Instituto de MATEMÁTICA E ESTATÍSTICA UFRGS

DEPARTAMENTO DE ESTATÍSTICA
E-mail: dest@mat.ufrgs.br

Trabalho de Conclusão de Curso

Estudo Comparativo de Abordagens Clássicas de Controle Estatístico Multivariado

Eduardo de Oliveira Correa

10 de janeiro de 2020

Eduardo de Oliveira Correa

Estudo Comparativo de Abordagens Clássicas de Controle Estatístico Multivariado

Trabalho de Conclusão apresentado à comissão de Graduação do Departamento de Estatística da Universidade Federal do Rio Grande do Sul, como parte dos requisitos para obtenção do título de Bacharel em Estatística.

Orientador: Prof. Dr. Danilo Marcondes Filho

Porto Alegre
Dezembro de 2019

Eduardo de Oliveira Correa

Estudo Comparativo de Abordagens Clássicas de Controle Estatístico Multivariado

Este Trabalho foi julgado adequado para obtenção dos créditos da disciplina Trabalho de Conclusão de Curso em Estatística e aprovado em sua forma final pela Orientador e pela Banca Examinadora.

Orientador:__________________________________
Prof. Dr. Danilo Marcondes Filho, UFRGS
Doutor pela Universidade Federal do Rio Grande do Sul, Porto Alegre, RS

Banca Examinadora:

Prof. Drª. Liane Werner, UFRGS
Doutora pela Universidade Federal do Rio Grande do Sul – Porto Alegre, RS

Porto Alegre
Dezembro de 2019

Resumo

Processos industriais geram inúmeras variáveis de interesse correlacionadas. Cartas de Controle Multivariadas (CCMs) têm sido amplamente utilizadas nessas situações, pois incorporam a estrutura de correlação das variáveis no monitoramento de tais processos. As CCMs tradicionais Qui-Quadrado (X^2) e a da Variância Generalizada (W) são utilizadas para o monitoramento simultâneo do vetor de médias e da matriz das covariâncias, respectivamente. Outra abordagem clássica utiliza CCMs baseadas na Análise de Componentes Principais (ACP). A ACP destaca-se por conseguir reduzir o número de variáveis do processo, preservando grande parte da informação contida das variáveis de interesse. No contexto do Controle Estatístico de Processo, a ACP é eficaz para o entendimento das fontes de variabilidade do processo e na identificação da natureza de eventuais distúrbios no processo.

A partir de um processo simulado com 4 variáveis e uma estrutura de covariância, a sensibilidade das duas abordagens será analisada diante de diferentes descontroles impostos no vetor de médias e na matriz de covariâncias do processo sob controle.

Os resultados mostram que para alterações no vetor de médias, tanto nas direções comuns de variabilidade, quanto nas direções opostas à variabilidade comum dos dados, evidenciamos uma maior sensibilidade das CCMs via ACP em relação às tradicionais. A carta W é mais sensíveis às alterações na estrutura de correlação das variáveis em relação às CCMs via ACP.

Palavras-Chave: Cartas de Controle Multivariadas, Análise de Componentes Principais, Carta de Controle Qui-Quadrado, Carta de Controle da Variância Generalizada.

Abstract

Industrial processes generate numerous correlated variables of interest. Multivariate Control Charts (CCMs) have been widely used in these situations as they incorporate the correlation structure of the variables in the monitoring of such processes. Traditional Chi-Square MCCs (X^2) and Generalized Variance (textit W) are used for simultaneous monitoring of the mean vector and covariance matrix, respectively. Another classic approach uses CCMs based on Principal Component Analysis (ACP). ACP stands out for being able to reduce the number of process variables, preserving much of the information contained in the variables of interest. In the context of Statistical Process Control, PCA is effective in understanding the sources of process variability and in identifying the nature of potential process disturbances.

From a simulated process with 4 variables and a covariance structure, the sensitivity of the two approaches will be analyzed in view of different impositions imposed on the mean vector and covariance matrix of the process under control.

The results show that for changes in the mean vector, both in the common directions of variability and in the opposite directions to the common variability of the data, we show a higher sensitivity of CCMs via ACP compared to traditional ones. The W chart is more sensitive to changes in the correlation structure of variables than CCMs via ACP.

Keywords: Multivariate Control Charts, Principal Components Analysis, Chi-Square Control Chart, Generalized Variance Control Chart.

Lista de Tabelas

Sumário

1 Introdução

A melhoria na qualidade de um processo produtivo tem se tornado uma importante estratégia de negócio. O crescente desenvolvimento tecnológico leva cada vez mais ao aumento nas exigências do consumidor. Para se firmarem em um mercado competitivo, as indústrias buscam formas de manter um alto padrão na qualidade de seus produtos, satisfazendo as necessidades dos clientes, melhorando a performance na produção do processo e diminuindo os custos. Processos industrias geram uma gama de dados sobre o seu desempenho. Medições em tempo real são feitas sobre as características que influenciam a qualidade do processo. Podemos utilizar técnicas procedentes do Controle Estatístico de Processo (CEP) para monitorar o desempenho desses processos.

De acordo com (Louzada, 2009), CEP é um conjunto de métodos utilizados para planejar, monitorar e aprimorar um processo produtivo. Em geral, o interesse inicial do CEP está ligado ao monitoralmento da variabilidade. (Montgomery, 2007) define variabilidade e qualidade como conceitos antagônicos; sendo a qualidade inversamente proporcional à variabilidade.

Dentre as sete ferramentas desenvolvidas para Controle Estatístico de Processo, destaca-se a Carta de Controle (CC), sendo a mais conhecida e utilizada no meio industrial. A CC é empregada para acompanhar se o processo está operando sobre controle estatístico, isto é, sem causas especiais. Causas especiais são variações inesperadas, anômalas ao processo, normalmente devido a falha individual, por exemplo, imprecisão no ajuste da máquina. Mesmo em um processo controlado espera-se encontrar alguma variabilidade natural. Essas variabilidades inerantes ao processo são denominadas causas comuns. As CCs permitem uma redução na variabilidade das características de interesse do processo.

Para construção de forma mais usual da CC deve-se registrar uma ou mais características obtidas através de medições em fases apropriadas do processo, dispor esses valores em ordem cronológica em um gráfico que possui duas linhas denominadas de limites de controle. Valores fora dos limites de controle podem dar indícios de que o processo está atuando fora do controle estatístico, essas observações devem ser analisadas e verificar o tipo de causa que está atuando sobre elas. Caso os valores estejam dentro dos limites de controle e não apresentem comportamento sistêmico, considera-se não haver evidências de que o processo esteja atuando fora do controle.

Na maioria dos cenários de monitoramento dos processos, as variáveis apresentam correlação. Então, ao aplicar as CCs univariadas, está se assumindo a hipótese de independência, quando na verdade não há, podendo gerar conclusões equivocadas e ineficiência no processo produtivo. As Cartas de Controle Multivariadas (CCMs) são extensões das CCs univariadas para o caso onde duas ou mais variáveis são monitoradas simultaneamente. Destacam-se as cartas X^2 e W para monitoramento do vetor de médias e da covariância das variáveis envolvidas, respectivamente. As CCMs têm sido amplamente utilizadas pelas indústrias, pois incorporam a estrutura de correlação das variáveis no monitoramento do processo.

Segundo (Montgomery, 2007), as CCMs foram concebidas para monitorar um número limitado de variáveis (definida por ele como não maior que 10) de qualidade do produto final. Uma estrutura com uma quantidade vasta de variáveis favorece o surgimento de colinearidade, em outras palavras, variáveis que apresentam, praticamente, informações duplicadas contidas nas demais, trazendo pouca informação. A estrutura colinear gerada a partir dessas redundâncias faz com que a estatística forneça sinalizações distorcidas acerca do estado do processo, bem como diagnósticos imprecisos de suas causas especiais. Para eliminar estas redundâncias, pode-se utilizar CCs baseadas em Análise de Componentes Principais (ACP).

A ACP tem por objetivo a redução da dimensionalidade de um conjunto de dados multivariados, através de combinações lineares independentes das variáveis originais, mantendo parte substancial da informação contida nos mesmos.

As CCMs utilizando ACP adotam método semelhante das CCMs tradicionais utilizando, em vez das variáveis do processo, os Componentes Principais (CPs) derivados delas. Dessa forma, simplifica-se o monitoramento do processo, uma vez que as informações contidas nas variáveis são projetas em um espaço de dimensões menores, facilitando a interpretação e utilização das CCs.

Este trabalho apresenta um estudo simulado para comparar o desempenho das CCM (X^2) e (W) e das CCMs basedas na ACP, a partir de diversos cenários com dados multivariados simulados.

