Roberto Celso Colacioppo

CONTROLE ESTATÍSTICO MULTIVARIADO DE PROCESSOS PARA OBSERVAÇÕES INDIVIDUAIS

Dissertação apresentada ao Instituto de Matemática, Estatística e Ciência da Computação da Universidade Estadual de Campinas para obtenção do título de Mestre em Estatística.

Orientador: Prof. Ademir José Petenate

Campinas - SP
IMECC / UNICAMP
2001

Controle Estatístico Multivariado de Processos para Observações Individuais

Este exemplar corresponde à redação final da dissertação devidamente corrigida e defendida por Roberto Celso Colacioppo e aprovada pela comissão julgadora.

Campinas, 15 de março de 2001

Prof. Dr.: Ademir José Petenate

Banca Examinadora:

1. Dr. Ademir José Petenate
2. Dra. Regina Célia de Carvalho Pinto Moran
3. Dr. Milton Mori

Dissertação apresentada ao Instituto de Matemática, Estatística e Computação Científica, UNICAMP, como requisito parcial para obtenção do título de MESTRE em ESTATÍSTICA

FICHA CATALOGRÁFICA ELABORADA PELA BIBLIOTECA DO IMECC DA UNICAMP

Colacioppo, Roberto Celso

C67c Controle estatístico multivariado de processos para observações individuais / Roberto Celso Colacioppo -- Campinas, [S.P. :s.n.], 2001.

Orientador : Ademir José Petenate

Dissertação (mestrado) - Universidade Estadual de Campinas, Instituto de Matemática, Estatística e Computação Científica.

1. Controle de qualidade – Métodos estatísticos. 2. Estatística industrial. 3. Análise de componentes principais. I. Petenate, Ademir José. II. Universidade Estadual de Campinas. Instituto de Matemática, Estatística e Computação Científica. III. Título.

Dissertação de Mestrado defendida em 15 de março de 2001 e aprovada

Pela Banca Examinadora composta pelos Profs. Drs.

Prof (a). Dr (a). ADEMIR JOSÉ PETENATE

Prof (a). Dr (a). REGINA CÉLIA DE CARVALHO PINTO MORAN

Prof (a). Dr (a). MILTON MORI

com ternura
para minha mulher Beatriz

Gostaria de expressar meus agradecimentos

- Aos meus pais Marisa e Roberto, que sempre me incentivaram e apoiaram;
- Ao Prof. Ademir pela grande amizade e pela enorme contribuição na minha formação e vida profissional;
- À Profa. Regina pela habilidosa influência e sabedoria que me inspirou para o tema;
- Ao Prof. Milton pela confiança depositada e pelo empenho em promover a Estatística nos cursos da Engenharia Química;
- Aos amigos que fiz no mestrado, em especial ao Rodrigo e Bruno pela união na batalha pela qualificação;
- À Compaq, na pessoa de César Pucci, que propiciou intercâmbio com consultores da API (Associates in Process Improvement);
- Ao CNPQ pela ajuda financeira.

as coisas

As coisas têm peso, massa, volume, tamanho, tempo, forma, cor, posição, textura, duração, densidade, cheiro, valor, consistência, profundidade, contorno, temperatura, função, aparência, preço, destino, idade, sentido. As coisas não têm paz.

(Arnaldo Antunes)

Controle Estatístico Multivariado de Processos para Observações Individuais

Resumo

Propomos nesta dissertação um roteiro de técnicas simples de serem implementadas para auxílio no entendimento da variabilidade de processos cujas observações são individuais e multivariadas. A motivação básica para o uso do controle estatístico multivariado de processos (CEMP) vem do fato de se ter que levar em conta a estrutura de correlação dos dados para se responder adequadamente a pergunta: "o processo está sob controle?". No início do texto, o CEMP é revisto enquanto técnica utilizada para subgrupos racionais e uma ilustração de seu uso é mostrada tanto para a fase I (exame retrospectivo para se estimar os parâmetros do processo) como para a fase II (análise de futuros subgrupos). Um tratamento especial deve ser dedicado quando os subgrupos têm tamanho 1 (observações individuais), principalmente na fase I em que as estimativas dos parâmetros não são independentes das próprias observações a serem testadas, além disso, a estimativa usual da matriz de covariância pode ser inflacionada caso tenhamos causas assinaláveis no conjunto de dados inicial. Por essa razão, estimadores robustos dessa matriz, similares ao de amplitudes móveis do caso univariado, são estudados e limites adequados do gráfico de controle são desenvolvidos. Mostramos que nesses casos eles apresentam um bom desempenho para causas especiais tipo degrau e rampa. Por outro lado, sinais provocados por observações aberrantes são mais difíceis de serem detectados, principalmente se o afastamento não é tão evidente, e alternativas são sugeridas baseadas em técnicas de exploração da estrutura interna dos dados como Análise de Componentes Principais. Abordamos, em seguida, uma técnica de fácil implementação e interpretação para diagnóstico das causas assinaláveis. Ao final, utilizamos essas técnicas para análise de dados de um processo de montagem de cabinas de caminhões.

ÍNDICE

LISTA DE FIGURAS

LISTA DE TABELAS

Introdução

Gráficos de Controle e Melhoria de Processos

Em 1924, o Diretor de Engenharia de Inspeção da Bell Telephone Laboratories enviou um memorando ao Dr. Walter A. Shewhart com o seguinte pedido:

"... o desenvolvimento de uma forma aceitável de relatório de inspeção que possa ser modificado de tempos em tempos, a fim de fornecer à primeira vista a maior quantidade de informação acurada."

Shewhart respondeu apresentando um gráfico com valores de uma estatística plotados ao longo do tempo entre linhas que denominou de limites de controle cuja função era separar dois tipos de variação: aquelas inerentes ao processo e causadas por eventos de difícil diagnóstico, daquelas variações excessivas e causadas por um evento surgido numa circunstância específica e rastreável. Ele deu o nome de causas assinaláveis a esse último tipo de variação. Deming popularizou mais tarde essa distinção de tipos de variação como Causas Comuns e Causas Especiais. Estava assim criado o Gráfico de Controle e estabelecido um conceito fundamental para o estudo e melhoria de processos.

Somente na Segunda Grande Guerra, por explícita necessidade, essa técnica começou a ganhar destaque nos Estados Unidos. Harold Hotelling em 1947 foi quem primeiro publicou o desenvolvimento de gráficos de controle para o caso multivariado. Ele usou sua teoria para melhorar a acurácia no lançamento de bombas pelos aviões da marinha americana. O pós-guerra foi um período de abundância para os EUA e de renascimento para o Japão que a partir de 1950 começava usar sistematicamente técnicas estatísticas e conceitos de controle da qualidade para levantar sua economia, que 30 anos mais tarde atormentou os empresários norte-americanos com a enxurrada de produtos de boa qualidade vindos do Japão. Nos últimos 20 anos tem havido mudanças consideráveis no

cenário econômico mundial e isso foi acompanhado por um crescimento significativo de aplicação dos métodos estatísticos bem como pelo desenvolvimento de novas técnicas e de novas áreas de aplicação, como por exemplo, o setor de serviços.

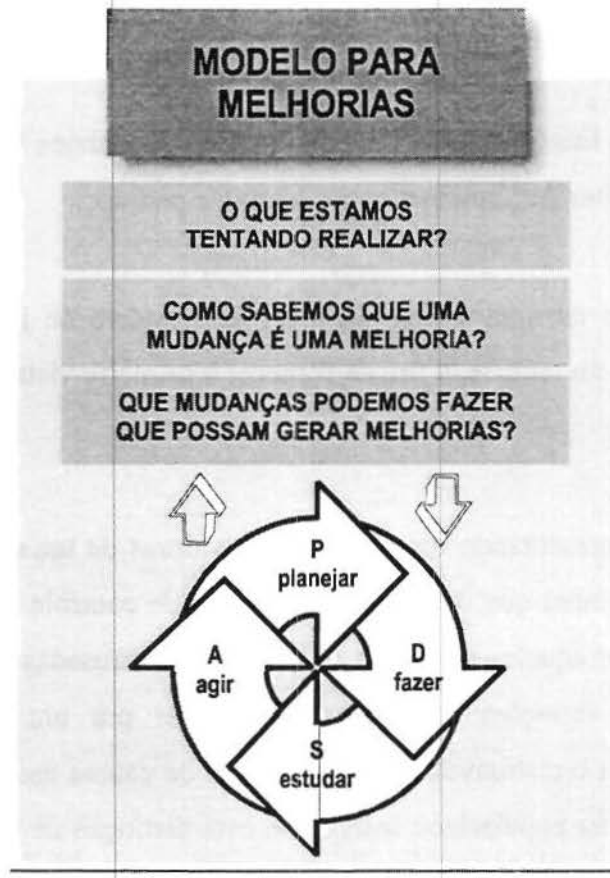

Figura I – Modelo para Melhorias proposto por Langley et alli (1996)

Estamos atravessando uma terceira revolução industrial, a revolução da informação, em que dados estão mais abundantes e acessíveis. As organizações que tirarem melhor proveito deles, aprendendo sobre os processos, serão mais eficazes na realização de mudanças necessárias para se obter melhorias e estarão em vantagem na competição. Nesse contexto, o termo "controle" do gráfico de controle parece não adequado, uma vez que o uso mais comum dessa técnica é investigar a variação e avaliar o impacto de mudanças. Em API (1998) os autores sugerem o nome "gráficos de aprendizado", mas o nome que Shewhart escolheu deve permanecer pelo uso comum.

Langley et alli (1996) propuseram um modelo para melhorias (Figura I) combinando 3 questões fundamentais ao ciclo PDSA de Deming (planejar, fazer, estudar e agir) onde o uso e aprendizado com dados é uma habilidade essencial a ser desenvolvida pelos grupos para apoiar a realização de melhorias. Os gráficos de controle estão entre as ferramentas mais eficientes nesse aprendizado.

Motivação para Uso dos Gráficos de Controle Multivariados

A qualidade de um processo é usualmente medida através do nível conjunto de várias características correlacionadas. Podemos adotar gráficos de controles usuais em separado para cada característica ou adotar uma abordagem multivariada para esse caso.

Os Gráficos de Controle Multivariados são a representação visual do resultado para o teste da razão de verossimilhança sobre o vetor de médias do processo a cada amostragem. Sua construção é separada em duas fases. A primeira (fase I) consiste da análise de dados passados com o objetivo de verificar se o processo estava sob controle estatístico quando os primeiros subgrupos foram obtidos, bem como "limpar" os dados (retirar os pontos suspeitos de estarem fora de controle), e estimar a distribuição do processo para teste de pontos futuros. A fase seguinte (fase II) consiste em usar o gráfico de controle para detectar qualquer fuga do processo em relação à sua distribuição estimada na fase I.

Três razões se destacam a favor dos Gráficos de Controle Multivariados:

i) Produzem uma única resposta à pergunta: "O processo está sob controle?";

ii) É mantida a probabilidade especificada de considerarmos o processo fora de controle quando na verdade ele não está (erro tipo I);

iii) Levam em conta as relações entre as variáveis.

As duas últimas razões podem ser melhor compreendidas através de um exemplo apresentado por Jackson (1985). São dados fictícios mas representam um caso bem comum

numa planta química. Ele se remete a uma situação em que são feitas análises de rotina da concentração de um produto num determinado ponto do processo. As amostras a serem testadas são divididas em duas alíquotas e analisadas por dois métodos diferentes, 1 e 2. A Tabela I mostra os valores adotados na explicação.

Tabela I – Dados usados por Jackson (1985) para ilustração do método

	método 1	**método 2**
	10.0	10.7
	10.4	9.8
	9.7	10.0
	9.7	10.1
	11.7	11.5
	11.0	10.8
	8.7	8.8
	9.5	9.3
	10.1	9.4
	9.6	9.6
	10.5	10.4
	9.2	9.0
	11.3	11.6
	10.1	9.8
	8.5	9.2
média	10.0	10.0
Desvio padrão	.89	.86

Aplicando nesta amostra os procedimentos usuais dos gráficos de controle univariados para cada método, temos os seguintes limites de controle mantendo um erro tipo I de 5%:

Método 1: $10.0 \pm 1.9 = (8.1, 11.9)$; *Método 2*: $10.0 \pm 1.8 = (8.2, 11.8)$

Portanto, se o processo estiver sob controle, amostras analisadas pelo método 1 terão 5% dos valores fora desses limites em média. Isso ocorrendo da mesma forma para o método 2. Contudo, se os dois métodos fossem independentes estatisticamente, a probabilidade de que ambos os métodos produzam pontos sob controle é de $0{,}95^2 = 0{,}90$, ou seja, estamos trabalhando com um erro tipo I de 10% e não de 5%.

Para explicar como os gráficos multivariados levam em conta as relações entre as variáveis, veja o que acontece quando se plota um método contra o outro na Figura II. Esse gráfico mostra que há uma correlação entre as variáveis. Nele já se encontram o limite de controle bivariado em forma de elipse, assim como os limites marginais (pontilhados) baseados na amostra.

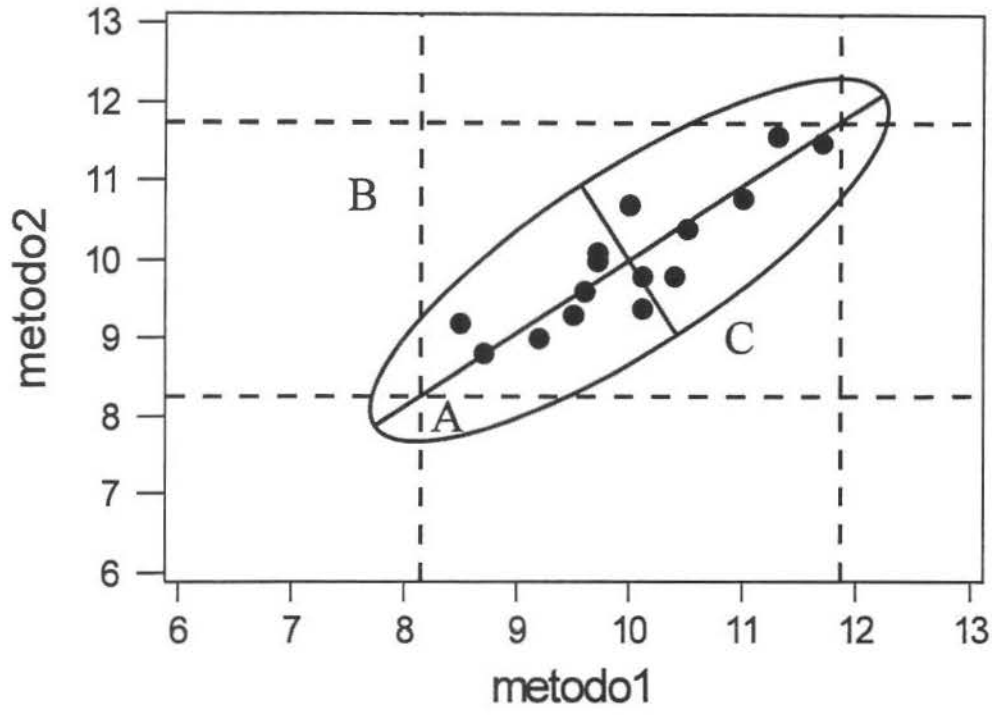

Figura II – Gráfico para comparação das abordagens univariada e multivariada para o controle estatístico

Note que as regiões de controle não são as mesmas para as duas abordagens e que pontos na região "A" seriam considerados fora de controle por um dos métodos, mas na verdade estão sob controle segundo a abordagem multivariada.

Pontos na região "B" seriam considerados sob controle por um método, mas na verdade houve desvio do processo. A região "C" mostra que apesar de ambos os métodos indicarem uma situação de controle estatístico, esta não ocorre. Os eixos da elipse representam um movimento de rotação e translação adequado para melhor explicar a variabilidade dos dados. São frutos da Análise de Componentes Principais, técnica da Análise Estatística Multivariada que é muito útil como complemento aos Gráficos de Controle Multivariados.

Os procedimentos para a fase I na construção de Gráficos de Controle Multivariados têm recebido menos atenção na literatura que os da fase II. Especialmente para dados individuais, isto é, subgrupos de tamanho unitário. Contudo, são muito freqüentes os casos em que grupos de trabalho tenham coletado dados multivariados individuais e necessitem de ferramentas para dar início a um esforço de melhoria.

Objetivos e Organização da Dissertação

Esta dissertação discute os gráficos de controle para observações individuais com destaque para a fase I. O objetivo é apresentar procedimentos de detecção e diagnóstico de sinais nessa situação atentando para que possuam as seguintes propriedades: i) eficácia na detecção e diagnóstico; ii) simplicidade de implantação e iii) facilidade de interpretação. Essas são qualidades importantes para que possam ser adotados por grupos com conhecimento introdutório de estatística, auxiliados por um pacote estatístico ergonômico como, por exemplo, o MINITAB.

Trazemos no Capítulo 1 a forma multivariada dos Gráficos de Controle de Shewhart proposta por Hotelling acompanhada da discussão dos resultados importantes para seu desenvolvimento.

No Capítulo 2 abordamos o caso especial em que os subgrupos têm tamanho 1. O estimador usual da matriz de covariância tem pouco desempenho na detecção de causas especiais tipo degrau e rampa, e por isso, estimadores alternativos são discutidos. Sinais devidos a observações aberrantes são mais difíceis de serem detectados, principalmente se o afastamento não é tão evidente, e alternativas são sugeridas baseadas em técnicas de exploração da estrutura interna dos dados como a Análise de Componentes Principais. Discutimos a seguir um método simples e objetivo de interpretação de sinais para ambas as fases.

Finalmente, um exemplo de aplicação para melhoria do processo de montagem de carroçarias de caminhões é apresentado no Capítulo 3 onde as técnicas abordadas tiveram um bom desempenho na análise de dados iniciais de um esforço de melhoria.

Todos os procedimentos computacionais utilizados neste trabalho foram feitos com o pacote estatístico MINITAB (Release 13.0) com exceção de gráficos de probabilidade para a distribuição Gama que foram realizado com o pacote STATISTICA (Kernel Release 5.5 A).

Capítulo 1

Gráficos de Shewhart Multivariados

Neste Capítulo apresentamos o desenvolvimento dos gráficos equivalentes aos de Shewhart para o caso multivariado. Iniciamos com a apresentação dos principais resultados usados nesse desenvolvimento, em seguida aplicamos esses resultados para quando se dispõe de subgrupos racionais, construindo os gráficos multivariados para as fases I e II e mostrando exemplos. Gráficos de controle para a variabilidade serão discutidos em seguida. O caso de subgrupos de tamanho 1 é tratado no Capítulo 2 por necessidade de procedimentos específicos.

1.1 Resultados Importantes para o Desenvolvimento das Cartas de Controle Multivariadas

1.1.1 Representação dos Subgrupos Racionais

As cartas de controle multivariadas são desenvolvidas a partir de medições do resultado do processo para múltiplas variáveis. As medições são apresentadas, geralmente, para subgrupos de itens coletados de tal forma que estejam sob a atuação de somente causas comuns. São chamados na literatura de Subgrupos Racionais. Os subgrupos são periodicamente amostrados podendo aparecer causas especiais entre esses momentos. Para cada tipo de processo, podemos definir operacionalmente o item amostrado e as medições a serem realizadas[1]. Em industrias de peças e componentes, por exemplo, os itens podem ser as próprias peças confeccionadas ou partes de um conjunto de peças e é relativamente fácil de obtermos subgrupos de tamanho maiores que 1. Já em indústrias de processo (químicas, petroquímicas, de mineração e outras) podemos considerar um item como sendo uma

[1] Veja uma discussão sobre Definições Operacionais em Deming (1990) cap. 9

porção de material coletado instantaneamente de determinada corrente ou de um lote homogêneo de algum produto. Nesse último caso, temos geralmente subgrupos de tamanho 1 devido às restrições de amostragem.

Considere um subgrupo amostrado do processo contendo n itens e que em cada item foram medidas p variáveis de interesse. Essas medidas podem ser organizada na forma de uma matriz $p \times n$ que aqui denominaremos de $\mathbf{X}$:

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{p1} & x_{p2} & \cdots & x_{pn} \end{bmatrix}$$

1.1.2 Estatísticas Descritivas

Muitas informações desses dados podem ser obtidas através das estatísticas descritivas básicas abaixo relacionadas:

- Vetor Média do Subgrupo:

$$\bar{\mathbf{x}} = \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \vdots \\ \bar{x}_p \end{bmatrix}$$

Onde a média do subgrupo para a variável j é dada por

$$\bar{x}_j = \frac{1}{n} \sum_{i=1}^{n} x_{ij}$$

- Matriz de variâncias e covariâncias do subgrupo

$$\mathbf{S}_n = \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1p} \\ s_{21} & s_{22} & \cdots & s_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ s_{p1} & s_{p2} & \cdots & s_{pp} \end{bmatrix}$$

em que a variância amostral de uma determinada variável j é dada por

$$s_{jj} = \frac{1}{n}\sum_{i=1}^{n}(x_{ij} - \bar{x}_j)^2$$

e a covariância das variáveis j e k é

$$s_{jk} = s_{kj} = \frac{1}{n}\sum_{i=1}^{n}(x_{ij} - \bar{x}_j)(x_{ik} - \bar{x}_k)$$

O n da notação $\mathbf{S}_n$ serve como lembrete de que usamos n como divisor dos elementos de s_{jk}. Veremos adiante que o divisor $n-1$ será adequado para obtermos uma matriz de covariância não viciada para Σ a qual adotaremos simplesmente $\mathbf{S}$ como notação.

1.1.3 Distâncias e Formas Quadráticas

A maioria das técnicas do controle estatístico multivariado do processo está baseada no conceito de distância. A distância estatística, no espaço dos itens, de um ponto arbitrário $P=(x_1,x_2,...,x_p)$ até um ponto fixo $Q=(y_1,y_2,...,y_p)$, interpretada em unidades de desvio padrão é dada por

$$d(P,Q)=\sqrt{\begin{array}{l}a_{11}(x_1-y_1)^2+a_{22}(x_2-y_2)^2+\cdots+a_{pp}(x_{p1}-y_p)^2+2a_{12}(x_1-y_1)(x_2-y_2)\\+2a_{13}(x_1-y_1)(x_3-y_3)+\cdots+2a_{p-1,p}(x_{p-1}-y_{p-1})(x_p-y_p)\end{array}}$$

cujos a_{ij}'s são números tais que a distância seja não negativa para quaisquer pares de pontos. Esses a_{ij}'s podem ser acondicionados numa matriz quadrada da forma

$$\mathbf{A}=\begin{bmatrix}a_{11} & a_{12} & \cdots & a_{1p}\\ a_{12} & a_{22} & \cdots & a_{2p}\\ \vdots & \vdots & \ddots & \vdots\\ a_{1p} & a_{2p} & \cdots & a_{pp}\end{bmatrix}$$

E podemos expressar a distância quadrática entre os pontos P e Q acima numa forma matricial, mais agradável de ser trabalhada.

$$0\le[d(P,Q)]^2=(\mathbf{x}-\mathbf{y})'\mathbf{A}(\mathbf{x}-\mathbf{y})$$

Nota-se que a matriz $\mathbf{A}_{pxp}$ deve ser uma matriz simétrica positiva definida. Portanto distâncias são determinadas por formas quadráticas positivas definidas como acima.

Quaisquer formas quadráticas podem ser escritas como $\mathbf{x}'\mathbf{A}\mathbf{x}$, onde $\mathbf{x}$ é um vetor coluna de p elementos e $\mathbf{A}$ é uma matriz simétrica $p \times p$. Uma forma quadrática é chamada de positiva definida se para todos os valores de $\mathbf{x}$ exceto $\mathbf{x}=\mathbf{0}$ ela é positiva. Uma que é nunca negativa, mas pode zerar para valores não nulos de $\mathbf{x}$ é chamada de positiva semidefinida. Quaisquer matrizes de somas de quadrados e produtos cruzados, de variâncias e covariâncias, ou de correlações, são positivas definidas ou ao menos positivas semidefinidas.

A decomposição espectral de uma matriz simétrica é usada para se obter resultados envolvendo formas quadráticas. A decomposição espectral de uma matriz $\mathbf{A}_{pxp}$ é dada por:

$$\underset{(p\times p)}{\mathbf{A}} = \lambda_1 \underset{(p\times 1)(1\times p)}{\mathbf{e}_1\,\mathbf{e}_1'} + \lambda_2 \underset{(p\times 1)(1\times p)}{\mathbf{e}_2\,\mathbf{e}_2'} + \cdots + \lambda_p \underset{(p\times 1)(1\times p)}{\mathbf{e}_p\,\mathbf{e}_p'}$$

onde $\lambda_1, \lambda_2, \ldots, \lambda_p$ são os autovalores de $\mathbf{A}$ e $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_p$ são os autovetores normalizados associados, isto é, $\mathbf{e}_i'\mathbf{e}_i = 1,\ i = 1,2,\ldots,p$ e $\mathbf{e}_i'\mathbf{e}_j = 0\ \ i \neq j$.

Podemos expressar $\mathbf{A}$ da seguinte forma:

$$\mathbf{A} = \sum_{i=1}^{p} \lambda_i \mathbf{e}_i \mathbf{e}_i' = \mathbf{P}\Lambda\mathbf{P}'$$

onde $\mathbf{P} = \left[\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_p\right]$ é a matriz cujas colunas são os autovetores de $\mathbf{A}$ e trata-se de uma matriz ortonormal com as propriedades i) a soma dos quadrados dos elementos de quaisquer colunas ou linhas é igual a 1; ii) o determinante de $\mathbf{P}$, $|\mathbf{P}|$, é igual a mais ou menos 1; iii) a inversa de $\mathbf{P}$, $\mathbf{P}^{-1}$, é igual a sua transposta $\mathbf{P}'$.

