autovetor normalizado ou não. O autovetor não normalizado, a_j , foi obtido a partir de uma das soluções do sistema de equações dado por:

$$[\rho - \lambda_j I] a_j = \phi,$$

em que

$a_j = [a_{j1} \ a_{j2}]$ = autovetor a_j não normalizado do CP_j ; e

ϕ = vetor nulo de dimensão 2×1 .

O autovetor a_j^* normalizado foi dado por:

$$a_j^* = \begin{bmatrix} a_{j1}^* \\ a_{j2}^* \end{bmatrix} = \frac{1}{\sqrt{a_{j1}^2 + a_{j2}^2}} \begin{bmatrix} a_{j1} \\ a_{j2} \end{bmatrix},$$

em que $a_{j1}^{*2} + a_{j2}^{*2} = 1$.

Em seguida, foram obtidos os escores, ou seja, os valores de cada componente originados da seguinte combinação linear:

$$cp_{1r} = a_{11}^* z_{1r} + a_{12}^* z_{2r}; \text{ e}$$

$$cp_{2r} = a_{21}^* z_{1r} + a_{22}^* z_{2r},$$

em que

z_{1r} = valor padronizado de Y_1 na ordem r ;

z_{2r} = valor padronizado de Y_2 na ordem r ; e

$r = 1, 2, \dots, 1.000$.

A partir dos escores, foram obtidos os valores da variável a ser plotada no gráfico de controle 2 (V_2) proposto, a partir da seguinte expressão:

$$v_{2r} = \frac{\sum_{i=1}^k \lambda_i CP_i}{\sum_{i=1}^k \lambda_i}, \text{ para } r = 1, 2, \dots, 1.000,$$

em que k = número de componentes principais.

A partir dos 1.000 valores de V_2 e para cada conjunto de dados, foi construído o gráfico de controle 2 composto de 1.000 subgrupos racionais com observações individuais ($n = 1$).

Os limites de controle, inferior e superior, foram obtidos de acordo com o processo sob controle estatístico, a partir da seguinte expressão:

$$LC = LM \pm k \sqrt{V(V_2)},$$

em que

LM = linha média; e

$V(V_2)$ = variância da variável V_2 a ser plotada no gráfico de controle 2 (GC_2).

A variância da variável V_2 a ser plotada foi definida como:

$$\begin{aligned} V(V_2) &= V\left(\frac{\lambda_1 CP_1 + \lambda_2 CP_2}{\lambda_1 + \lambda_2}\right) \\ &= \frac{\lambda_1^3 + \lambda_2^3}{\lambda_1^2 + 2\lambda_1\lambda_2 + \lambda_2^2}. \end{aligned}$$

Assim como na proposta anterior, foi escolhido o valor $k = 3$, estabelecendo um intervalo de controle contendo 99,73% de todos os valores dos CPs associados a um processo sob controle estatístico.

Para as construções dos 20 gráficos de controle, os elementos da matriz ρ foram definidos por:

$$\rho = \begin{bmatrix} 1 & \rho_{12} \\ \rho_{12} & 1 \end{bmatrix},$$

em que $\rho_{12} = 0,1; 0,3; 0,5; 0,7$ e $0,9$.

Por outro lado, a LM foi definida por:

$$LM = E(V_2) = E\left(\frac{\lambda_1 CP_1 + \lambda_2 CP_2}{\lambda_1 + \lambda_2}\right) = \frac{1}{\lambda_1 + \lambda_2} (\lambda_1 \mu_{CP_1} + \lambda_2 \mu_{CP_2}).$$

Ribeiro Júnior (2012) explica que, no caso de se utilizar as variáveis padronizadas, a matriz de variâncias e covariâncias dos dados padronizados é equivalente à matriz de correlações dos dados originais e a média dos escores obtidos a partir destas variáveis padronizadas será zero.

4.2.3. Gráfico de Controle 3

Nesta abordagem, os CPs foram obtidos a partir da matriz dos coeficientes de variação (Σ^*), em que os dados, após serem padronizados com média zero e variância um, têm suas variâncias ponderadas pelos respectivos coeficientes de variação (CVs).

Dados os 1.000 pares de valores transformados para as duas variáveis-resposta (Y_1 e Y_2), em cada um dos 20 conjuntos de dados, o objetivo foi obter as novas variáveis (CP_1 e CP_2), tal que CP_j ($j=1,2$) seja a combinação linear das duas variáveis Y s. Os dois CPs obtidos a partir da matriz Σ^* foram dados por:

$$CP_1 = a_{11}^* Z_1^* + a_{12}^* Z_2^*; e$$

$$CP_2 = a_{21}^* Z_1^* + a_{22}^* Z_2^*,$$

em que

$a_j^* = [a_{j1}^* \ a_{j2}^*]$ = autovetor normalizado do CP_j , para $j = 1, 2$, obtido a partir da matriz Σ^* ;

$$Z_1^* = \frac{Y_1 - \mu_1}{\mu_1}; \text{ e}$$

$$Z_2^* = \frac{Y_2 - \mu_2}{\mu_2}.$$

Para os dois CPs, foram obtidos os dois autovalores (λ_j) por meio do determinante da expressão baseada na matriz Σ^* , como segue:

$$\Sigma^* - \lambda_j I = 0,$$

em que

$$\Sigma^* = \begin{bmatrix} \sigma_{01}^{*2} & \sigma_{12}^* \\ \sigma_{21}^* & \sigma_{02}^{*2} \end{bmatrix};$$

I = matriz identidade de ordem 2;

$\sigma_{01}^{*2} = CV_1^2$ = variância da variável transformada Z_1^* ;

$\sigma_{02}^{*2} = CV_2^2$ = variância da variável transformada Z_2^* ; e

$\sigma_{12}^* = \sigma_{21}^* = \rho_{12} CV_1 CV_2$ = covariância entre as variáveis transformadas Z_1^* e Z_2^* .

Conforme citado anteriormente, para cada autovalor obtido, λ_1 e λ_2 , corresponde um autovetor normalizado ou não. O autovetor não normalizado a_j , foi obtido a partir de uma das soluções do sistema de equações dado por:

$$[\Sigma^* - \lambda_j I] a_j = \phi,$$

em que

$a_j = [a_{j1} \ a_{j2}]$ = autovetor a_j não normalizado do CP_j; e

ϕ = vetor nulo de dimensão 2x1.

O autovetor a_j^* normalizado foi dado por:

$$a_j^* = \begin{bmatrix} a_{j1}^* \\ a_{j2}^* \end{bmatrix} = \frac{1}{\sqrt{a_{j1}^2 + a_{j2}^2}} \begin{bmatrix} a_{j1} \\ a_{j2} \end{bmatrix},$$

em que $a_{j1}^{*2} + a_{j2}^{*2} = 1$.

Em seguida, foram obtidos os escores, ou seja, os valores de cada componente originados da seguinte combinação linear:

$$\begin{aligned} cp_{1t} &= a_{11}^* z_{1t}^* + a_{12}^* z_{2t}^*; \text{ e} \\ cp_{2t} &= a_{21}^* z_{1t}^* + a_{22}^* z_{2t}^*, \end{aligned}$$

em que

z_{1t}^* = valor transformado de Y_1 na ordem t ;

z_{2t}^* = valor transformado de Y_2 na ordem t ; e

$t = 1, 2, \dots, 1.000$.

A partir dos escores obtidos, foram obtidos os valores da variável a ser plotada no gráfico de controle 3 (V_3) proposto, a partir da seguinte expressão:

$$v_{3t} = \frac{\sum_{i=1}^k \lambda_i CP_i}{\sum_{i=1}^k \lambda_i}, \text{ para } t = 1, 2, \dots, 1.000,$$

em que k = número de componentes principais.

A partir dos 1.000 valores de V_3 e para cada conjunto de dados, foi construído o gráfico de controle 3 composto de 1.000 subgrupos racionais com observações individuais ($n = 1$).

Os limites de controle, inferior e superior, foram obtidos, de acordo com o processo sob controle estatístico, a partir da seguinte expressão:

$$LC = LM \pm k \sqrt{V(V_3)},$$

em que

LM = linha média; e

$V(V_3)$ = variância da variável V_3 a ser plotada no gráfico de controle 3 (GC_3).

A variância da variável V_3 a ser plotada foi definida como:

$$\begin{aligned} V(V_3) &= V\left(\frac{\lambda_1 CP_1 + \lambda_2 CP_2}{\lambda_1 + \lambda_2}\right) \\ &= \frac{\lambda_1^3 + \lambda_2^3}{\lambda_1^2 + 2\lambda_1\lambda_2 + \lambda_2^2}. \end{aligned}$$

Assim como nas duas propostas anteriores, foi escolhido o valor $k = 3$, estabelecendo um intervalo de controle contendo 99,73% de todos os valores plotados associados a um processo sob controle estatístico.

Para as construções dos 20 gráficos de controle, os elementos da matriz Σ^* foram definidos por:

$$CV_1^2 = \left(\frac{\sigma_{01}}{\mu_{01}} \right)^2 ;$$

$$CV_2^2 = \left(\frac{\sigma_{02}}{\mu_{02}} \right)^2 ;$$

$\rho_{12} CV_1 CV_2$, fora da diagonal, para:

$$\rho_{12} = 0,1; 0,3; 0,5; 0,7 \text{ e } 0,9.$$

Já a LM foi definida por:

$$\begin{aligned} LM &= E(V_3) = \\ &= \frac{1}{\lambda_1 + \lambda_2} (\lambda_1 \mu_{CP_1} + \lambda_2 \mu_{CP_2}). \end{aligned}$$

As diferenças entre os limites de controle, inferior (LIC) e superior (LCS) e a linha média (LM), obtidos nos três gráficos de controle propostos, serão decorrentes das estimativas dos autovalores e das médias dos CPs, obtidas a partir das matrizes Σ , ρ e Σ^* , respectivamente.

4.2.4. Avaliação do desempenho

O desempenho de um determinado tipo de gráfico de controle pode ser medido por meio das probabilidades do alarme falso (α) e do alarme verdadeiro (Pd).

Como a probabilidade do alarme falso é a de detectar pontos fora de controle devido à causas aleatórias, esta foi a forma empregada para avaliar o desempenho dos gráficos propostos no primeiro cenário, onde o processo encontra-se sob controle estatístico. A proporção de pontos acusados pelos gráficos propostos como fora de controle dentre os 1.000 valores plotados foi utilizada para a estimativa da probabilidade do alarme falso, como:

$$\hat{\alpha} = \frac{\text{NPFC}_0}{1000},$$

em que

$\hat{\alpha}$ = estimativa da probabilidade do alarme falso; e

NPFC_0 = número de pontos acusados pelo gráfico como fora de controle estatístico de um processo sob controle.

Para os outros três cenários, onde os processos encontram-se fora de controle estatístico, a forma empregada para avaliar o desempenho dos gráficos propostos foi a probabilidade do alarme verdadeiro ou o poder do gráfico. Isto porque a probabilidade do alarme verdadeiro é a de detectar pontos fora dos limites de controle devido a causas especiais. A proporção de pontos acusados pelos gráficos propostos como fora de controle dentre os 1.000 valores plotados foi utilizada para a estimativa da probabilidade do alarme verdadeiro, como:

$$\hat{\text{Pd}} = \frac{\text{NPFC}_1}{1000},$$

em que

$\hat{\text{Pd}}$ = estimativa da probabilidade do alarme verdadeiro (poder); e

NPFC_1 = número de pontos acusados pelo gráfico como fora de controle estatístico de um processo fora de controle.

5. RESULTADOS E DISCUSSÃO

5.1. Processo sob controle

Os resultados obtidos para os autovalores e com as respectivas porcentagens de explicação, em cada uma das três matrizes abordadas no trabalho (Σ , ρ e Σ^*) e de acordo com a correlação existente entre as variáveis-resposta Y_1 e Y_2 , encontram-se na Tabela 4.