O presente estudo encontra-se reunido em 5 capítulos. No capítulo 2 será realizada uma revisão bibliográfica abordando as Cartas de Controle Multivariadas utilizadas no estudo e o método de Análise de Componentes Principais. O capítulo 3 apresenta a metodologia empregada no estudo. Para o capítulo 4 são aplicadas as CCMs tradicionais e via ACP, expondo os resultados e as considerações dos mesmos. Por fim, o capítulo 5 apresenta a conclusão do estudo.

2 Referencial Bibliográfico

Esta seção apresenta o referencial teórico que será empregado nos procedimentos metodológicos adotados neste trabalho. Serão descritas inicialmente as Cartas de Controle Multivariadas clássicas para a Média e Variância. A seguir, a técnica multivariada da Análise de Componentes Principais (ACP) será apresentada. Na parte final da seção serão descritas as Cartas de Controle clássicas derivadas da ACP.

2.1 Cartas de Controle Univariadas

As origens modernas das Cartas de Controle começam por Walter Shewhart na década de 1920, com a publicação de seu artigo *Quality Control Chart* para o jornal *Bell System Technical*.

Carta de Controle Univariada é uma representação gráfica de medições de uma caracaterística de qualidade (variável de interesse) através de sucessivas amostras espaçadas ao longo do tempo. O gráfico possui três linhas paralelas ao eixo das abscissas: O limite superior de controle (LSC); limite inferior de controle (LIC); e uma linha central (LC), representando o valor esperado da variável [(Montgomery, 2007)]. Para o estado do processo ser dito sob controle estatístico, os pontos amostrais, de acordo com o nível de significância pré-estabelacido, devem encontrar-se entre os limites de controle, com comportamento aleatório ou não sistemático.

A construção das CCs divide-se em duas fases com objetivos distintos: a fase I envolve a elaboração da CC. Um conjunto de dados é coletado, realiza-se uma análise retrospectiva para a construção dos limites de controle. O objetivo da fase I é colocar o processo em estado sob controle; a fase II consiste no monitoramento do processo já sob condições estáveis do processo. Um novo conjunto de dados é monitorado a partir dos limites atuantes da fase I.

Após o desenvolvimento por Shewhart, as CCs têm sido empregadas com diferentes modificações desde então [(Jackson, 2005)].

2.2 Cartas de Controle Multivariadas

As primeiras publicações na perspectiva multivariada foram feitas por Harold Hotelling (1947), utilizando abordagem multivariada em dados contendo informações sobre bombardeios durante a Segunda Guerra Mundial. Seu uso foi difundido, pois observou-se que a maioria das variáveis de monitoramento do processo não se comportam de forma independente, portanto, não sendo adequado tratá-las utilizando Cartas de Controle Univariadas. Uma vez que as variáveis possuem correlação umas com as outras, elas devem ser consideradas em conjunto e não separadamente [(Mason e Young, 2002)].

2.2.1 Carta de Controle Multivariada para Média

A CCM Qui-Quadrado χ^2 é a extensão multivariada da CC univariada de Shewhart. Consiste no procedimento empregado para monitorar o vetor de médias. Sua grande vantagem é permitir que toda a informação presente nas amostras disponíveis seja resumida em uma única variável estatística, que segue uma distribuição conhecida.

Considera-se p-características correlacionadas medidas simultaneamente compondo amostras p-variadas de tamanho n. Dado que o processo opera dentro do seu padrão usual, no contexto estatístico, representamos essas características por um vetor aleatório $\boldsymbol{X} = [X_1, X_2, \ldots, X_p]'$, possuindo distribuição Normal *p-variada* com vetor de médias $\boldsymbol{\mu}$ e matriz de covariância $\boldsymbol{\Sigma}$, de dimensão $p \times p$, mais explicitamente dados por:

$$\boldsymbol{\mu} \;=\; \begin{bmatrix} \mu_1 \\ \mu_2 \\ \ldots \\ \mu_p \end{bmatrix} \qquad \boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \sigma_{12}^2 & \ldots & \sigma_{1p}^2 \\ \sigma_{21}^2 & \sigma_2^2 & \ldots & \sigma_{2p}^2 \\ \ldots & \ldots & \ldots & \ldots \\ \sigma_{p1}^2 & \sigma_{p2}^2 & \ldots & \sigma_p^2 \end{bmatrix}$$

onde μ_i e σ_i^2 para $i, j = 1, ..., p.$ representam a média e a variância da *i-ésima* variável aleatória, respectivamente.

A matriz $\boldsymbol{\Sigma}$ traz nas suas entradas as covariâncias das variáveis, duas a duas, do vetor aleatório $\boldsymbol{X}$; sendo a diagonal principal composta pelas variâncias.

Suponha que o processo é monitorado através de amostras aleatórias recolhidas periodicamente. Consideraremos aqui amostras de tamanho unitário. Estas amostras se constituem em realizações do vetor aleatório $\boldsymbol{X}$, e que serão denotadas por $\boldsymbol{x}$. Observe que, dado processo sob controle, estas amostras seguem a distribuição do vetor aleatório $\boldsymbol{X}$, ou seja, seguem a distribuição normal *p-variada* com média $\boldsymbol{\mu}$ e matriz de covariância $\boldsymbol{\Sigma}$.

Dado que os verdadeiros parâmetros do processo são conhecidos, utiliza-se a Carta de Controle *Qui-Quadrado* para monitorar o vetor de médias do processo. A cada nova amostra $\boldsymbol{x}$ o monitoramento do processo é realizado pela seguinte estatística:

$$\chi^2 = (\boldsymbol{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \boldsymbol{\mu}) \tag{2.1}$$

onde χ^2 segue uma distribuição *Qui-Quadrado* com p graus de liberdade e representa a distância quadrada padronizada, p-dimensional, entre um vetor de observações $\boldsymbol{x}$ e o vetor de médias do processo $\boldsymbol{\mu}$. Os limites de controle são obtidos da seguinte maneira:

$$\text{LSC} = \chi^2_{p,\alpha} \quad \text{e}$$

$$\text{LIC} = 0$$

onde $\chi^2_{p,\alpha}$ representa o percentil da distribuição Qui-Quadrado obtido a partir da probabilidade α adotada.

Valores que ultrapassem $\chi^2_{p,\alpha}$ indicam que a amostra das características de interesse $\boldsymbol{x}$ difere significamente dos seus valores esperados $\boldsymbol{\mu}$. Nesse caso, há evidências que a média μ_i de pelo menos uma das características tenha se alterado.

Na prática, é necessário estimar $\boldsymbol{\mu}$ e $\boldsymbol{\Sigma}$ a partir de amostras preliminares, tomadas quando se assume que o processo está sob controle. Temos que $\overline{\boldsymbol{x}}$ e $\mathbf{S}$ representam, respectivamente, as estimativas para o vetor de médias $\boldsymbol{\mu}$ e a matriz de covariância $\boldsymbol{\Sigma}$ do processo. Dessa forma, substituímos na equação (2.1) $\boldsymbol{\mu}$ e $\boldsymbol{\Sigma}$ por $\overline{\boldsymbol{x}}$ e $\mathbf{S}$, nesta ordem. Neste caso, a expressão (2.1) é conhecida como estatística de Hotelling [(Montgomery, 2007)].

2.2.2 Carta de Controle Multivariada para Variância

Ao monitorar o vetor de médias do processo, os valores obtidos através da estatística de χ^2 podem sofrer alterações caso a variabilidade do processo seja modificada. Em geral, o interesse inicial ao empregar CCs é o de verificar se a variabilidade do conjunto de medições associadas ao processo está sob controle estatístico, isto é, se a variabilidade das medidas permanece constante ao longo do período observado. Somente após a verificação do controle da variabilidade faz sentido examinar se a média do processo está atuando sob controle estatístico. (Montgomery, 2007) discute um método apropriado para o monitoramento simultâneo da variabilidade do processo. Consiste na construção da Carta de Controle da Variância Generalizada (Carta W), uma extensão multivariada da Carta de Controle Univariada para a variância (Carta S^2).