$\Lambda = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_p \end{bmatrix}$ é a matriz diagonal contendo os autovalores de $\mathbf{A}$.

Com isso temos também que $\mathbf{A}^{-1} = (\mathbf{P}\Lambda\mathbf{P}')^{-1} = (\mathbf{P}')^{-1}(\Lambda)^{-1}(\mathbf{P})^{-1} = \mathbf{P}\Lambda^{-1}\mathbf{P}' = \sum_{i=1}^{p} \frac{1}{\lambda_i} \mathbf{e}_i \mathbf{e}_i'$

Veja agora que $\mathbf{x}'\mathbf{A}\mathbf{x}$, pode ser escrita da seguinte forma:

$$\mathbf{x}'\mathbf{A}\mathbf{x} = \mathbf{x}'\mathbf{P}\Lambda\mathbf{P}'\mathbf{x} = \mathbf{y}'\Lambda\mathbf{y} = \sum_{i=1}^{p} \lambda_i \mathbf{y}'\mathbf{y} = \sum_{i=1}^{p} \lambda_i y_i^2$$

como também $\mathbf{x}'\mathbf{A}^{-1}\mathbf{x}$, pode ser escrita da seguinte forma:

$$\mathbf{x}'\mathbf{A}^{-1}\mathbf{x} = \mathbf{x}'\mathbf{P}\Lambda^{-1}\mathbf{P}'\mathbf{x} = \mathbf{y}'\Lambda^{-1}\mathbf{y} = \sum_{i=1}^{p}\frac{1}{\lambda_i}\mathbf{y}'\mathbf{y} = \sum_{i=1}^{p}\frac{1}{\lambda_i}y_i^2$$

onde $y_i = \mathbf{e}_i'\mathbf{x} \quad i = 1,\cdots,p$

Se fizermos $\mathbf{A} = \mathbf{S}$, os y_i's são combinações lineares de $\mathbf{x}$ chamadas de componentes principais, com propriedades muito interessantes que serão exploradas mais adiante.

1.1.4 Vetores Aleatórios

Um vetor $\mathbf{X}$ cujos elementos são variáveis aleatórias é chamado de vetor aleatório. Os valores esperados para cada elemento do vetor aleatório constituem o vetor de médias

$$E(\mathbf{X}) = \begin{bmatrix} E(X_1) \\ E(X_2) \\ \vdots \\ E(X_p) \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix} = \boldsymbol{\mu}$$

As variâncias e covariâncias dessas p variáveis estarão contidas na matriz simétrica de variâncias e covariâncias como abaixo.

$$\begin{aligned} \boldsymbol{\Sigma} = Cov(\mathbf{X}) &= E(\mathbf{X}-\boldsymbol{\mu})(X-\boldsymbol{\mu})' \\ &= E(\mathbf{X}\mathbf{X}' - \boldsymbol{\mu}\mathbf{X}' - \mathbf{X}\boldsymbol{\mu}' + \boldsymbol{\mu}\boldsymbol{\mu}') \\ &= E(\mathbf{X}\mathbf{X}') - \boldsymbol{\mu}\boldsymbol{\mu}' \\ &= \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{bmatrix} \end{aligned}$$

onde

$$\sigma_{ii} = \sigma_i^2 = E(X_i - \mu_i)^2 \text{ e } \sigma_{ij} = \sigma_{ji} = E(X_i - \mu_i)(X_j - \mu_j)$$

1.1.5 Amostras Aleatórias

Para tomar decisões a respeito da estabilidade do processo, precisamos lançar mão de suposições sobre as observações multivariadas de que dispomos. O conceito de amostra aleatória será a primeira suposição a ser aplicada.

Suponha que os dados ainda não foram observados. Nessa situação, podemos tratar as medições como variáveis aleatórias. Uma amostra aleatória será formada se os vetores $\mathbf{X}_1$, $\mathbf{X}_2$,..., $\mathbf{X}_n$ representarem observações independentes provenientes de uma distribuição conjunta com função densidade dada por $f(\mathbf{x})=f(\mathbf{x}_1, \mathbf{x}_2,...,\mathbf{x}_p)$ que nesse caso será na verdade o produto $f(\mathbf{x}_1)f(\mathbf{x}_2)..f(\mathbf{x}_n)$ pela suposição de independência

Alguns resultados são estabelecidos a respeito das distribuições amostrais de $\overline{\mathbf{X}}$ e $\mathbf{S}$ sem fazermos outras suposições sobre a origem da distribuição conjunta das variáveis.

Resultado 1.1 - Seja $\mathbf{X}_1$, $\mathbf{X}_2$, ..., $\mathbf{X}_n$ uma amostra aleatória de tamanho n de uma distribuição conjunta com média μ e matriz de covariância Σ. Então $E(\overline{\mathbf{X}}) = \mu$ e $Cov(\overline{\mathbf{X}}) = \frac{1}{n}\Sigma$ ou seja, $\overline{\mathbf{X}}$ é um estimador não viciado para μ e sua matriz de covariância é $\frac{1}{n}\Sigma$. A respeito da matriz de covariância amostral $\mathbf{S}_n$, temos que $E(\mathbf{S}_n) = \frac{n-1}{n}\Sigma$, e $\mathbf{S} = \frac{n}{n-1}\mathbf{S}_n$ é um estimador não viciado para Σ.

Prova: veja Johnson & Wichern (1998)

1.1.6 A Distribuição Normal Multivariada

A maioria das técnicas usadas no controle estatístico multivariado de processo é baseada na suposição de multinormalidade, ou seja, de que os dados têm distribuição normal multivariada. Essa suposição, além de ser matematicamente atraente, tem dois argumentos a seu favor. O primeiro é que distribuições normais são bons modelos em muitas situações industriais. Outro fato é que estaremos lidando com amostras cujas estatísticas estarão sob o efeito do teorema central do limite.

A função densidade de probabilidade (f.d.p) normal multivariada é uma generalização da univariada para duas ou mais dimensões. Uma distribuição normal univariada de média μ e variância σ^2 tem a seguinte f.d.p.:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \qquad -\infty < x < +\infty$$

Note que o termo $\left(\frac{x-\mu}{\sigma}\right)^2$ pode ser lido como a distância quadrática de x para μ em desvios padrão e podemos escrever esse mesmo termo da seguinte forma:

$$\left(\frac{x-\mu}{\sigma}\right)^2 = (x-\mu)(\sigma^2)^{-1}(x-\mu)$$

que pode ser generalizado para p dimensões da seguinte maneira:

$$(\mathbf{x}-\boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x}-\boldsymbol{\mu})$$

A f.d.p. normal multivariada é obtida através da generalização acima e do ajuste da constante na f.d.p univariada para a obtenção de volumes quando da sua integração para o cálculo de probabilidades. Veja a f.d.p normal multivariada abaixo.

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} e^{-\frac{1}{2}(\mathbf{x}-\mu)'\Sigma^{-1}(\mathbf{x}-\mu)} \qquad -\infty < x < +\infty, \qquad i = 1,2,\ldots,p$$

Dessa expressão podemos ver que os contornos de densidade de probabilidade constante são gerados quando o termo da exponencial for constante, isto é,

$$f(\mathbf{x}) = cte \Leftrightarrow (\mathbf{x}-\mu)' \Sigma^{-1}(\mathbf{x}-\mu) = c^2$$

Pode-se deduzir que os contornos têm a forma de um elipsóide centrado em μ cujos eixos estão na direção dos autovetores de Σ e seus tamanhos são proporcionais às raízes quadradas dos autovalores de Σ, ou seja os eixos são

$$\pm c\sqrt{\lambda_i}\mathbf{e}_i \text{ onde } \Sigma\mathbf{e}_i = \lambda_i\mathbf{e}_i \quad i = 1,2,\cdots,p$$

Outros resultados importantes são verdadeiros para um vetor aleatório $\mathbf{X}$ que tem distribuição normal multivariada, (Johnson & Wichern, 1998):

- Combinações lineares dos componentes de $\mathbf{X}$ são distribuídos normalmente;
- Qualquer subconjunto de componentes de $\mathbf{X}$ tem distribuição normal multivariada;
- Covariância zero implica que os componentes correspondentes são independentes;
- Distribuições condicionais dos componentes são normais multivariadas;

1.1.7 Distribuições Amostrais Importantes

Primitivamente, o cálculo dos limites de controle para um gráfico de controle multivariado envolve a escolha adequada de c^2 e para isso precisamos conhecer a distribuição de $(\mathbf{x}-\mu)' \Sigma^{-1}(\mathbf{x}-\mu)$ que resulta do seguinte teorema:

Teorema 1.1 - Se o vetor de m componentes $\mathbf{Y}$ é distribuído de acordo com uma $N(\nu,\mathbf{T})$ não singular, então $\mathbf{Y'T^{-1}Y}$ é distribuído de acordo com uma distribuição χ^2 não central com m graus de liberdade e parâmetro de não-centralidade $\nu'\mathbf{T}^{-1}\nu$. Se $\nu=\mathbf{0}$, a distribuição é uma χ^2 central.

Prova, Anderson (1958):

Seja $\mathbf{C}$ uma matriz não singular tal que $\mathbf{CTC'=I}$ e defina $\mathbf{Z=CY}$. Então $\mathbf{Z}$ é normalmente distribuído com média $E(\mathbf{Z})=\mathbf{C}E(\mathbf{Y})=\mathbf{C}\nu=\lambda$ e matriz de covariância $E(\mathbf{Z}-\lambda)(\mathbf{Z}-\lambda)' = E\mathbf{C}(\mathbf{Y}-\nu)(\mathbf{Y}-\nu)'\mathbf{C'}=\mathbf{CTC'}=\mathbf{I}$.

Então $\mathbf{Y'T^{-1}Y}=\mathbf{Z'(C')^{-1}T^{-1}C^{-1}Z}=\mathbf{Z'(CTC')^{-1}Z}=\mathbf{Z'Z}$ que é uma soma de quadrados dos componentes de $\mathbf{Z}$. Similarmente temos que $\nu'\mathbf{T}^{-1}\nu=\lambda'\lambda$. Portanto $\mathbf{Y'T^{-1}Y}$ é distribuído como $\sum_{i=1}^{m} Z_i^2$ onde Z_1, ..., Z_m são independentes e normalmente distribuídos com médias $\lambda_1,...,\lambda_m$, respectivamente, e variâncias 1. Por definição, essa distribuição é uma χ^2 não central com parâmetro de não centralidade $\sum_{i=1}^{m}\lambda_i^2$ Se $\lambda_1=\cdots=\lambda_m=0$ a distribuição é central.

Ao tratarmos um subgrupo de tamanho n como uma amostra aleatória, estaremos interessados nos elipsóides de contorno para $\overline{\mathbf{X}}$. Vejamos, então, como são distribuídos os vetores amostrais $\overline{\mathbf{X}}$ e $\mathbf{S}$ no caso de multinormalidade.

Resultado 1.2 - Seja $\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_n$ uma amostra aleatória de tamanho n de uma distribuição normal p-variada com média μ e matriz de covariância Σ. Então

- $\overline{\mathbf{X}}$ tem distribuição $N_p\left(\mu, \frac{1}{n}\Sigma\right)$;

- $(n-1)\mathbf{S}$ tem distribuição Wishart com $n-1$ graus de liberdade, isto é

$$(n-1)\mathbf{S} \sim W_p(n-1, \Sigma) = \sum_{j=1}^{n-1} \mathbf{Z}_j \mathbf{Z}_j' \quad \text{onde} \quad \mathbf{Z}_j \sim N_p(\mathbf{0}, \Sigma)$$

- $\overline{\mathbf{X}}$ e $\mathbf{S}$ são independentes

Aqui temos o primeiro resultado que irá nos levar ao cálculo do limite de controle para cartas de controle multivariadas de médias. Desde que $\sqrt{n}(\overline{\mathbf{X}} - \mu) \sim N_p(\mathbf{0}, \Sigma)$ segue do Teorema 1.1 que $n(\overline{\mathbf{X}} - \mu)' \Sigma^{-1}(\overline{\mathbf{X}} - \mu)$ tem uma distribuição χ^2 com p graus de liberdade. Note que esse resultado só tem lugar quando a matriz de variâncias e covariâncias for conhecida. Na seção 1.2 desenvolveremos o caso em que ela é desconhecida.

O teorema central do limite confere robustez para $\overline{\mathbf{X}}$ em relação à suposição de multinormalidade no caso de amostras grandes em relação a p. O teorema pode ser formulado da seguinte forma:

Resultado 1.3 - (teorema central do limite) - Seja $\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_n$ observações independentes de qualquer população com média μ e matriz de covariância finita Σ. Então $\sqrt{n}(\overline{\mathbf{X}} - \mu)$ é distribuído aproximadamente como uma $N_p(\mathbf{0}, \Sigma)$ para grandes amostras ($n>>p$). *Prova.* Veja Anderson (1958)

1.2 Controle de Processo e Inferência Estatística

Verossimilhança é um conceito fundamental na teoria da estatística. Esse conceito pode ser usado para controle de processos resultando em procedimentos de decisão para situações em que populações estão sujeitas a mudanças repentinas de seus parâmetros. Esses procedimentos deveriam evitar o disparo de sinais quando o comportamento das observações do processo pudessem ser descritos por um vetor de parâmetros pertinente a um conjunto ω desejável, o que se caracterizaria como um falso alarme. Por outro lado, caso o vetor de parâmetros sofresse uma mudança e vier a pertencer a um indesejável conjunto Ω, então os procedimentos deveriam sinalizar assim que possível. Portanto, o problema de controlar processos se apresenta como sucessivos testes de hipóteses sobre os seus parâmetros. Apresentamos a seguir o desenvolvimento de testes sobre os parâmetros μ e Σ no caso multivariado.

1.2.1 A Estatística T^2 de Hotelling

Considere o problema de se testar as seguintes hipóteses

$$H_0: \mu = \mu_0 \text{ contra } H_1: \mu \neq \mu_0$$

Se dispusermos de uma amostra aleatória $X_1, X_2, ..., X_n$ a estatística do teste será

$$t = \frac{\overline{X} - \mu_0}{s/\sqrt{n}} \text{ onde } \overline{X} = \frac{1}{n}\sum_{j=1}^{n} X_j \text{ e } s^2 = \frac{1}{n-1}\sum_{j=1}^{n}\left(X_j - \overline{X}\right)^2$$

A estatística t tem distribuição *t-student* com n-1 graus de liberdade e rejeitamos H_0 se $|t| > t_{n-1}(\alpha/2)$ em que $t_{n-1}(\alpha/2)$ é o (100α)-ésimo percentil superior de tal distribuição e α

é o tamanho do teste considerado. Agora, rejeitar H_0 quando $|t|$ é grande é equivalente a rejeitá-la quando t^2 for elevado.

Podemos escrever t^2 como a distância quadrática de $\overline{X}$ ao valor μ_0 do teste em unidades de desvio padrão de $\overline{X}$, isto é, $s/\sqrt{n}$

$$t^2 = \frac{(\overline{X} - \mu_0)^2}{s^2/n} = n(\overline{X} - \mu_0)(s^2)^{-1}(\overline{X} - \mu_0)$$

Uma generalização natural dessa distância quadrática para o caso multivariado será

$$T^2 = n(\overline{\mathbf{X}} - \mu_0)' \mathbf{S}^{-1}(\overline{\mathbf{X}} - \mu_0)$$

Essa estatística é chamada de T^2 de Hotelling em homenagem a Harold Hotelling, quem primeiro obteve sua distribuição amostral. Veremos sua dedução formal e a sua distribuição, já que tal estatística tem papel central nos gráficos de controle multivariados.

A estatística T^2 pode ser derivada a partir do teste da razão de verossimilhança para $H_0: \mu = \mu_0$.

Para tanto, partiremos da dedução dos estimadores de máxima verossimilhança (e.m.v.) para μ e Σ no caso de multinormalidade cujos resultados aplicaremos ao teste da razão de verossimilhança.

1.2.2 Os Estimadores de Máxima Verossimilhança para μ e Σ

Suponha que tenhamos uma amostra de tamanho n de uma distribuição normal p-variada com média μ e matriz de covariância Σ cujos valores obtidos foram $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$. A função

verossimilhança na qual os vetores $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ são fixos é uma função de $\boldsymbol{\mu}$ e $\boldsymbol{\Sigma}$, que aqui são variáveis pertencentes ao espaço paramétrico (restrito a matrizes positivas definidas para $\boldsymbol{\Sigma}$). A função tem a seguinte forma

$$L = \prod_{i=1}^{n} n(\mathbf{x}_i | \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{\frac{1}{2}pn} |\boldsymbol{\Sigma}|^{\frac{1}{2}n}} \exp\left[-\frac{1}{2} \sum_{i=1}^{n} (\mathbf{x}_i - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x}_i - \boldsymbol{\mu}) \right]$$

O máximo de L será no mesmo ponto para $\log L$, já que esta é uma função crescente de L. Tomando o logaritmo de L temos

$$\log L = -\frac{1}{2} pn \log(2\pi) - \frac{1}{2} n \log|\boldsymbol{\Sigma}| - \frac{1}{2} \sum_{i=1}^{n} (\mathbf{x}_i - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x}_i - \boldsymbol{\mu})$$

Procuremos, portanto, pelo máximo dessa função no espaço paramétrico cujo resultado nos dará os estimadores de máxima verossimilhança para $\boldsymbol{\mu}$ e $\boldsymbol{\Sigma}$.

Seja $\mathbf{A}$ a matriz de somas de quadrados e produtos cruzados dos desvios da média

$$\mathbf{A} = \sum_{i=1}^{n} (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})'$$

Pode-se mostrar que

$$\mathbf{A} = \sum_{i=1}^{n} (\mathbf{x}_i - \boldsymbol{\mu})(\mathbf{x}_i - \boldsymbol{\mu})' - n(\bar{\mathbf{x}} - \boldsymbol{\mu})(\bar{\mathbf{x}} - \boldsymbol{\mu})'$$

Usando o fato de que tr$\mathbf{CD}$ = tr$\mathbf{DC}$, podemos escrever o último termo de $\log L$ da seguinte maneira

$$\begin{aligned}\sum_{i=1}^{n}(\mathbf{x}_i-\mu)'\Sigma^{-1}(\mathbf{x}_i-\mu) &= \mathrm{tr}\sum_{i=1}^{n}(\mathbf{x}_i-\mu)'\Sigma^{-1}(\mathbf{x}_i-\mu)\\ &= \mathrm{tr}\sum_{i=1}^{n}\Sigma^{-1}(\mathbf{x}_i-\mu)(\mathbf{x}_i-\mu)'\\ &= \mathrm{tr}\Sigma^{-1}\mathbf{A}+\mathrm{tr}\Sigma^{-1}n(\bar{\mathbf{x}}-\mu)(\bar{\mathbf{x}}-\mu)'\\ &= \mathrm{tr}\Sigma^{-1}\mathbf{A}+n(\bar{\mathbf{x}}-\mu)'\Sigma^{-1}(\bar{\mathbf{x}}-\mu)\end{aligned}$$

Assim, $\log L$ ficará, então

$$\log L = -\frac{1}{2}pn\log(2\pi)-\frac{1}{2}n\log|\Sigma|-\frac{1}{2}\mathrm{tr}\Sigma^{-1}\mathbf{A}-\frac{1}{2}n(\bar{\mathbf{x}}-\mu)'\Sigma^{-1}(\bar{\mathbf{x}}-\mu)$$

Já podemos afirmar que o valor de μ que maximiza L é $\bar{\mathbf{x}}$ pois desde que Σ é positiva definida, Σ^{-1} também será, assim, $n(\bar{\mathbf{x}}-\mu)'\Sigma^{-1}(\bar{\mathbf{x}}-\mu)\geq 0$, assumindo zero se e somente se $\bar{\mathbf{x}}=\mu$

Para maximizar o segundo e terceiro termos de $\log L$ usaremos o seguinte resultado

Lema 1.1 Se $\mathbf{D}$ é positiva definida de ordem p, o máximo de

$$f(\mathbf{G}) = -n\log|\mathbf{G}|-\mathrm{tr}\mathbf{G}^{-1}\mathbf{D}$$

com respeito à matriz positiva definida $\mathbf{G}$ existe, ocorre em $\mathbf{G}=\frac{1}{n}\mathbf{D}$ e tem o valor

$$f\left(\frac{1}{n}\mathbf{D}\right) = -pn\log n - n\log|\mathbf{D}|-pn$$

Prova. Ver Anderson (1958).

Portanto log L é maximizada para $\Sigma = \frac{1}{n}\mathbf{A}$. Assim, podemos expressar o seguinte resultado

Teorema 1.2 Se $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n$ constituem uma amostra de uma distribuição normal p-variada $N_p(\mu,\Sigma)$ com $p < n$, os estimadores de máxima verossimilhança de μ e Σ são

$$\hat{\mu} = \bar{\mathbf{x}} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i \quad \text{e} \quad \hat{\Sigma} = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})' = \frac{(n-1)}{n}\mathbf{S}$$

e o máximo para a função verossimilhança será

$$\max_{\mu,\Sigma} L = \frac{1}{(2\pi)^{\frac{1}{2}np}\left|\hat{\Sigma}\right|^{\frac{1}{2}n}}\exp\left[-\frac{1}{2}np\right]$$

1.2.3 O Teste da Razão de Verossimilhança para a Média

O critério da razão de verossimilhança para o teste de $H_0: \mu = \mu_0$ é dado por

$$\lambda = \frac{\max_{\Sigma} L(\mu_0, \Sigma)}{\max_{\mu,\Sigma} L(\mu, \Sigma)} \quad \text{onde } L \text{ é a função verossimilhança}$$

Essa razão tem como numerador o máximo de L no espaço paramétrico restrito pela hipótese nula (ω), já no denominador o espaço é integral (Ω), observando que Σ deva ser positiva definida. Portanto, os estimadores de máxima verossimilhança são aqueles que maximizam o denominador, ou seja

$$\hat{\mu}_{\Omega} = \bar{\mathbf{x}}$$

$$\hat{\Sigma}_{\Omega} = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})'$$

Pelo Lema 1.1, em ω a função verossimilhança é maximizada em

$$\hat{\Sigma}_{\omega} = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{x}_i - \mu_0)(\mathbf{x}_i - \mu_0)'$$

O critério da razão de verossimilhança fica, portanto

$$\lambda = \frac{\max_{\Sigma} L(\mu_0, \Sigma)}{\max_{\mu,\Sigma} L(\mu, \Sigma)} = \frac{\frac{1}{(2\pi)^{\frac{1}{2}pn}\left|\hat{\Sigma}_{\omega}\right|^{\frac{1}{2}n}} \exp\left[-\frac{1}{2}\sum_{i=1}^{n}(\mathbf{x}_i - \mu_0)'\hat{\Sigma}_{\omega}^{-1}(\mathbf{x}_i - \mu_0)\right]}{\frac{1}{(2\pi)^{\frac{1}{2}pn}\left|\hat{\Sigma}_{\Omega}\right|^{\frac{1}{2}n}} \exp\left[-\frac{1}{2}\sum_{i=1}^{n}(\mathbf{x}_i - \mu)'\hat{\Sigma}_{\Omega}^{-1}(\mathbf{x}_i - \mu)\right]}$$

No desenvolvimento dos e.m.v. vimos como achar os máximos para a função verossimilhança normal, e aplicando agora tanto para o numerador quanto para o denominador teremos

$$\lambda = \frac{\frac{1}{(2\pi)^{\frac{1}{2}np}\left|\hat{\Sigma}_{\varpi}\right|^{\frac{1}{2}n}} \exp\left[-\frac{1}{2}np\right]}{\frac{1}{(2\pi)^{\frac{1}{2}np}\left|\hat{\Sigma}_{\Omega}\right|^{\frac{1}{2}n}} \exp\left[-\frac{1}{2}np\right]} = \frac{\left|\hat{\Sigma}_{\Omega}\right|^{\frac{1}{2}n}}{\left|\hat{\Sigma}_{\omega}\right|^{\frac{1}{2}n}} =$$

$$= \frac{\left|\sum_{i=1}^{n}(\mathbf{x}_i - \mu)(\mathbf{x}_i - \mu)'\right|}{\left|\sum_{i=1}^{n}(\mathbf{x}_i - \mu_0)(\mathbf{x}_i - \mu_0)'\right|} = \frac{\left|\mathbf{A}\right|^{\frac{1}{2}n}}{\left|\mathbf{A} + (\bar{\mathbf{x}} - \mu_0)(\bar{\mathbf{x}} - \mu_0)'\right|^{\frac{1}{2}n}}$$

Agora, pode-se mostrar que para uma matriz **C** não singular,

$$|\mathbf{C}+\mathbf{y}\mathbf{y}'| = |\mathbf{C}|\left(1+\mathbf{y}'\mathbf{C}^{-1}\mathbf{y}\right)$$

Portanto, temos que

$$\begin{aligned}\lambda^{\frac{2}{n}} &= \frac{|\mathbf{A}|}{\left|\mathbf{A}+\left[\sqrt{n}\left(\bar{\mathbf{x}}-\mu_0\right)\right]\left[\sqrt{n}\left(\bar{\mathbf{x}}-\mu_0\right)\right]'\right|} \\ &= \frac{1}{1+n\left(\bar{\mathbf{x}}-\mu_0\right)'\mathbf{A}^{-1}\left(\bar{\mathbf{x}}-\mu_0\right)} \\ &= \frac{1}{1+T^2/n-1}\end{aligned}$$

A região crítica para o teste da razão de verossimilhança é dada por

$$\lambda \le \lambda_0 \text{ e } P\left(\lambda \le \lambda_0 | H_0\right) = \alpha$$

Portanto, equivalente a

$$T^2 \ge T_0^2$$

onde

$$T_0^2 = (n-1)\left(\lambda_0^{-2/n}-1\right) \text{ e } P\left(T^2 \ge T_0^2 | H_0\right) = \alpha$$

1.2.4 A Distribuição de T^2

Como já dissemos, T^2 tem uma distribuição conhecida e é dada pelo seguinte teorema

Teorema 1.3 - Seja $T^2 = \mathbf{Y}'\mathbf{S}\mathbf{Y}$, onde $\mathbf{Y}$ é distribuído de acordo com uma $N_p(\nu,\Sigma)$ e $n\mathbf{S}$ é independentemente distribuída como $\sum_{i=1}^{n}\mathbf{Z}_i\mathbf{Z}_i'$ com $\mathbf{Z}_1$, ..., $\mathbf{Z}_n$ independentes cada qual com distribuição $N_p(0,\Sigma)$. Então $(T^2/n)\left[(n-p+1)/p\right]$ é distribuído como uma F não central com p e n-p+1 graus de liberdade e parâmetro de não centralidade $\nu'\Sigma^{-1}\nu$. Se $\nu = 0$, a distribuição é uma F central.