Tabela 4 – Autovalores obtidos pelas matrizes Σ , ρ e Σ^* e as respectivas porcentagens de explicação de cada componente principal de acordo com a correlação

ρ_{12}	Matriz Σ				Matriz ρ				Matriz Σ^*			
	λ_1	λ_2	% λ_1	% λ_2	λ_1	λ_2	% λ_1	% λ_2	λ_1	λ_2	% λ_1	% λ_2
0,1	901,12	98,88	90	10	1,10	0,90	55	45	110,00	90,00	55	45
0,3	910,00	90,00	91	9	1,30	0,70	65	35	130,00	70,00	65	35
0,5	927,20	72,80	93	7	1,50	0,50	75	25	150,00	50,00	75	25
0,7	951,77	48,23	95	5	1,70	0,30	85	15	170,00	30,00	85	15
0,9	982,60	17,40	98	2	1,90	0,10	95	5	190,00	10,00	95	5

No caso da menor correlação entre as variáveis-resposta ($\rho_{12} = 0,1$), o primeiro CP, de acordo com a matriz Σ , foi responsável pela explicação de 90% da variabilidade total, ficando clara a influência da escala, por haver predominância da variável-resposta Y_1 de maior variância. Neste caso, a porcentagem de explicação foi superestimada. Já pelas matrizes ρ e Σ^* , o efeito da escala foi contornado, ficando o primeiro CP responsável por pouco mais que 50%, devido ao baixo valor da correlação. À medida em que a correlação entre as variáveis-resposta aumentou, o percentual de explicação do primeiro CP também aumentou, como esperado.

No presente estudo, as explicações dos CP₁ e CP₂ obtidos pelas matrizes ρ e Σ^* foram exatamente as mesmas, dado que Y_1 e Y_2 apresentaram o mesmo CV, igual a 0,10. Caso contrário, a variável que apresentasse o maior CV teria maior influência pela matriz Σ^* .

O aumento da correlação aumentou a importância do primeiro CP nos três casos. Portanto, quando se utilizou a matriz Σ , a magnitude dos autovalores dos CPs

esteve diretamente relacionada à magnitude das variâncias e também das covariâncias entre as variáveis estudadas.

Os resultados dos autovetores normalizados, que ponderam os valores das variáveis-resposta Y_1 e Y_2 nos respectivos CPs, também foram proporcionais aos resultados dos respectivos autovalores, conforme apresentados na Tabela 5.

Tabela 5 – Autovetores normalizados dos componentes principais CP₁ e CP₂, obtidos pelas matrizes Σ , ρ e Σ^*

ρ_{12}	Matriz Σ				Matriz ρ				Matriz Σ^*			
	CP ₁		CP ₂		CP ₁		CP ₂		CP ₁		CP ₂	
	a_{11}^*	a_{12}^*	a_{21}^*	a_{22}^*	a_{11}^*	a_{12}^*	a_{21}^*	a_{22}^*	a_{11}^*	a_{12}^*	a_{21}^*	a_{22}^*
0,1	0,999	0,037	0,037	0,999	0,707	0,707	-0,707	0,707	0,707	0,707	-0,707	0,707
0,3	0,994	0,11	0,11	0,994	0,707	0,707	-0,707	0,707	0,707	0,707	-0,707	0,707
0,5	0,984	0,178	0,178	0,984	0,707	0,707	-0,707	0,707	0,707	0,707	-0,707	0,707
0,7	0,971	0,239	0,239	0,971	0,707	0,707	-0,707	0,707	0,707	0,707	-0,707	0,707
0,9	0,956	0,293	0,293	0,956	0,707	0,707	-0,707	0,707	0,707	0,707	-0,707	0,707

A partir dos autovalores e autovetores obtidos, foram caracterizados os dois componentes principais, em termos de média e desvio-padrão, para o processo sob controle (Tabela 6).

Tabela 6 – Médias e desvios-padrão dos componentes principais obtidos pelas matrizes Σ , ρ e Σ^*

ρ_{12}	Matriz Σ				Matriz ρ				Matriz Σ^*			
	μ_{CP1}	σ_{CP1}	μ_{CP2}	σ_{CP2}	μ_{CP1}	σ_{CP1}	μ_{CP2}	σ_{CP2}	μ_{CP1}	σ_{CP1}	μ_{CP2}	σ_{CP2}
0,1	303,57	30,00	88,71	9,94	0	1,04	0	0,95	0	10,49	0	9,49
0,3	309,19	30,12	66,25	9,49	0	1,14	0	0,84	0	11,40	0	8,37
0,5	313,00	30,47	44,87	8,54	0	1,22	0	0,71	0	12,25	0	7,07
0,7	315,20	30,84	25,28	6,94	0	1,30	0	0,55	0	13,04	0	5,48
0,9	315,20	30,84	25,28	6,94	0	1,38	0	0,32	0	13,78	0	3,16

Em função das médias e dos desvios-padrão obtidos para os dois componentes principais, foram obtidas as mesmas medidas para as variáveis V_1 , V_2 e V_3 plotadas nos gráficos de controle GC₁, GC₂ e GC₃, respectivamente (Tabela 7).

De acordo com o processo sob controle estatístico, foram obtidos a linha média (LM) e os limites de controle inferior (LIC) e superior (LSC), para as construções dos gráficos de controle GC₁, GC₂ e GC₃, tanto para o processo sob controle como para os processos fora de controle estatístico (Tabela 8).

Tabela 7 – Médias e desvios-padrão das variáveis plotadas nos gráficos

ρ_{12}	GC ₁	GC ₂	GC ₃	GC ₁	GC ₂	GC ₃
	μ_{V_1}	σ_{V_1}	μ_{V_2}	ρ_{12}	μ_{V_1}	σ_{V_1}
0,1	282,3	27,1	0	0,72	0	7,2
0,3	287,3	27,5	0	0,80	0	8,0
0,5	293,5	28,2	0	0,93	0	9,3
0,7	301,2	29,4	0	1,11	0	11,1
0,9	310,1	30,3	0	1,31	0	13,1

Tabela 8 – Linha média e limites de controle obtidos nos três gráficos de controle

ρ_{12}	GC ₁			GC ₂			GC ₃		
	LIC	LM	LSC	ρ_{12}	LIC	LM	LSC	ρ_{12}	LIC
0,1	201,1	282,3	363,5	-2,2	0	2,2	-21,5	0	21,5
0,3	204,9	287,3	369,7	-2,4	0	2,4	-23,9	0	23,9
0,5	208,9	293,5	378,1	-2,8	0	2,8	-28,0	0	28,1
0,7	213,1	301,2	389,3	-3,3	0	3,3	-33,3	0	33,4
0,9	219,2	310,1	401,1	-3,9	0	3,9	-39,2	0	39,3

Na Figura 4, os gráficos de controle gerados no cenário onde o processo encontrava-se sob controle estatístico empregando a matriz Σ .

Também chama a atenção o fato de o gráfico proposto apresentar a probabilidade do alarme falso constante trabalhando com observações individuais em cada subgrupo racional, mostrando que, de forma semelhante aos gráficos de Shewhart univariados, o alarme falso independe do tamanho de n.

Observou-se neste caso que o gráfico de controle 1 apontou o processo como sob controle estatístico, pois a maioria dos pontos ficou localizada entre os limites de controle. Os pontos que ficaram além dos limites de controle são os alarmes falsos. Logo, conclui-se que a matriz Σ foi adequada para este cenário.

Na Figura 5, os gráficos de controle gerados no cenário onde o processo encontra-se sob controle estatístico empregando a matriz ρ .

Assim como no gráfico de controle 1, observou-se neste cenário que o gráfico de controle 2 apontou o processo como sob controle estatístico, pois a maioria dos pontos ficou localizada entre os limites de controle. Os pontos que ficaram além dos limites de controle são os alarmes falsos. Logo, conclui-se que a matriz ρ foi adequada para este cenário. Como os dados foram padronizados com média zero e variância um, os escores obtidos apresentaram média zero e, conseqüentemente, a média da variável V_2 também foi zero.

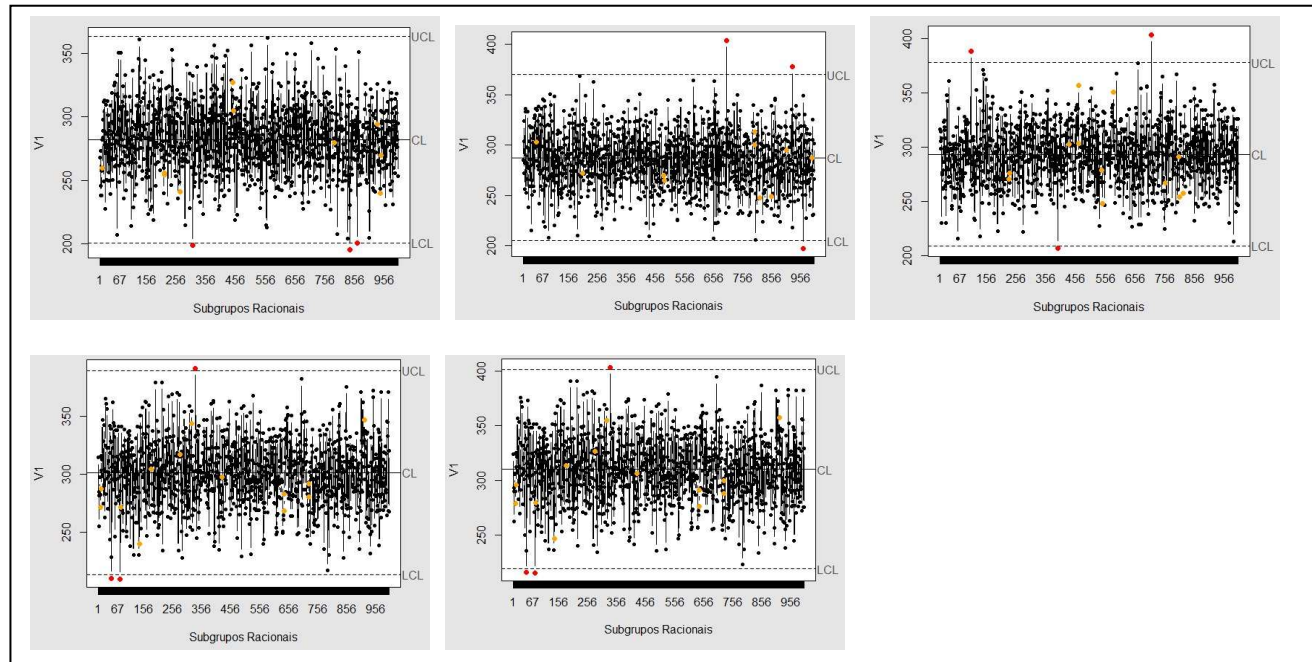


Figura 4 – Gráfico de Controle 1 – Processo sob controle estatístico, com correlações de 0,1, 0,3, 0,5, 0,7 e 0,9, da esquerda para direita e de cima para baixo, respectivamente.