Considere novamente um conjunto de observações $\mathbf{x}$ do vetor aleatório $\mathbf{X}$ que possui distribuição normal p-variada com vetor de médias $\boldsymbol{\mu}$ e matriz de covariâncias $\boldsymbol{\Sigma}$. Assumimos agora que grupos de m vetores de observação $\mathbf{x}$ são recolhidos do processo com determinada frequência. A Carta de Controle da *Variância Generalizada* é empregada para monitorar a estrutura de covariância em $\boldsymbol{\Sigma}$ através das matrizes de covariâncias amostrais $\mathbf{S}$, obtidas de cada grupo de amostras. (Montgomery, 2007) descreve a estatística para monitorar a estrutura de covariâncias de futuras amostras baseada no determinante de matrizes de covariância, como se segue:

$$W_i = -pm + pmln(m) - mln(|\boldsymbol{A}_i|/|\boldsymbol{\Sigma}|) + tr(\boldsymbol{\Sigma}^{-\mathbf{1}}\boldsymbol{A}_i) \tag{2.2}$$

onde W segue uma distribuição *Qui-Quadrado* com $p(p+1)/2$ graus de liberdade, $\boldsymbol{\Sigma}$ é a matriz de covariância populacional, $\boldsymbol{A}_i = (m-1)\boldsymbol{S}$, $\boldsymbol{S}$ é a matriz de covariância do grupo de m observações $\boldsymbol{x}$,tr é o operador de traço e ln significa logaritmo natural. Os limites de controle são dados por:

$$\text{LSC} = \chi^2_{p(p+1)/2,\alpha} \quad \text{e}$$

$$\text{LIC} = 0$$

onde $\chi^2_{p(p+1)/2,\alpha}$ representa o percentil da distribuição Qui-Quadrado obtido a partir da probabilidade α adotada.

Valores de W que ultrapassam o LSC indicam que a estrutura de covariância presente nas amostras desse grupo, difere significativamente da estrutura presente em $\mathbf{\Sigma}$. Nesse caso, há evidências que pelo menos uma variância ou covariância de $\mathbf{\Sigma}$ tenha se alterado. Assim como na Carta χ^2, na pratica, é necessário estimar $\mathbf{\Sigma}$ a partir de amostras preliminares, tomadas quando se assume que o processo está sob controle. Dessa forma, substituímos na equação (2.2) $\mathbf{\Sigma}$ por $\boldsymbol{S_0}$, esta última, obtida a partir de um conjunto de vetores amostrais de $\boldsymbol{x}$.

2.3 Análise de Componentes Principais

Análise de Componentes Principais (ACP) foi proposta inicialmente por Karl Pearson (1901) baseada na teoria da álgebra linear. É uma técnica estatística de análise multivariada que visa explicar a estrutura da matriz de variância-covariância, com objetivo principal de reduzir um determinado conjunto de dados através da combinação linear de todas as p variáveis originais correlacionadas em k novas variáveis não-correlacionadas, sendo $k < p$.

Mediante as combinações lineares, novas coordenadas são obtidas por intermédio de rotação no sistema original $X_1, X_2, \ldots, X_p$ como eixo das coordenadas. Os novos eixos representam as direções de máxima variabilidade e fornecem uma descrição parcimoniosa da estrutura de covariância ou correlação. Não se faz necessário assumir na ACP a suposição de normalidade, mesmo sendo uma técnica derivada de uma população normal multivariada.

Considere $\boldsymbol{X} = [X_1, X_2, \ldots, X_p]'$ um vetor aleatório *p-variado* com média $\boldsymbol{\mu}$ e matriz de covariância $\mathbf{\Sigma}$. Sendo $\mathbf{\Sigma}$ uma matriz $p \times p$ simétrica e não singular ($|\mathbf{\Sigma}| \neq 0$), pode ser reduzida para uma matriz diagonal $\mathbf{\Sigma_c}$, multiplicando por uma particular matriz ortogonal e normalizada, denominada de matriz ortonormal $\mathbf{U}$

Temos assim:

$$\mathbf{U}'\mathbf{\Sigma}\mathbf{U} = \mathbf{\Sigma}_{\boldsymbol{c}} \tag{2.3}$$

onde

$$\mathbf{\Sigma}_{\boldsymbol{c}} = \begin{bmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & \lambda_p \end{bmatrix}$$

Os elementos $\lambda_1, ..., \lambda_p$ representam os autovalores de ordem decrescente e são obtidos através da *equação característica*:

$$|\mathbf{\Sigma} - \boldsymbol{\lambda}\mathbf{I}| = 0 \tag{2.4}$$

A matriz $\boldsymbol{U}_{p\times p}$, traz nas suas colunas os autovetores $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_p$, obtidos da solução da seguinte equação:

$$[\mathbf{\Sigma} - \boldsymbol{\lambda}\mathbf{I}]\mathbf{u_j} = 0 \tag{2.5}$$

onde $\mathbf{u_j}$ representa o autovetor ortonormal, sendo $\mathbf{U}^{'}\mathbf{U} = \mathbf{I}$. Sendo $\mathbf{I}$ uma matriz identidade de mesma ordem, isto é, $\langle \mathbf{u_i}, \mathbf{u_j} \rangle = 0$; para $i \neq j$, e $\langle \mathbf{u_i}, \mathbf{u_j} \rangle = 1$, para $i = j$. Os autovetores $\mathbf{u_j}$ realizam a rotação das variáveis originais $\boldsymbol{X}$ nas direções ortogonais de variabilidade. Os novos eixos ortognais são denominados Componetes Principais (CPs), onde $\mathbf{u_j}$ e λ_j representam, respectivamente, o coeficiente e a variância do *j-ésimo* CP.

Encontrando os autovalores e autovetores da matriz de covariância populacional, pode-se realizar a redução de p variáveis correlacionadas $(X_1, X_2, ..., X_p)$ em k variáveis não-correlacionadas $(Z_1, Z_2, ..., Z_k)$.

Dado um vetor aleatório $\boldsymbol{X}$, o *j-ésimo* componente principal é definido por:

$$Z_j = \mathbf{u_j}^{'} X, \quad j = 1, 2, \ldots, k, \tag{2.6}$$

onde Z_j possui média $\boldsymbol{u_j^{'}\mu}$ e variância λ_j, tendo em vista o vetor aleatório $\boldsymbol{X}$ centrado na sua média (ou seja, considerando a variável aleatória $\boldsymbol{X}$ - $\boldsymbol{\mu}$) a variável aleatória Z_j apresenta média zero e variância λ_j.

ACP dispõe propriedade cuja equação (2.6) pode ser invertida, declarando as variáveis originais como função dos CPs, da seguinte forma:

$$\hat{\boldsymbol{X}}_{\boldsymbol{(p\times 1)}} = \mathbf{U}_{(p\times k)}\boldsymbol{Z}_{\boldsymbol{(k\times 1)}}, \tag{2.7}$$

onde $\mathbf{U}_{(p\times k)}$ representa a matriz cuja as colunas contêm k autovetores associados aos k maiores autovalores. O vetor $\boldsymbol{Z}_{\boldsymbol{(k\times 1)}}$ traz os k CPs descritos na equação (2.6). Observe que se utilizarmos a matriz $\mathbf{U}_{(p\times p)}$, ou seja, contendo todos os p autovetores obtidos via (2.5) e o vetor $\boldsymbol{Z}_{\boldsymbol{(p\times 1)}}$ contendo os p CPs, a variável aleatória $\boldsymbol{X}$ seria reconstruída via (2.7) sem erro.

Dessa forma, o vetor aleatório $\boldsymbol{X}$ pode ser escrito em função dos CPs da seguinte forma:

$$\boldsymbol{X}_{\boldsymbol{(p\times 1)}} = \mathbf{U}_{(p\times k)}\boldsymbol{Z}_{\boldsymbol{(k\times 1)}} + \mathbf{E}, \tag{2.8}$$

onde

$$\boldsymbol{E} = \boldsymbol{X} - \hat{\boldsymbol{X}} \tag{2.9}$$

representa o vetor de erro da reconstrução do vetor aleatório $\boldsymbol{X}$ a partir do número k de CPs. A partir da equação (2.9), podemos obter o erro quadrático absoluto (Q) fazendo:

$$Q = \mathbf{E}'\mathbf{E} \tag{2.10}$$

Segundo (Jolliffe, 2011), deve-se determinar um critério para a escolha da quantidade de CPs retidos na análise. Geralmente poucos CPs conseguem explicar parte substancial da estrutura de covariância $\mathbf{\Sigma}$ das variáveis originais. Um método de escolha simples disponível na literatura é o critério de Kaiser. Considere a ACP empregada no vetor aleatório $\boldsymbol{X}$ contendo cada uma das variáveis aleatórias X_i padronizadas, isto é, $(X_i - \mu_i)/\sigma_i$. A variável aleatória padronizada possui média zero e variância unitária. O critério consiste em selecionar os CPs cujo autovalores possuem valor maior que a unidade $(\lambda_j \geqslant 1)$. A ideia é eleger as variáveis Z_j, cuja variância λ_j seja no mínimo igual a variância das variáveis originais X_i. Uma discussão ampla sobre métodos de seleção dos CPs pode ser encontrada em (Jolliffe, 2011) e (Johnson e Wichern, 2007).