Prova. Anderson (1958)

Esse é um teorema chave para o estabelecimento de limites dos os gráficos de controle para o caso multivariado cuja construção é baseada no teste de hipóteses para a média.

Aplicando diretamente esse teorema temos que o teste de hipóteses $H_0{:}\mu = \mu_0$ versus $H_1{:}\mu \neq \mu_0$,ao nível de significância α, rejeita H_0 em favor de H_1 se

$$T^2 = n\left(\overline{\mathbf{X}} - \mu_0\right)' \mathbf{S}^{-1}\left(\overline{\mathbf{X}} - \mu_0\right) > \frac{(n-1)p}{(n-p)} F_{p,n-p}(\alpha)$$

onde a F é central, pois sob H_0 temos que $\sqrt{n}\left(\overline{\mathbf{X}} - \mu_0\right) \sim N_p(0,\Sigma)$ e $(n-1)\mathbf{S} = \sum_{i=1}^{n-1}\mathbf{Z}_i\mathbf{Z}_i'$, cujos $\mathbf{Z}_i$'s têm distribuição $N_p(0,\Sigma)$.

1.2.5 A Região de Confiança para a Média

Os limites de controle nos gráficos multivariados serão derivados de cálculos de regiões de confiança para o vetor de médias. Veremos com isso, que o uso de p gráficos de controle individuais é conceitualmente errado pois inflaciona-se o erro tipo I projetado para o controle. Erro tipo I, nesse caso, será a probabilidade de considerarmos um ponto fora de controle quando, na verdade, ele pertence à distribuição original do processo.

Uma região de confiança é uma região de valores verossímeis para o parâmetro desconhecido. Do teorema 1.3, podemos afirmar que, para uma amostra de tamanho n de uma população $N_p(\mu, \Sigma)$,

$$P\left[n(\overline{\mathbf{X}} - \mu)' \mathbf{S}^{-1} (\overline{\mathbf{X}} - \mu) \leq \frac{(n-1)p}{(n-p)} F_{p,n-p}(\alpha) \right] = 1 - \alpha$$

Assim a 100(1-α)% região de confiança para μ de uma distribuição normal p-variada é o conjunto de todos os valores de μ tais que

$$n(\overline{\mathbf{X}} - \mu)' \mathbf{S}^{-1} (\overline{\mathbf{X}} - \mu) \leq \frac{(n-1)p}{(n-p)} F_{p,n-p}(\alpha)$$

Para mais de 4 componentes as regiões elipsóides acima não podem ser representadas num gráfico, porém, podemos calcular seus eixos e comprimentos relativos.

Tendo $\overline{\mathbf{x}}$ como centro, os eixos do elipsóide de confiança são

$$\pm \left[\sqrt{\lambda_i} \sqrt{\frac{p(n-1)}{n(n-p)} F_{p,n-p}(\alpha)} \right] \mathbf{e}_i \quad \text{onde} \quad \mathbf{S}\mathbf{e}_i = \lambda_i \mathbf{e}_i \quad i = 1,2,\cdots,p$$

1.2.6 Informações Sobre os Componentes Individuais

As regiões de confiança levam em conta a estrutura de correlação entre as variáveis e dão acesso correto ao conhecimento conjunto do comportamento das mesmas. No entanto, nenhuma informação pode ser obtida em relação aos componentes individuais. Isso torna-se um revés na aplicação dessa teoria ao controle de processos pois é necessário identificar

qual ou quais variáveis são as responsáveis por uma causa assinalável ocorrida no uso de cartas multivariadas para eliminá-la.

Uma primeira tentativa de análise das variáveis individuais pode ser feita através de declarações simultâneas de confiança, em que os intervalos univariados devem ser alargados para manter o mesmo critério de confiança das regiões estudadas anteriormente. Vejamos o que acontece caso não os alargarmos.

Os intervalos univariados mencionados são construídos independentemente, sem levar em conta a distribuição conjunta das variáveis e sua estrutura de correlação. Podemos expressar esses intervalos como segue:

$$\bar{x}_i - t_{n-1}(\alpha/2)\sqrt{\frac{s_{ii}}{n}} \leq \mu_i \leq \bar{x}_i + t_{n-1}(\alpha/2)\sqrt{\frac{s_{ii}}{n}} \qquad i = 1,2,\cdots,p$$

Nesse caso, sabemos que o i-ésimo intervalo tem $1\text{-}\alpha$ de probabilidade de cobrir o verdadeiro valor μ_i, antes de ser retirada a amostra. Porém não sabemos qual a probabilidade de todos os intervalos conjuntamente cobrirem seus verdadeiros parâmetros. Sabemos que não será $1\text{-}\alpha$, veja por exemplo, o caso especial em que as variáveis são independentes. Antes de obtermos a amostra, temos que

$$\begin{aligned} P[\text{todos os } p \text{ - intervalos conterem seus parâmetros}] &= (1-\alpha)(1-\alpha)\cdots(1-\alpha) \\ &= (1-\alpha)^p \end{aligned}$$

Assim, para 10 variáveis controladas, ajustando-se α = 0,0027 como freqüentemente é usado pelas empresas, 1- α = 0,9973 e a probabilidade acima será $(0,9973)^{10}$ = 0,9733 Nesse caso, quando se trabalha com testes de hipóteses, o erro tipo I é inflacionado de 0,27% para 2,6%, causando perdas por paradas excessivas na produção sem necessidade.

As declarações simultâneas de confiança alargam, como dissemos anteriormente, os intervalos individuais para manter o erro tipo I especificado. Vejamos como são construídos.

Se $\mathbf{X}$ é distribuído como uma $N_p(\mu, \Sigma)$, uma combinação linear $Z = l'\mathbf{X}$ tem distribuição normal com média e variância $\mu_Z = E(Z) = l'\mu$ e $\sigma_Z^2 = \text{Var}(Z) = l'\Sigma l$ respectivamente. Uma variável i em particular, por exemplo, é uma combinação linear de $\mathbf{X}$ com $l' = [0, \cdots, 0, l_i, 0, \cdots, 0]$ em que $l_i = 1$. Dispondo de uma amostra aleatória $\mathbf{X}_1$, $\mathbf{X}_2$, ..., $\mathbf{X}_n$ Poder-se-ia construir um intervalo de confiança (1-α)% para uma particular combinação linear escolhida dentre as infinitas possíveis escolhas de l. O intervalo seria

$$l'\bar{\mathbf{x}} - t_{n-1}(\alpha/2)\sqrt{\frac{l'\mathbf{S}l}{n}} \le l'\mu \le l'\bar{\mathbf{x}} + t_{n-1}(\alpha/2)\sqrt{\frac{l'\mathbf{S}l}{n}}$$

Ou seja, esse intervalo corresponde ao conjunto de valores de $l'\mu$ tal que

$$|t| = \left|\frac{\sqrt{n}l'(\bar{\mathbf{x}} - \mu)}{\sqrt{l'\mathbf{S}l}}\right| \le t_{n-1}(\alpha/2) \text{ ou ainda, } t^2 = \frac{n[l'(\bar{\mathbf{x}} - \mu)]^2}{l'\mathbf{S}l} \le t_{n-1}^2(\alpha/2)$$

Dessa maneira, poderíamos construir mais intervalos para outras escolhas de l, mas a confiança associada com todos eles tomados na mesma amostra não seria 1-α como já discutimos.

É razoável pensar em aumentar a constante $t_{n-1}(\alpha/2)$ para um valor em que possamos construir intervalos para mais de um parâmetro.

Estaremos em busca do valor máximo de t^2 assumido quando qualquer escolha de l seja possível, ou seja, queremos determinar

$$\max_{l} t^2 = \max_{l} \frac{n\left[l'(\bar{\mathbf{x}} - \boldsymbol{\mu})\right]^2}{l'\mathbf{S}l}$$

Pode-se demonstrar (veja Johnson & Wichern, 1998) que esse valor é justamente T^2 de Hotelling. Portanto os intervalos simultâneos para todas as p variáveis terão a forma

$$\bar{x}_i - \sqrt{\frac{p(n-1)}{(n-p)} F_{n,n-p}(\alpha)} \sqrt{\frac{s_{ii}}{n}} \leq \mu_i \leq \bar{x}_i + \sqrt{\frac{p(n-1)}{(n-p)} F_{n,n-p}(\alpha)} \sqrt{\frac{s_{ii}}{n}} \quad i = 1,2,\cdots,p$$

Note, que esses intervalos conjuntamente produzem um coeficiente de confiança maior ou igual a 1-α. Muito provavelmente maior, portanto acaba-se trabalhando de maneira muito conservativa com eles. O método Bonferroni para comparações múltiplas melhora essa situação para os casos em que não se têm muitos intervalos a construir.

1.2.7 O Método Bonferroni para Comparações Múltiplas

O método Bonferroni é o primeiro método prático para tentar identificar as causas de um ponto fora de controle em gráficos multivariados.

Suponha que queiramos realizar intervalos de confiança conjuntamente para m combinações lineares $l_1'\boldsymbol{\mu}, l_2'\boldsymbol{\mu}, \cdots, l_m'\boldsymbol{\mu}$, antes de coletar os dados. Seja C_i a declaração de confiança a respeito do valor de $l_i'\boldsymbol{\mu}$ com $P[C_i \text{ verdadeira}] = 1\text{-}\alpha_i$, $i = 1,2,\cdots,m$, assim,

$$\begin{aligned} P[\text{todas as } C_i \text{ verdadeiras}] &= 1 - P[\text{pelo menos uma } C_i \text{ falsa}] \\ &\geq 1 - \sum_{i=1}^{m} P(C_i \text{ falsa}) = 1 - \sum_{i=1}^{m} \left[1 - P(C_i \text{ verdadeira})\right] \\ &= 1 - (\alpha_1 + \alpha_2 + \cdots + \alpha_m) \end{aligned}$$

É razoável se escolher $\alpha_i = \frac{\alpha}{m}$ $i = 1,2,\cdots,m$ para facilitar o controle do erro tipo I no caso prático. Portanto para as nossas p variáveis, os intervalos individuais dados pelo método de Bonferroni serão:

$$\bar{x}_i - t_{n-1}(\alpha/2p)\sqrt{\frac{s_{ii}}{n}} \le \mu_i \le \bar{x}_i + t_{n-1}(\alpha/2p)\sqrt{\frac{s_{ii}}{n}} \qquad i = 1,2,\cdots,p$$

1.2.8 O Teste da Razão de Verossimilhança para a Matriz de Covariância

Não somente podem haver mudanças bruscas no vetor de médias, mas a variabilidade pode ser alvo de tais mudanças. Veremos que esquemas de controle para a variabilidade podem ser desenvolvidos e um deles baseia-se no teste de hipótese de que a matriz de covariância é igual a uma dada matriz. O seguinte resultado será usado nesse desenvolvimento.

Teorema 1.4 Dados $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n$ como observações p-variadas de uma N($\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$), o critério da razão de verossimilhança para testar a hipótese, onde $\boldsymbol{\Sigma}_0$ é especificada é:

$$\lambda_1 = \left(\frac{e}{n}\right)^{\frac{np}{2}} \left|\mathbf{A}\boldsymbol{\Sigma}_0^{-1}\right|^{\frac{n}{2}} e^{-\frac{1}{2}tr\left(\mathbf{A}\boldsymbol{\Sigma}_0^{-1}\right)} \text{ onde } \mathbf{A} = \sum_{i=1}^{n}(\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})'$$

A distribuição assintótica de $-2\ln\lambda_1$ será uma χ^2 com $\frac{1}{2}p(p+1)$ graus de liberdade

Prova: Anderson (1959)

1.2.9 A Variância Generalizada

Nos critérios dos testes da razão de verossimilhança aparece o determinante de Σ, $|\Sigma|$. Esse escalar é conhecido como variância generalizada da distribuição multivariada. Para uma amostra, temos $|\mathbf{S}|$ por similaridade. Tratam-se, de certa forma, de medidas de dispersão dos dados. O seguinte resultado nos traz uma interpretação da variância generalizada:

Teorema 1.5 Seja $|\mathbf{S}|$ a variância generalizada amostral onde $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n$, são n vetores de uma amostra. Então $|\mathbf{S}|$ é proporcional à soma dos quadrados dos volumes de todos os paralelepípedos formados usando como arestas principais p vetores com extremidades $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n$ de um lado e $\bar{\mathbf{x}}$ do outro, e o fator de proporcionalidade é $1/(n-1)^p$.

Isso quer dizer que a variância generalizada para um conjunto fixo de dados é proporcional ao quadrado do volume gerado pelos p vetores desvio. Intuitivamente sabemos que quanto maior o tamanho desses vetores e mais abertos forem os ângulos entre eles maior será a dispersão dos dados e a variância generalizada segue esse princípio. Porém, em Johnson & Wichern (1998) há um alerta para uma fraqueza básica da variância generalizada com respeito à estrutura de correlação das variáveis. Veja o exemplo que ele traz. As matrizes $\mathbf{S}_1$, $\mathbf{S}_2$ e $\mathbf{S}_3$ abaixo têm o mesmo valor como determinante, porém a estrutura de correlação completamente distintas:

$$\mathbf{S}_1 = \begin{bmatrix} 5 & 4 \\ 4 & 5 \end{bmatrix} \quad \mathbf{S}_2 = \begin{bmatrix} 5 & -4 \\ -4 & 5 \end{bmatrix} \qquad \mathbf{S}_3 = \begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix}$$

Anderson (1959) traz um resultado básico quanto à distribuição da variância generalizada. Vejamo-lo:

Teorema 1.6 A distribuição da variância generalizada $|\mathbf{S}|$ de uma amostra $\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_n$ proveniente de uma distribuição $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ é a mesma da distribuição de $|\boldsymbol{\Sigma}|/(n-1)^p$ vezes o produto de p fatores independentes, sendo que a distribuição do i-ésimo fator é χ^2 com $n-i$ graus de liberdade.

Para $p = 1$ ou 2, as distribuições exatas de $|\mathbf{S}|$ podem ser obtidas mas para altos valores de p, as integrais envolvidas são pouco manipuláveis. Seguem então os resultados para $V = |\mathbf{A}|/|\boldsymbol{\Sigma}|$ no caso $p = 1$ e 2

Para $p = 1$, temos facilmente que $V \sim \chi^2_{(n-1)}$ (caso univariado). Para $p = 2$, temos que $2V^{\frac{1}{2}} \sim \chi^2_{(2n-4)}$. Por outro lado, os momentos de $|\mathbf{S}|$ podem ser obtidos, vejamos como isso se dá:

Sabendo que $|\mathbf{S}| = |\mathbf{A}|/(n-1)^p$ e usando o teorema 1.6 temos que $|\mathbf{A}| = |\boldsymbol{\Sigma}| \cdot \chi^2_{n-1} \cdot \chi^2_{n-2} \cdot \cdots \cdot \chi^2_{n-p}$. Desde que o h-ésimo momento de uma distribuição χ^2_m é $2^h \Gamma(\frac{1}{2}m + h)/\Gamma(\frac{1}{2}m)$ e também que o momento do produto de variáveis independentes é o produto dos momentos das variáveis, o h-ésimo momento de $|\mathbf{A}|$ é

$$|\boldsymbol{\Sigma}|^h \prod_{i=1}^{p} \left(2^h \frac{\Gamma[\frac{1}{2}(n-i)+h]}{\Gamma[\frac{1}{2}(n-i)]} \right).$$

Assim,

$$E|\mathbf{A}| = |\boldsymbol{\Sigma}| \prod_{i=1}^{p} (n-i)$$

$$E|\mathbf{A}|^2 = |\Sigma|^2 \prod_{i=1}^{p} (n-i-2)(n-i)$$

Portanto, desde que $Var|\mathbf{A}| = E|\mathbf{A}|^2 - (E|\mathbf{A}|)^2$, temos que[2]

$$Var(|\mathbf{A}|) = |\Sigma|^2 \prod_{i=1}^{p} (n-i)\left[\prod_{j=1}^{p}(n-j+2) - \prod_{j=1}^{p}(n-j)\right]$$

Alternativamente, a distribuição assintótica da variância generalizada é dada pelo seguinte resultado:

Teorema 1.7 Seja **S** uma matriz de covariância p x p com n graus de liberdade. Então $\sqrt{n}(|\mathbf{S}|/|\Sigma| - 1)$ tem distribuição assintótica normal com média 0 e variância $2p$.

Esses resultados acima serão explorados nas alternativas de controle da variabilidade mais adiante.

---

[2] Contrastando com o que está apresentado em Anderson (1959) - pp.170 equação (20) - a função gama não aparece nessa expressão.

1.3 Gráficos de Controle para Subgrupos Racionais

A construção de gráficos de controle é separada em duas fases. A primeira (fase I) consiste da análise de dados passados com o objetivo de verificar se o processo estava sob controle estatístico quando os primeiros subgrupos foram obtidos, bem como retirar os pontos suspeitos de estarem fora de controle e estimar a distribuição do processo para teste de pontos futuros. A fase seguinte (fase II) consiste em usar o gráfico de controle para detectar qualquer fuga do processo em relação à sua distribuição estimada na fase I.

Nesta seção, veremos como aplicar a teoria exposta anteriormente para uso dos gráficos de controle cuja construção é diferenciado para as fases I e II.

1.3.1 Análise de Dados Passados (Fase I)

Considere que obtivemos k subgrupos independentes de tamanho n ($n > p$) provenientes de distribuições $N_p(\mu_i, \Sigma)$ $i = 1,2,\cdots,k$ em que os vetores das médias sejam desconhecidos e cujas matrizes de covariâncias são desconhecidas mas iguais. Nesta primeira fase, precisamos de testar a hipótese $H_0: \mu_1 = \mu_2 = \cdots = \mu_k$. Seja $\overline{\mathbf{X}}_i$ a média amostral do i-ésimo subgrupo e $\mathbf{S}_i$ sua matriz de covariância amostral. Seja $\overline{\mu} = \frac{1}{k}\sum_{i=1}^{k}\mu_i$ o vetor que representa a média das médias das distribuições dos subgrupos, $\overline{\overline{\mathbf{X}}} = \frac{1}{k}\sum_{i=1}^{k}\overline{\mathbf{X}}_i$ o vetor amostral de médias globais e $\overline{\mathbf{S}} = \frac{1}{k}\sum_{i=1}^{k}\mathbf{S}_i$ um estimador não viciado para Σ representando a variabilidade dentro dos subgrupos.

As distribuições das estatísticas serão as seguintes:

$$\overline{\mathbf{X}}_i \sim N_p\left(\mu_i, \frac{1}{n}\Sigma\right), \ \overline{\overline{\mathbf{X}}} \sim N_p\left(\overline{\mu}, \frac{1}{kn}\Sigma\right), \ (n-1)\mathbf{S}_i = \sum_{j=1}^{n-1}\mathbf{Z}_j\mathbf{Z}_j' \sim W_p(n-1,\Sigma)$$

Portanto, podemos derivar as distribuições das seguintes expressões:

$$\overline{\mathbf{X}}_i - \overline{\overline{\mathbf{X}}} \sim N_p\left(\mu_i - \overline{\mu}, \frac{1}{kn/k-1}\Sigma\right)$$

$$\sqrt{kn/k-1}\left(\overline{\mathbf{X}}_i - \overline{\overline{\mathbf{X}}}\right) \sim N_p\left(\sqrt{kn/k-1}(\mu_i - \overline{\mu}), \Sigma\right)$$

$$k(n-1)\overline{\mathbf{S}} = \sum_{i=1}^{k}\sum_{j=1}^{n-1}\mathbf{Z}_j\mathbf{Z}_j' \sim W_p(k(n-1),\Sigma)$$

Aplicando o Teorema 1.3 temos que

$$\frac{\frac{kn}{k-1}\left(\overline{\mathbf{X}}_i - \overline{\overline{\mathbf{X}}}\right)'\overline{\mathbf{S}}^{-1}\left(\overline{\mathbf{X}}_i - \overline{\overline{\mathbf{X}}}\right)}{k(n-1)}\left[\frac{k(n-1)-p+1}{p}\right] \sim F\left(p, k(n-1)-p+1, \tau^2{}_i\right)$$

Cujo parâmetro de não centralidade é dado por

$$\tau_i^2 = \frac{kn}{k-1}(\mu_i - \overline{\mu})'\Sigma^{-1}(\mu_i - \overline{\mu})$$

Assim

$$T_i^2 = n\left(\overline{\mathbf{X}}_i - \overline{\overline{\mathbf{X}}}\right)'\overline{\mathbf{S}}^{-1}\left(\overline{\mathbf{X}}_i - \overline{\overline{\mathbf{X}}}\right) \sim \frac{p(k-1)(n-1)}{k(n-1)-p+1}F\left(p, k(n-1)-p+1, \tau^2{}_i\right)$$

Então, H_0 é testada plotando os valores de $T_i^2 \quad i = 1,2,\cdots,k$ num gráfico de controle cujo limite superior é dado por

$$\text{LSC}_{fase1} = \frac{p(k-1)(n-1)}{k(n-1)-p+1} F_{p,k(n-1)-p+1}(\alpha) \qquad (1.1)$$

Caso ocorra pontos plotados acima do LSC_{fase1}, esses pontos são retirados do conjunto de dados e é recalculado um novo LSC_{fase1} para teste dos pontos remanescentes. Esse processo é realizado até que todos os pontos restantes estejam abaixo do LSC_{fase1} calculado por último. Assumimos, a partir daí que esses pontos restantes são amostras aleatórias de uma distribuição $N_p(\mu, \Sigma)$ e essas informações serão usadas para a construção do gráfico de controle para futuras observações do processo.

1.3.2 Controle do Processo (Fase II)

Temos, agora, uma estimativa dos parâmetros da distribuição do processo dada pelos subgrupos amostrais que não foram descartados na fase I. Suponha que descartamos a subgrupos, restando $m = k - a$. As estimativas do vetor de médias e da matriz de covariância serão

$$\hat{\mu} = \frac{1}{m}\sum_{i=1}^{m} \bar{\mathbf{X}}_i \text{ e } \hat{\Sigma} = \frac{1}{m}\sum_{i=1}^{m} \mathbf{S}_i$$

Seja, agora, um subgrupo amostral futuro de tamanho n obtido independentemente dos primeiros subgrupos. Suponha que a distribuição de cada item desse subgrupo seja $N_p(\mu_f, \Sigma)$. Testaremos a hipótese $H_0: \mu_f = \mu$ baseados no teste da razão de verossimilhança e o gráfico de controle será a representação visual de sucessivos testes dos subgrupos que serão obtidos do processo. Vejamos a estatística do teste, note que não será

a mesma da fase I devido ao fato de que os novos subgrupos não participaram das estimativas de μ e Σ.

As distribuições das estatísticas amostrais serão as seguintes:

$$\overline{\mathbf{X}}_f \sim N_p\left(\mu_f, \frac{1}{n}\Sigma\right) \quad , \quad \hat{\mu} \sim N_p\left(\mu, \frac{1}{mn}\Sigma\right)$$

$$m(n-1)\hat{\Sigma} = \sum_{j=1}^{m}\sum_{i=1}^{n-1}\mathbf{Z}_i\mathbf{Z}_i' \sim W_p\left(m(n-1), \Sigma\right)$$

$$\overline{\mathbf{X}}_f - \hat{\mu} \sim N_p\left(\mu_f - \mu, \frac{1}{mn/m+1}\Sigma\right)$$

$$\sqrt{mn/m+1}\left(\overline{\mathbf{X}}_f - \hat{\mu}\right) \sim N_p\left(\sqrt{mn/m+1}\left(\mu_f - \mu\right), \Sigma\right)$$

Aplicando o Teorema 1.3 temos que

$$\frac{\frac{mn}{m+1}\left(\overline{\mathbf{X}}_f - \hat{\mu}\right)' \hat{\Sigma}^{-1}\left(\overline{\mathbf{X}}_f - \hat{\mu}\right)}{m(n-1)}\left[\frac{m(n-1)-p+1}{p}\right] \sim F\left(p, m(n-1)-p+1, \tau^2{}_f\right)$$

Cujo parâmetro de não centralidade é dado por

$$\tau_f^2 = \frac{mn}{m+1}\left(\mu_f - \mu\right)' \Sigma^{-1}\left(\mu_f - \mu\right)$$

Assim a estatística do teste será

$$T^2 = n\left(\overline{\mathbf{X}}_f - \hat{\mu}\right)' \hat{\Sigma}^{-1}\left(\overline{\mathbf{X}}_f - \hat{\mu}\right) \sim \frac{p(m+1)(n-1)}{m(n-1)-p+1} F\left(p, m(n-1)-p+1, \tau^2{}_f\right)$$

O limite superior de controle do gráfico será, portanto

$$\mathrm{LSC}_{faseII} = \frac{p(m+1)(n-1)}{m(n-1)-p+1} F_{p,m(n-1)-p+1}(\alpha) \qquad (1.2)$$

Quando um ponto excede esse limite, dizemos que o processo está fora de controle. Alguma causa especial, também chamada de causa assinalável, que não pertence ao sistema de causas comuns do processo deve estar atuando e tem de ser investigada.