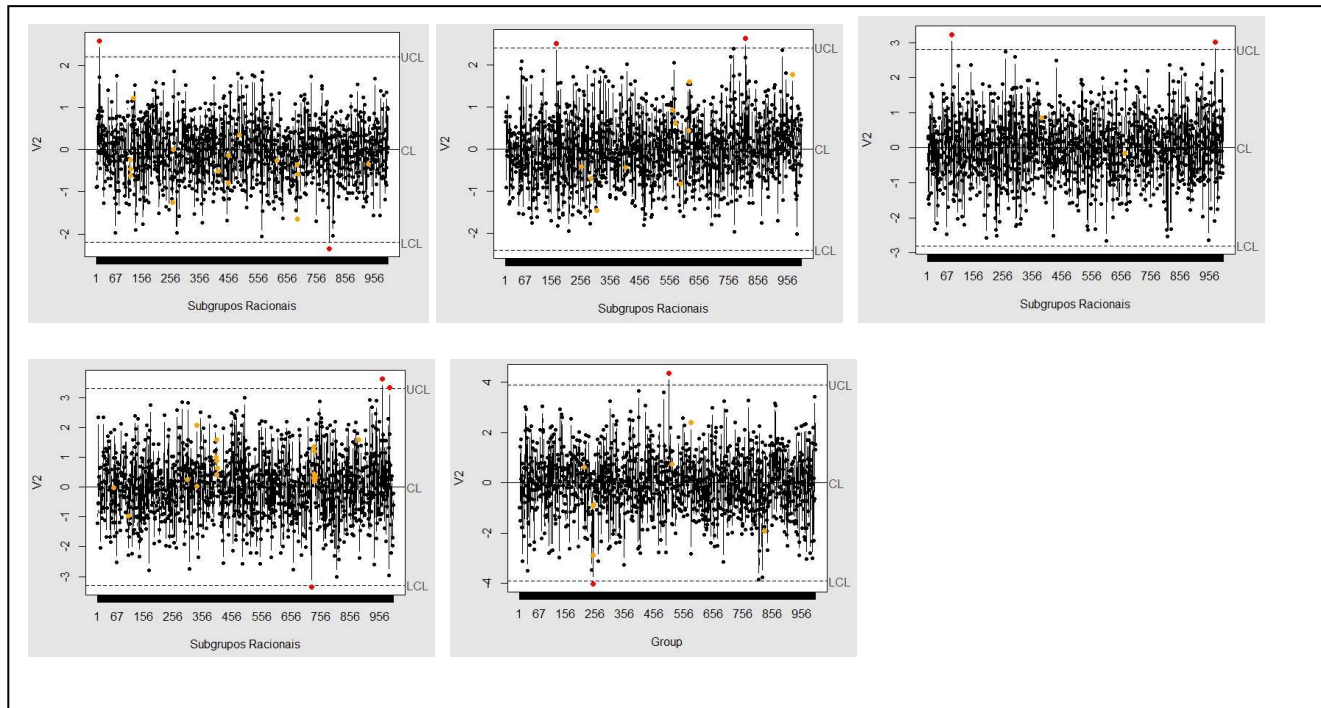


Figura 5 – Gráfico de Controle 2 – Processo sob controle estatístico, com correlações de 0,1, 0,3, 0,5, 0,7 e 0,9, da esquerda para direita e de cima para baixo, respectivamente.

Assim como nos gráficos de controle 1 (Figura 4) e controle 2 (Figura 5), observou-se neste cenário que o gráfico de controle 3 (Figura 6) apontou o processo como sob controle estatístico, pois a maioria dos pontos ficou localizada entre os limites de controle. Os pontos que ficaram além dos limites de controle são os alarmes falsos. Logo, conclui-se que a matriz Σ^* foi adequada para este cenário. Como os dados foram transformados, os escores obtidos também apresentaram média zero e, conseqüentemente, a média da variável V_3 também foi zero.

As variáveis plotadas nos gráficos, V_1 , V_2 e V_3 apresentaram comportamento independente, sendo variáveis aleatórias (v.a.) normais, em todas as diferentes correlações entre as variáveis-resposta e os gráficos propostos apresentaram comportamento similar ao gráfico de controle univariado de Shewhart. Portanto, quando o processo encontra-se sob controle estatístico, pode-se utilizar quaisquer uma das três matrizes propostas para o monitoramento.

Nos três gráficos de controle propostos, as variáveis V_1 , V_2 e V_3 a serem plotadas, contemplam 100% da variação total existente nas variáveis-resposta estudadas. Nas suas formulações, há as ponderações dos escores dos CPs pelos seus respectivos autovalores.

Portanto, as utilizações das variáveis V_1 , V_2 e V_3 implicam na construção de um gráfico de controle univariado, mas com a diferença de incorporar todas as variáveis-resposta de interesse com as suas devidas correlações. Tal fato substitui a construção de um gráfico de controle multivariado, sem a perda da informação multivariada.

A probabilidade do alarme falso proporcionada pelos gráficos propostos não dependeu das correlações existentes entre as variáveis-resposta Y_1 e Y_2 e nem da matriz empregada para a obtenção dos CPs, Σ (GC₁), ρ (GC₂) e Σ^* (GC₃), conforme é mostrado na Tabela 9.

Pode-se notar que a probabilidade do alarme falso manteve-se a mesma do valor teórico esperado de 0,0027 (COSTA *et al.*, 2004; SOUZA *et al.*, 2008; RIBEIRO JÚNIOR, 2013). Conforme descrito na seção 3.2, ocorre uma distorção na probabilidade do alarme falso no monitoramento de duas ou mais variáveis simultaneamente, ficando o valor consideravelmente maior que o valor teórico. Esta distorção não ocorreu nas propostas apresentadas.

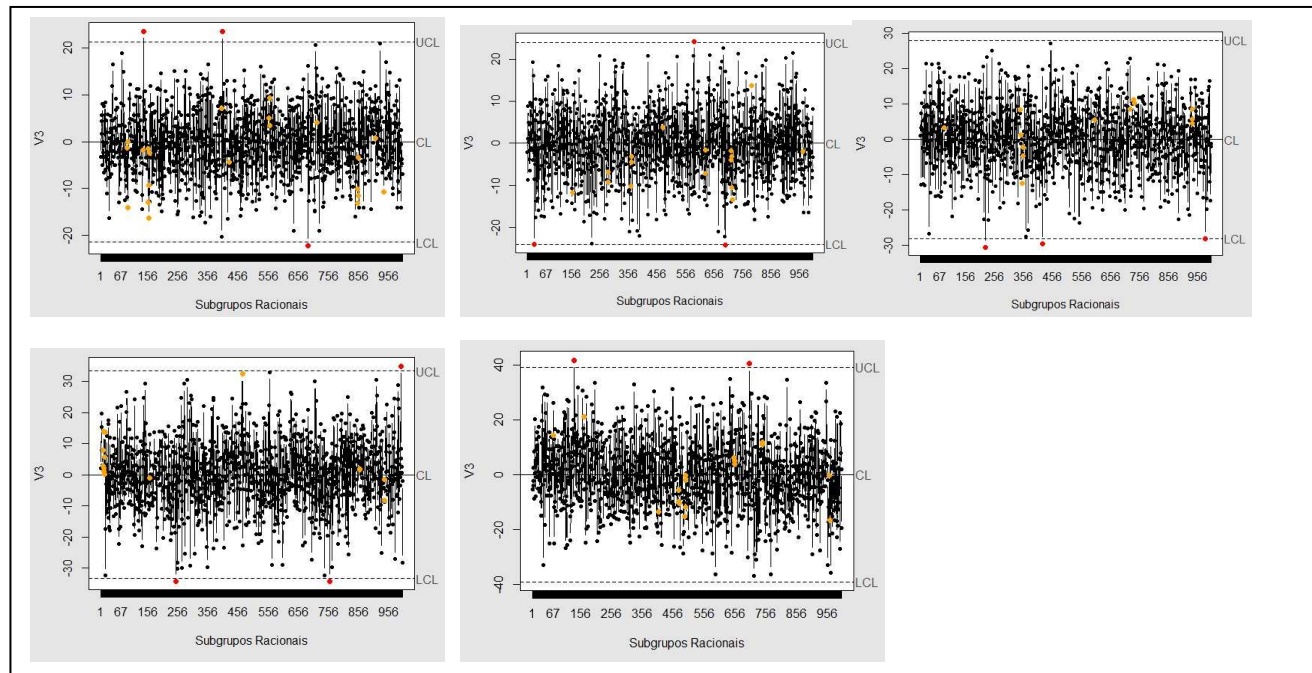


Figura 6 – Gráfico de Controle 3 – Processo sob controle estatístico, com correlações de 0,1, 0,3, 0,5, 0,7 e 0,9, da esquerda para direita e de cima para baixo, respectivamente.

Tabela 9 – Estimativa da probabilidade do alarme falso em função das correlações existentes entre Y_1 e Y_2 , em cada um dos três gráficos de controle propostos

ρ_{12}	GC ₁	GC ₂	GC ₃
0,1	0,0027	0,0027	0,0027
0,3	0,0027	0,0027	0,0027
0,5	0,0028	0,0027	0,0027
0,7	0,0027	0,0027	0,0027
0,9	0,0027	0,0027	0,0027
Média	0,0027	0,0027	0,0027

5.2. Processos Fora de Controle Estatístico

5.2.1. Deslocamento da Média

Conforme descrito nas seções 4.2.1, 4.2.2 e 4.2.3, as matrizes empregadas para a obtenção dos CPs, Σ (GC₁), ρ (GC₂) e Σ^* (GC₃) foram as mesmas para o processo sob controle e para os processos fora de controle. Portanto, em todos os cenários os autovalores e autovetores apresentaram-se iguais.

A partir dos autovalores e autovetores obtidos, que foram comuns aos processos sob e fora de controle, foram caracterizados os dois componentes principais, em termos de média e desvio-padrão, para o processo em que ocorreu deslocamento da média (Tabela 10).

Tabela 10 – Médias e desvios-padrão dos componentes principais obtidos pelas matrizes Σ , ρ e Σ^*

ρ_{12}	Matriz Σ				Matriz ρ				Matriz Σ^*			
	μ_{CP1}	σ_{CP1}	μ_{CP2}	σ_{CP2}	μ_{CP1}	σ_{CP1}	μ_{CP2}	σ_{CP2}	μ_{CP1}	σ_{CP1}	μ_{CP2}	σ_{CP2}
0,1	333,92	30,00	97,58	9,92	0,00	1,05	0,00	0,95	0,00	9,53	0,00	8,62
0,3	340,19	30,20	72,89	9,48	0,00	1,14	0,00	0,84	0,00	10,36	0,00	7,60
0,5	344,34	30,46	49,35	8,54	0,00	1,22	0,00	0,71	0,00	11,13	0,00	6,43
0,7	346,76	30,87	27,80	6,95	0,00	1,30	0,00	0,55	0,00	11,85	0,00	4,98
0,9	347,75	31,35	8,66	4,17	0,00	1,38	0,00	0,32	0,00	12,54	0,00	2,88

Em função das médias e dos desvios-padrão obtidos para os dois componentes principais, foram obtidas as mesmas medidas para as variáveis V_1 , V_2 e V_3 plotadas nos gráficos de controle GC₁, GC₂ e GC₃, respectivamente (Tabela 11).

Tabela 11 – Médias e desvios-padrão das variáveis plotadas nos gráficos

ρ_{12}	GC ₁		GC ₂		GC ₃	
	μ_{V_1}	σ_{V_1}	μ_{V_2}	σ_{V_2}	μ_{V_3}	σ_{V_3}
0,1	310,5	27,1	0	0,7	0	7,0
0,3	316,0	27,5	0	0,8	0	7,2
0,5	322,9	28,3	0	0,9	0	8,5
0,7	331,4	29,4	0	1,0	0	10,1
0,9	341,8	30,8	0	1,3	0	11,3

De acordo com o processo sob controle estatístico, foram obtidos a linha média (LM) e os limites de controle inferior (LIC) e superior (LSC), para as construções dos gráficos de controle GC₁, GC₂ e GC₃, tanto para o processo sob controle como para os processos fora de controle estatístico. Portanto, a linha média e os limites de controle são iguais aos do processo sob controle.