2.3.1 ACP no contexto do CEP

A ACP é uma técnica estatística amplamente utilizada no contexto do Controle Estatístico do Processo (CEP). Destacamos brevemente três vantagens importantes do emprego de Cartas de Controle baseadas em CPs: (i) monitoramento de um número reduzido de variáveis em cenários onde existe um número grande de variáveis de processo correlacionadas; (ii) entendimento das causas de variabilidade do processo, viabilizando o diagnóstico de disturbios do processo; (iii) análise acurada de amostras discrepantes (*outliers* multivariados).

Uma boa gama de processos industriais são controlados por censores que geram inúmeras amostras frequentes de uma grande quantidade de variáveis. Neste contexto, a utilização das Cartas de Controle χ^2 e W pode provocar em um monitoramento distorcido de tais processos. Isso ocorre pelo fato dos dados gerados apresentarem forte correlação, fazendo com que o determinante da matriz de covariância $\boldsymbol{\Sigma}$ se aproxime de zero, acarretando assim, problemas no cálculo da inversa da matriz de covariâncias $\boldsymbol{\Sigma}$, operação necessária para o cálculo da estatística χ^2 [Eq. (2.1)]. Adicionalmente, este fato também ocasiona inconsistência na estatística da Variância Generalizada [Eq. (2.2)], visto que, essa mede a razão de duas matrizes de covariância. Através da ACP, podemos reduzir o número original de variáveis e usar a matriz de covariância $\boldsymbol{\Sigma_c}$ [Eq. (2.3)] de um número reduzido de CPs no lugar da matriz original $\boldsymbol{\Sigma}$ para computar as estatísticas descritas em (2.1) e (2.2). Visto que os CPs retidos são os de maiores autovalores (isto é, com as maiores variâncias, preservando informação substancial do conjunto total de variáveis originais) e são não-correlacionados, a matriz $\boldsymbol{\Sigma_c}$ é diagonal (e com determinante significativamente maior do que zero) e portanto estavelmente inversível.

Uma outra utilidade da ACP no contexto de CEP está no fato de que através da utilização dos CPs, podemos isolar as fontes importantes de variabilidade do processo. Isso ocorre porque cada CP retido na ACP representa uma combinação linear do conjunto total de variáveis de processo, onde podemos identificar as variáveis com maior peso no escore gerado por este CP. Dessa maneira, a partir de um ponto fora dos limites de controle apontado por Cartas de Controle utilizadas em CPs, podemos identificar subconjuntos de variáveis mais prováveis de serem as responsáveis pelos descontroles.

Finalmente, através da ACP podemos identificar amostras discrepantes em um processo multivariado. No caso multivariado, um outlier pode ser resultado de um valor extremo apenas para um dos CPs ou resultante de pequenos erros ordenados em direções diversas, sendo este último, um problema relacionado a correlação e não de localização (média) ou dispersão (variância). Utilizar ACP como método de detecção facilita a procura por direções extremas. (Jolliffe, 2011) afirma que os primeiros componentes permitem detectar *outliers* com variâncias e covariâncias inflacionadas. Estes *outliers* também são extremos nas variáveis originais, podendo ser detectados diretamente. Por outro lado, os *outliers* que não se encontram visíveis nas variáveis originais (os que perturbam a correlação entre variáveis) são detectados pelos CPs não-retidos. A presença de descontroles para os CPs retidos e não-retidos pode ser testada através das Cartas χ^2 e Q (resíduos), respectivamente. Mais sobre detecção de *outliers* baseado em ACP pode ser visto em (Hawkins, 1974), (Gnanadesikan, 2011), (Jolliffe, 2011) e (Barnett e Lewis, 1974).

2.3.2 Cartas de controle utilizando ACP

Nesta subseção descrevemos uma abordagem de controle alternativa à abordagem clássica para monitoramento do vetor de médias e da matriz de covariância do processo. Esta abordagem utiliza Cartas de Controle baseadas na ACP.

Suponha novamente p-características correlacionadas medidas simultaneamente compondo amostras p-variadas de tamanho unitário ($n = 1$). Em um processo sob controle, supõe-se que essas características são descritas por um vetor aleatório $\boldsymbol{X} = [X_1, X_2, \ldots, X_p]'$, possuindo distribuição normal *p-variada* com vetor de médias $\boldsymbol{\mu}$ e matriz de covariância $\boldsymbol{\Sigma}$, de dimensão $p \times p$. Novamente, para viabilizar o monitoramento deste processo, amostras aleatórias são recolhidas periodicamente, constituindo-se em realizações do vetor aleatório $\boldsymbol{X}$.

Dado a correlação entre as p-características consideramos a utilização da ACP para capturar parte substancial da estrutura de correlação do vetor $\boldsymbol{X}$ utilizando um número menor de variáveis de análise (Componentes Principais). Considere a ACP empregada no vetor aleatório $\boldsymbol{X}$ contendo cada uma das variáveis aleatórias X_i padronizadas (isto é, com média zero e variância unitária). Inicialmente, encontramos os autovalores λ_j e autovetores $\boldsymbol{u_j}$, através da diagonalização da matriz $\boldsymbol{\Sigma}$ [Eqs. (2.3), (2.4) e (2.5)]. No passo seguinte, determinamos o número k ($k < p$) de CPs que serão considerados na análise. Neste trabalho destacamos o critério de Kaiser, descrito na seção 2.3. Dessa forma, o CP cuja variância for de no mínimo 1, será retido (isto é, com autovalor $\lambda_j \geqslant 1$). O monitoramento do processo via ACP é realizado a cada novo vetor de observações $\boldsymbol{x}$ produzindo os k CPs [Eq.(2.6)] e o respectivo resíduo gerado [Eqs.(2.6), (2.7), (2.9) e (2.10)].

As amostras se constituem em realizações do vetor aleatório $\boldsymbol{Z}$, e que serão denotadas por $\boldsymbol{z}$. Observe que, dado processo sob controle, estas amostras seguem a distribuição do vetor aleatório $\boldsymbol{Z}$, ou seja, seguem a distribuição normal *k-variada* com média $\boldsymbol{\mu}$ e matriz de covariância $\boldsymbol{\Sigma}_{\boldsymbol{c}_{(p\times k)}}$. O monitoramento futuro de vetores *p-variados* para a carta de Qui-Quadrado, em sua forma original, foi visto pela equação (2.1). Utilizando a técnica de ACP para a Carta de *Qui-Quadrado* esta equação pode ser redefinida como:

$$\chi^2_{CP} = \boldsymbol{z}'\boldsymbol{\Sigma}_{\boldsymbol{c}}^{-1}\boldsymbol{z}\,, \quad k = 1, 2, \ldots \tag{2.11}$$

ou

$$\chi^2_{CP} = \sum_{j=1}^{k} \frac{Z_j^2}{\lambda_j}, \quad j = 1, 2, \ldots, k, \tag{2.12}$$

onde Z_j é obtido na equação (2.6) e o vetor $\boldsymbol{z}$ traz em suas entradas medidas dos k CPs retidos pela ACP. Os limites de controle são dados por:

$$\text{LSC} = \frac{p(n-1)}{n-p} F_{p,n-p} \quad \text{e}$$

$$\text{LIC} = 0$$

onde n é o número de observações do modelo e χ^2_k representa a distância quadrada no plano definido pelos K CPs retidos no modelo. Segundo (Jackson, 2005), a distribuição χ^2 está diretamente relacionada com a distribuição F.

Em um processo gerado é possível observar a ocorrência de eventos que não estão presentes na distribuição de referência, provocando alterações na estrutura de covariância das variáveis do processo. Os resíduos da análise, isto é, os componentes que não foram retidos no modelo, conseguem captar estas pertubações geradas fornecendo informações suplementares que auxiliam em uma sintetização mais robusta do modelo. O monitoramento dos resíduos dá-se pela Carta de Contole Q.

Dado novamente um conjunto de observações $\boldsymbol{x}$ do vetor aleatório $\boldsymbol{X}$ que possui distribuição normal p-variada com vetor de médias $\boldsymbol{\mu}$ e matriz de covariâncias $\boldsymbol{\Sigma}$, assumimos que grupos de n vetores de observação $\boldsymbol{x}$ são recolhidos do processo com determinada frequência. Considerando a ACP empregada no vetor aleatório $\boldsymbol{X}$ contendo cada uma das variáveis aleatórias X_i padronizadas, cujos os autovalores λ_j e autovetores $\mathbf{u_j}$ são encontrados através da diagonalização da matriz $\boldsymbol{\Sigma}$ [Eqs. (2.3), (2.4), (2.5)]. Visto que determinamos através do critério de Kaiser (Seção 2.3) o número k ($k < p$) de CPs que serão retidos no modelo, agora devemos considerar à análise as informações dos CPs não-retidos no modelo, ou seja, que não obtiveram variância igual ou maior que a unidade (autovalor $\lambda_j \geqslant 1$).