1.3.3 Exemplo Ilustrativo de Aplicação (Fase I)

Simulamos 20 subgrupos de tamanho $n = 10$ de uma distribuição normal bivariada com parâmetros:

$$\mathbf{X}_i = \begin{bmatrix} X_{i1} \\ X_{i2} \end{bmatrix} \sim N_2(\mu_i, \Sigma) \text{ com } \mu_i = \begin{bmatrix} 10 \\ 100 \end{bmatrix} \; i = 1,2,\cdots,20 \text{ e } \Sigma = \begin{bmatrix} 4 & 7 \\ 7 & 25 \end{bmatrix}$$

Incluímos mais um subgrupos gerado com a mesma matriz de covariância, mas o vetor de médias recebeu os seguintes valores

$$\mu_{21} = \begin{bmatrix} 10 + 2\sigma_{\overline{X}_1} \\ 100 - 2\sigma_{\overline{X}_2} \end{bmatrix} = \begin{bmatrix} 10 + 4/\sqrt{10} \\ 100 - 10/\sqrt{10} \end{bmatrix} = \begin{bmatrix} 11.2649 \\ 96.8377 \end{bmatrix}$$

As médias de cada subgrupo, bem como a matriz de covariância e o cálculo de T^2 são apresentadas na Tabela 1.1. Vejamos os resultados importantes obtidos com essa amostra:

$$\hat{\mu} = \frac{1}{k}\sum_{i=1}^{k}\overline{\mathbf{X}}_i = \begin{bmatrix}10.05\\99.93\end{bmatrix}, \quad \hat{\Sigma} = \frac{1}{k}\sum_{i=1}^{k}\mathbf{S}_i = \begin{bmatrix}3.63 & 6.51\\6.51 & 27.44\end{bmatrix} \text{ e } \hat{\Sigma}^{-1} = \begin{bmatrix}.4784 & -.1134\\-.1134 & .0633\end{bmatrix}$$

Baseado num α de 0,0027, o LSC para o gráfico T^2 será dado por

$$\text{LSC}_{fase1} = \frac{2(21-1)(10-1)}{21(10-1)-2+1} F_{2,21(10-1)-2+1}(0{,}0027) = 11{,}69$$

Tabela 1.1 - Resultados para os 21 subgrupos

subgrupo	$\overline{X}_1$	$\overline{X}_2$	s_{11}	s_{22}	s_{12}	T^2
1	10.41	100.20	2.19	29.70	5.31	0.44
2	10.35	101.86	2.60	20.25	3.51	1.48
3	11.16	102.58	1.80	8.88	2.99	3.67
4	9.93	100.41	3.09	21.90	6.80	0.34
5	8.83	97.55	2.70	26.63	2.88	4.12
6	8.21	96.53	4.31	21.72	4.55	9.32
7	10.37	99.07	3.43	21.81	3.04	1.59
8	10.23	100.12	2.05	38.07	7.64	0.09
9	9.39	97.55	2.49	29.48	6.91	2.12
10	9.99	100.39	2.21	25.00	5.23	0.21
11	9.60	100.73	6.96	37.09	12.72	2.22
12	10.96	102.44	3.18	25.00	6.94	2.75
13	10.87	101.37	4.13	18.58	4.84	1.82
14	9.61	100.13	5.75	38.44	11.87	1.15
15	9.36	99.34	2.88	30.03	2.67	1.61
16	9.94	101.60	5.08	25.60	7.29	2.25
17	10.88	101.74	5.33	31.92	10.86	1.96
18	9.50	98.16	6.92	43.16	12.07	1.23
19	10.56	101.32	1.84	37.70	7.56	0.86
20	9.69	98.64	5.29	22.18	8.01	0.63
21	11.27	96.90	2.06	23.04	3.00	21.28

O gráfico de controle para o caso bivariado pode ser representado tanto por um gráfico T^2, em que se tem a idéia do comportamento do processo no tempo, mas também é permitido plotar-se uma variável contra a outra num gráfico de dispersão, em que os limites de controle são visualizados como uma elipse[3].

[3] Veja um método simples de construção de elipses em Tracy et alli (1995b)

Os gráficos de controle são mostrados nas Figuras 1.1 e 1.2 abaixo seguidos de comentários. Fizemos também os gráficos de controle separados para cada variável encontrados nas Figuras 1.3 e 1.4

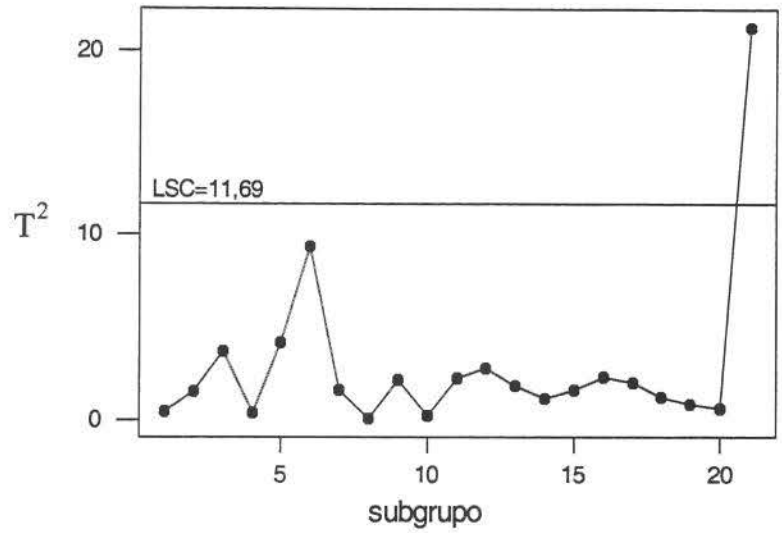

Figura 1.1 – Gráfico T^2 em seqüência temporal (21 subgrupos)

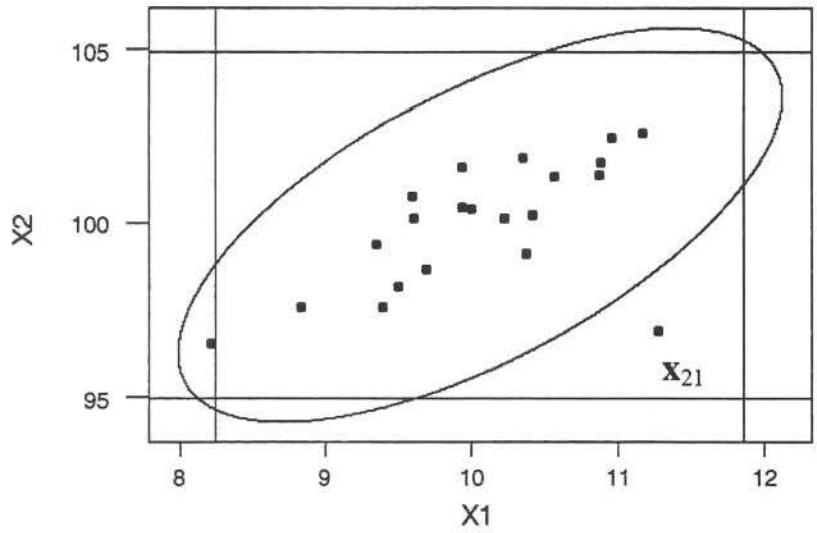

Figura 1.2 – Gráfico de dispersão (21 subgrupos)

Na Figura 1.1 notamos que todos os pontos com exceção do último (cuja distribuição foi alterada) estão abaixo do LSC calculado. No gráfico de dispersão da Figura 1.2, notamos como cada ponto se posiciona em relação à elipse calculada para os dados em questão. O subgrupo 21 aparece fora de controle por não respeitar a estrutura de correlação entre as variáveis. Na verdade, para processos em que temos somente 2 variáveis de interesse, essa

representação dos subgrupos é o melhor método de identificação e diagnóstico das causas especiais. Vemos de imediato se o sinal foi disparado por não seguir a estrutura de correlação observada pela distância na direção do eixo menor da elipse, ou por apresentar variabilidade excessiva verificada no eixo maior da elipse.

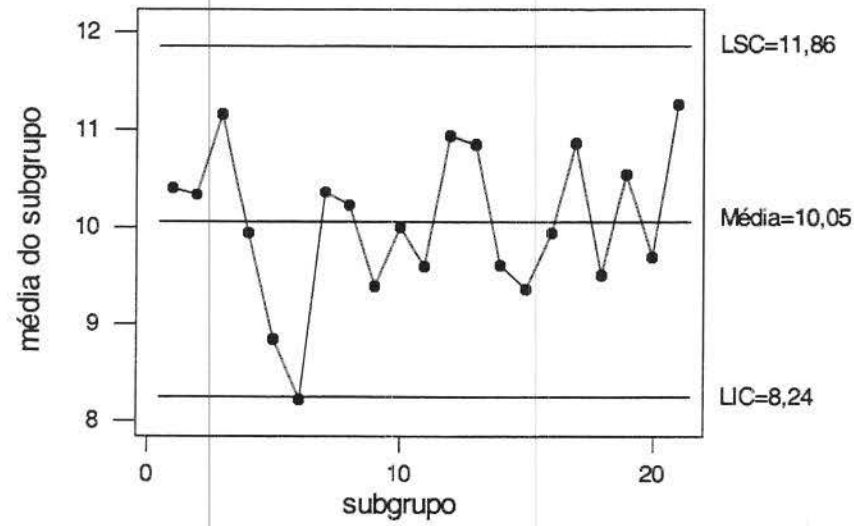

Figura 1.3 - Gráfico de controle individual para a variável X1 (21 subgrupos)

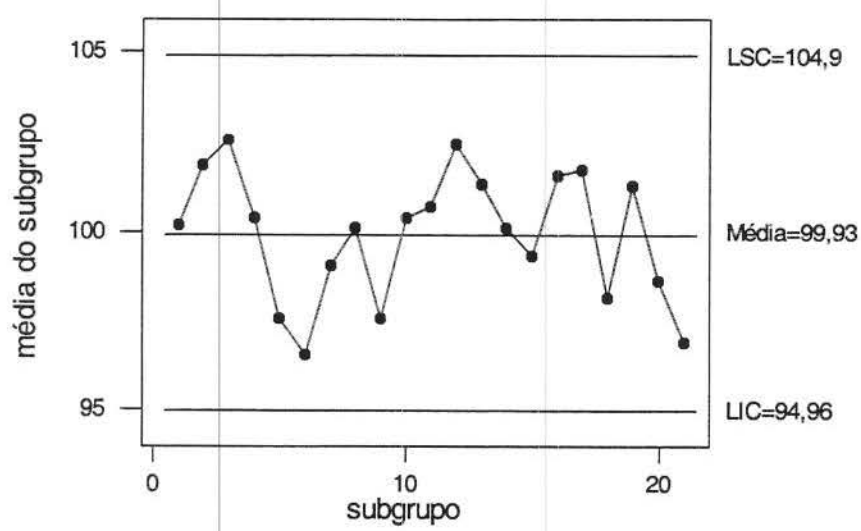

Figura 1.4 - Gráfico de controle individual para a variável X1 (21 subgrupos)

Observe, acima, que o último ponto encontra-se sob controle nas duas cartas calculadas separadamente. Por outro lado, o subgrupo 6 está abaixo do limite inferior em X1. Essas discordâncias são devido à presença de correlação entre as variáveis somado à inflação do erro tipo I quando se constróem gráficos separados.

Retirado o subgrupo 21 dos dados, os estimadores amostrais são recalculados como segue:

$$\hat{\mu} = \frac{1}{m}\sum_{i=1}^{m}\bar{\mathbf{X}}_i = \begin{bmatrix} 9.99 \\ 100.09 \end{bmatrix}, \quad \hat{\Sigma} = \frac{1}{m}\sum_{i=1}^{m}\mathbf{S}_i = \begin{bmatrix} 3.71 & 6.68 \\ 6.51 & 27.67 \end{bmatrix} \text{ e } \hat{\Sigma}^{-1} = \begin{bmatrix} .4768 & -.1152 \\ -.1152 & .0640 \end{bmatrix}$$

O novo limite de controle será

$$\text{LSC}_{fase1} = \frac{2(20-1)(10-1)}{20(10-1)-2+1} F_{2,20(10-1)-2+1}(.0027) = 11.68$$

Os gráficos com esses novos valores são mostrados nas figuras 1.5 e 1.6 abaixo. Percebe-se que não há mais subgrupos rejeitados, assim partiremos para o controle do processo com as informações "limpas" obtidas.

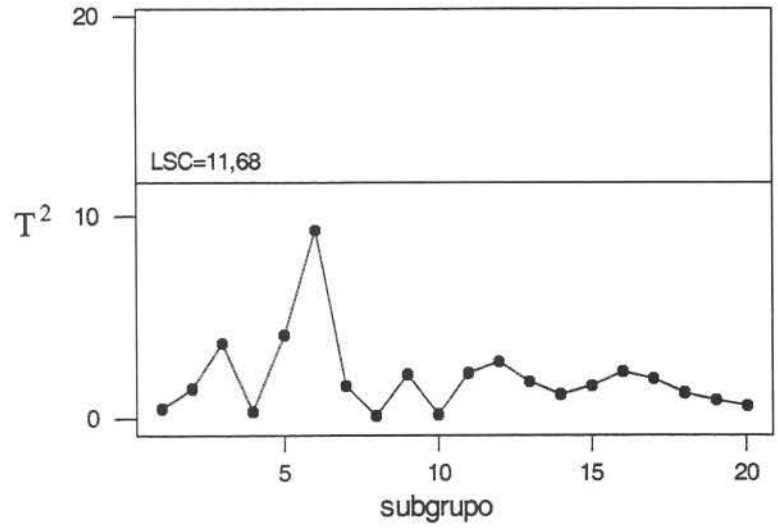

Figura 1.5 – Gráfico T^2 retirando-se o último subgrupo

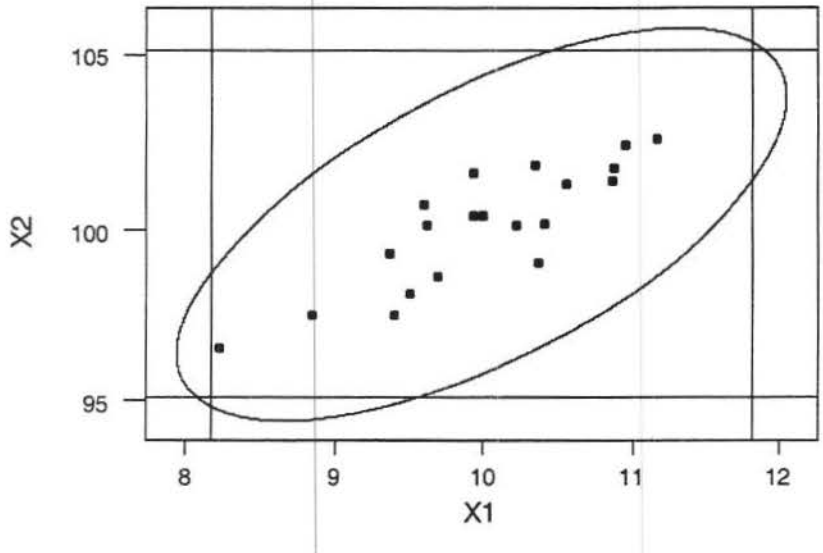

Figura 1.6 - Gráficos multivariados retirando-se o último subgrupo

1.3.4 Exemplo Ilustrativo de Aplicação - Fase II

Como dissemos, a fase II consiste no teste de subgrupos de observações tomadas no futuro para efeito de controle do processo. Neste momento já temos as estimativas de μ e Σ do processo, portanto basta calcularmos o limite superior de controle fase II e o comparar o valor de T^2 calculado para cada novo subgrupo com esse valor. A seqüência desses testes será plotada num gráfico para melhor visualização da situação "entre subgrupos" do processo. No nosso exemplo, mostraremos os subgrupos no gráfico X1 x X2 para interpretarmos causas especiais.

Usando o procedimento do cálculo de LSC_{faseII} dado em 1.3.2, temos que este limite vale 12,98 para o processo simulado em questão. A seguir simulamos cinco condições diferentes do processo, em cada uma, foram gerados 10 subgrupos de tamanho 10. Ao todo, portanto, simulamos 50 subgrupos de tamanho 10. Cada situação é descrita na Tabela 1.2 abaixo.

Tabela 1.2 - Condições de processo simuladas

Condição	subgrupos	média de X1	média de X2
I	1-10	original (μ_1)	original (μ_2)
II	11-20	$\mu_1 + 2\sigma_1/\sqrt{10}$	original (μ_2)
III	21-30	$\mu_1 - \sigma_1/\sqrt{10}$	$\mu_2 + \sigma_2/\sqrt{10}$
IV	31-40	$\mu_1 + 2\sigma_1/\sqrt{10}$	$\mu_2 - 2\sigma_2/\sqrt{10}$
V	41-50	$\mu_1 - 3\sigma_1/\sqrt{10}$	$\mu_2 - 3\sigma_2/\sqrt{10}$

Os resultados das simulações são mostrados em gráficos. O gráfico T^2 indica que realmente temos problemas em todas as condições que fogem da distribuição original. Note que a condições II e III parecem similares quanto ao cálculo de T^2, mas veja a posição bastante diferente dos subgrupos em relação à elipse de controle.

Outro ponto a se notar é que a condição IV aparenta ser pior que a V, isto é, deslocamentos de 2 desvios em sentidos opostos na direção perpendicular ao da correlação entre as variáveis são mais detectáveis que deslocamentos de 3 desvios em cada variável na direção da correlação entre elas. Isso é explicado quando comparamos os parâmetros de não-centralidade $n(\boldsymbol{\mu}^* - \boldsymbol{\mu}_0)'\Sigma^{-1}(\boldsymbol{\mu}^* - \boldsymbol{\mu}_0)$ nos dois casos que são respectivamente 26,67 para a condição IV e 10,59 para a V.

Ainda com respeito às condições IV e V, é interessante notar como há diferenças de julgamento entre o gráfico T^2 e os gráficos dos componentes individuais. No caso da condição IV, O gráfico T^2 acusa 9 entre os 10 pontos como fora de controle, enquanto os gráficos dos componentes individuais acusam somente 2 entre os 10 sendo pouco sensível para essa condição.

Na condição V, O gráfico T^2 acusa 4 dos 10 pontos como fora de controle (parece bom, pois nessa situação esperaríamos menos que 50% dos pontos fora de controle devido à

presença de correlação) enquanto os gráficos individuais acusam 7 dos 10 pontos como assinaláveis podendo indicar uma inflação do erro tipo I.

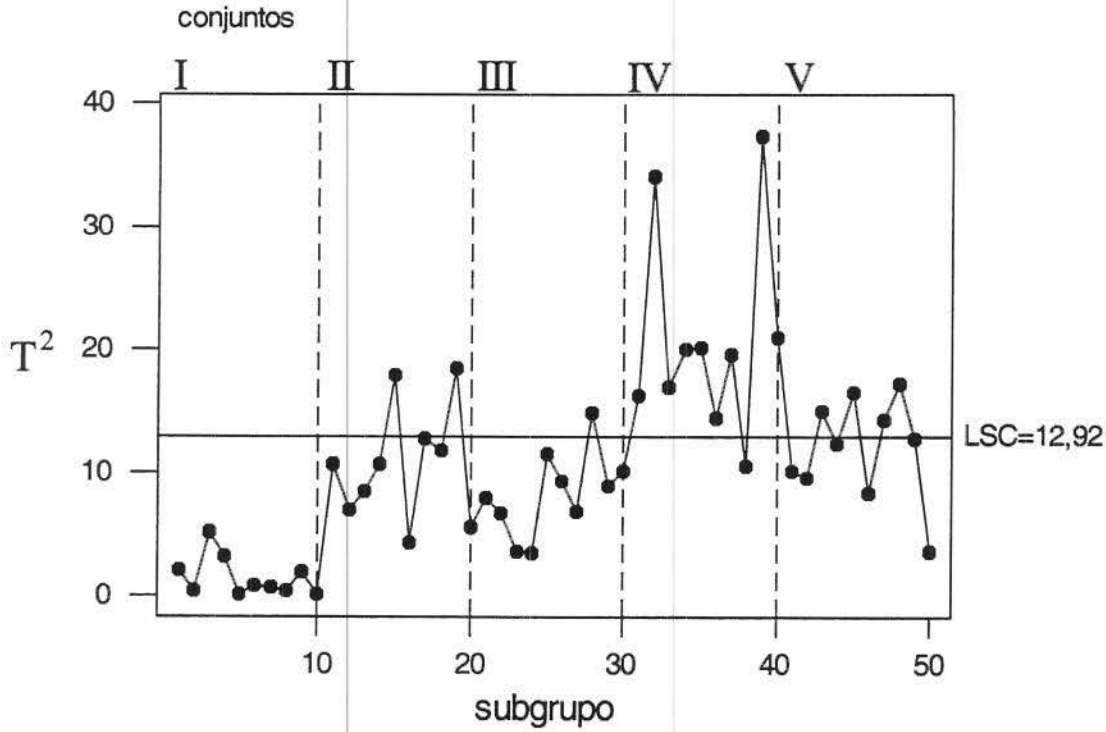

Figura 1.7 - Gráfico T^2 para as condições simuladas

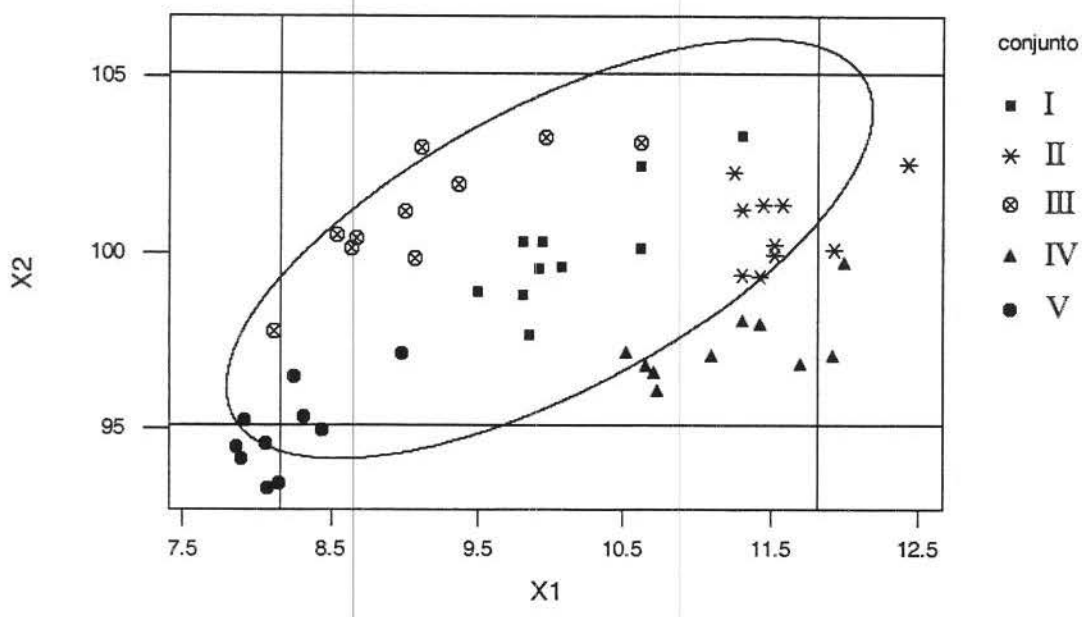

Figura 1.8 - Elipse de controle para as cinco condições simuladas

1.3.5 Controle da Variabilidade

Em algumas situações o interesse pode ser o monitoramento da variabilidade do processo, por exemplo, no uso conjunto de controle estatístico e controle automático de processos em plantas químicas e de processo, alterações na variabilidade de abertura de um conjunto de válvulas caracterizam uma causa assinalável que nem sempre pode ser corrigida pelo controle automático. Alt (1985) apresenta uma boa introdução ao problema com duas abordagens.

A primeira é baseada nos resultados das seção 1.2.8 definindo-se $W_{1,i} = -2\ln\lambda_1$ como a estatística do gráfico de controle da variabilidade. Desenvolvendo o resultado em 1.2.8 temos:

$$W_{1,i} = -pn + pn\ln(n) - n\ln(|\mathbf{A}_i|/|\Sigma|) + tr(\Sigma^{-1}\mathbf{A}_i)$$

O limite superior desse gráfico será $LSC_{W_1} = \chi^2_{p(p+1)/2}\ (\alpha)$

A segunda forma apresentada é baseada nos resultados de distribuição da variância generalizada vistos na seção 1.2.9. Quando temos somente 2 características de qualidade em estudo, podemos usar o Teorema 1.6 para a confecção do gráfico de controle para a estatística $W_{2,i} = 2V^{\frac{1}{2}}$ e o limite superior para esse gráfico será

$$LSC_{W_2} = \chi^2_{2n-4}\ (\alpha)$$

Para mais de 2 características usamos os resultados sobre os momentos de $|\mathbf{S}|$ dados na seção 1.2.9. em que vimos que:

$$E|\mathbf{A}| = |\Sigma| \prod_{i=1}^{p} (n-i)$$

portanto,

$$E|\mathbf{S}| = \frac{1}{(n-1)^p} |\Sigma| \prod_{i=1}^{p} (n-i)$$

Também vimos que

$$Var(|\mathbf{A}|) = |\Sigma|^2 \prod_{i=1}^{p} (n-i) \left[\prod_{j=1}^{p} (n-j+2) - \prod_{j=1}^{p} (n-j) \right]$$

assim,

$$Var(|\mathbf{S}|) = |\Sigma|^2 \frac{1}{(n-1)^{2p}} \prod_{i=1}^{p} (n-i) \left[\prod_{j=1}^{p} (n-j+2) - \prod_{j=1}^{p} (n-j) \right]$$

e os parâmetros do gráfico de controle para $|\mathbf{S}|$ serão baseados em $E|\mathbf{S}| \pm 3\sqrt{Var(|\mathbf{S}|)}$:

Limite Superior de Controle: $LSC = |\Sigma|(b_1 + 3\sqrt{b_2})$,

Linha Central: $LC = b_1|\Sigma|$,

Limite Inferior de Controle $LIC = |\Sigma|(b_1 - 3\sqrt{b_2})$

onde

$$b_1 = \frac{1}{(n-1)^p}\prod_{i=1}^{p}(n-i) \text{ e } b_2 = \frac{1}{(n-1)^{2p}}\prod_{i=1}^{p}(n-i)\left[\prod_{j=1}^{p}(n-j+2)-\prod_{j=1}^{p}(n-j)\right]$$

É sabido, porém, que o poder dos testes relativos à matriz de covariância são muito baixos e contribuem negativamente para a eficácia dos gráficos sugeridos para a variabilidade. Na prática, causas especiais que afetam a matriz de covariância, distorcendo a estrutura de correlação ou aumentando a variabilidade, com freqüência são detectadas pelos procedimentos de controle do vetor de médias. Nos casos em que a variabilidade diminui, somente em longo prazo isso será descoberto.