A seguir, nas Figuras 7, 8 e 9, os gráficos de controle gerados no cenário onde o processo encontra-se fora de controle estatístico, com deslocamento da média.

Observou-se neste caso que o gráfico de controle 1 apontou o processo como fora de controle estatístico, pois vários pontos ficaram localizados além dos limites de controle. A maioria dos pontos que estava além dos limites ficou acima do limite superior de controle e nenhum ponto ficou abaixo do limite inferior de controle, indicando que a média foi deslocada para cima, ou seja, a média foi maior que a média de controle, conforme esperado. Conclui-se, então, que a matriz Σ foi adequada para este cenário.

Neste caso, o gráfico de controle 2 apontou, erroneamente, o processo como sob controle estatístico, pois a maioria dos pontos encontraram-se entre os limites de controle. Esta característica foi uma consequência da padronização dos dados, pois eles apresentaram média zero e variância um. Assim, a variável a ser plotada V_2 também apresentou média zero, conforme indicou o gráfico. Conclui-se, então, que a matriz ρ não foi adequada para este cenário.

Neste caso, de modo semelhante ao anterior, o gráfico de controle 3 apontou, erroneamente, o processo como sob controle estatístico, pois vários pontos ficaram localizados além dos limites de controle. Esta característica foi uma consequência da transformação dos dados, pois eles apresentaram média zero. Assim, a variável a ser plotada V_3 também apresentou média zero, conforme indicou o gráfico. Conclui-se, então, que a matriz Σ^* não foi adequada para este cenário.

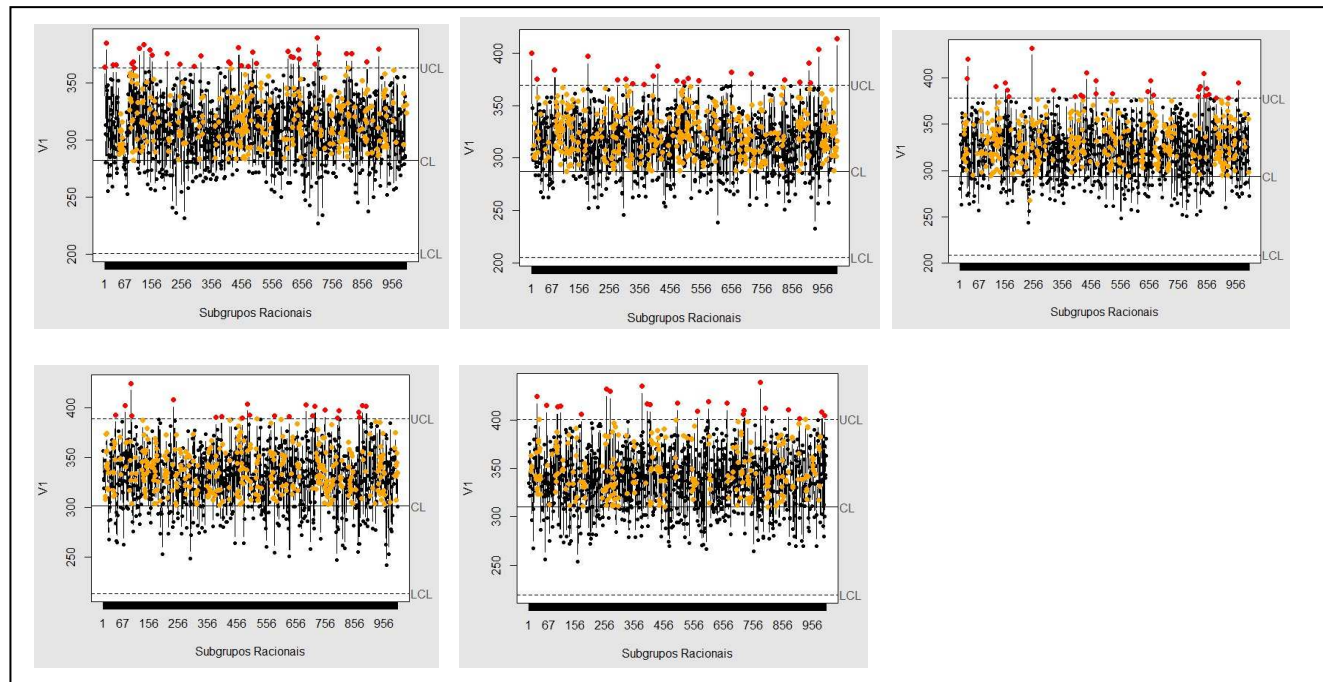


Figura 7 – Gráfico de Controle 1 – Processo com deslocamento da média de controle, com correlações de 0,1, 0,3, 0,5, 0,7 e 0,9, da esquerda para direita e de cima para baixo, respectivamente.

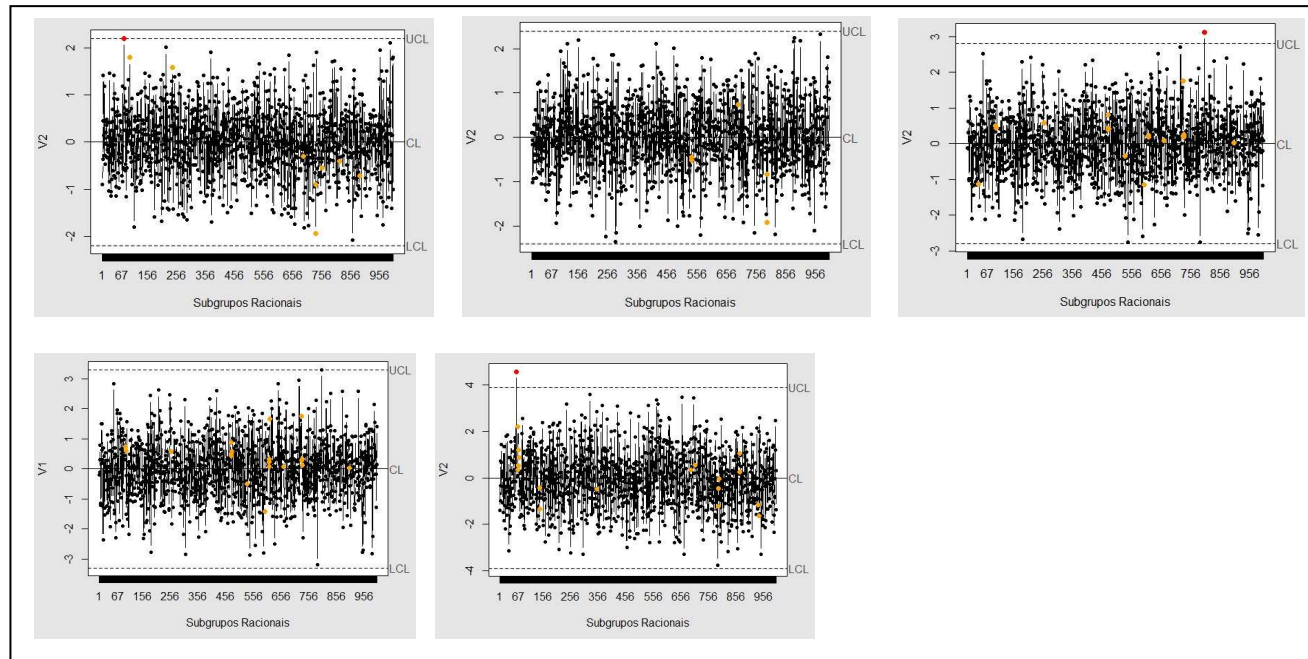


Figura 8 – Gráfico de Controle 2 – Processo com deslocamento da média de controle, com correlações de 0,1, 0,3, 0,5, 0,7 e 0,9, da esquerda para direita e de cima para baixo, respectivamente.

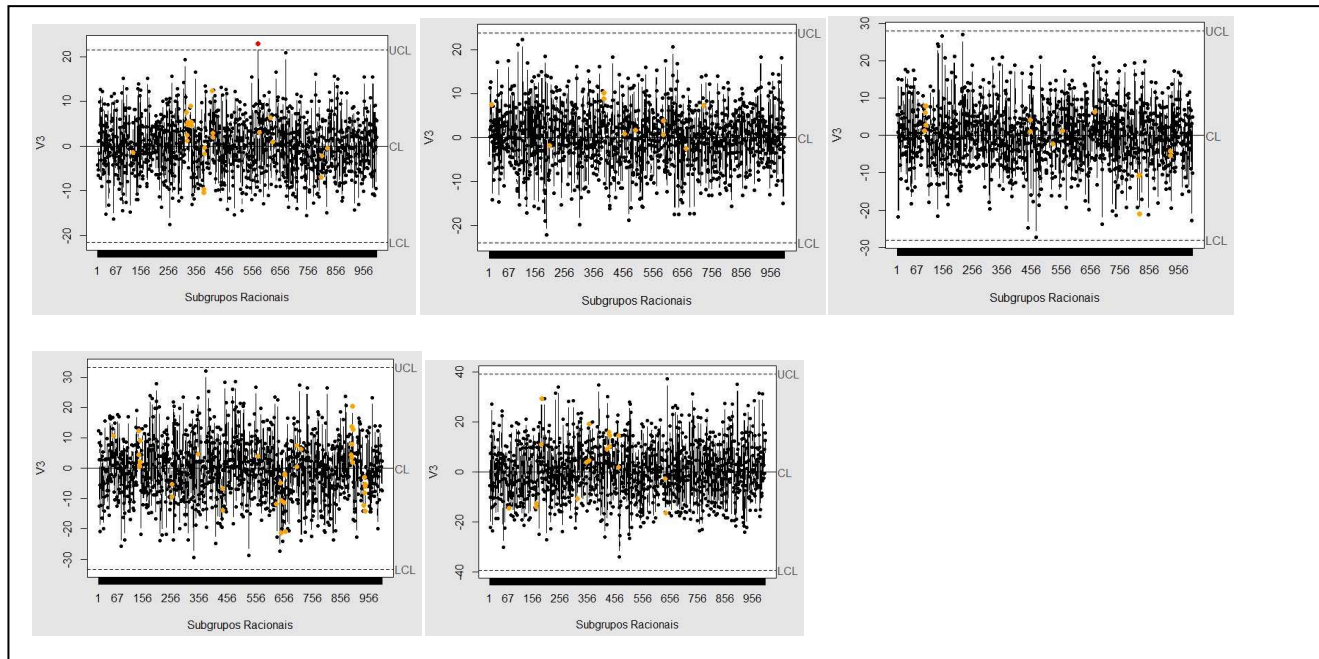


Figura 9 – Gráfico de Controle 3 – Processo com deslocamento da média de controle, com correlações de 0,1, 0,3, 0,5, 0,7 e 0,9, da esquerda para direita e de cima para baixo, respectivamente.

A medida empregada para avaliação do gráfico proposto foi a probabilidade do alarme verdadeiro, pois o objetivo era apontar os pontos como fora de controle em função das causas especiais. É desejável a obtenção de um alto valor para o Pd dos três gráficos ($pd \geq 0,90$).

Conforme apresentado na Tabela 12, o GC₁ apresentou baixo valor para o Poder e os GC₂ e GC₃ apresentaram poderes próximos a zero, não dependendo das correlações existentes entre as variáveis-respostas Y₁ e Y₂. Estes baixos valores indicaram a falta de detecção dos gráficos propostos (β), ou seja, eles não foram poderosos o suficiente para apontarem como fora de controle a maioria dos pontos que realmente estavam fora de controle.