O monitoramento do processo via ACP é realizado a cada novo vetor de observações $\boldsymbol{x}$ originando os k CPs [Eq.(2.6)] e o seu devido resíduo gerado [Eqs.(2.6), (2.7), (2.9) e (2.10)]. Devemos considerar para esta Carta o vetor de observações $\boldsymbol{x}$ dos CPs não-retidos, e através dessas observações, aplicar a Carta de Controle para o monitoramento de seus respectivos resíduos.

Consideraremos aqui amostras de tamanho unitário. Estas amostras se constituem em realizações do vetor aleatório $\boldsymbol{Z}$, e que serão denotadas por $\boldsymbol{z}$. Dado que o processo apresenta distribuição Normal com vetor de médias $\boldsymbol{\mu}$ e matriz de covariância $\boldsymbol{\Sigma}$ conhecidas, os resíduos possuem distribuição Gaussiana exata $\mathcal{N}(0,1)$ [(Saporta e Niang, 2009)].

O monitoramento dos resíduos para uma observação $\boldsymbol{x}$ é feito através da equação (2.10). Podemos reescrevê-la da seguinte forma:

$$Q = \sum_{j=1}^{p-k} \frac{Z_j^2}{\lambda_j} \sim chi^2_{p-k} \tag{2.13}$$

(Jackson e Mudholkar, 1979) descrevem a fórmula empregada para o cálculo do limite superior para a Carta de monitoramento dos resíduos:

$$\theta_1 = \textstyle\sum_{i=k+1}^{p} \lambda_i \quad \theta_2 = \sum_{i=k+1}^{p} \lambda_i^2 \quad \theta_3 = \sum_{i=k+1}^{p} \lambda_i^3 \quad h_0 = 1 - \dfrac{2\theta_1\theta_3}{3\theta_2^2}$$

onde a quantidade

$$c = \theta_1 \frac{\frac{Q}{\theta_1}^{h} - \frac{\theta_2 h_0 (h_0 - 1)}{\theta_1^2} - 1}{\sqrt{2\theta_2 h_0^2}} \sim \mathcal{N}(0,1)$$

Então, os limites de controle são dados por:

$$LSC = \theta_1 \frac{\sqrt{2\theta_2 h_0^2}}{\theta_1} + \frac{\theta_2 h_0 (h_0 - 1)}{\theta_1^2} + 1^{\,1/h_0} \tag{2.14}$$

$$\text{LIC} = 0$$

onde c_α é o desvio normal de uma área α sob a cauda superior da distribuição se h_0 é positivo ou sob a cauda inferior se h_0 é negativo. Resíduos que apresentam uma quantidade significativa (excedendo o esperado) de valores maiores que o resultante da equação (2.14) indicam que o modelo com k CPs pode não ser o melhor ajuste para as variáveis [(Jackson, 2005)].

3 Metodologia

Este trabalho apresenta um estudo por simulação, criando variáveis e valores fictícios, com o único propósito de demonstrar o poder de detecção das Cartas de Controle Multivariadas tradicionais e via ACP, descritas na seção 2, quando descontroles são impostos em diferentes cenários na estrutura da média ou variabilidade dos dados. As simulações foram realizadas no software R, através das bibliotecas *mvtnorm* e *qcc*.

Considere um processo que segue uma distribuição Normal 4-variada, operando a partir da fase II, em outras palavras, atuando sob controle estatístico. Mais explicitamente, a distribuição dos dados do processo é definida como:

$$\boldsymbol{X} \sim \mathcal{N}_4(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \tag{3.1}$$

Dado que o estudo é realizado por simulação, podemos determinar os parâmetros como conhecidos, portanto, o vetor de médias e a matriz de covariâncias são considerados como parâmetros populacionais, definidos por:

$$\boldsymbol{\mu} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \qquad \boldsymbol{\Sigma} = \begin{bmatrix} 1 & 0.9 & 0.05 & 0.05 \\ 0.9 & 1 & 0.05 & 0.05 \\ 0.05 & 0.05 & 1 & 0.9 \\ 0.05 & 0.05 & 0.9 & 1 \end{bmatrix}, \tag{3.2}$$

Consideraremos no monitoramento vetores de médias amostrais 4-variados com observações de tamanho $n = 1$. Para avaliação do desempenho da carta de controle χ^2, em cada cenário simulado de descontrole são gerados 5000 vetores de médias e replicados 500 vezes. Novas amostras são descontroladas no vetor de médias. Para cada amostra os escores referentes à estatística χ^2 são comparados ao limite de controle correspondente [Eq. (2.1)]. Considere descontroles realizados para dois cenários distintos: (I) monitoramento do processo simulando o descontrole no vetor de médias para duas variáveis fortemente correlacionadas, mantendo as outras duas variáveis com média constante em 0. Escolheu-se deslocar, para esse caso, a média da primeira (μ_1) e da segunda (μ_2) variável, dado a alta correlação entre elas ($\Sigma_{1,2} = 0.9$); (II) monitoramento do processo simulando o descontrole para duas variáveis fracamente correlacionadas, mantidas as médias constantes em 0 para as demais variáveis. Foram escolhidas as médias da primeira (μ_1) e quarta (μ_4) variável, com correlação $\Sigma_{1,4} = 0.05$. Temos assim:

$$\text{Cenário (I)} \qquad\qquad \text{Cenário (II)}$$

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ 0 \\ 0 \end{bmatrix} \qquad \boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ 0 \\ 0 \\ \mu_4 \end{bmatrix} \tag{3.3}$$

onde $\mu_1 = \mu_2 = \mu_4 = 0,\ 0.1,\ 0.2,\ 0.3,\ 0.4,\ ...,\ 2.9,\ 3$. Para cada cenário foi utilizado intervalo de 0.1 na média das variáveis descontrolada.

Para o monitoramento da variabilidade consideramos a matriz de covariância do processo sob controle via equação (3.2). Para avaliar o desempenho da Carta de Controle W, em cada cenário simulado de descontrole, foram gerados 100 vetores de médias de 50 amostras e computada a matriz de covariâncias associada a cada vetor. Os cenários foram replicados 500 vezes. Essa estrutura de covariância é descontrolada para novas amostras. A cada amostra, os escores referentes à estatística W são comparados ao limite de controle correspondente [Eq. (2.2)]. Dois cenários escolhidos a serem simulados: (III) monitoramento do processo simulando o descontrole na estrutura de covariância em duas variáveis altamente correlacionadas, mantendo as demais constantes. Escolhemos, para esse caso, a primeira e a segunda variável, cuja correlação é $\Sigma_{1,2} = 0.9$; (IV) monitoramento do processo simulando o descontrole na estrutura de covariância em duas variáveis fracamente correlacionadas, mantendo as demais constantes. Para este caso, escolhemos a primeira e quarta variável ($\Sigma_{1,4} = 0.05$). Temos assim:

$$\text{Cenário (III)} \qquad\qquad \text{Cenário (IV)}$$

$$\boldsymbol{\Sigma} = \begin{bmatrix} 1 & \Sigma_{1,2} & 0.05 & 0.05 \\ \Sigma_{1,2} & 1 & 0.05 & 0.05 \\ 0.05 & 0.05 & 1 & 0.9 \\ 0.05 & 0.05 & 0.9 & 1 \end{bmatrix} \qquad \boldsymbol{\Sigma} = \begin{bmatrix} 1 & 0.9 & 0.05 & \Sigma_{1,4} \\ 0.9 & 1 & 0.05 & 0.05 \\ 0.05 & 0.05 & 1 & 0.9 \\ \Sigma_{1,4} & 0.05 & 0.9 & 1 \end{bmatrix}, \tag{3.4}$$

onde $\Sigma_{1,2} = \Sigma_{1,4} =$ -0.95, -0.90, -0.85,..., 0.85, 0.9, 0.95. Para cada cenário foi utilizado intervalo de 0.05 no coeficiente de correlação das variáveis descontroladas.

Os limites de controle para as Cartas χ_2 e W são construidos utilizando probabilidade de alarme falso de $\alpha = 0.05$ (percentil de 5% na cauda superior da distribuição em ambas as cartas)

4 Estudo de Caso Simulado

Nesta seção apresentamos os resultados obtidos das simulações nos cenários expostos na seção 3. Para cada tabela os resultados das cartas são apresentados em função da média e do desvio padrão entre as replicações. As linhas grifadas em cinza representam o processo atuando sob controle estatístico.

4.1 Obtenção dos Componentes Principais

Considere novamente o processo com distribuição e os parâmetros de localização e dispersão definidos, respectivamente, nas equações (3.1) e (3.2).