A questão do controle da variabilidade tem merecido pouca atenção na literatura em parte devido a essas razões.

Capítulo 2

Gráficos de Controle para Observações Individuais Multivariadas

Há processos em que a técnica de subagrupamento é inadequada de se aplicar, por exemplo em plantas químicas e de processos onde o material amostral pode ser coletado de um vaso com conteúdo homogêneo ou de uma corrente contínua do processo. Nesses casos a suposição de independência dos itens estaria muito comprometida caso se definisse como amostra um subgrupo de material do vaso ou da corrente. Em indústrias de peças e componentes, a suposição de independência dos itens pode ser adequada, mas às vezes fica muito caro para se obter medições múltiplas de subgrupos de itens. Isso se deve pelo elevado tempo de ciclo da medição e ou pelo seu custo. Observações individuais são muito comuns nesses casos.

Os procedimentos para observações individuais são diferentes dos apresentados no capítulo anterior, principalmente para a fase I. Neste Capítulo veremos o desenvolvimento desses procedimentos, passando rapidamente pela fase II e depois entrando em detalhes da fase I estudando estimadores de Σ com melhor desempenho, veremos também alguns métodos alternativos para a detecção de observações aberrantes e por último, um método simples mas poderoso de interpretação de sinais em ambas as fases.

2.1 Fase II: Controle para Futuras Observações Individuais

Começaremos o estudo dos gráficos de controle para observações individuais pela fase II pelo fato dos resultados serem muito próximos daqueles apresentados para subgrupos racionais.

O objetivo é calcular os parâmetros do gráfico de controle para testar se o processo mantém sua média para uma observação futura tendo realizado os trabalhos da fase I, isto é, já temos uma amostra de tamanho m do processo sob controle. Considere esse conjunto de observações iniciais $\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_m$ cuja média e matriz de covariância são dadas por $\overline{\mathbf{X}}$ e $\mathbf{S}$, e uma futura observação $\mathbf{X}_f$. Se o processo permanece sob controle para $\mathbf{X}_f$, então essas variáveis tem a seguinte distribuição: $\mathbf{X}_i \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, assim, $\overline{\mathbf{X}} \sim N_p\left(\boldsymbol{\mu}, \frac{1}{m}\boldsymbol{\Sigma}\right)$ e $(m-1)\mathbf{S} \sim W_p(m-1, \boldsymbol{\Sigma})$. Suponha, agora, que $\overline{\mathbf{X}}$, $\mathbf{S}$ e $\mathbf{X}_f$ sejam independentes, então, $\mathbf{X}_f - \overline{\mathbf{X}} \sim N_p\left(\mathbf{0}, \frac{m+1}{m}\boldsymbol{\Sigma}\right)$ e $\sqrt{\frac{m}{m+1}}(\mathbf{X}_f - \overline{\mathbf{X}}) \sim N_p(\mathbf{0}, \boldsymbol{\Sigma})$. Assim a estatística T^2 de Hotelling a ser definida será:

$T_f^2 = \frac{m}{m+1}(\mathbf{X}_f - \overline{\mathbf{X}})' \mathbf{S}^{-1} (\mathbf{X}_f - \overline{\mathbf{X}})$, portanto, aplicando-se o Teorema 1.3, temos que

$$\left(\frac{T^2}{m-1}\right)\left(\frac{m-p}{p}\right) \sim F(p, (m-1)-p+1)$$

Por facilidade de cálculo, a estatística que irá ser controlada no gráfico e que é normalmente usada na literatura é

$$Q_f = \frac{m+1}{m} T_f^2 = (\mathbf{X}_f - \overline{\mathbf{X}})' \mathbf{S}^{-1} (\mathbf{X}_f - \overline{\mathbf{X}})$$

assim teremos que o limite superior de controle para um gráfico em que serão plotados os resultados da estatística Q_f para futuras observações será:

$$LSC_{faseII} = \frac{p(m-1)(m+1)}{m(m-p)} F_{p,m-p}(\alpha) \qquad (2.1)$$

2.2 Fase I para Individuais

No caso de observações individuais, não é obvia a maneira de se estimar a variabilidade do processo na fase I. Os gráficos de controle vistos no Capítulo 2 representam a comparação da variabilidade *entre* e *dentro* de cada subgrupo. Agora os subgrupos contém somente uma observação e tal comparação ficará prejudicada. Um procedimento comumente aceito no caso univariado consiste em se usar o valor absoluto das diferenças entre observações sucessivas como amplitudes e daí estimar-se a variabilidade *dentro*. Esse método é chamado de Amplitude Móvel. O procedimento usualmente recomendado (veja Alt 1988; Jackson 1985; Tracy, Young & Mason 1992; Lowry & Montgomery 1995) para estimar a matriz de covariância do processo no caso multivariado é fazer uso de todas as observações colhidas na fase I combinadas numa amostra única (correspondendo ao estimador usual **S**). Woodall & Sullivan (1996) mostraram que esse procedimento não é efetivo na detecção de desvios na média nessa fase por causa da estimativa inflacionada da matriz de covariância pelas causas especiais. Eles compararam muitos métodos alternativos e recomendaram um procedimento análogo ao uso de amplitudes móveis do caso univariado.

Outra particularidade do caso de observações individuais é que na fase I, a estatística T^2 é proporcional a uma distribuição beta. Por tudo isso, vale a pena estudarmos os detalhes dos procedimentos envolvendo a fase I para observações individuais.

2.2.1 Estimação de Sigma na Fase I para Individuais

Suponha que dispomos de m observações sucessivas p-variadas de um processo. Nosso objetivo inicial é verificar se há causas assinaláveis presentes nessa fase e obter uma estimativa 'limpa' dos parâmetros do processo. No caso univariado, é largamente reconhecido que a variabilidade estimada através do esquema de amplitudes móveis é mais robusta que a variância amostral na presença de desvios da média na fase I. Essa robustez vem do fato de que para pares sucessivos de observações é razoável que a média tenha a tendência de se manter a mesma. Assim, as causas especiais mais comuns como um pulso, uma rampa ou um degrau, são passíveis de serem detectadas.

Alt (1985) afirma que "o esquema de amplitudes móveis para o caso multivariado é intratável". Recentemente, porém, com o trabalho de Woodall & Sullivan (1996), esse esquema foi consolidado para o caso multivariado. Eles estudaram cinco diferentes estimadores da matriz de covariância e, através de simulações, os compararam com o desempenho relativo à "verdadeira" matriz de covariância que gerou os dados de simulação. Três deles merecem destaque.

O primeiro estimador é o usual **S** que aqui será chamado de $\mathbf{S}_1$. Sua obtenção é da forma $\mathbf{S}_1 = \frac{1}{m}\sum_{i=1}^{m}(\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})'$. Observando sua expressão, nota-se que se houver um ou mais pontos fora de controle, essa variação excessiva será incorporada à estimativa de $\boldsymbol{\Sigma}$.

Os outros dois são abordagens multivariadas do estimador de amplitudes móveis do caso univariado. O estimador $\mathbf{S}_2$ é formado a partir do particionamento dos dados em grupos disjuntos com duas observações independentes cada, descartando a última observação se necessário. $\mathbf{S}_2$ é calculado da seguinte forma:

$$\mathbf{S}_2 = \frac{1}{2}\frac{\mathbf{Y}'\mathbf{Y}}{\left\lfloor \frac{m}{2} \right\rfloor},$$

onde $\mathbf{Y}_{\left\lfloor \frac{m}{2} \right\rfloor \times p}$ é a matriz contendo os vetores linha das diferenças $\mathbf{y}_i = \mathbf{x}_{2i} - \mathbf{x}_{2i-1}$, $i = 1, \cdots, \left\lfloor \frac{m}{2} \right\rfloor$ em que $\lfloor\ \rfloor$ denota a função "maior inteiro menor ou igual a".

Esse estimador será de posto completo desde que $m \geq 2p+1$, pois trata-se da soma das estimativas da matriz de covariância para $\lfloor m/2 \rfloor$ grupos de tamanho 2. Essas estimativas serão independentes e distribuídas cada uma como Wishart com 1 grau de liberdade e nesse caso serão de posto incompleto. Porém, a soma dessas matrizes Wishart será também Wishart com $\lfloor m/2 \rfloor$ graus de liberdade e terá posto completo. Portanto, a distribuição de $\lfloor m/2 \rfloor \mathbf{S}_2$ será Wishart com $\lfloor m/2 \rfloor$ graus de liberdade.

O terceiro estimador de sigma, $\mathbf{S}_3$, trata-se do equivalente praticamente direto ao de amplitudes móveis é dado pela diferença entre cada par sucessivo de observações. Seja a diferença entre sucessivas observações $\mathbf{v}_i = \mathbf{x}_{i+1} - \mathbf{x}_i$, $i = 1,...,m-1$, o estimador para sigma será a metade da matriz de covariância amostral dessas diferenças,

$$\mathbf{S}_3 = \frac{1}{2}\frac{\mathbf{V}'\mathbf{V}}{(m-1)}$$

Sua distribuição não é fácil de ser obtida como a de $\mathbf{S}_2$ mas uma propriedade desse estimador é ser não viciado pois o produto matricial de cada diferença sucessiva terá como valor esperado $2\boldsymbol{\Sigma}$ então para a média ponderada do produto de $m-1$ sucessivas diferenças entre si terá valor esperado de $2(m-1)\boldsymbol{\Sigma}$ ou seja, $E(\mathbf{V}'\mathbf{V}) = 2(m-1)\boldsymbol{\Sigma}$. Veja isso como acontece conforme mostram Woodall & Sullivan (1996) em notação matricial.

Seja a matriz $\mathbf{E}_{(m-1)\times m}$ que auxiliará no estabelecimento das diferenças sucessivas,

$$\mathbf{E} = \begin{bmatrix} -1 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & -1 & 1 \end{bmatrix}$$

. Trabalharemos com a matriz $\mathbf{X}_{n\times p}$ na forma vetorizada, isto é, com as colunas de $\mathbf{X}'$ empilhadas: $(\mathbf{X}')^{vet} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_m \end{bmatrix}$, a forma vetorizada da matriz $\mathbf{V}$ pode ser definida pelo produto de Kronecker da seguinte maneira: $(\mathbf{V}')^{vet} = (\mathbf{E} \otimes \mathbf{I}_p)(\mathbf{X}')^{vet}$. Assim, como o valor esperado para a matriz de covariância de $(\mathbf{X}')^{vet}$ é $(\mathbf{I}_m \otimes \boldsymbol{\Sigma})$ então o valor esperado para a matriz de covariância de $(\mathbf{V}')^{vet}$ será:

$$(\mathbf{E}\otimes\mathbf{I}_p)(\mathbf{I}_m\otimes\Sigma)(\mathbf{E}'\otimes\mathbf{I}_p)=(\mathbf{E}\mathbf{E}')\otimes\Sigma=\begin{bmatrix} 2\Sigma & -\Sigma & \mathbf{0} & \cdots & \mathbf{0} \\ -\Sigma & 2\Sigma & -\Sigma & \cdots & \mathbf{0} \\ \mathbf{0} & -\Sigma & 2\Sigma & \cdots & \vdots \\ \vdots & \vdots & \vdots & \ddots & -\Sigma \\ \mathbf{0} & \mathbf{0} & \cdots & -\Sigma & 2\Sigma \end{bmatrix}_{((m-1)\times p)\times((m-1)\times p)}$$

cujos elementos da diagonal representam o valor esperado para o produto de cada linha de $\mathbf{V}$ consigo própria e o valor esperado de $\mathbf{V}\mathbf{V}'$ será a soma dessa diagonal.

2.2.2 Gráficos de Controle na Fase I para Individuais

O gráfico de controle na fase I para individuais será a plotagem sucessiva da estatística $T_{j,i}^2=(\mathbf{x}_i-\bar{\mathbf{x}})'\mathbf{S}_j^{-1}(\mathbf{x}_i-\bar{\mathbf{x}})$, $i=1,...,m$. A dificuldade em se trabalhar na fase I com os estimadores $\mathbf{S}_j$, $j=1,\cdots,3$ recai sobre a distribuição da estatística $T_{j,i}^2$ ao utilizá-los. Não se pode aplicar diretamente o Teorema 1.3 pois ao se comparar retrospectivamente cada observação, ocorre que elas próprias compuseram a estimativa dos parâmetros, então a exigência de independência entre $\mathbf{x}_i$ e $\hat{\Sigma}$ é violada.

Gnanadesikan & Kettenring (1972) quando estudaram a detecção de pontos aberrantes multivariados, usaram várias estatísticas que medem a contribuição de observações individuais para efeitos multivariados específicos. Uma delas foi justamente a distância quadrática generalizada da i-ésima observação à média amostral ($T_{1,i}^2$), cuja distribuição exata, eles desenvolveram a partir de um teorema em Wilks ([1962], p.562) e é proporcional a uma distribuição Beta da seguinte forma:

$$\frac{m}{(m-1)^2}T_{1,i}^2 \sim B\left(\frac{p}{2},\frac{m-p-1}{2}\right) \qquad (2.2)$$

Tracy, Young & Mason (1992) concluíram que o uso de aproximações pelas distribuições F e χ^2 geram erros muito importantes para tamanhos de amostra pequenos ($m<100$), e recomendam, é claro, o uso da distribuição exata.

Similarmente, é possível, derivar a distribuição de $T^2_{2,i}$ que é apresentada a seguir:

$$\frac{m}{(m-1)\lfloor m/2 \rfloor} T^2_{2,i} \sim B\left(\frac{p}{2}, \frac{\lfloor m/2 \rfloor - p}{2}\right) \tag{2.3}$$

A distribuição de $T^2_{3,i}$ não é de fácil acesso devido aos grupos formadores da matriz de covariância serem dependentes entre si. Uma aproximação da distribuição marginal de $T^2_{3,i}$ é apresentada por Woodall & Sullivan (1996) sugerida por Neil H. Timm da universidade de Pittsburgh (EUA) em comunicação privada a esses autores. Veja-a abaixo.

$$\frac{m}{(m-1)^2} T^2_{3,i} \sim B\left(\frac{p}{2}, \frac{f-p-1}{2}\right) \text{ onde } f = \frac{2(m-1)^2}{3m-4}$$

Constatamos, porém, que essa forma produz superestimativas dos limites de controle. Acreditando ser um defeito de impressão na referência acima, e como não temos acesso àquela comunicação privada, decidimos por outra forma em coerência ao desenvolvimento anterior das distribuições de T^2 com os outros estimadores. Usaremos a seguinte expressão da distribuição marginal aproximada de $T^2_{3,i}$:

$$\frac{m}{(m-1)(f-1)} T^2_{3,i} \sim B\left(\frac{p}{2}, \frac{f-p-1}{2}\right) \tag{2.4}$$

onde $f = \dfrac{2(m-1)^2}{3m-4}$

Esses mesmos autores compararam o desempenho dos estimadores. Os resultados foram bastante distintos na presença de causas especiais do tipo degrau e rampa. Os estimadores $\mathbf{S}_2$ e $\mathbf{S}_3$ apresentaram comportamento quase que similar ao uso da "verdadeira" matriz de covariância na detecção dessas causas especiais, enquanto o estimador $\mathbf{S}_1$ apresentou-se praticamente insensível às elas. Obviamente a situação em que eles têm o mesmo desempenho é na ausência de causas especiais na fase I. Com isso, Woodall & Sullivan (1996) recomendam a utilização de $\mathbf{S}_3$ que apresentou ligeira vantagem de desempenho em relação ao $\mathbf{S}_2$.

A seguir mostraremos exemplos ilustrativos comparando o desempenho dos estimadores $\mathbf{S}_1$ e $\mathbf{S}_3$ para as situações fora de controle mais comuns. Para isso precisamos de calcular os limites de controle fase I dos gráficos de T^2 usando ambos os estimadores.

O limite superior de controle para quando se usa o estimador $\mathbf{S}_1$ será:

$$LSC_{\mathbf{S}_1,fasel} = \frac{(m-1)^2}{m} B_{p/2,m-p-1/2}(\alpha) \qquad (2.5)$$

Para $\mathbf{S}_3$, o de diferenças sucessivas, o limite superior de controle é dado por:

$$LSC_{\mathbf{S}_3,fasel} = \frac{(m-1)(f-1)}{m} B_{p/2,f-p-1/2}(\alpha) \qquad (2.6)$$

onde $f = \dfrac{2(m-1)^2}{3m-4}$

2.3 Comparação do Desempenho dos Estimadores da Matriz de Covariância na Fase I

Uma comparação efetiva do desempenho dos estimadores seria a determinação da probabilidade de detecção de causas especiais (o poder do gráfico) para cada estimador em função do parâmetro de não centralidade $(\boldsymbol{\mu}_i - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_i - \boldsymbol{\mu})$. Esse trabalho foi realizado e é exposto por Woodall & Sullivan (1996). Eles realizaram simulações extensivamente e compuseram gráficos do poder do gráfico de controle contra o parâmetro de não centralidade para todos os estimadores que foram propostos estudando algumas situações comuns de instabilidade do processo. Através desses gráficos fica claro que $\mathbf{S}_1$ é muito inferior aos demais para situações de degrau e rampa, porém todos têm desempenho similar para pontos aberrantes distribuídos pelos dados. O que não é evidente no trabalho acima citado é de que maneira isso acontece pois eles não incluíram exemplos ilustrativos no trabalho. A seguir mostramos tal ilustração e teceremos comentários de como se comportam os estimadores $\mathbf{S}_1$ e $\mathbf{S}_3$ nas situações de instabilidade mais comuns.

2.3.1 Dados Adotados para a Ilustração

O desempenho dos estimadores será comparado através da simulação de $m = 100$ observações em 3 dimensões com a matriz de covariância e correlação originais dadas por:

$$\boldsymbol{\Sigma} = \begin{pmatrix} 16 & 15 & -20 \\ 15 & 25 & -18 \\ -20 & -18 & 49 \end{pmatrix} \qquad \boldsymbol{\rho} = \begin{pmatrix} 1 & 0{,}75 & -0{,}71 \\ 0{,}75 & 1 & -0{,}51 \\ -0{,}71 & -0{,}51 & 1 \end{pmatrix}$$

Note que temos duas variáveis com correlação positiva entre si e correlação negativa com uma terceira. Poderíamos utilizar para essa comparação qualquer matriz simétrica positiva definida para sigma e qualquer vetor de médias como parâmetros pois dada uma causa especial, a probabilidade de detecção desse sinal no gráfico de controle depende somente

do parâmetro de não-centralidade $(\boldsymbol{\mu}_i - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_i - \boldsymbol{\mu})$. Optamos, sem perda de generalidade, por um vetor de médias nulo e pela matriz sigma acima.

Os dados simulados (sem causas assinaláveis) são mostrados em forma gráfica na Figura 2.1 abaixo:

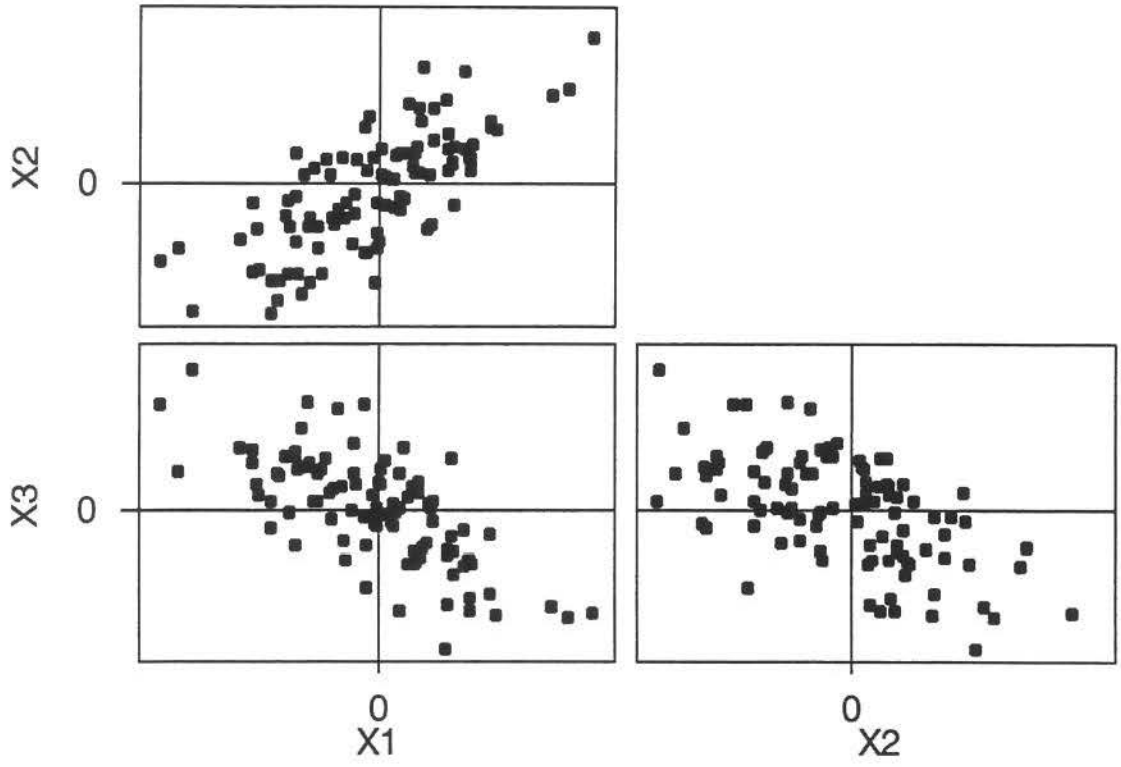

Figura 2.1 – Diagramas de dispersão em matriz dos dados simulados

Os limites de controle fase I para $\alpha = 0{,}0027$, como também as linhas correspondentes ao percentil 50 ($p_{0,5}$) são apresentados abaixo:

Estimador $\mathbf{S}_1$: $LSC_{fasel} = 13{,}38$; $p_{0,5} = 2{,}37$

Estimador $\mathbf{S}_3$: $LSC_{fasel} = 13{,}05$; $p_{0,5} = 2{,}39$

Os gráficos de controle fase I com esses dados 'limpos' podem ser observados a seguir.

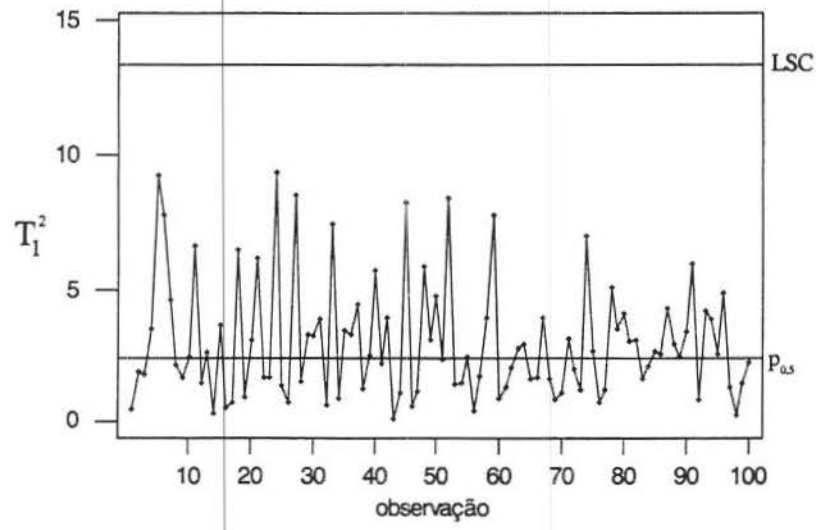

Figura 2.2 – Gráfico T^2 dos dados simulados utilizando $\mathbf{S}_1$

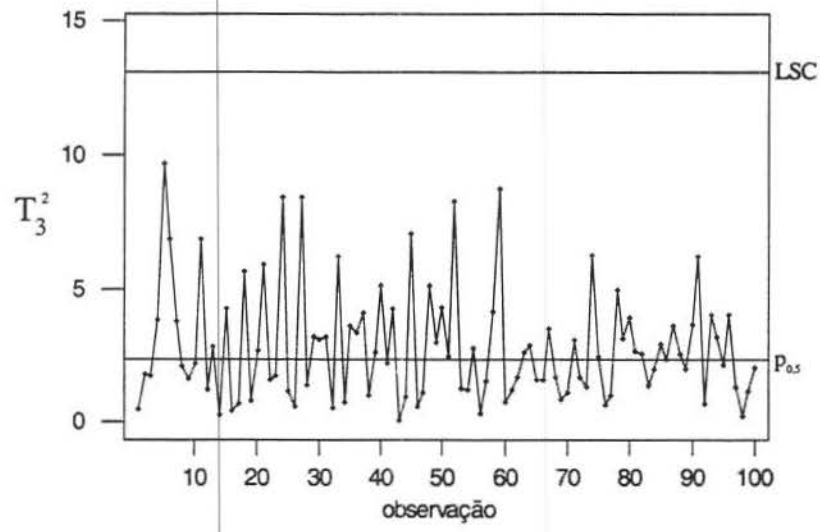

Figura 2.3 – Gráfico T^2 dos dados simulados utilizando $\mathbf{S}_3$

O que se nota é que na ausência de causas especiais, não há sinal em nenhum dos gráficos como era de se esperar.