Tabela 12 – Probabilidade do alarme verdadeiro observada em função das correlações existentes entre Y₁ e Y₂ em cada um dos gráficos propostos

ρ_{12}	GC ₁	GC ₂	GC ₃
0,1	0,03	0,0021	0,0010
0,3	0,03	0,0025	0,0010
0,5	0,03	0,0028	0,0010
0,7	0,03	0,0016	0,0009
0,9	0,03	0,0029	0,0005
Média	0,03	0,0024	0,0009

Pode-se notar que a probabilidade do alarme verdadeiro nos três gráficos manteve-se muito próxima do valor do alarme verdadeiro do gráfico univariado de Shewhart, em função de k, n e δ .

Conforme mencionado nas seções 4.1 e 4.2.1, em todas as três propostas apresentadas empregou-se observações individuais em cada subgrupo racional. Além disto, neste cenário específico, o deslocamento da média foi da ordem de um desvio-padrão de controle. Comparando estas informações com os gráficos de controle de Shewhart, observa-se uma considerável semelhança, pois nestes gráficos o Poder é baixo quando n e δ também são baixos. Nos gráficos de Shewhart, quando tem-se n = 1 e $\delta = 1$, para k = 3, o poder é de apenas 0,0228 (RIBEIRO JÚNIOR, 2013).

As variáveis plotadas nos gráficos, V_1 , V_2 e V_3 , apresentaram comportamento similar em todas as diferentes correlações entre as variáveis-resposta e os gráficos propostos apresentaram comportamento similar ao gráfico de controle univariado de Shewhart. Portanto, quando o processo encontra-se fora de controle estatístico, com deslocamento da média de controle, é recomendável utilizar apenas a matriz Σ proposta para o monitoramento.

Como as variáveis V_2 e V_3 são ponderadas pelos valores das variáveis padronizadas Z_s e transformadas Z_s^* , com médias iguais a zero, os gráficos de controle GC2 e GC3 apresentaram poderes ou probabilidades dos alarmes verdadeiros próximas às probabilidades dos alarmes falsos.

Portanto, concluiu-se que as duas variáveis anteriores não podem ser utilizadas para o monitoramento da média. Isto implica somente a indicação da variável V_1 para tal finalidade.

Com o objetivo de aumentar o Poder, uma alternativa adotada foi plotar a variável V_1 nos gráficos de controle da Média Móvel Ponderada Exponencialmente (EWMA) e gráfico de controle da Soma Acumulada (CUSUM), ambos univariados. Os resultados encontrados para o poder são apresentados nas Tabelas 13 e 14. Segundo Ribeiro Júnior (2013), nestes gráficos a decisão sobre o processo é baseada na informação acumulada dos subgrupos racionais anteriores e não apenas no último. Desta forma, acumulando informações sobre o processo, consegue-se maior rapidez na sinalização de menores desajustes na média.

Conforme pode-se observar na Tabela 13, no EWMA o GC₁ mostrou-se altamente poderoso, enquanto os GC₂ e GC₃ permaneceram com poderes iguais a zero. No gráfico CUSUM o comportamento foi similar ao EWMA, conforme a Tabela 14, também mostrando-se um gráfico poderoso.

Tabela 13 – Probabilidade do alarme verdadeiro observada em função das correlações existentes entre Y_1 e Y_2 plotando a variável V_1 no gráfico de controle EWMA

Correlação	GC ₁
0,1	0,995
0,3	0,997
0,5	0,992
0,7	0,987
0,9	0,991

Tabela 14 – Probabilidade do alarme verdadeiro observada em função das correlações existentes entre Y_1 e Y_2 plotando a variável V_1 no gráfico de controle CUSUM

Correlação	GC1
0,1	0,996
0,3	0,997
0,5	0,992
0,7	0,987
0,9	0,991

5.2.2. Aumento da variabilidade

Conforme descrito nas seções 4.2.1, 4.2.2 e 4.2.3, as matrizes empregadas para a obtenção dos CPs, Σ (GC₁), ρ (GC₂) e Σ^* (GC₃) foram as mesmas para o processo sob controle e para os processos fora de controle. Portanto, em todos os cenários os autovalores e autovetores apresentaram-se iguais.

A partir dos autovalores e autovetores obtidos, que foram comuns aos processos sob e fora de controle, foram caracterizados os dois componentes principais, em termos de média e desvio-padrão, para o processo em que ocorreu aumento na variabilidade (Tabela 15).

Tabela 15 – Médias e desvios-padrão dos componentes principais obtidos pelas matrizes Σ , ρ e Σ^*

ρ_{12}	Matriz Σ				Matriz ρ				Matriz Σ^*			
	μ_{CP1}	σ_{CP1}	μ_{CP2}	σ_{CP2}	μ_{CP1}	σ_{CP1}	μ_{CP2}	σ_{CP2}	μ_{CP1}	σ_{CP1}	μ_{CP2}	σ_{CP2}
0,1	303,62	60,01	88,71	19,85	0	1,04	0	0,95	0	21,00	0	18,97
0,3	309,25	60,28	66,21	18,99	0	1,14	0	0,84	0	22,79	0	22,79
0,5	313,08	60,87	44,86	17,06	0	1,22	0	0,71	0	24,51	0	14,13
0,7	315,19	61,71	25,27	13,89	0	1,30	0	0,55	0	26,04	0	10,95
0,9	316,07	62,70	7,86	8,35	0	1,38	0	0,32	0	27,56	0	6,33

Em função das médias e dos desvios-padrão obtidos para os dois componentes principais, foram obtidas as mesmas medidas para as variáveis V_1 , V_2 e V_3 plotadas nos gráficos de controle GC₁, GC₂ e GC₃, respectivamente (Tabela 16).

Tabela 16 – Médias e desvios-padrão das variáveis plotadas nos gráficos

ρ_{12}	GC ₁		GC ₂		GC ₃	
	μ_{V_1}	σ_{V_1}	μ_{V_2}	σ_{V_2}	μ_{V_3}	σ_{V_3}
0,1	282,3	54,1	0	0,7	0	14,4
0,3	287,3	55,0	0	0,8	0	14,9
0,5	295,1	56,5	0	0,9	0	18,7
0,7	301,3	58,8	0	1,1	0	22,2
0,9	310,8	61,6	0	1,3	0	26,2

De acordo com o processo sob controle estatístico, foram obtidos a linha média (LM) e os limites de controle inferior (LIC) e superior (LSC), para as construções dos gráficos de controle GC₁, GC₂ e GC₃, tanto para o processo sob controle como para os processos fora de controle estatístico. Portanto, a linha média e os limites de controle são iguais aos do processo sob controle.

A seguir, nas Figuras 10, 11 e 12, os gráficos de controle gerados no cenário onde o processo encontra-se fora de controle estatístico, com aumento da variabilidade.

Observou-se neste caso que o gráfico de controle 1 apontou o processo como fora de controle estatístico, pois vários pontos ficaram localizados além dos limites de controle. A maioria dos pontos que saíram de controle ficou acima do limite superior de controle e abaixo do limite inferior de controle, indicando que a variabilidade aumentou, conforme esperado. Conclui-se, então, que a matriz Σ foi adequada para este cenário.

Neste caso, o gráfico de controle 2 apontou, erroneamente, o processo como sob controle estatístico, pois a maioria dos pontos encontraram-se entre os limites de controle. Esta característica foi uma consequência da padronização dos dados, pois eles apresentaram média zero e variância um. Assim, a variável a ser plotada V_2 também apresentou média zero, conforme indicou o gráfico. Conclui-se, então, que a matriz ρ não foi adequada para este cenário.

Neste caso, o gráfico de controle 3 apontou o processo como fora de controle estatístico, pois vários pontos encontraram-se fora dos limites de controle. A transformação dos dados não influenciou no poder do gráfico, pois eles apresentaram média zero e a variância ponderada pelo respectivo CV. Assim, a variável a ser plotada V_3 também apresentou média zero, conforme indicou o gráfico. Conclui-se, então, que a matriz Σ^* foi adequada para este cenário.

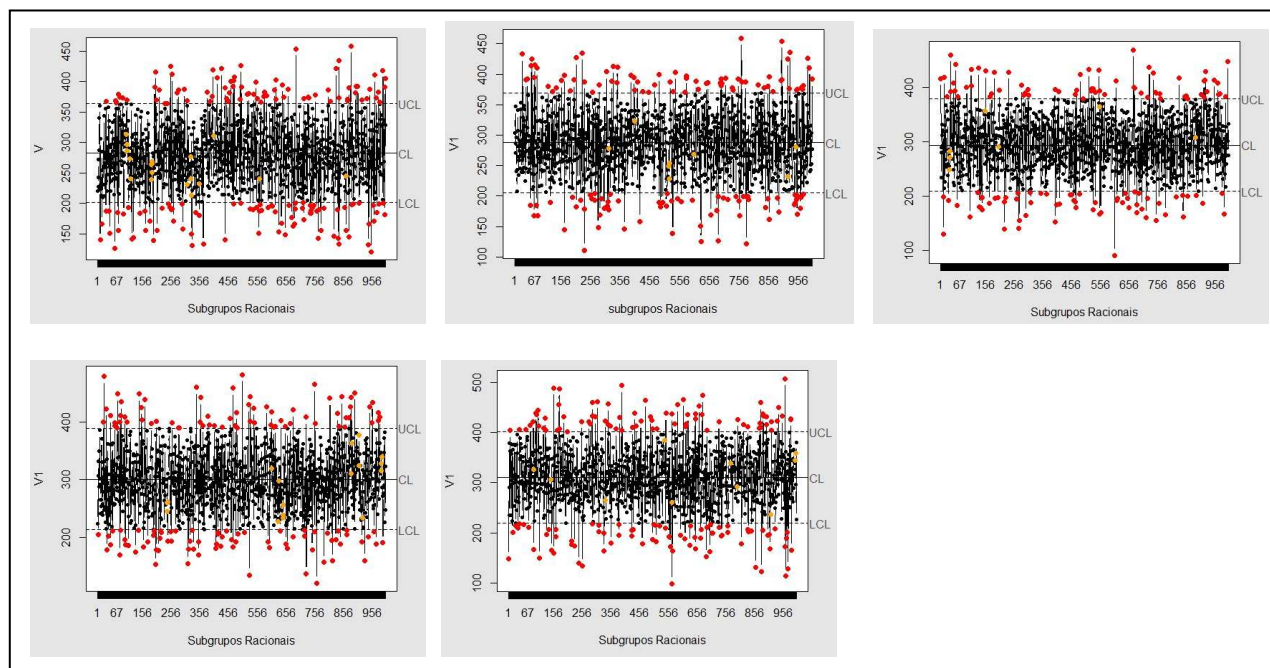


Figura 10 – Gráfico de Controle 1 – Processo com aumento da variabilidade, com correlações de 0,1, 0,3, 0,5, 0,7 e 0,9, da esquerda para direita e de cima para baixo, respectivamente.