Para obtenção dos CPs deve-se, primeiramente, calcular os autovalores [Eq. (2.4)] e autovetores [Eq. (2.5)]. A tabela Tabela (4.1) mostra a variância dos CPs e a proporção da variabilidade original dos dados contida em cada CP. Os autovetores obtidos para cada CP estão representados na tabela Tabela (4.2)

Tabela 4.1: Autovalores e percentual de variabilidade do processo para cada CP.

$\boldsymbol{Z_i}$	$\boldsymbol{\lambda_i}$	**Proporção de $\boldsymbol{\lambda_i}$ (%)**	**Proporção Acumulada de $\boldsymbol{\lambda_i}$ (%)**
1	2	50	50
2	1.8	45	95
3	0.1	2.5	97.5
4	0.1	2.5	100

Fonte: Autor.

Tabela 4.2: Autovetores obtidos para cada CP.

$\boldsymbol{Z_i}$	$\mathbf{u}_1$	$\mathbf{u}_2$	$\mathbf{u}_3$	$\mathbf{u}_4$
1	-0.5	0.5	0.71	0.01
2	-0.5	0.5	-0.71	-0.01
3	-0.5	-0.5	0.01	-0.71
4	-0.5	-0.5	-0.01	0.71

Fonte: Autor.

Destacamos o critério de Kaiser, descrito na seção 2.3, como método para escolha do número de CPs que serão retidos para utilização na Carta χ^2 via ACP. Dessa forma, os dois primeiros CPs apresentaram variância de no mínimo a unidade ($\lambda_j \geqslant 1$), representando 95% da variabilidade original dos dados [Ver Tabela (4.1)]. O monitoramento desta Carta será realizado pela equação (2.11).

Visto que determinamos os dois primeiros CPs a serem retidos no modelo, consideraremos, por consequência, os outros dois CPs não-retidos a serem utilizados no monitoramento para a Carta dos Resíduos (Q). Este monitoramento se dá pela estatística (2.13).

4.2 Resultados

4.2.1 Cenário I

A Tabela (4.3) apresenta os resultados das simulações para o cenário (I), em que os descontroles são impostos no vetor de médias para duas variáveis altamente correlacionadas ($\Sigma_{1,2} = 0.9$).

Tabela 4.3: Cenário (I): percentual médio e desvio padrão de amostras perturbadas em μ_1 e μ_2 do vetor de médias $\boldsymbol{\mu}$, utilizando probabilidade de alarme falso de $\alpha = 0.5$ na cauda superior de cada distribuição.

	χ^2		W		χ^2_{CP}		Q	
Descontrole	Média	Desvio Padrão	Média	Desvio Padrão	Média	Desvio Padrão	Média	Desvio Padrão
0.0	0.0486	0.0081	0.0850	0.0609	0.0489	0.0077	0.0509	0.0053
0.1	0.0484	0.0069	0.0575	0.0438	0.0522	0.0061	0.0483	0.0068
0.2	0.0516	0.0072	0.0825	0.0520	0.0526	0.0077	0.0498	0.0071
0.3	0.0557	0.0068	0.0600	0.0553	0.0584	0.0090	0.0519	0.0078
0.4	0.0590	0.0074	0.0925	0.0712	0.0646	0.0063	0.0531	0.0070
0.5	0.0656	0.0059	0.0825	0.0545	0.0722	0.0065	0.0530	0.0054
0.6	0.0663	0.0070	0.0500	0.0513	0.0773	0.0064	0.0512	0.0083
0.7	0.0772	0.0106	0.0625	0.0582	0.0902	0.0109	0.0525	0.0097
0.8	0.0867	0.0098	0.0725	0.0380	0.1040	0.0076	0.0526	0.0072
0.9	0.0949	0.0094	0.0625	0.0582	0.1174	0.0087	0.0540	0.0080
1.0	0.1084	0.0128	0.0550	0.0536	0.1377	0.0146	0.0504	0.0069
1.1	0.1256	0.0115	0.0975	0.0658	0.1618	0.0097	0.0536	0.0086
1.2	0.1352	0.0090	0.0875	0.0510	0.1754	0.0119	0.0492	0.0063
1.3	0.1566	0.0124	0.0625	0.0425	0.2036	0.0115	0.0512	0.0069
1.4	0.1752	0.0100	0.0675	0.0654	0.2296	0.0151	0.0520	0.0067
1.5	0.1958	0.0129	0.0650	0.0651	0.2606	0.0119	0.0495	0.0063
1.6	0.2231	0.0125	0.0625	0.0510	0.2922	0.0130	0.0516	0.0062
1.7	0.2512	0.0133	0.0900	0.0641	0.3268	0.0154	0.0531	0.0077
1.8	0.2776	0.0144	0.0700	0.0616	0.3635	0.0137	0.0516	0.0073
1.9	0.3081	0.0091	0.0700	0.0410	0.4000	0.0129	0.0488	0.0056
2.0	0.3430	0.0195	0.0650	0.0587	0.4370	0.0168	0.0554	0.0088
2.1	0.3700	0.0131	0.0625	0.0483	0.4751	0.0134	0.0512	0.0064
2.2	0.4022	0.0162	0.0850	0.0651	0.5136	0.0197	0.0516	0.0074
2.3	0.4370	0.0143	0.0650	0.0630	0.5491	0.0182	0.0482	0.0068
2.4	0.4808	0.0165	0.0725	0.0697	0.5947	0.0160	0.0520	0.0077
2.5	0.5155	0.0174	0.0650	0.0401	0.6286	0.0187	0.0534	0.0083
2.6	0.5508	0.0163	0.0500	0.0429	0.6658	0.0202	0.0531	0.0065
2.7	0.5861	0.0107	0.0575	0.0373	0.6990	0.0108	0.0516	0.0037
2.8	0.6199	0.0122	0.0800	0.0594	0.7317	0.0105	0.0503	0.0064
2.9	0.6568	0.0146	0.0575	0.0406	0.7668	0.0103	0.0508	0.0081
3.0	0.6965	0.0149	0.0650	0.0489	0.7991	0.0147	0.0504	0.0065

Fonte: Autor.

Observamos que a Carta χ^2 via ACP (χ^2_{CP}) apresenta maior sensibilidade na detecção em todas as faixas de distúrbios gerados em relação a Carta χ^2. Isto se deve ao fato de que os descontroles foram impostos numa direção comum de variabilidade dos dados (isto é, em duas variáveis fortemente correlacionadas), direção esta capturada pela ACP. A medida que os descontroles no vetor de médias aumentam, afastando-se do processo sob controle estatístico, a maior eficiência da Carta χ^2_{CP} se torna ainda mais visível.

A Carta Q serve como uma Carta suporte à Carta χ^2_{CP}, auxiliando na detecção de descontroles quando eles não ocorrem na distribuição de referência. Uma vez que notamos o bom desempenho da Carta χ^2_{CP} para este cenário, os resíduos permanecem sob controle estatístico, indicando que para este caso, o modelo com dois CPs retidos se torna um bom ajuste para as variáveis.

A Carta W mantem o número de distúrbios detectados próximos ao processo operando sob controle ($\alpha = 0.05$), uma vez que não houve modificações na estrutura de correlação dos dados.

4.2.2 Cenário II

Para a Tabela (4.4), são mostrados os resultados dos descontroles impostos no vetor de médias para duas variáveis com baixa correlação ($\rho = 0.05$).

Tabela 4.4: Cenário (II): percentual médio e desvio padrão de amostras perturbadas em μ_1 e μ_4 do vetor de médias $\boldsymbol{\mu}$, utilizando probabilidade de alarme falso de $\alpha = 0.5$ na cauda superior de cada distribuição.