2.3.2 *Causa Especial Tipo Degrau*

Vejamos o desempenho dos dois métodos para uma causa assinalável do tipo degrau. Um degrau significa uma mudança de $\boldsymbol{\delta} = (\boldsymbol{\mu}^* - \boldsymbol{\mu}_0)$ repentina e que se mantém no vetor de médias podendo afetar uma ou mais variáveis, tanto no sentido das correlações como de maneira oposta a elas. Um degrau pode ser modelado como:

$$\boldsymbol{\mu}_i = \begin{cases} \boldsymbol{\mu}_0 & i = 1,\ldots,k \\ \boldsymbol{\mu}_0 + \boldsymbol{\delta} & i = k+1,\ldots,m \end{cases}$$

no nosso caso, o vetor de médias dos dados sob controle será $\boldsymbol{\mu}_0 = \mathbf{0}$ e o parâmetro de não-centralidade das observações sujeitas à causa especial será $\boldsymbol{\delta}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\delta}$.

Seja nosso exemplo com $m = 100$, $k = 50$ e $\boldsymbol{\delta}' = [28 \quad 35 \quad -48]$, isto é, trata-se de um degrau de perto de 7 sigmas em cada variável no mesmo sentido das correlações, obtendo-se um parâmetro de não-centralidade de cerca de 64. Uma representação gráfica dessa situação é mostrada abaixo.

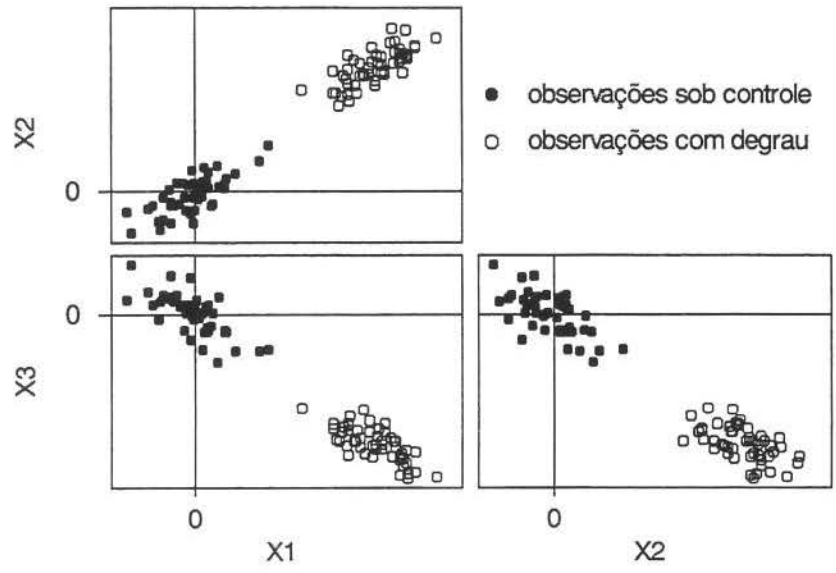

Figura 2.4 – Dados com causa especial tipo degrau no mesmo sentido da estrutura de correlação

Os gráficos de controle fase I com cada estimador são mostrados a seguir:

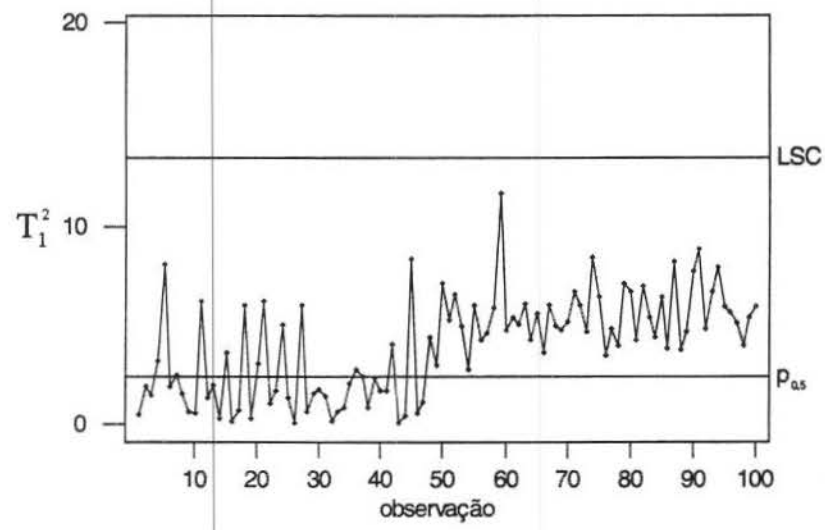

Figura 2.5 – Gráfico T^2 utilizando $\mathbf{S}_1$ dos dados da Figura 2.4

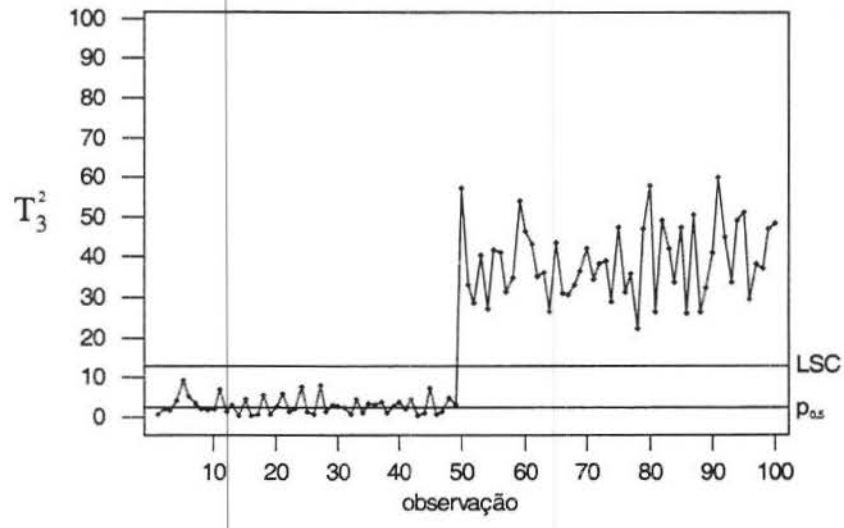

Figura 2.6 – Gráfico T^2 utilizando $\mathbf{S}_3$ dos dados da Figura 2.4

É evidente a diferença de desempenho dos dois procedimentos. Os cálculos com o estimador usual $\mathbf{S}_1$ não detectam o degrau a não ser por uma ligeira elevação da média de T^2 que nesse caso seria percebida somente com pelo menos 10 observações. Já o gráfico para amplitudes móveis não deixa dúvida, detectando desde a primeira observação fora de controle.

Podemos nos perguntar se o caso em que o degrau segue o oposto das correlações ocorre o mesmo resultado. Fizemos, então, os gráficos para um degrau de $\boldsymbol{\delta}' = [12 \quad 15 \quad 21]$. Note que nesse caso o desvio em X_3 segue o contrário das correlações com X_1 e X_2. Outro ponto a se notar é a magnitude bem menor de desvio em cada variável, isso foi feito para que se conservasse o parâmetro de não-centralidade em torno de 64. Segue a representação gráfica dessa situação.

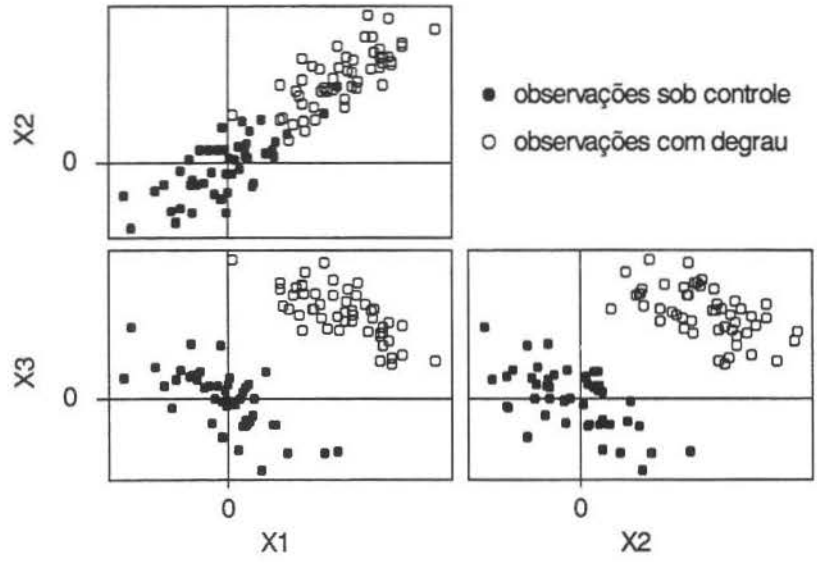

Figura 2.7 – Dados com causa especial tipo degrau no sentido oposto ao da estrutura de correlação

Note abaixo que os gráficos de controle seguem praticamente o mesmo resultado do caso anterior de degrau a favor das correlações:

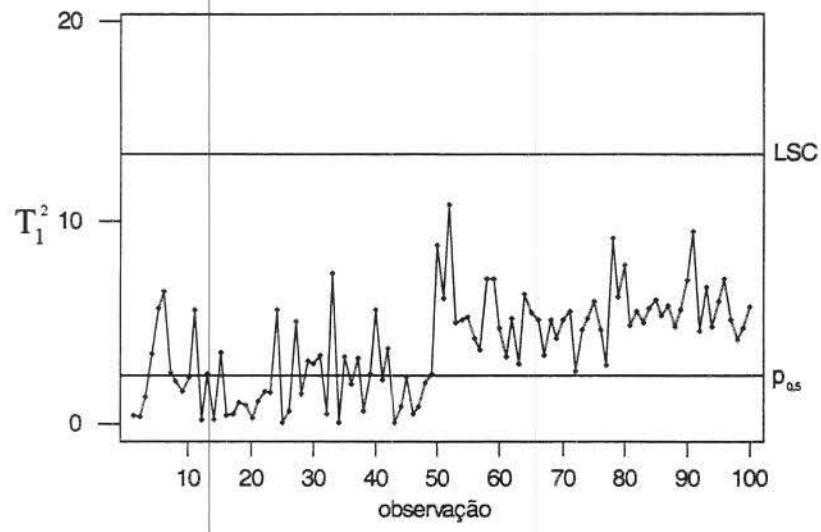

Figura 2.8 – Gráfico T^2 utilizando $\mathbf{S}_1$ dos dados da Figura 2.7

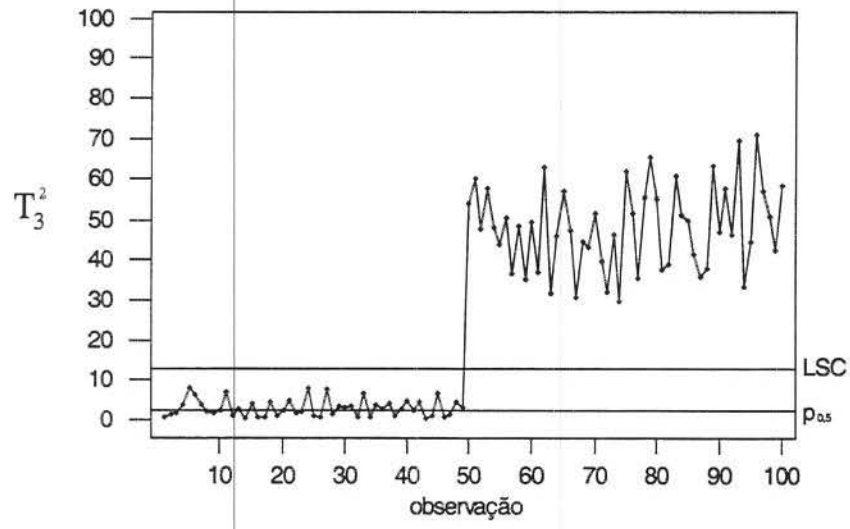

Figura 2.9 – Gráfico T^2 utilizando $\mathbf{S}_3$ dos dados da Figura 2.7

Nem sempre o usuário desses procedimentos tem na fase I o valor de referência ou média esperada das variáveis. Nesses casos o vetor de médias deve ser estimado diretamente dos dados e quando há um degrau, o gráfico de T^2 traz indicações não conclusivas de quais observações estão realmente fora de controle. Veja, por exemplo, essa última simulação de degrau quando analisada usando-se a média das observações. Note que ao utilizar o

estimador de diferenças todo o conjunto de dados parece suspeito e, novamente, o estimador $\mathbf{S}_1$ resultou em insensibilidade.

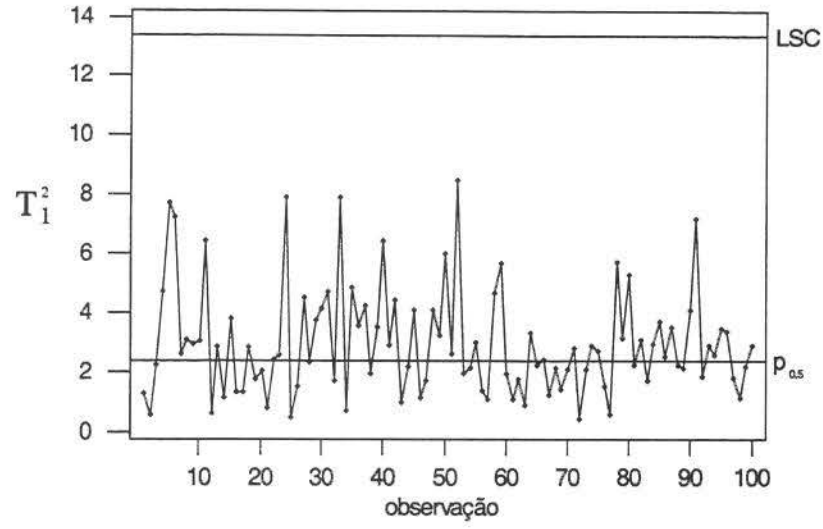

Figura 2.10 – Gráfico T^2 utilizando $\mathbf{S}_1$ dos dados da Figura 2.4 – média desconhecida

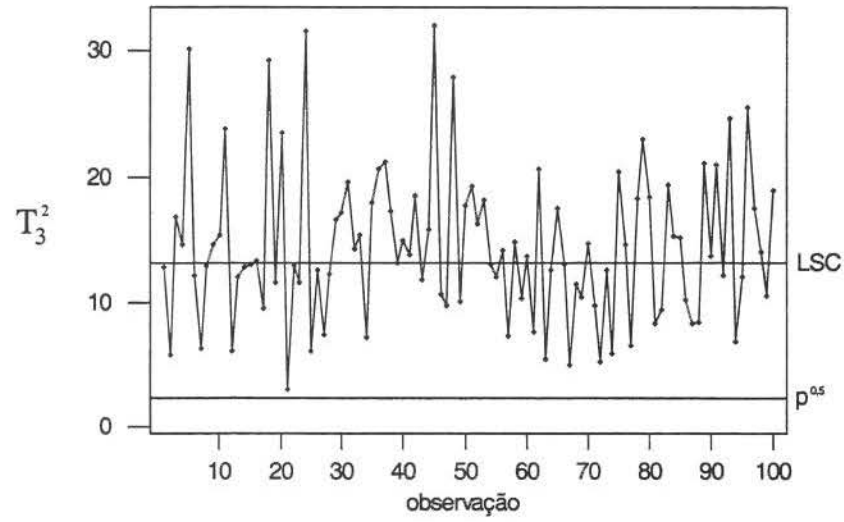

Figura 2.11 – Gráfico T^2 utilizando $\mathbf{S}_3$ dos dados da Figura 2.4 – média desconhecida

Alternativamente ao vetor amostral de médias, nesses casos podemos usar estimações robustas do parâmetro de locação como (i) o vetor de medianas das observações, (ii) o vetor das estimativas de Hodges-Lehmann, isto é, a mediana das médias de pares de observações

ou (iii) o vetor de médias α-aparadas[4], isto é, o vetor de médias das observações remanescentes após omitir-se uma proporção α dos menores e maiores valores de cada variável.

2.3.3 Causa Especial Tipo Rampa

Outra ocorrência possível com dados de um processo é uma rampa, isto é, um desvio gradual na média das variáveis, também chamada de tendência. Suponha uma rampa em que o vetor de médias sofra um desvio de mesmo tamanho a cada observação, esse estado pode ser modelado como se segue.

$$\boldsymbol{\mu}_i = \boldsymbol{\mu}_0 + \frac{i-1}{m-1}\boldsymbol{\delta} \ , \ 1 = 1,\ldots,m$$

onde $\boldsymbol{\mu}_0$ é o vetor sob controle de médias e $\boldsymbol{\delta}$ é o vetor diferença entre o primeira e a e a última observação da rampa.

Woodall & Sullivan (1996) mostraram com surpresa que para essa situação, o estimador das diferenças móveis produz gráficos de controle com poder maior que a própria matriz de covariância original da simulação. Por outro lado, o estimador $\mathbf{S}_1$ mostrou-se muito pouco sensível a esse tipo de causa especial. Simulamos situações de rampa com desvios a favor e contra o sentido de correlações e os resultados mostraram-se muito semelhantes, vejamos portanto o caso a favor da correlação e, como ilustração, a comparação de desempenho quando $\boldsymbol{\delta}' = [28 \quad 35 \quad -48]$ que corresponde a um parâmetro de não-centralidade de 64. O gráfico produzido pelo estimador de amplitudes móveis (Figura 2.13) já detecta a rampa entre a 20ª e 30ª observações enquanto que o do estimador $\mathbf{S}_1$, Figura 2.12, vai sinalizar somente nas últimas 10 observações.

---

[4] Do inglês α-*trimmed means*

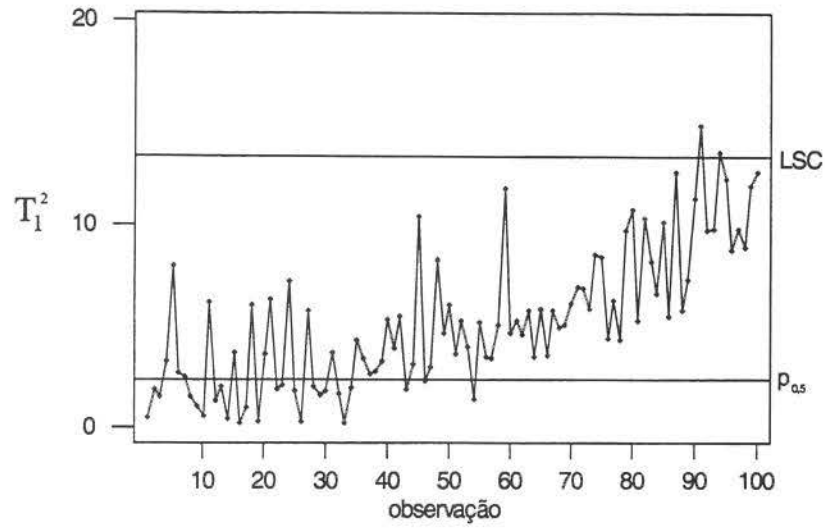

Figura 2.12 – Gráfico T^2 utilizando $\mathbf{S}_1$ - dados com sinal tipo rampa

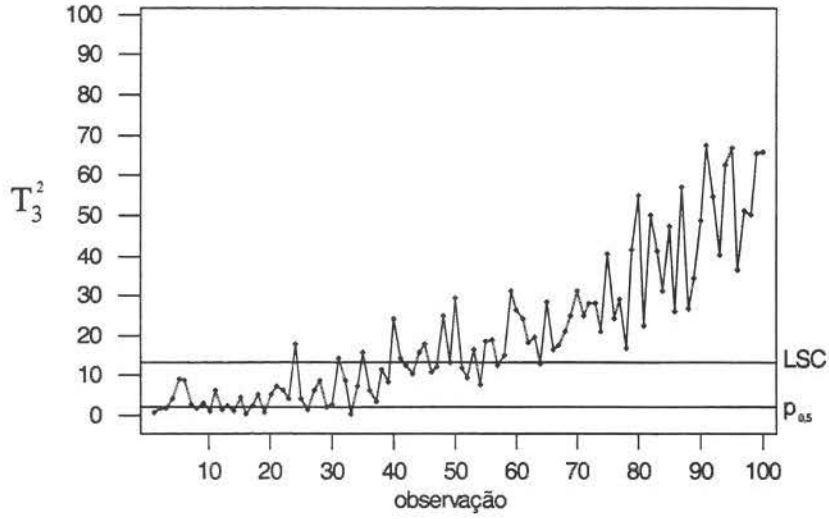

Figura 2.13 – Gráfico T^2 utilizando $\mathbf{S}_3$ - dados com sinal tipo rampa

2.3.4 Observações Aberrantes

Uma situação de difícil detecção na fase I para dados individuais é a presença de observações aberrantes distribuídas de maneira aleatória entre as m observações. Veremos que agora o estimador de diferenças sucessivas $\mathbf{S}_3$ terá desempenho semelhante ao estimador $\mathbf{S}_1$.

Com os dados originais do item 2.3.1, alteramos as observações 10, 30, 50, 70 e 90 para valores relativamente altos com respeito à não-centralidade. Especificamente, os valores das observações são: $\mathbf{x}_{10} = [12 \quad 15 \quad -21]$, $\mathbf{x}_{30} = [-6 \quad 7{,}5 \quad 10{,}5]$, $\mathbf{x}_{50} = [6 \quad -7{,}5 \quad 10{,}5]$, $\mathbf{x}_{70} = [-6 \quad -7{,}5 \quad -10{,}5]$ e $\mathbf{x}_{90} = [-12 \quad -15 \quad 21]$, e cujos parâmetros de não-centralidade são respectivamente [11,9]; [18,2]; [32,9]; [15,9]; [11,9]. O gráfico da Figura 2.14 traz a posição relativa dessas observações em relação aos dados.

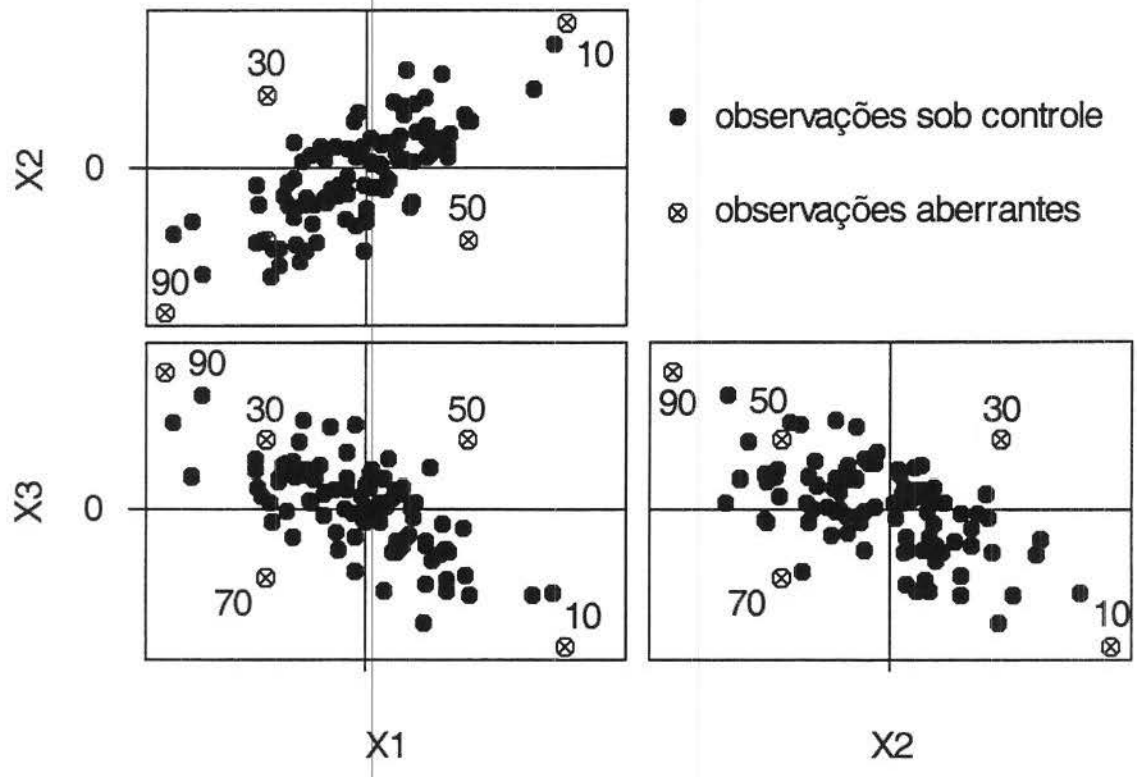

Figura 2.14 – Dados com observações aberrantes

Veja nos gráficos de controle a seguir que ambos os estimadores produzem resultados muito semelhantes quanto à magnitude dos valores da estatística plotada. É fácil imaginar porque o estimador $\mathbf{S}_3$ tem relativo baixo desempenho, essas observações aberrantes encontram-se distribuídas separadamente ao longo do tempo, cada uma provoca deformação em duas diferenças sucessivas inflacionando a estimativa da matriz de covariância. Era de se esperar até um desempenho pior que o de $\mathbf{S}_1$ nesse caso.