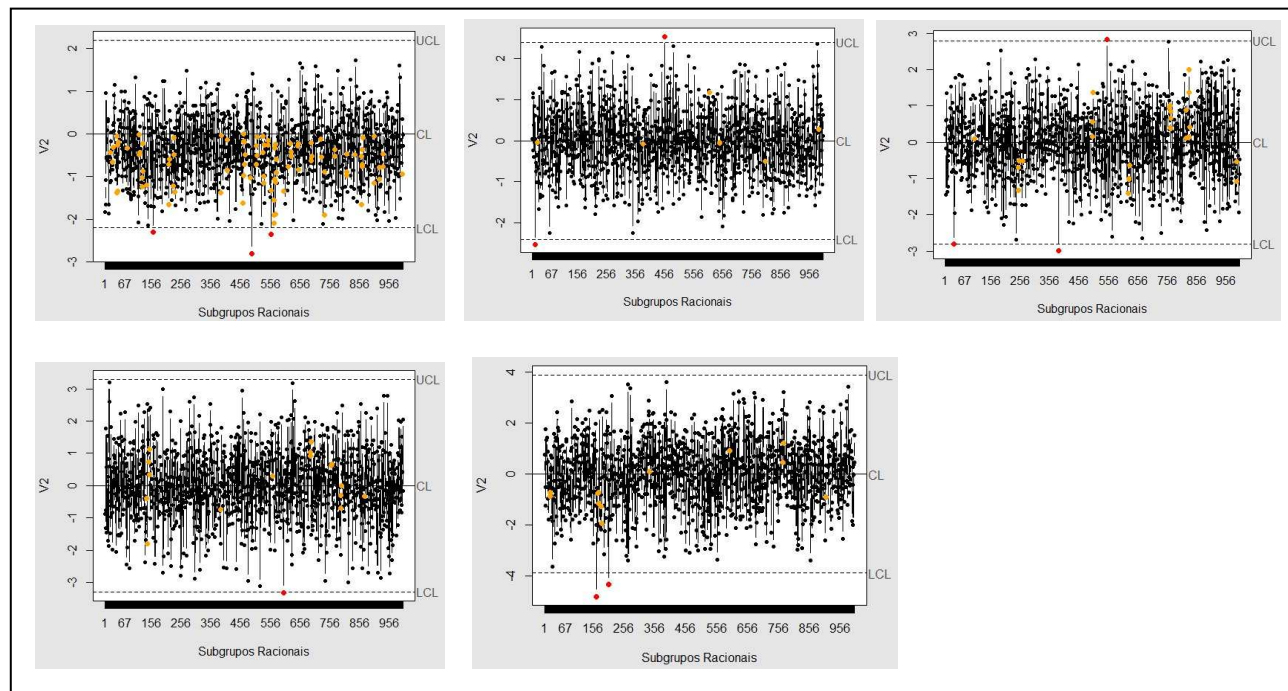


Figura 11 – Gráfico de Controle 2 – Processo com aumento da variabilidade, com correlações de 0,1, 0,3, 0,5, 0,7 e 0,9, da esquerda para direita e de cima para baixo, respectivamente.

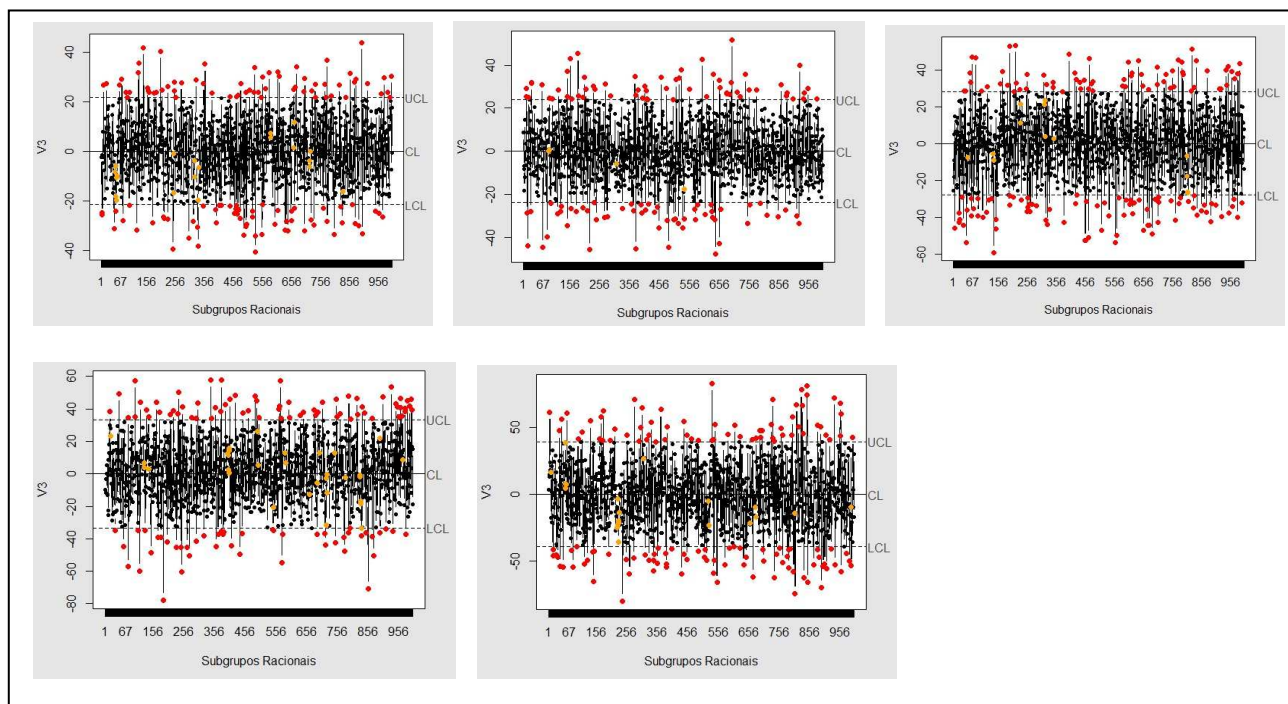


Figura 12 – Gráfico de Controle 3 – Processo com aumento da variabilidade, com correlações de 0,1, 0,3, 0,5, 0,7 e 0,9, da esquerda para direita e de cima para baixo, respectivamente.

Neste caso, assim como na proposta anterior, a medida empregada para avaliação do gráfico proposto foi a probabilidade do alarme verdadeiro, pois o objetivo também era apontar os pontos como fora de controle em função das causas especiais. É desejável a obtenção de um alto valor para o Pd dos três gráficos ($Pd \geq 0,90$).

Conforme apresentado na Tabela 17, os gráficos de controle GC_1 e GC_3 apresentaram baixos valores para o Poder, não dependendo das correlações existentes entre as variáveis-respostas Y_1 e Y_2 e o GC_2 apresentou poder igual a zero, pois os dados foram padronizados com média zero e variância um. Estes baixos valores indicaram a falta de detecção dos gráficos propostos (β), ou seja, eles não foram poderosos suficientes para apontarem como fora de controle a maioria dos pontos que realmente estava fora de controle.

Tabela 17 – Probabilidade do alarme verdadeiro observada em função das correlações existentes entre Y_1 e Y_2 em cada um dos gráficos propostos

ρ_{12}	GC_1	GC_2	GC_3
0,1	0,13	0	0,13
0,3	0,13	0	0,13
0,5	0,13	0	0,13
0,7	0,13	0	0,13
0,9	0,14	0	0,13
Média	0,13	0	0,13

Para o aumento da variabilidade, foi proposto um desvio-padrão equivalente ao dobro do desvio-padrão de controle, ou seja, um fator de aumento do desvio padrão de controle de 2 ($\gamma=2$). Comparando a proposta com os gráficos de controle de Shewhart, R e S, também observa-se uma considerável semelhança, pois nestes gráficos o Poder é baixo quando tem-se $n \leq 5$ e $1 < \gamma \leq 2$. Eles são eficientes somente para detectar causas especiais associadas ao aumento da variabilidade nas razões de $\gamma > 2$ e, mesmo assim, à custa de altos valores de n (RIBEIRO JÚNIOR, 2013).

5.2.3. Deslocamento da média e aumento da variabilidade

Conforme descrito nas seções 4.2.1, 4.2.2 e 4.2.3, as matrizes empregadas para a obtenção dos CPs, Σ (GC_1), ρ (GC_2) e Σ^* (GC_3) foram as mesmas para o processo sob controle e para os processos fora de controle. Portanto, em todos os cenários os autovalores e autovetores apresentaram-se iguais.

A partir dos autovalores e autovetores obtidos, que foram comuns aos processos sob e fora de controle, foram caracterizados os dois componentes principais, em termos de média

e desvio-padrão, para o processo em que ocorreu deslocamento da média e aumento na variabilidade simultaneamente (Tabela 18).

Tabela 18 – Médias e desvios-padrão dos componentes principais obtidos pelas matrizes Σ , ρ e Σ^*

ρ_{12}	Matriz Σ				Matriz ρ				Matriz Σ^*			
	μ_{CP1}	σ_{CP1}	μ_{CP2}	σ_{CP2}	μ_{CP1}	σ_{CP1}	μ_{CP2}	σ_{CP2}	μ_{CP1}	σ_{CP1}	μ_{CP2}	σ_{CP2}
0,1	333,73	60,04	97,57	19,89	0	1,05	0,00	0,95	0	19,06	0,00	17,26
0,3	340,09	60,34	72,88	18,95	0	1,14	0,00	0,84	0	20,73	0,00	15,20
0,5	344,36	60,91	49,37	17,06	0	1,22	0,00	0,71	0	22,25	0,00	12,87
0,7	346,67	61,67	27,84	13,89	0	1,31	0,00	0,55	0	23,70	0,00	9,96
0,9	347,77	62,69	8,65	8,33	0	1,38	0,00	0,32	0	25,03	0,00	5,75

Em função das médias e dos desvios-padrão obtidos para os dois componentes principais, foram obtidas as mesmas medidas para as variáveis V_1 , V_2 e V_3 plotadas nos gráficos de controle GC_1 , GC_2 e GC_3 , respectivamente (Tabela 19).

Tabela 19 – Médias e desvios-padrão das variáveis plotadas nos gráficos

ρ_{12}	GC ₁		GC ₂		GC ₃	
	μ_{V1}	σ_{V1}	μ_{V2}	σ_{V2}	μ_{V3}	σ_{V3}
0,1	310,5	54,2	0	0,7	0	13,0
0,3	316,1	55,0	0	0,8	0	14,5
0,5	322,8	56,5	0	0,9	0	17,0
0,7	331,4	58,7	0	1,1	0	20,2
0,9	341,8	61,7	0	1,3	0	23,8

De acordo com o processo sob controle estatístico, foram obtidos a linha média (LM) e os limites de controle inferior (LIC) e superior (LSC), para as construções dos gráficos de controle GC_1 , GC_2 e GC_3 , tanto para o processo sob controle como para os processos fora de controle estatístico. Portanto, a linha média e os limites de controle são iguais aos do processo sob controle.

A seguir, nas Figuras 13, 14 e 15, os gráficos de controle gerados no cenário onde o processo encontra-se fora de controle estatístico, com deslocamento da média de controle e aumento da variabilidade, simultaneamente.