	χ^2		W		χ^2_{CP}		Q	
Descontrole	Média	Desvio Padrão	Média	Desvio Padrão	Média	Desvio Padrão	Média	Desvio Padrão
0.0	0.0494	0.0068	0.0500	0.0363	0.0504	0.0052	0.0520	0.0059
0.1	0.0538	0.0041	0.0500	0.0397	0.0499	0.0080	0.0592	0.0081
0.2	0.0686	0.0090	0.0525	0.0444	0.0498	0.0065	0.0822	0.0074
0.3	0.1036	0.0090	0.0650	0.0516	0.0546	0.0078	0.1237	0.0082
0.4	0.1488	0.0104	0.0775	0.0444	0.0551	0.0068	0.1940	0.0125
0.5	0.2156	0.0101	0.0450	0.0426	0.0606	0.0072	0.2766	0.0126
0.6	0.3031	0.0141	0.0575	0.0520	0.0649	0.0073	0.3828	0.0152
0.7	0.4035	0.0165	0.0775	0.0697	0.0694	0.0079	0.4954	0.0171
0.8	0.5274	0.0204	0.0725	0.0752	0.0759	0.0094	0.6194	0.0183
0.9	0.6340	0.0146	0.0725	0.0444	0.0834	0.0123	0.7268	0.0116
1.0	0.7387	0.0146	0.0700	0.0594	0.0897	0.0104	0.8220	0.0115
1.1	0.8293	0.0085	0.0775	0.0550	0.0998	0.0079	0.8934	0.0107
1.2	0.8968	0.0114	0.0775	0.0716	0.1080	0.0093	0.9402	0.0054
1.3	0.9399	0.0077	0.0575	0.0520	0.1218	0.0105	0.9658	0.0053
1.4	0.9686	0.0055	0.0725	0.0595	0.1300	0.0101	0.9838	0.0038
1.5	0.9858	0.0037	0.0850	0.0516	0.1460	0.0116	0.9924	0.0029
1.6	0.9930	0.0030	0.0500	0.0397	0.1578	0.0129	0.9968	0.0018
1.7	0.9968	0.0013	0.0550	0.0536	0.1709	0.0097	0.9990	0.0009
1.8	0.9989	0.0008	0.0700	0.0497	0.1908	0.0140	0.9998	0.0004
1.9	0.9997	0.0006	0.0875	0.0666	0.2062	0.0129	1	0
2.0	1	0	0.0725	0.0658	0.2264	0.0127	1	0
2.1	1	0	0.1000	0.0585	0.2434	0.0141	1	0
2.2	1	0	0.0550	0.0359	0.2558	0.0119	1	0
2.3	1	0	0.0600	0.0553	0.2951	0.0131	1	0
2.4	1	0	0.0475	0.0499	0.3134	0.0148	1	0
2.5	1	0	0.0700	0.0523	0.3300	0.0095	1	0
2.6	1	0	0.0750	0.0500	0.3538	0.0149	1	0
2.7	1	0	0.0850	0.0709	0.3842	0.0145	1	0
2.8	1	0	0.0700	0.0571	0.4094	0.0176	1	0
2.9	1	0	0.0725	0.0472	0.4340	0.0174	1	0
3.0	1	0	0.0650	0.0489	0.4628	0.0210	1	0

Fonte: Autor.

A partir dos resultados apresentados, notamos acentuada sensibilidade na detecção de descontroles da Carta χ^2 se comparada à mesma no cenário (I). Isto se deve à baixa correlação nas variáveis descontroladas, dado que esses descontroles estão em direções opostas às direções comuns de variabilidade. Dessa maneira, mesmo pequenas alterações são detectadas com frequência relativa alta. Em outras palavras, a Carta χ^2 possui melhor desempenho em direções opostas de variabilidade do que em direções comuns de variabilidade, quando comparada a ela mesma. O oposto ocorre à Carta χ^2_{CP}, onde apresenta maior senbilidade na detecção de distúrbios em direções comuns de variabilidade, não apresentando um desempenho satisfatório para este cenário se comparada com sua própria Carta no cenário (I), ou mesmo com a Carta χ^2 para este cenário.

Com o baixo desempenho da Carta χ^2_{CP}, por consequência, a Carta Q consegue captar esses distúrbios não detectados. Com isso, notamos uma alta quantidade de resíduos nesse cenário, sinalizando que o modelo com dois CPs retidos não é suficiente para descrever estas alterações em direções opostas às direções capturadas pela ACP. Assim como no cenário (I), as Cartas via ACP continuam se mostrando mais eficientes no cenário (II) comparadas as Cartas Tradicionais. Neste cenário, a Carta Q demonstrou maior sensibilidade perante as demais para detectar descontroles no vetor de médias em direções opostas à variabilidade. A interpretação do desempenho da carta W no cenário (I) não se altera no cenário (II).

4.2.3 Cenário III

A Tabela (4.5) exibe os resultados das simulações para o cenário (III), em que os descontroles são impostos na estrutura de correlação em duas variáveis altamente correlacionadas ($\Sigma_{1,2} = 0.9$).

Tabela 4.5: Cenário (III): percentual médio e desvio padrão de amostras perturbadas em $\Sigma_{1,2}$ da matriz de correlação $\mathbf{\Sigma}$, utilizando probabilidade de alarme falso de $\alpha = 0.05$ na cauda superior de cada distribuição.

	χ^2		W		χ^2_{CP}		Q	
Descontrole	Média	Desvio Padrão	Média	Desvio Padrão	Média	Desvio Padrão	Média	Desvio Padrão
-0.95	0.5464	0.0061	1	0	0.0146	0.0016	0.6235	0.0050
-0.90	0.5377	0.0066	1	0	0.0146	0.0018	0.6153	0.0058
-0.85	0.5333	0.0071	1	0	0.0153	0.0014	0.6113	0.0066
-0.80	0.5312	0.0073	1	0	0.0151	0.0016	0.6089	0.0071
-0.75	0.5262	0.0058	1	0	0.0156	0.0019	0.6041	0.0062
-0.70	0.5195	0.0061	1	0	0.0154	0.0013	0.5968	0.0067
-0.65	0.5148	0.0065	1	0	0.0162	0.0019	0.5899	0.0069
-0.60	0.5077	0.0055	1	0	0.0165	0.0020	0.5843	0.0051
-0.55	0.5043	0.0070	1	0	0.0171	0.0012	0.5816	0.0079
-0.50	0.4954	0.0069	1	0	0.0169	0.0014	0.5726	0.0061
-0.45	0.4909	0.0062	1	0	0.0176	0.0016	0.5666	0.0053
-0.40	0.4824	0.0057	1	0	0.0178	0.0016	0.5603	0.0051
-0.35	0.4765	0.0061	1	0	0.0182	0.0023	0.5523	0.0068
-0.30	0.4728	0.0073	1	0	0.0191	0.0025	0.5491	0.0072
-0.25	0.4646	0.0047	1	0	0.0196	0.0021	0.5401	0.0051
-0.20	0.4531	0.0063	1	0	0.0204	0.0022	0.5279	0.0061
-0.15	0.4459	0.0049	1	0	0.0204	0.0019	0.5213	0.0062
-0.10	0.4358	0.0083	1	0	0.0211	0.0016	0.5111	0.0076
-0.05	0.4284	0.0083	1	0	0.0227	0.0018	0.5006	0.0073
0.00	0.4165	0.0082	1	0	0.0229	0.0022	0.4911	0.0094
0.05	0.4078	0.0073	1	0	0.0236	0.0027	0.4837	0.0067
0.10	0.3949	0.0079	1	0	0.0255	0.0017	0.4679	0.0078
0.15	0.3798	0.0097	1	0	0.0264	0.0031	0.4529	0.0078
0.20	0.3703	0.0074	1	0	0.0274	0.0023	0.4428	0.0071
0.25	0.3561	0.0071	1	0	0.0292	0.0024	0.4285	0.0076
0.30	0.3411	0.0050	1	0	0.0298	0.0025	0.4119	0.0046
0.35	0.3271	0.0064	1	0	0.0301	0.0020	0.3965	0.0060
0.40	0.3094	0.0060	1	0	0.0327	0.0019	0.3774	0.0055
0.45	0.2924	0.0061	1	0	0.0346	0.0026	0.3571	0.0076
0.50	0.2708	0.0065	1	0	0.0359	0.0025	0.3335	0.0076
0.55	0.2487	0.0048	1	0	0.0363	0.0031	0.3104	0.0046
0.60	0.2259	0.0081	0.9995	0.0022	0.0390	0.0035	0.2826	0.0060
0.65	0.1990	0.0057	0.9990	0.0045	0.0411	0.0027	0.2514	0.0050
0.70	0.1721	0.0042	0.9875	0.0107	0.0429	0.0034	0.2179	0.0050
0.75	0.1398	0.0044	0.9335	0.0221	0.0445	0.0024	0.1788	0.0049
0.80	0.1092	0.0034	0.7270	0.0478	0.0473	0.0019	0.1385	0.0039
0.85	0.0774	0.0038	0.2540	0.0483	0.0470	0.0017	0.0946	0.0040
0.90	0.0494	0.0037	0.0735	0.0239	0.0503	0.0033	0.0499	0.0028
0.95	0.0324	0.0023	0.6090	0.0488	0.0518	0.0036	0.0229	0.0022

Fonte: Autor.

Conforme descrito na subseção 2.2.2, alterações na estrutura de variabilidade do processo também pode ocasionar descontroles no vetor de médias. Dessa forma, observamos que, mesmo o vetor de médias sendo mantido constante em todas as faixas de descontrole, a Carta χ^2 consegue captar esses descontroles modificados na estrutura de correlação. A medida que a correlação das variáveis descontroladas diminui, se afastandando das direções não opostas de variabilidade, a Carta χ^2 melhora seu desempenho de detecção.