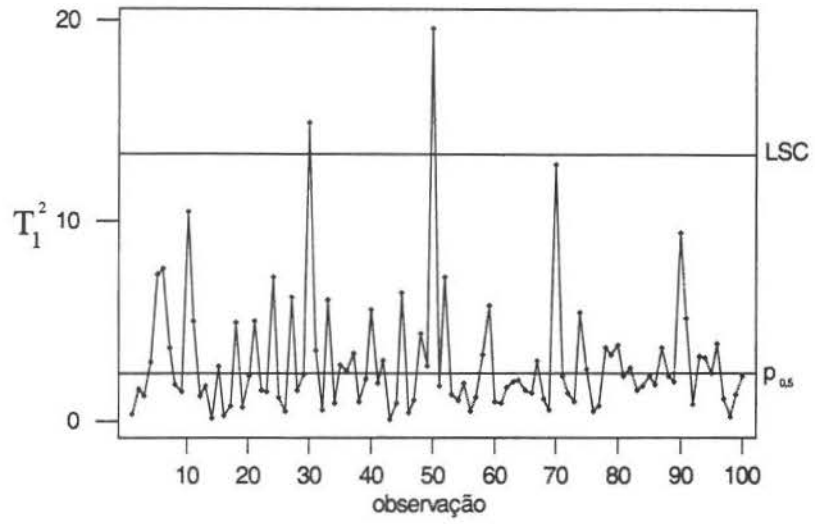

Figura 2.15 – Gráfico T^2 utilizando $\mathbf{S}_1$ para dados da Figura 2.14

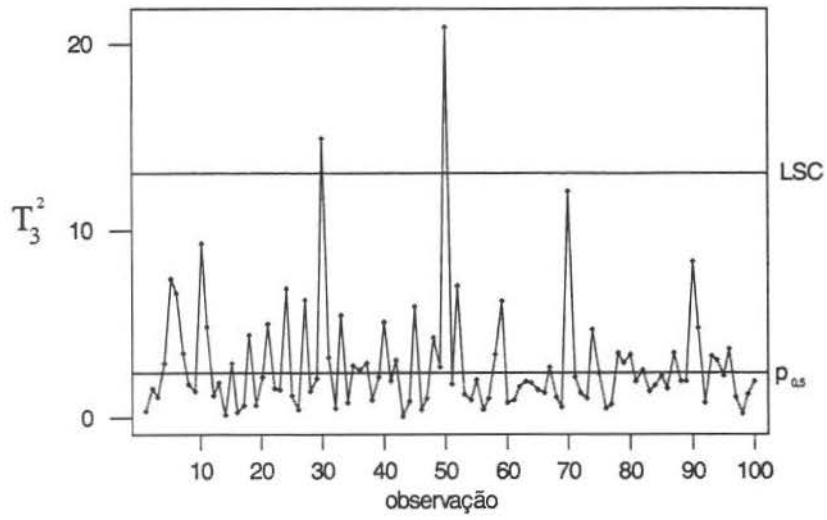

Figura 2.16 – Gráfico T^2 utilizando $\mathbf{S}_3$ para dados da Figura 2.14

2.4 Alternativas para Detecção de Observações Aberrantes na Fase I

As conseqüências de haver observações aberrantes numa amostra multivariada são mais complexas que no caso univariado. Uma razão é que a distorção provocada pela observação aberrante pode atingir não somente as medidas de locação e escala como também a estrutura de correlação dos dados. Outra razão é a maior dificuldade de caracterização da observação aberrante. Uma terceira razão é a variedade de tipos de aberrações que podem ocorrer, por exemplo, um erro grosseiro numa das variáveis, ou espalhado em todas as componentes.

Em geral, técnicas estatísticas para se detectar e explicar fugas da condição sob controle diferem em efetividade dependendo da natureza da fuga. É prática comum dos usuários aplicarem um verdadeiro arsenal de técnicas para se chegar a um diagnóstico abrangente. Duas características fundamentais dessas técnicas é que sejam simples suficiente para que grandes bancos de dados possam ser processados e com resultados fáceis de serem interpretadas pelos usuários. Há duas categorias de métodos, aqueles de análise da estrutura interna dos dados e aqueles que dependem de alguma estrutura externa. As aplicações de controle estatístico normalmente não exploram os dados com estrutura externas, portanto iremos nos deter à primeira categoria, especificamente veremos técnicas de agrupamento de observações e de decomposição da estatística T^2 procurando descrever e resumir o corpo de observações.

Woodall & Sullivan (1996) quando se depararam com a ineficácia dos estimadores de sigma na fase I para observações aberrantes estudaram um método de agrupamento iterativo de aumento do poder dos gráficos de controle. O método sugerido é proposto por Atkinson & Mulira (1993) e se chama Gráfico de Estalactite devido ao seu aspecto visual. Tem uma boa eficiência na detecção de observações aberrantes, porém pecam nas características simplicidade de cálculos e facilidade de interpretação dos resultados.

Entre as mais importantes técnicas de decomposição da estatística T^2 está a análise de componentes principais. A maioria dos autores (Jackson 1979, 1980, 1985; Kourti & MacGregor 1996; Johnson & Wichern 1998) enfatizam os aspectos de redução da dimensão e de interpretação do corpo de dados colocando grande atenção às primeiras componentes que retêm a maioria da variabilidade. Por outro lado, com objetivos de detecção de observações aberrantes, atenção é dada às últimas componentes em Gnanadesikan & Kettenring (1972) por conterem informações de possíveis singularidades devido à falta de ajuste aos subespaços gerados pelas primeiras componentes.

Não podemos nos esquecer, contudo, das técnicas fundamentais de análise descritiva das variáveis marginais e dos procedimentos do item 2.2 que deveria ser sempre o primeiro passo de toda investigação.

2.4.1 Gráfico de Estalactite

O método do gráfico de estalactite proposto por Atkinson & Mulira (1993) consiste na seleção de um subgrupo inicial pequeno contendo $p+1$ das observações o qual será usado, nesse primeiro passo, para se estimar os parâmetros para o cálculo de $T_i^2 \quad i=1,\cdots,m$. No próximo passo forma-se um novo subgrupo de $p+2$ observações com os menores valores da estatística de Hottelling do passo anterior. Isso segue até o passo em que o subgrupo contenha todas as observações. Note que uma observação que estava no subgrupo no passo anterior pode não estar lá no passo seguinte. A cada passo, as observações com valores excessivamente elevados de T^2 são marcadas como observações aberrantes em potencial. Assim ao final teremos uma matriz com a situação de cada observação em cada passo. Essas informações gerarão o gráfico de estalactite.

As observações aberrantes tenderão a serem marcadas até que o momento em que elas passam a pertencer ao subgrupo de cálculo, isto é, nos últimos passos do método. A partir daí, elas tendem a não serem marcadas mais. Eventualmente, nos primeiros passos, observações que não são aberrantes podem ser marcadas e espera-se que isso não se

perdure durante o processo. Atkinson & Mulira (1993) dizem que não é necessário procurar por um subgrupo inicial sem observações suspeitas pois elas tendem a sair do subgrupos nos passos seguintes. Eles dão um exemplo em que o subgrupo inicial contém somente observações aberrantes e mesmo assim o método funciona bem.

Resta, ao usuário do método, desenvolver a habilidade de distinguir as observações aberrantes no gráfico de estalactite. Vejamos por exemplo o gráfico de estalactite da Figura 2.17 que é resultado do processamento dos dados do item 2.3.4.

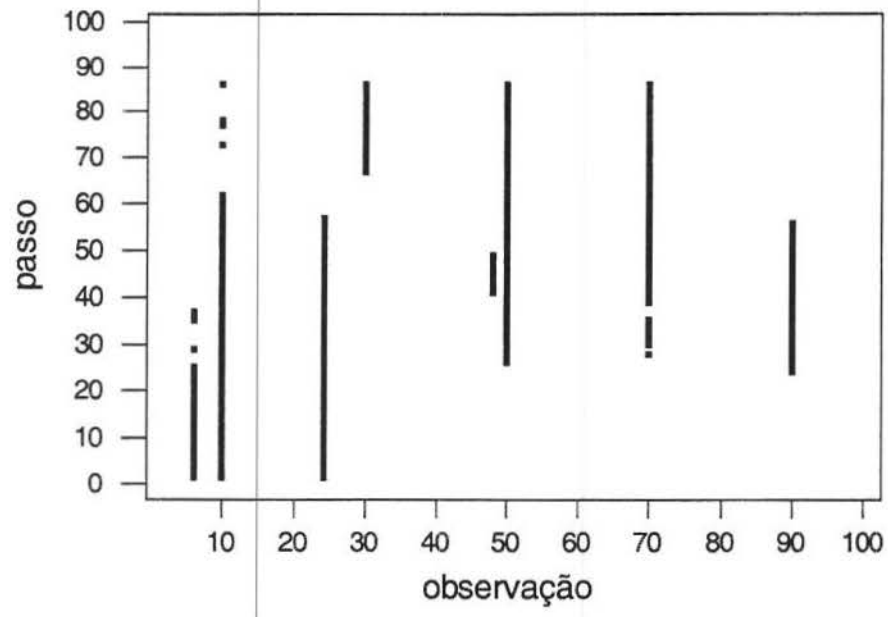

Figura 2.17 – Gráfico de Estalactite para os dados da Figura 2.14

Nesse gráfico vemos que inicialmente as observações 6, 10 e 24 são marcadas. Ao meio do processo, a observação 6 deixa de ser marcada e surgem, além da 10 e 24, as observações 48 (brevemente), 50, 70 e 90. Ao final restam somente a 10 (sem muita "firmeza"), 30, 50 e 70.

Sem dúvida, esse método aumenta o poder de detecção de observações aberrantes. No nosso caso, ele foi capaz de detectar 4 dos 5 pontos aberrantes implantados nos dados.

Sentimo-nos inseguros quanto para afirmar, ao aplicar esse método, quais pontos são aberrantes diante de uma massa de dados desconhecida. Ou seja, não há um critério objetivo (um teste) que leve à tal decisão. O seu poder, por isso, não pode ser sequer avaliado sendo um método puramente gráfico. Sua implantação em termos de cálculos também tem problemas. O critério de marcação das observações não tem uma definição clara nas referências.

O método é sem dúvida muito promissor se esses problemas forem resolvidos e se sua aplicação e interpretação tornarem-se mais confortáveis para o usuário.

2.4.2 Resíduos de Componentes Principais

Conforme vimos no item 1.1.3, as formas quadráticas podem ser decompostas e o ponto de partida da análise de componentes principais amostrais consiste na transformação das variáveis da matriz $\mathbf{X}$ num novo conjunto de variáveis $\hat{\mathbf{Y}}$ dada por:

$$\hat{\mathbf{Y}} = \hat{\mathbf{P}}'(\mathbf{X} - \overline{\mathbf{X}})$$

com as seguintes propriedades:

(i) Cada variável de $\hat{\mathbf{Y}}$ é uma combinação linear de $\mathbf{X}$ ou seja, $\hat{\mathbf{y}}_i = \hat{\mathbf{e}}_i'(\mathbf{X} - \overline{\mathbf{X}})$, $i = 1, \ldots, p$

(ii) A soma dos quadrados dos coeficientes é unitária, ou seja, $\hat{\mathbf{e}}_i'\hat{\mathbf{e}}_i = 1, \quad i = 1, \ldots, p$

(iii) De todas as possíveis combinações lineares desse tipo, $\hat{\mathbf{y}}_1$ tem a maior variância

(iv) De todas as possíveis combinações lineares desse tipo com correlação nula com $\hat{\mathbf{y}}_1$, $\hat{\mathbf{y}}_2$ tema maior variância. Similarmente, $\hat{\mathbf{y}}_3$ tem a maior variância entre as combinações com correlação nula com $\hat{\mathbf{y}}_1$ e $\hat{\mathbf{y}}_2$, e assim por diante até que se complete todas as p combinações.

Dessa maneira, um novo conjunto de p variáveis é definido, sem correlação entre si e arranjados em ordem decrescente de variância. A principal idéia desse método é que as primeiras poucas componentes retenham a maioria da variabilidade presente nos dados originais e por isso é largamente utilizado para se reduzir dimensões em estudos multivariados.

O seguinte resultado esclarece melhor quem são as componentes principais em termos da matriz **S** e sua decomposição espectral:

Resultado 2.1: Seja **S** a matriz de covariância $p \times p$ com os pares de autovalores e autovetores $(\hat{\lambda}_1, \hat{\mathbf{e}}_1), (\hat{\lambda}_2, \hat{\mathbf{e}}_2), \ldots, (\hat{\lambda}_p, \hat{\mathbf{e}}_p)$ onde $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \cdots \geq \hat{\lambda}_p \geq 0$. A i-ésima componente principal é dados por:

$\hat{y}_i = \hat{\mathbf{e}}_i'(\mathbf{x} - \bar{\mathbf{x}}) = \hat{e}_{1i}(x_1 - \bar{x}_1) + \hat{e}_{2i}(x_2 - \bar{x}_2) + \cdots + \hat{e}_{pi}(x_p - \bar{x}_p), \quad i = 1, \ldots, p$ como variância amostral de $(\hat{y}_i) = \hat{\mathbf{e}}_i \mathbf{S} \hat{\mathbf{e}}_i = \hat{\lambda}_i$, $i = 1, \ldots, p$ e covariância amostral de

$(\hat{y}_i, \hat{y}_k) = \hat{\mathbf{e}}_i' \mathbf{S} \hat{\mathbf{e}}_k = 0$, $i \neq k$, também é verdade que

a variância amostral total $\sum_{i=1}^{p} s_{ii} = \sum_{i=1}^{p} \hat{\lambda}_i$

Prova: Johnson & Wichern (1998)

Usando-se variáveis padronizadas como ponto inicial, as definições e descrições precedentes corresponderiam à análise de componentes principais da matriz de correlações **R**.

Visto como método de ajuste de subespaços lineares, essa técnica pode ser usada para detectar e descrever possíveis singularidades nos dados. Neste caso o interesse seria nas projeções dos dados nas coordenadas, não somente das primeiras como também das últimas componentes principais, correspondentes aos maiores e menores autovalores, respectivamente.

As primeiras componentes são especialmente sensíveis a observações aberrantes que inflacionam de maneira imprópria as variâncias e covariâncias. As últimas componentes podem trazer informações sobre observações com falta de ajuste aos subespaços gerados pelas primeiras componentes, por exemplo observações que se posicionam muito contra a estrutura de correlação dos dados. Uma rotulagem dessas componentes podem levar ao usuário da técnica a descrições do porquê as observações foram assinaladas melhorando a qualidade do diagnóstico.

O uso de formas quadráticas da família de T^2 podem contribuir para a análise. Essa estatística é membro de uma classe geral de formas quadráticas $(\mathbf{x}_i - \bar{\mathbf{x}})' \mathbf{S}^b (\mathbf{x}_i - \bar{\mathbf{x}})$ cujo expoente b controla a ênfase sobre quais componentes principais estaremos colocando. Segundo Gnanadesikan & Kettenring (1972), quando b aumenta acima de +1, mais e mais ênfase é colocada nas primeiras poucas componentes e quando b decresce abaixo de –1, a ênfase é dada mais às últimas componentes principais.

Essa afirmação pode ser mostrada através da observação da decomposição das formas quadráticas como discutido em 1.1.3. Veja que quando $b = -1$ temos que $(\mathbf{x}_i - \bar{\mathbf{x}})' \mathbf{S}^{-1} (\mathbf{x}_i - \bar{\mathbf{x}}) = \sum_{i=1}^{p} \frac{1}{\hat{\lambda}_i} \hat{y}_i^2$, assim, vemos que se trata de uma soma de quadrados ponderados pelo inverso da variância de cada componente, portanto cada componente

contribui essencialmente com o mesmo peso. Um ponto aberrante detectado por essa estatística pode ter inflacionado quaisquer das componentes, sendo um indicador bem abrangente. Gnanadesikan & Kettenring (1972) sugerem que se utilize também estatísticas chamadas nesta dissertação de Q^2 (em que $b=0$) e D^2 (em que $b=+1$) como suporte à análise.

Veja que $D^2=(\mathbf{x}_i-\bar{\mathbf{x}})'\mathbf{S}(\mathbf{x}_i-\bar{\mathbf{x}})=\sum_{i=1}^{p}\hat{\lambda}_i\hat{y}_i^2$, o que significa que o peso praticamente estará sobre as primeiras componentes e indicará pontos aberrantes na direção de maior variabilidade.

Um indicador de problemas afetando as últimas componentes principais pode ser a estatística $U^2=\sum_{i=p-q+1}^{p}\frac{1}{\hat{\lambda}_i}\hat{y}_i^2$. Trata-se da soma dos quadrados ponderados dos escores das últimas q componentes principais.

Temos, assim, várias opções de manipulação dos dados para procurar por pontos aberrantes. Nossa sugestão de roteiro de técnicas a serem usadas é:

a) Análise descritiva das variáveis marginais. Um passo fundamental;

b) Gráficos de controle multivariados para individuais (procedimentos dos item 2.2)

c) Gráficos de dispersão em matriz das variáveis originais bem como dos escores das componentes principais quando conveniente;

d) Gráficos de individuais para cada componente principal;

e) Gráficos de probabilidade para as estatísticas D^2 e U^2 para procurar problemas nas primeiras e últimas componentes em separado. Gnanadesikan & Kettenring (1972) sugerem gráficos de probabilidade tipo Gama com parâmetros ajustáveis para isso.

f) Retirada de pontos considerados sinais e procura por outros nos pontos remanescentes através dessas mesmas técnicas.

2.4.3 Exemplo Ilustrativo de Detecção de Observações Aberrantes

Vejamos a seguir esses métodos aplicados à situação do item 2.3.4. A situação é visualizada através do gráfico de dispersão em matriz (Figura 2.14). A análise de componentes principais da matriz de correlação desses dados tem como resultado os autovalores e coeficientes da Tabela 2.1 abaixo.

Note que a componente 1 trata-se de um contraste entre a média de X1 e X2 contra X3, refletindo a estrutura de correlação e representando portanto o eixo de maior dispersão dos dados. Já a segunda componente pontua observações que não seguem a estrutura de correlação por ser uma média ponderada de todas as variáveis com ênfase em X2 e X3. A terceira componente vai pontuar as observações que não seguem o padrão da correlação entre X1 e X2 por ser um contraste entre essas variáveis.

Tabela 2.1 – Resultado da análise de componentes principais

	$\hat{\lambda}_1$	$\hat{\lambda}_2$	$\hat{\lambda}_3$
Autovalor	82,252	16,713	6,098
Proporção	0,783	0,159	0,058
	coeficientes		
Variável	PC1	PC2	PC3
X1	-0,413	-0,255	0,874
X2	-0,491	-0,746	-0,449
X3	0,767	-0,615	0,183

Veja o gráfico de dispersão das pontuações nas componentes principais na Figura 2.18 abaixo e observe a posição relativa das observações aberrantes implantadas nos dados. Note o grande destaque das observações 30, 50 e 70 no gráfico da componente 2 contra a componente 3.

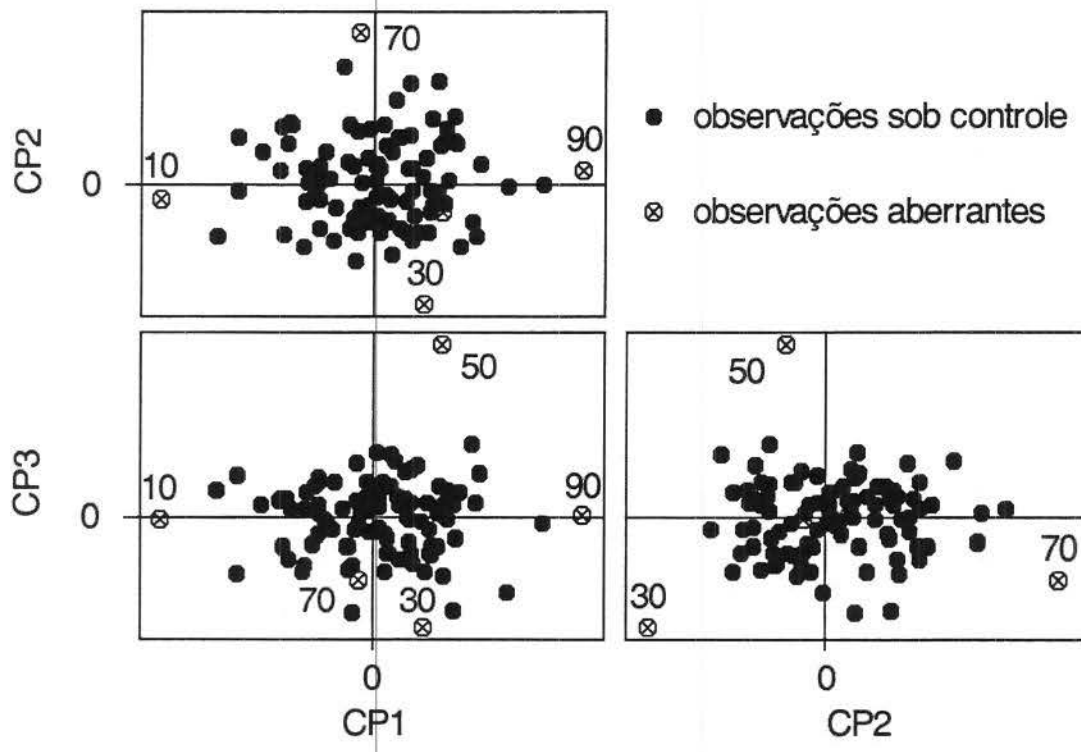

Figura 2.18 – Gráfico de dispersão dos escores das componentes principais

Na Figura 2.19 as cinco observações aberrantes são detectadas pelos gráficos de individuais das componentes principais em separado.

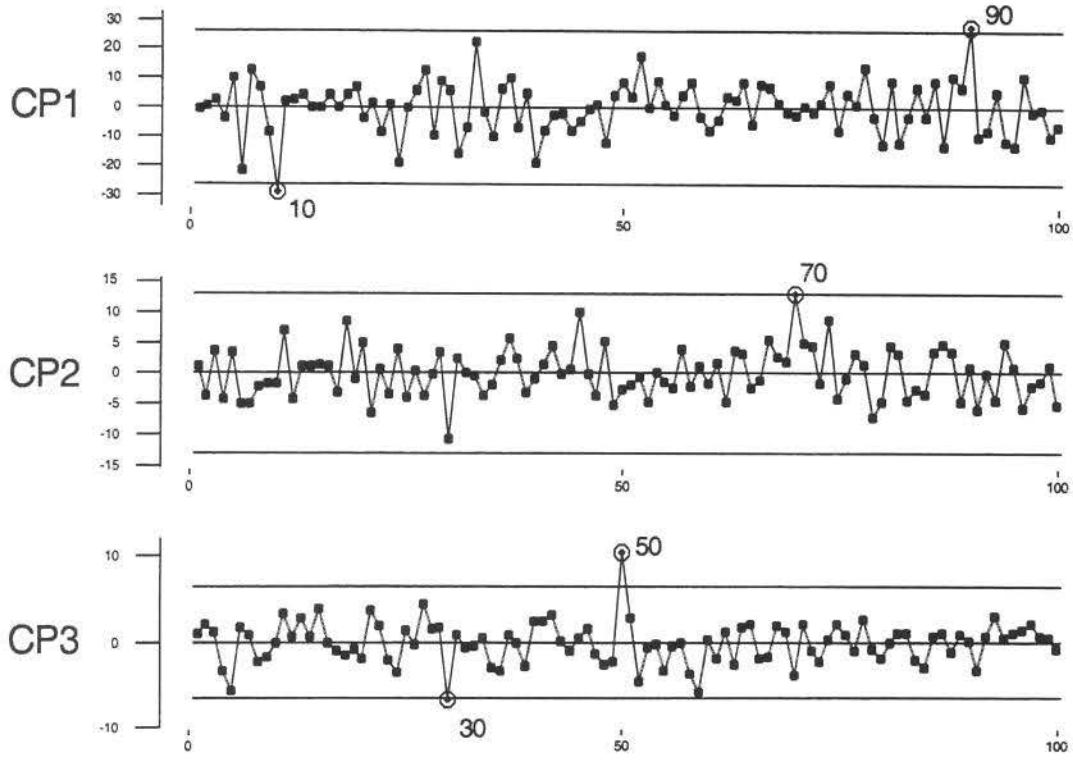

Figura 2.19 – Gráficos de controle individuais para os escores das componentes principais

As estatísticas D^2 e U^2 foram calculadas e seus gráficos de probabilidade (Q-Q Plots) são mostrados abaixo.

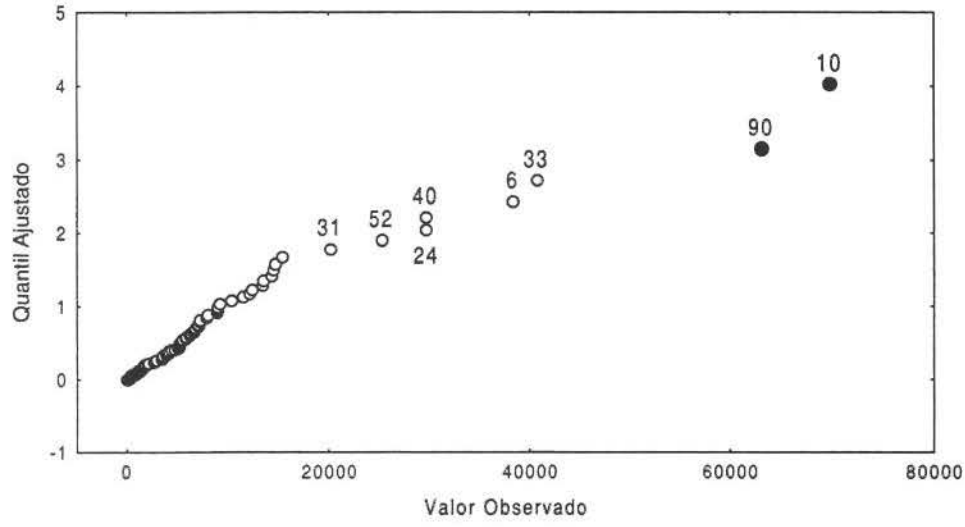

Figura 2.20 – Q-Q Plot (distribuição Gama) para a estatística D^2

O gráfico de D^2 da Figura 2.20 evidencia as observações 90 e 10 como possíveis pontos aberrantes que estão afastados da média ao longo da direção de maior variabilidade. Essa é uma informação muito importante para os usuários desde que representa possivelmente uma mudança do processo como um todo, sendo necessário um ajuste global para que volte ao estado de controle estatístico.