Observou-se neste caso que o gráfico de controle 1 apontou o processo como fora de controle estatístico, pois vários pontos ficaram localizados além dos limites de controle. A maioria dos pontos que saíram de controle ficou acima do limite superior de controle indicando o deslocamento da média para cima, mas também tivemos a ocorrência de pontos

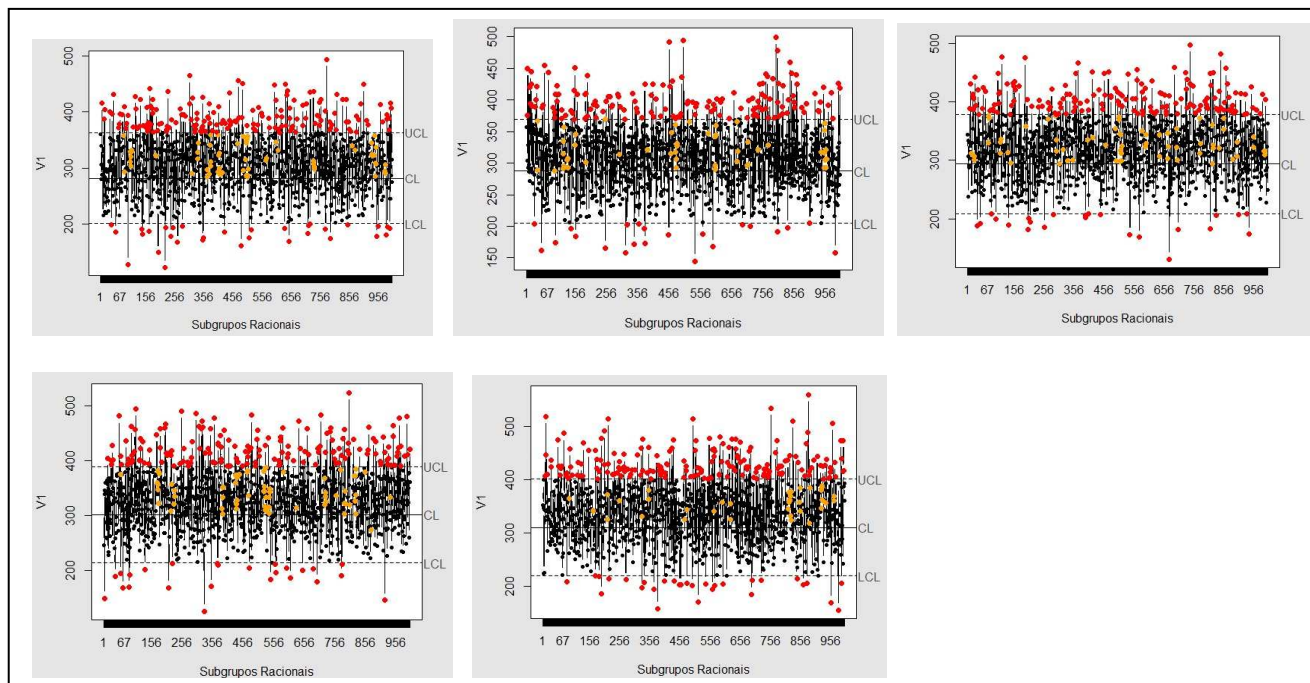


Figura 13 – Gráfico de Controle I – Processo com deslocamento da média e aumento da variabilidade, com correlações de 0,1, 0,3, 0,5, 0,7 e 0,9, da esquerda para direita e de cima para baixo, respectivamente.

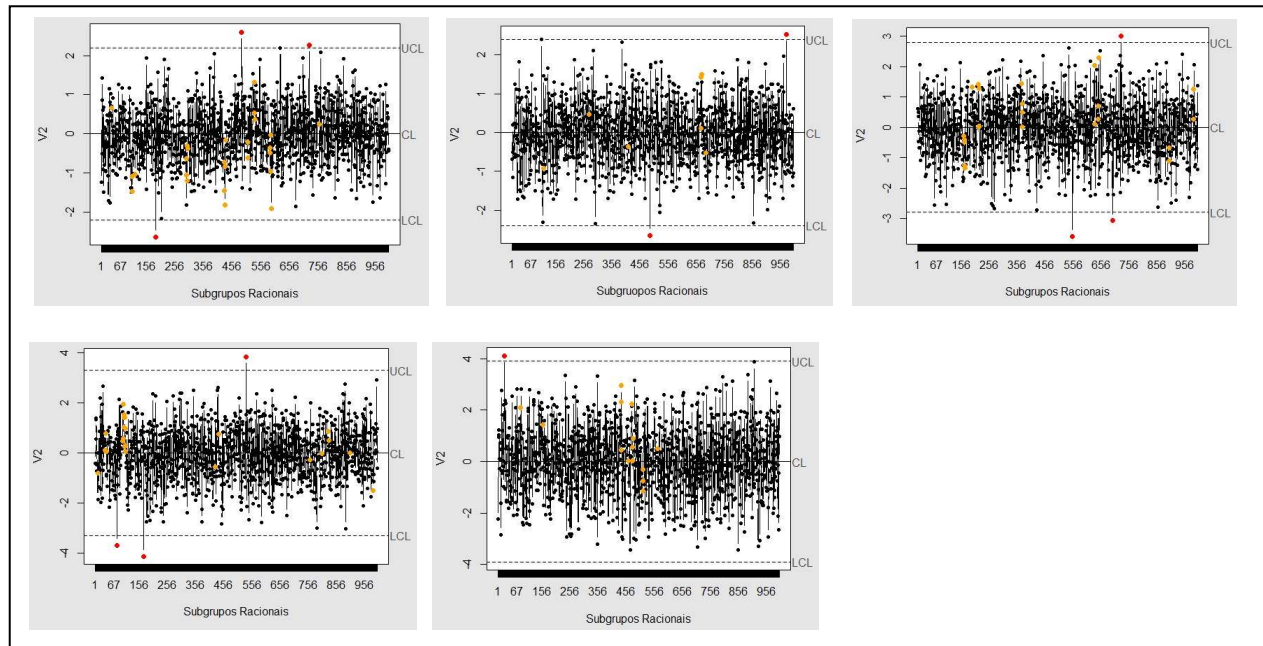


Figura 14 – Gráfico de Controle 2 – Processo com deslocamento da média e aumento da variabilidade, com correlações de 0,1, 0,3, 0,5, 0,7 e 0,9, da esquerda para direita e de cima para baixo, respectivamente.

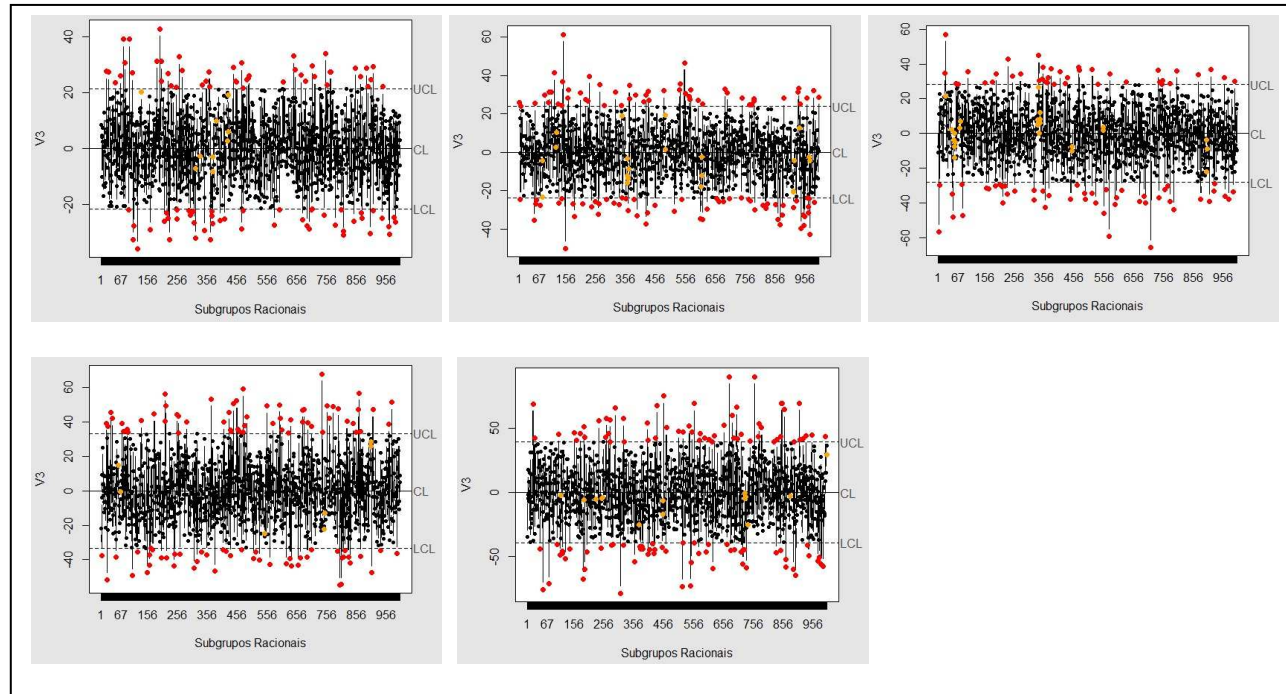


Figura 15 – Gráfico de Controle 3 – Processo com deslocamento da média e aumento da variabilidade, com correlações de 0,1, 0,3, 0,5, 0,7 e 0,9, da esquerda para direita e de cima para baixo, respectivamente.

abaixo do limite inferior de controle, indicando que a variabilidade aumentou, conforme esperado. Conclui-se, então, que a matriz Σ foi adequada para este cenário.

O gráfico de controle 2 apontou, erroneamente, o processo como sob controle estatístico, pois a maioria dos pontos ficaram localizados entre os limites de controle. Esta característica foi uma consequência da padronização dos dados, pois eles apresentaram média igual a zero e variância igual a um. Assim, a variável a ser plotada V_2 também apresentou média zero, conforme indicou o gráfico. Conclui-se, então, que a matriz ρ não foi adequada para este cenário.

No gráfico de controle GC_3 foi possível verificar a existência de aumento da variabilidade apenas, pois os dados foram transformados, ficando com média zero, daí a impossibilidade de detectar o deslocamento da média.

Conforme mostrado na Tabela 20, os gráficos de controle GC_1 e GC_3 apresentaram baixos valores para o Poder, não dependendo das correlações existentes entre as variáveis-respostas Y_1 e Y_2 e o gráfico de controle GC_2 apresentou poder igual a zero. O GC_1 apresentou o maior Poder porque detectou tanto o deslocamento da média de controle quanto o aumento na variabilidade, enquanto o Poder do GC_3 foi um pouco menor pelo fato deste detectar apenas aumento da variabilidade. Já o GC_2 teve poder igual a zero porque os dados foram padronizados com média zero e variância um.

Tabela 20 – Probabilidade do alarme verdadeiro observada em função das correlações existentes entre Y_1 e Y_2 em cada um dos gráficos propostos

ρ_{12}	GC_1	GC_2	GC_3
0,1	0,19	0,00	0,10
0,3	0,19	0,00	0,10
0,5	0,19	0,00	0,10
0,7	0,18	0,00	0,10
0,9	0,19	0,00	0,10
Média	0,19	0,00	0,10

6. CONCLUSÕES

A criação de uma nova variável que pondera os escores dos componentes principais pelos seus respectivos autovalores consegue incorporar, de forma univariada, todas as múltiplas características estudadas com as devidas correlações.

A construção do gráfico de controle por meio dos escores, ponderados pelos autovalores, possibilita a redução de uma análise multivariada para uma análise univariada.

As correlações entre as variáveis-resposta não altera o desempenho dos gráficos de controle propostos porque a incorporam em uma única variável.

Os efeitos dos deslocamentos da média e, ou, dos aumentos das variabilidades das variáveis-resposta são os mesmos daqueles provocados nas análises dos gráficos de controle univariados.

A matriz de variâncias e covariâncias pode ser empregada para o monitoramento da média e da variabilidade, enquanto a matriz de coeficientes de variação somente pode ser empregada para o monitoramento da variabilidade.

As probabilidades dos alarmes falsos nos três gráficos propostos foram baixas e não diferiram entre si, sendo semelhantes às apresentadas pelos gráficos de controle univariados de Shewhart.

As probabilidades dos alarmes verdadeiros foram baixas, com superioridade para o GC₁ (que empregou a matriz de variâncias e covariâncias) no cenário em que ocorreu deslocamento da média de controle. No cenário em que ocorreu apenas aumento da variabilidade as probabilidades foram exatamente iguais para os GC₁ (que empregou a matriz de variâncias e covariâncias) e GC₃ (que empregou a matriz de coeficientes de variação), e zero para o GC₂ (que empregou a matriz de correlações). Já no cenário em que ocorreram ambas alterações, a probabilidade do alarme verdadeiro foi superior para o GC₁ quando comparada com o GC₃ e zero no GC₂.

REFERÊNCIAS BIBLIOGRÁFICAS

AGUIAR, S. **Integração das Ferramentas da Qualidade ao PDCA e ao Programa Seis Sigma**. Belo Horizonte: Desenvolvimento Gerencial, 2002.