Os resultados demonstrados para este cenário comprovam a alta sensibilidade na detecção dos descontroles à Carta W, uma vez que pequenas alterações em relação a correlação de referência são prontamente detectadas em 100% das amostras. Este resultado é esperado, visto que está é a única Carta de controle que monitora diretamente a estrutura de correlação das variáveis.

A Carta χ^2_{CP} apresenta baixo desempenho, visto que os descontrole foram impostos na correlação e não diretamente no vetor de médias. Como esperado, a Carta Q irá capturar os descontroles não identificados pela Carta χ^2_{CP}, porém sem desempenho tão robusto quanto a Carta W.

4.2.4 Cenário IV

Tabela 4.6: Cenário (IV): percentual médio e desvio padrão de amostras perturbadas em $\Sigma_{1,4}$ da matriz de correlação $\mathbf{\Sigma}$, utilizando probabilidade de alarme falso de $\alpha = 0.05$ na cauda superior de cada distribuição.

	χ^2		W		χ^2_{CP}		Q	
Descontrole	Média	Desvio Padrão	Média	Desvio Padrão	Média	Desvio Padrão	Média	Desvio Padrão
-0.95	0.8441	0.0046	1	0	0.4674	0.0073	0.8589	0.0042
-0.90	0.8276	0.0048	1	0	0.4339	0.0060	0.8427	0.0047
-0.85	0.8084	0.0051	1	0	0.3930	0.0084	0.8256	0.0052
-0.80	0.7905	0.0041	1	0	0.3486	0.0079	0.8082	0.0046
-0.75	0.7597	0.0062	1	0	0.3040	0.0076	0.7812	0.0056
-0.70	0.7346	0.0082	1	0	0.2656	0.0076	0.7568	0.0081
-0.65	0.6956	0.0059	1	0	0.2260	0.0056	0.7218	0.0061
-0.60	0.6573	0.0074	1	0	0.1878	0.0047	0.6856	0.0069
-0.55	0.6120	0.0064	1	0	0.1570	0.0040	0.6413	0.0069
-0.50	0.5543	0.0060	1	0	0.1269	0.0053	0.5865	0.0073
-0.45	0.4944	0.0066	1	0	0.1058	0.0046	0.5292	0.0072
-0.40	0.4181	0.0073	1	0	0.0876	0.0040	0.4544	0.0062
-0.35	0.3360	0.0066	1	0	0.0757	0.0041	0.3724	0.0069
-0.30	0.2537	0.0081	1	0	0.0646	0.0040	0.2882	0.0064
-0.25	0.1817	0.0081	1	0	0.0578	0.0038	0.2088	0.0066
-0.20	0.1212	0.0045	1	0	0.0544	0.0029	0.1401	0.0047
-0.15	0.0794	0.0041	1	0	0.0514	0.0035	0.0900	0.0049
-0.10	0.0671	0.0041	1	0	0.0506	0.0031	0.0729	0.0036
-0.05	0.0564	0.0039	0.8135	0.0502	0.0497	0.0032	0.0606	0.0025
0.00	0.0532	0.0022	0.2135	0.0428	0.0515	0.0029	0.0533	0.0028
0.05	0.0510	0.0029	0.0595	0.0265	0.0492	0.0034	0.0519	0.0030
0.10	0.0507	0.0029	0.2245	0.0338	0.0488	0.0029	0.0526	0.0032
0.15	0.0573	0.0031	0.8375	0.0470	0.0494	0.0025	0.0616	0.0040
0.20	0.0663	0.0029	1	0	0.0519	0.0025	0.0722	0.0034
0.25	0.0806	0.0036	1	0	0.0512	0.0038	0.0910	0.0046
0.30	0.1236	0.0043	1	0	0.0524	0.0026	0.1427	0.0030
0.35	0.1830	0.0056	1	0	0.0577	0.0031	0.2099	0.0066
0.40	0.2538	0.0051	1	0	0.0620	0.0026	0.2878	0.0047
0.45	0.3380	0.0064	1	0	0.0723	0.0034	0.3743	0.0068
0.50	0.4208	0.0080	1	0	0.0863	0.0033	0.4558	0.0089
0.55	0.4930	0.0071	1	0	0.1039	0.0030	0.5278	0.0065
0.60	0.5558	0.0068	1	0	0.1255	0.0047	0.5900	0.0068
0.65	0.6128	0.0078	1	0	0.1553	0.0049	0.6422	0.0083
0.70	0.6610	0.0093	1	0	0.1890	0.0048	0.6882	0.0076
0.75	0.7000	0.0064	1	0	0.2255	0.0048	0.7245	0.0070
0.80	0.7328	0.0057	1	0	0.2649	0.0074	0.7563	0.0049
0.85	0.7620	0.0059	1	0	0.3049	0.0087	0.7829	0.0062
0.90	0.7878	0.0056	1	0	0.3494	0.0049	0.8064	0.0063
0.95	0.8114	0.0047	1	0	0.3889	0.0070	0.8278	0.0045

Fonte: Autor.

Podemos notar que a Carta χ^2 apresenta um desempenho superior a mesma no cenário (III), demonstrando resultados semelhantes obtidos nos cenários de monitoramento em que os descontroles ocorreram no vetor de médias, ou seja, maior eficiência em cenários que os distúrbios ocorrem em diferentes direções de variabilidade.

Assim como no cenário (III), notamos ótimo desempenho à Carta W, detectando 100% das amostras a partir de pequenas modificações na correlação original das variáveis.

O bom desempenho da Carta Q se mostra útil para investigar a adequabilidade do número de CPs retidos no modelo. Assim como no cenário (III), a quantidade de resíduos detectados excedendo o esperado indicam que a Carta χ^2_{CP} com os dois primeiros CPs retidos não se mostra o melhor ajuste para as variáveis, o que é de se esperar, visto que a Carta χ^2_{CP} monitora diretamente o vetor de médias dos dados.

5 Considerações Finais

Este trabalho apresentou um estudo do desempenho das Cartas de Controle Multivariadas Clássicas e com projeção de dados via Análise de Componentes Principais. Através de um estudo simulado utilizando quatro variáveis apresentando uma estrutura de covariância, exibindo correlações fortes e fracas, diferentes cenários foram investigados incluindo diversos descontroles impostos no vetor de médias e na matriz de covariâncias.

Em relação à Carta χ^2, verificamos que descontroles impostos nas direções comuns de variabilidade são detectados com menor sensibilidade quando comparados aos descontroles impostos fora das direções comuns. Aferimos boa sensibilidade à Carta W na detecção de descontroles na estrutura de covariância, independente do tamanho da correlação entre as variáveis no processo sob controle.

A Carta Q apresentou bom desempenho em cenários envolvendo distúrbios nas direções contrárias de variabilidade, independentemente da estrutura em que os descontroles foram ocasionados.

Para alterações no vetor de médias, nas direções comuns de variabilidade, evidenciamos a maior sensibilidade da Carta χ^2_{CP} sobre a Carta χ^2 na detecção desses descontroles. A carta W é mais sensível em alterações na estrutura de covariância das variáveis em relação as Cartas utilizando ACP.

Referências Bibliográficas

Barnett, V. e Lewis, T. (1974). *Outliers in statistical data.* Wiley.

Genz, A., Bretz, F., Hothorn, T., e Hothorn, M. T. (2007). The mvtnorm package.

Gnanadesikan, R. (2011). *Methods for statistical data analysis of multivariate observations*, volume 321. John Wiley & Sons.

Hawkins, D. M. (1974). The detection of errors in multivariate data using principal components. *Journal of the American Statistical Association*, 69(346):340–344.

Jackson, J. E. (2005). *A user's guide to principal components*, volume 587. John Wiley & Sons.

Jackson, J. E. e Mudholkar, G. S. (1979). Control procedures for residuals associated with principal component analysis. *Technometrics*, 21(3):341–349.

Johnson, R. e Wichern, D. (2007). Applied multivariate statistical analysis. *INC., New Jersey.*

Jolliffe, I. (2011). *Principal component analysis.* Springer.

Louzada, F. (2009). *Controle Estatístico de Processo: Uma abordagem prática para cursos de Engenharia e Administração.* LTC.

Marcondes Filho, D. (2001). Monitoramento de processos em bateladas através de cartas de controle multivariadas utilizando análise de componentes principais multidirecionais. *Master's thesis, Universidade Federal do Rio Grande do Sul.*

Mason, R. L. e Young, J. C. (2002). *Multivariate statistical process control with industrial applications*, volume 9. Siam.

Montgomery, D. C. (2007). *Introduction to statistical quality control.* John Wiley & Sons.

Saporta, G. e Niang, N. (2009). Principal component analysis: application to statistical process control. *Data analysis*, volume 1.