ALVES, C. C. **O método de equação integral com quadratura gaussiana para otimizar os parâmetros do gráfico de controle multivariado de somas acumuladas**. 2009. 196 f. Tese (Doutorado em Engenharia de Produção) – Centro Tecnológico, Universidade Federal de Santa Catarina, Universidade Federal de Santa Catarina, Florianópolis, 2009.

BARBOSA, L.; LOPES, P. S.; REGASSI, A. J.; GUIMARÃES, S. E. F.; TORRES, R. A. de. Seleção de variáveis de desempenho de suínos por meio da análise de componentes principais. **Arquivo Brasileiro de Medicina Veterinária e Zootecnia**, Belo Horizonte, v. 57, n. 6, p. 805-810, 2005.

BUSTOS, O.H.; ORGAMBIDE, A.C.F. Simulação estocástica: teoria e algoritmos. In: SIMPÓSIO NACIONAL DE PROBABILIDADE E ESTATÍSTICA, 10., 1992, São Paulo, **Anais ...** São Paulo: 1992. 152 p.

CAMPANA, A. C. M.; RIBEIRO JÚNIOR, J. I.; NASCIMENTO, M. Uma proposta de transformação de dados para análise de componentes principais. **Revista Brasileira de Biometria**, v. 28, p. 1-15, 2010.

CARVALHO, J. L. A. **Reengenharia de processos na indústria farmacêutica**. Lisboa, 2010. 231 f. Dissertação (Mestrado em Engenharia e Gestão Industrial) – Programa de Pós-Graduação em Engenharia e Gestão Industrial, Universidade Nova de Lisboa, 2010.

COSTA, A. F. B.; EPPRECHT, E. K.; CARPINETTI, L. C. R. **Controle estatístico de qualidade**. São Paulo: Atlas, 2004. 334 p.

FERREIRA, D. F. **Estatística multivariada**. 2. ed. rev. e ampl. – Lavras: Ed. UFLA, 2011. 676 p.

FILHO, D. M. **Monitoramento de processos em batelada através de cartas de controle multivariadas baseadas na análise de componentes principais multidirecionais (ACPM)**. 2001. 204 f. Dissertação (Mestrado em Engenharia de Produção), Universidade Federal do Rio Grande do Sul, Porto alegre, 2001.

HAIR, J. F.; ANDERSON, R. E.; TATHAM, R. L.; BLACK, W. **Análise multivariada de dados**. 5. ed. Porto Alegre: Bookman, 2005. 600 p.

HENNING, E.; ALVES, C. C.; ZVIRTES, L.; ARAÚJO, N. G. A. Um estudo comparativo entre o desempenho de gráficos de controle estatístico multivariados com a aplicação da análise de componentes principais. In: SIMPÓSIO BRASILEIRO DE PESQUISA OPERACIONAL, 42., 2010. **Anais...** Bento Gonçalves, 2010.

HENNING, E.; ARAÚJO, N. G. A.; ALVES, C. C.; ZVIRTES, L. Aplicação de gráficos de controle multivariados baseados na projeção de dados via análise de componentes principais. **Produção em Foco**, v. 1, n. 1, p. 41-66 jan./jun. 2011.

HOTELLING, H. Multivariate quality control-illustrated by the air testing of sample bombsights. In: **Techniques of statistical analysis**. New York: McGraw Hill, 1947. p. 111-184.

JACKSON, J. E. **A user guide to principal componentes**. New York: Wiley. 1991.

JALALI, S. M. J.; SHAFIEEZADEH, M.; NAIINI, A. S. Using knowledge management in DMAIC methodology of six sigma projects of the 2008. In: PROCEEDINGS Of The 2008 1st INTERNATIONAL CONFERENCE ON INFORMATION TECHNOLOGY, IT. Gdask, Polônia, 2008.

JOHNSON, R. A.; WICHERN, D. W. **Applied multivariate statistical analysis**. 5. ed. New Jersey: Prentice Hall, 2002. 767 p.

LOWRY, A. C.; MONTGOMERY, D. C. A review of multivariate control charts. **IIE Transactions**, v. 27, p. 800-810, 1995.

MACHADO, M. A. G. **Gráficos de controle para o monitoramento de processos multivariados**. 2009. 158 f. Tese (Doutorado em Engenharia Mecânica) – Faculdade de Engenharia do Campus de Guaratinguetá, Universidade Estadual Paulista, Guaratinguetá, 2009.

MAGALHÃES, R. R.; CYMROT, R. Técnicas alternativas de monitoramento e controle estatístico de processos aos gráficos de controle de Shewhart. In: JORNADA DE INICIAÇÃO CIENTÍFICA, 3., Mackenzie, 2007.

MARDIA, K.V.; KENT, J.T.; BIBBY, J.M. **Multivariate analysis**. 6. ed. Londres: Academic Press, 1997. 518 p.

MONTGOMERY, D. C. **Introdução ao controle estatístico da qualidade**. 4. ed. Rio de Janeiro: LTC, 2004. 513 p.

MONTGOMERY, D. C.; RUNGER, G. C. **Estatística aplicada e probabilidade para engenheiros**. 4. ed. Rio de Janeiro: LTC Editora, 2009. 493p.

PEÑA, D. **Análisis de datos multivariantes**. España: McGraw-Hill Interamericana de España, 2002. 539 p.

PHAM, H. **Springer handbook of engineering statistics**. New Jersey: Rutgers the State University of New Jersey, 2006.

PRADO, D. **Teoria das filas e da simulação**. Série Pesquisa Operacional. Vol. 2. Belo Horizonte, MG: Desenvolvimento Gerencial, 1999. 122 p.

PYSDEK, T.; KELLER, P. A. **O seis sigma: guia do profissional, um guia complete para Green Belts, Black Belts e gerentes em todos os níveis**. 3. ed. Rio de Janeiro, RJ: Alta Books, 2011. 560 p.

RIBEIRO JÚNIOR, J. I. **Métodos estatísticos aplicados à melhoria da qualidade**. Viçosa: UFV, 2012. 385 p.

RIBEIRO JÚNIOR, J. I. **Métodos estatísticos aplicados ao controle da qualidade**. Viçosa: UFV, 2013. 274 p.

SOUZA, L. M.; RIBEIRO JÚNIOR, J. I.; REIS, G. M.; M. S. Eficiência dos gráficos de controle Xbarra, EWMA e CUSUM. **Revista Eletrônica Produção & Engenharia**, v.1, n.1, p. 81-94. 2008.

ANEXO A – ROTINA DO R UTILIZADA PARA ANALISAR O CENÁRIO ONDE O PROCESSO ESTAVA SOB CONTROLE ESTATÍSTICO

```
# Carregando os pacotes empregados (qcc e MASS)
library(qcc)
library(MASS)
#Definindo o vetor de médias que chamamos de Mu
Mu=c(a,b) #a=média de Y1 e b=média de Y2
#Definindo a matriz utilizada para simulação dos dados que chamamos de Matriz
Matriz=cbind(c(d,e),c(e,f)) #d=variância de Y1; e=covariância entre Y1 e Y2 e f=variância
de Y2
#Simulando os dados que chamamos de Dados
Dados=mvrnorm(n=1000,mu=Mu,Sigma=Matriz)
#Definindo a matriz utilizada para obtenção dos autovalores e autovetores que chamamos de
S
S=matrix(c(g,h,i,j),2,2) #elementos das matrizes
#Obtenção dos autovalores e autovetores normalizados que chamamos de autos
autos=eigen(S)
#Obtenção da porcentagem de explicação de cada componente principal
percent_expli=autos$values/sum(diag(S))
#Obtenção dos escores
dad=as.matrix(Dados)
escores=dad%*%autos$vector
#Obtenção dos escores com sinal positivo que chamamos de positivos
positivos=escores*(-1)
#Obtenção dos 1.000 valores da variável a ser plotada denominada VP
CP1=positivos[,-2]
CP2=positivos[,-1]
VP=((auto1*CP1)+(auto2*CP2))/(auto1+auto2) #Onde auto1=autovalor 1 e auto2=
autovalor 2
#Obtenção da linha média do gráfico de controle (média dos valores plotados) denominada
LM
LM=(1/(auto1+auto2))*((auto1*mean(CP1)+auto2*mean(CP2)))
#Obtenção da variância dos valores plotados para obtenção dos limites de controle
denominada VAR
```

$$\text{VAR} = ((\text{auto1}^2) * \text{var}(\text{CP1}) + (\text{auto2}^2) * \text{var}(\text{CP2}) + 2 * \text{auto1} * \text{auto2} * \text{cov}(\text{CP1}, \text{CP2})) / (\text{auto1}^2 + 2 * \text{auto1} * \text{auto2} + \text{auto2}^2)$$

#Obtenção do limite superior de controle denominado LSC

$$\text{LSC} = \text{LM} + 3 * \text{sqrt}(\text{VAR})$$

#Obtenção do limite inferior de controle denominado LIC

$$\text{LIC} = \text{LM} - 3 * \text{sqrt}(\text{VAR})$$

#Obtenção do alarme falso denominado alfa

$$\text{alfa} = (\text{sum}(\text{CPs} > \text{LSC}) + \text{sum}(\text{CPs} < \text{LIC})) / 1000$$

#Obtenção do gráfico de controle

`qcc(VP,type="xbar.one",nsigmas=3,center=LM,limits=c(LIC,LSC),xlab="Subgrupos`

`Racionais",ylab="k")` #Onde $k = V_1$ ou V_2 ou V_3

ANEXO B – ROTINA DO R UTILIZADA PARA ANALISAR O CENÁRIO ONDE O PROCESSO ESTAVA FORA DE CONTROLE ESTATÍSTICO

```
# Carregando os pacotes empregados (qcc e MASS)
library(qcc)
library(MASS)
#Definindo o vetor de médias que chamamos de Mu
Mu=c(a,b)      #a=média de Y1 e b=média de Y2
#Definindo a matriz utilizada para simulação dos dados que chamamos de Matriz
Matriz=cbind(c(d,e),c(e,f)) #d=variância de Y1; e=covariância entre Y1 e Y2 e f=variância
de Y2
#Simulando os dados que chamamos de Dados
Dados=mvrnorm(n=1000,mu=Mu,Sigma=Matriz)
#Definindo a matriz utilizada para obtenção dos autovalores e autovetores que chamamos de
S
S=matrix(c(g,h,i,j),2,2) #elementos das matrizes
#Obtenção dos autovalores e autovetores normalizados que chamamos de autos
autos=eigen(S)
#Obtenção da porcentagem de explicação de cada componente principal
percent_expli=autos$values/sum(diag(S))
#Obtenção dos escores
dad=as.matrix(Dados)
escores=dad%*%autos$vector
#Obtenção dos escores com sinal positivo que chamamos de positivos
positivos=escores*(-1)
#Obtenção dos 1.000 valores da variável a ser plotada denominada VP
CP1=positivos[,-2]
CP2=positivos[,-1]
VP=((auto1*CP1)+(auto2*CP2))/(auto1+auto2) #Onde auto1=autovalor 1 e auto2=
autovalor 2
#Obtenção da linha média do gráfico de controle denominada LM
LM=l          #l=linha média obtida no cenário em que o processo estava sob
controle
#Obtenção do limite superior de controle denominado LSC
LSC=m        #m=LSC obtido no cenário em que o processo estava sob controle
#Obtenção do limite inferior de controle denominado LIC
LIC=n        #n=LIC obtido no cenário em que o processo estava sob controle
```

#Obtenção do alarme verdadeiro denominado poder

poder=(sum(CPs>LSC)+sum(CPs<LIC))/1000

#Obtenção do gráfico de controle

qcc(VP,type="xbar.one",nsigmas=3,center=LM,limits=c(LIC,LSC),xlab="Subgrupos Racionais",ylab="k") #Onde k=V₁ ou V₂ ou V₃.