Outras observações também são suspeitas e estão assinaladas no gráfico – são observações em que em se examinando os registros de bordo do processo pode-se ter alguma decisão se devem ou não serem assinaladas como pontos aberrantes. Esse gráfico é muito útil, portanto, para elaboração de diagnóstico sobre a massa inicial de observações focada nas primeiras componentes principais.

Observando o gráfico U^2 na Figura 2.21, vemos que ele assinala claramente as observações que fogem da estrutura de correlação implantadas na massa de dados. É portanto um detetor importante de anomalias de processo provocadas por discordâncias entre variáveis, não querendo dizer que o processo propriamente dito esteja fora de controle.

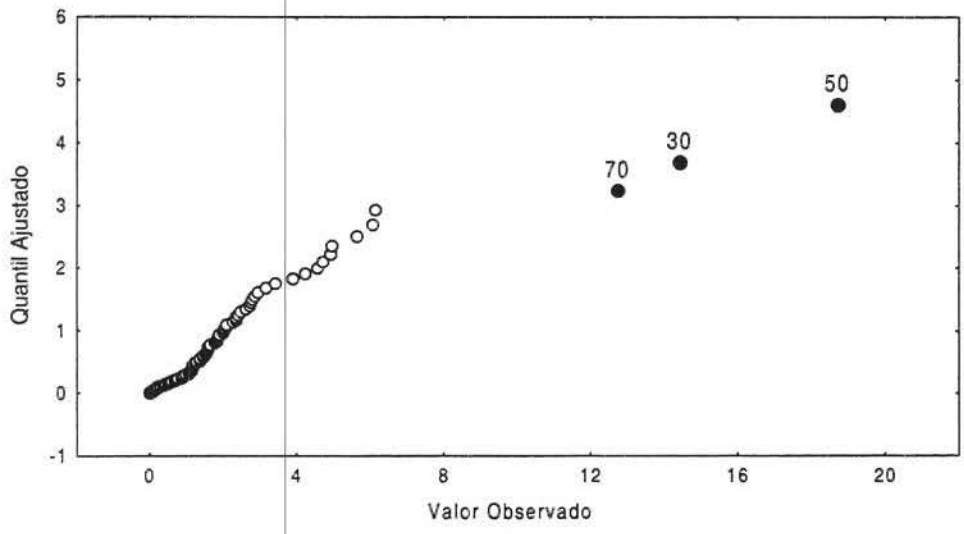

Figura 2.21 – Q-Q Plot (distribuição Gama) para a estatística U^2

Os usuários baseiam-se em observações do processo para controlá-lo e, nesse caso, procurarão por razões do porquê dessas discordâncias, antes de atuarem no processo como um todo.

Essa discussão já entra na questão de identificação das causas prováveis de observações fora de controle e alguns meios serão abordados na próxima seção.

2.5 Interpretação de Sinais

Gráficos de controle multivariados podem detectar um evento não usual acontecendo no processo, porém não dão informações das razões dele ocorrer. Essa é uma crítica antiga dessa metodologia. Por outro lado, o diagnóstico da causa raiz de um ponto fora de controle não é obvio nem no caso univariado, precisando de conhecimento do processo e uma investigação cuidadosa. Na conclusão final, é importante usar toda informação disponível, inclusive aquela que não está nos gráficos, mas aparece nos registros de processo.

No caso multivariado essa tarefa é complexa, mesmo tendo evidências de que uma variável em especial é a causadora do sinal, a investigação de outras variáveis não poderia ser deixada de lado. A certeza da certeza pode levar as pessoas de um processo a atuarem de maneira inadequada nele.

Considere, por exemplo, um processo bivariado com as variáveis centradas em zero, de variâncias iguais a 1 e correlação positiva entre si. A região de controle é uma elipse centrada em zero e inclinada 45 graus como mostra a Figura 2.22.

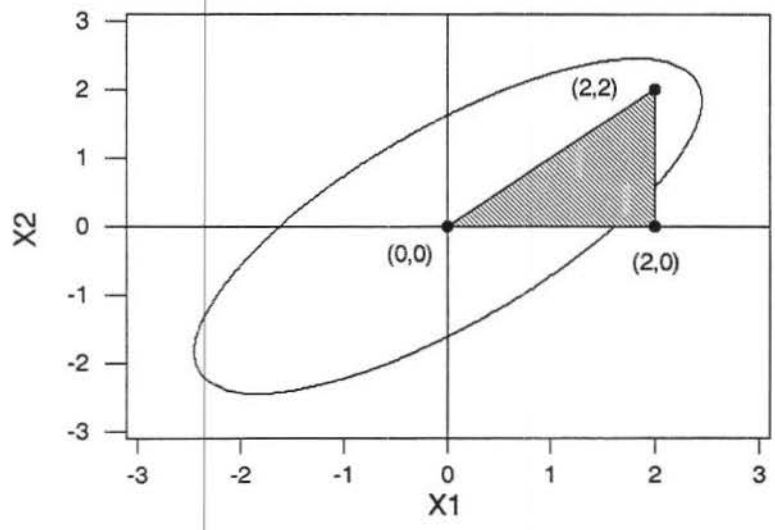

Figura 2.22 - Ilustração para diagnóstico de pontos fora de controle

Veja o que aconteceria se o vetor (2,0) fosse detectado como um sinal e uma interpretação tivesse que ser realizada. Em princípio, quaisquer pontos nas arestas do triângulo formado pelos pontos (0,0), (2,0) e (2,2) são possíveis de estarem representando o "verdadeiro valor do processo" no momento. A bem da verdade, quaisquer pontos dentro ou fora desse triângulo também teriam essa condição. Porém, dados perto do ponto (0,0) são mais prováveis, então as suspeitas nesse caso recaem sobre a variável X1, mas podendo ser, contudo, um diagnóstico falso.

Muitos métodos de interpretação de sinais foram propostos na literatura, mas a maioria tem desvantagens. Uma idéia óbvia é consultar os gráficos de controle das marginais, mas sinais que foram provocados por discordância da estrutura de correlação podem passar sem diagnóstico. Alt (1985) e outros autores sugerem isso com auxílio do método Bonferroni de comparações múltiplas (veja item 1.2.7).

Fuchs & Benjamini (1994) propuseram um método alternativo chamado de "Gráfico do Perfil Multivariado[5]" em que no lugar dos pontos do gráfico T^2, são colocados pequenos gráficos de barras representando a situação de estatísticas univariadas.

Técnicas de decomposição da estatística T^2 têm sido elaboradas na procura de subconjuntos de variáveis que expliquem as situações fora de controle. Murphy (1987) sugere particionar as p variáveis em dois conjuntos, um contendo aquelas variáveis que intuitivamente achamos que tenham problemas e as restantes formando o outro grupo. Trata-se de uma análise discriminante que incorpora as informações de correlação através da matriz de covariância. Um problema com esse procedimento é que quanto mais variáveis há no processo, haverá mais possibilidades de seleção das variáveis levando a erros quase certamente.

---

[5] Do inglês: "Multivariate Profile Chart"

Tracy et alli (1995a, 1997) decompuseram T^2 em p componentes independentes cada uma contendo informações das variáveis que contribuem para a acusa especial. Uma possível decomposição será, por exemplo:

$$T^2 = T_1^2 + \sum_{j=2}^{p} T_{j.1,2,\ldots,j-1}^2$$

onde T_1^2 é o quadrado da estatística t univariada e

$$T_{j.1,2,\ldots,j-1}^2 = \frac{\left(x_j - \bar{x}_{j.1,2,\ldots,j-1}\right)^2}{s_{j.1,2,\ldots,j-1}^2}$$

é o quadrado da j-ésima variável ajustada pelas estimativas da média e desvio padrão da distribuição condicional de x_j dado $x_1, x_2, \ldots, x_{j-1}$. A desvantagem desse método está na grande quantidade de cálculos a serem feitos por ser um método permutacional.

O uso de componentes principais foi bastante explorado como técnica de diagnóstico. Jackson (1979) (1980) e (1985) e Kourti & McGregor (1996) trazem discussões a respeito. Porém, a maioria dos autores menciona que a falta de interpretação prática das componentes limita a efetividade do método. Esse ceticismo talvez seja exagerado, já que o uso das componentes principais pode ser muito útil como parte complementar do trabalho de diagnóstico, tendo em mente que nenhuma das técnicas deveria ser usada como panacéia.

Runger, Alt & Montgomery (1996) propuseram um método muito simples e prático que pode perfeitamente complementar as informações trazidas com a decomposição das observações em componentes principais. Esse método busca a contribuição individual das variáveis para minimizar a estatística T^2 e será visto na seção seguinte.

2.5.1 *Método das Contribuições das Marginais*

A contribuição de uma variável para um sinal no gráfico de controle pode ser medida através da minimização da estatística T^2 obtida pela mudança em uma única variável. As variáveis que requerem uma mudança muito grande para minimizar T^2 serão importantes para o sinal. A comparação das grandezas dessas mudanças em cada evento especial levará ao conjunto de variáveis que merecem atenção no diagnóstico.

Seja a situação bivariada descrita no início da seção anterior e considere uma correlação de 0,75 entre as variáveis. Partindo-se de um ponto aberrante dado por (1,-2) poderíamos caminhar na direção X2 no sentido ascendente até que a estatística T^2 fosse minimizada. Isso acontecerá no ponto (1, 0,75). Por outro lado, se caminhássemos na direção X1 em sentido negativo achamos o mínimo de T^2 em (-1,5, -2). Como as duas variáveis têm a mesma variabilidade, as distâncias percorridas podem ser comparadas, isto é, mudamos 2,5 unidades em X1 para minimizar T^2 (mantendo X2 constante) e mudamos 2,75 unidades em X2 para minimizar T^2 (mantendo X1 constante). Isso indica que X2 contribui relativamente mais que X1 para esse evento assinalável, tornando-se a primeira variável suspeita.

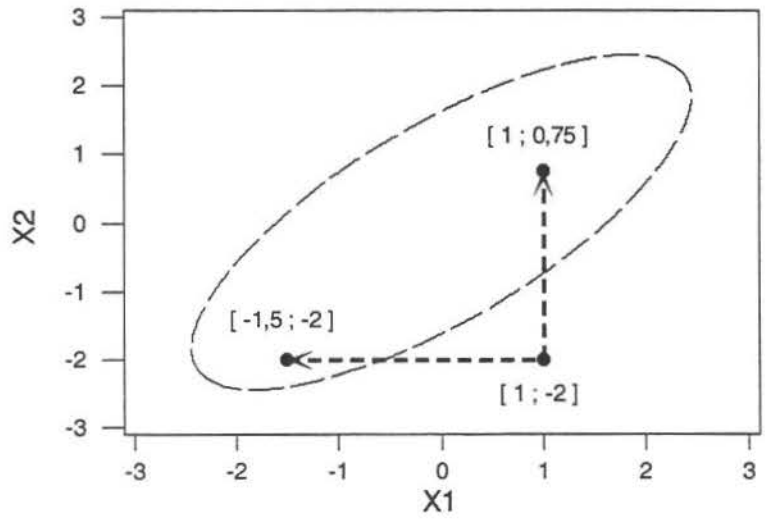

Figura 2.23 - Ilustração de aplicação do Método das Contribuições das Marginais

Para formalizar essa técnica seja $\mathbf{X}$ o vetor aleatório da observação a ser estudado com seus componentes X_i, $i = 1, \cdots, p$, podemos assumir que esse vetor tem média μ e matriz de covariância Σ. Será plotado no gráfico de controle o valor de $T^2 = (\mathbf{X} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{X} - \boldsymbol{\mu})$.

Seja u_i o vetor unitário na direção da i-ésima coordenada. Para medir a contribuição de X_i para T^2, determina-se c_i para minimizar

$$T_i^2 = \left(\mathbf{X} - \boldsymbol{\mu} - \frac{c_i \mathbf{u}_i}{\left(\mathbf{u}_i' \boldsymbol{\Sigma}^{-1} \mathbf{u}_i \right)^2} \right)' \boldsymbol{\Sigma}^{-1} \left(\mathbf{X} - \boldsymbol{\mu} - \frac{c_i \mathbf{u}_i}{\left(\mathbf{u}_i' \boldsymbol{\Sigma}^{-1} \mathbf{u}_i \right)^2} \right)$$

Podemos observar que c_i poderá ser comparado para todo i desde que está normalizado.

Usando-se

$$\mathbf{Z} = \boldsymbol{\Sigma}^{-\frac{1}{2}} (\mathbf{X} - \boldsymbol{\mu}) \text{ e } \mathbf{v}_i = \frac{\boldsymbol{\Sigma}^{-\frac{1}{2}} \mathbf{u}_i}{\left(\mathbf{u}_i' \boldsymbol{\Sigma}^{-1} \mathbf{u}_i \right)^2}$$

temos que

$$T_i^2 = (\mathbf{Z} - c_i \mathbf{v}_i)' (\mathbf{Z} - c_i \mathbf{v}_i)$$

e esta expressão pode ser interpretada como a soma de quadrados dos resíduos no modelo de regressão de $\mathbf{Z}$ sobre $\mathbf{v}_i$. Assim,

$$T_i^2 = \mathbf{Z}'(\mathbf{I} - \mathbf{P}_i)\mathbf{Z} = \mathbf{Z}'\mathbf{Z} - \mathbf{Z}'\mathbf{v}_i\mathbf{v}_i'\mathbf{Z} = \mathbf{Z}'\mathbf{Z} - c_i^2$$

Portanto, uma variável i será importante para diagnóstico se $c_i^2 = T^2 - T_i^2$ for grande.

O método para o cálculo de c_i^2 é bastante simples. Runger (1996) mostra que T_i^2 é igual ao valor da estatística T^2 calculada a partir das $p-1$ variáveis omitindo-se X_i.

Vejamos na seção seguinte os resultados desse método em conjunto com informações das componentes principais para diagnóstico dos pontos aberrantes do item 2.3.4.

2.5.2 Diagnóstico Ilustrativo de Pontos Aberrantes

Para o diagnóstico dos cinco pontos aberrantes implantados nos dados simulados do item 2.3.4, precisamos inicialmente dividi-los entre pontos detectados pela estatística D^2 (primeiras componentes principais) e detectados pela estatística U^2 (últimas componentes). Os pontos 10 e 90 foram detectados pela estatística D^2, portanto assinalam que o processo teve um deslocamento global na direção de maior variabilidade. O sentido do deslocamento é facilmente identificável, portanto essas já são informações suficientes para um diagnóstico confiável, ou seja, os responsáveis pelo processo devem agir no processo como um todo para que ele volte ao estado de estabilidade.

Já os pontos 30, 50, e 70 aparecem como tendo altos escores nas últimas componentes principais e não se tendo uma interpretação direta dessas componentes podemos aplicar o método de contribuições para obtermos o diagnóstico. Veja os resultados no gráfico da Figura 2.24 abaixo.

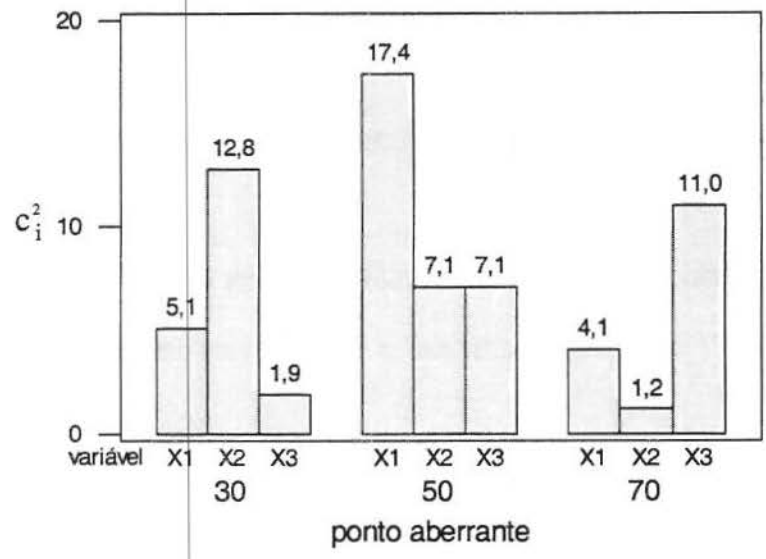

Figura 2.24 – Valores das contribuições das marginais para as observações aberrantes da Figura 2.14

O diagnóstico para o ponto 30 aponta X2 como a variável que provavelmente provocou essa situação (note o maior valor de c^2). Veja que na Figura 2.14 o ponto 30 aparece mal posicionado nos dois diagramas de dispersão em que uma das coordenadas é X2. No diagrama de dispersão entre X1 e X3, o ponto 30 aparece imerso na massa de dados. Para o ponto 50, a variável X1 deve ser a responsável e para o ponto 70, a variável X3 deveria ser investigada.

Após sabermos qual (ou quais) as variáveis que mais contribuem para o aparecimento da causa especial através do método exposto acima, podemos visualizar mais facilmente como explicar isso através dos diagramas de dispersão, da mesma maneira que para o ponto 30.

Seguindo o princípio de facilidade de cálculo e interpretação, esse método aliado ao roteiro de técnicas vistas até aqui, formam um conjunto de ferramentas úteis às pessoas do processo na investigação e aprendizado da variabilidade do processo, passo sempre presente e fundamental no alcance de metas de melhoria.

Capítulo 3

Exemplo de Aplicação: Processo de Montagem de Carroçaria de Caminhão.

3.1 Contexto e Dados para Análise[6]

Atualmente na montagem de carroçarias de veículos, peças estampadas em aço são unidas entre si formando sua estrutura. Medidas tridimensionais dessa estrutura têm de ser controladas para que o restante de peças possam ser adequadamente montadas em etapas posteriores do processo. Em particular, peças móveis como portas e capuz do motor estão recebendo especificações muito rígidas nas folgas no intuito de aumentar tanto a confiabilidade do produto quanto a qualidade percebida pelos consumidores.

No caso de caminhões, o processo de montagem da cabina pode ser muito resumidamente descrito no fluxograma da Figura 3.1. Uma característica de qualidade desse processo é a união perfeita das peças fixas para que se minimize os custos de ajustes das peças móveis.

Um dispositivo de medição tridimensional é usado para se obter dados de vários pontos da cabina pré montada (local 1 na Figura 3.1). Tratam-se de medidas do desvio dos valores nominais em unidades de comprimento para as cotas X, Y e Z em cada ponto especificado.

---

[6] Os dados apresentados neste Capítulo são de propriedade de uma indústria montadora de caminhões que autorizou somente a divulgação de informações gerais essenciais para o entendimento do uso das técnicas apresentadas.

As medições são caras e demoradas e nesse caso uma cabina é retirada do processo em intervalos regulares de tempo para ser medida caracterizando um processo de medições individuais multivariadas.

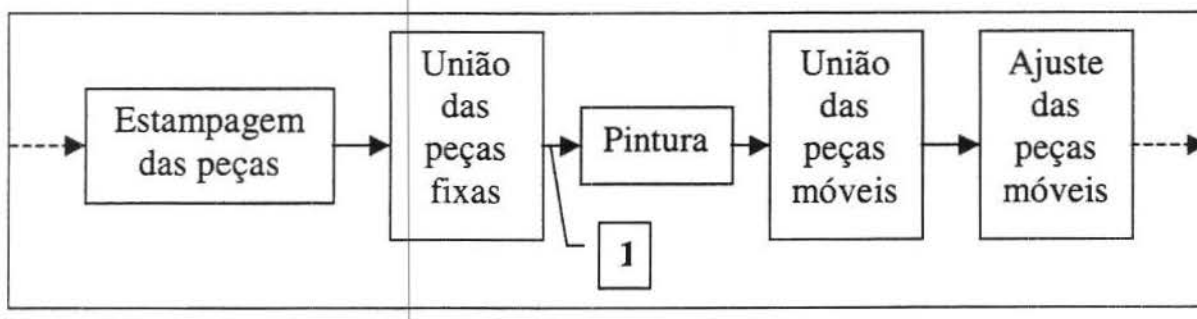

Figura 3.1 – Parte do processo de montagem de cabinas de caminhão

Estudando-se em particular o sub-processo de montagem do capuz do motor, 4 pontos no assento do capuz são importantes para controle. Em cada ponto as cotas X e Y são de maior interesse já que Z, a altura, é de fácil ajuste. Os pontos são mostrados esquematicamente na Figura 3.2. As 8 variáveis de interesse serão denominadas XFD, XFE, XTD, XTE, YFD, YFE, YTD, YTE e o significado dessas siglas pode ser compreendido na Tabela 3.1

Tabela 3.1 – Significado das siglas usadas como nomes das variáveis em estudo

Variável	Ponto	Direção de medição	Posição	Lado
XFD	A	Transversal	Frente	Direito
XFE	B	Transversal	Frente	Esquerdo
XTD	C	Transversal	Traseira	Direito
XTE	D	Transversal	Traseira	Esquerdo
YFD	A	Longitudinal	Frente	Direito
YFE	B	Longitudinal	Frente	Esquerdo
YTD	C	Longitudinal	Traseira	Direito
YTE	D	Longitudinal	Traseira	Esquerdo

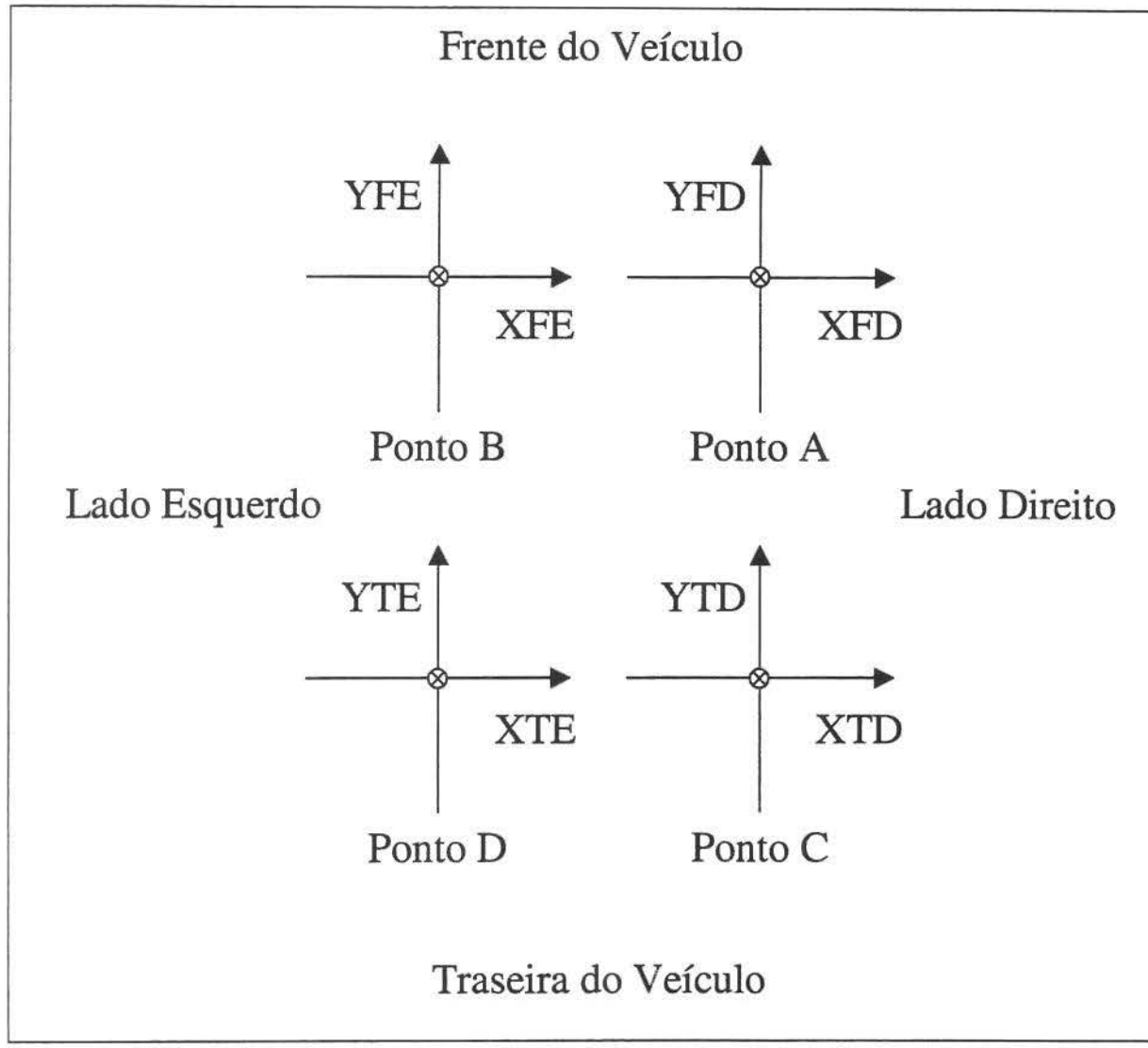

Figura 3.2 – Esquema de posicionamento dos pontos de medição

O grupo de trabalho iniciou os esforços de melhoria com o objetivo de entender a situação atual do processo em relação à variabilidade do assento do capuz para poder agir preventivamente e reduzir custos de ajuste bem como ganhar produtividade na linha de montagem. Para isso, 43 observações desse processo foram obtidas durante 2 meses de coleta de dados para um determinado tipo de caminhão.

Os dados encontram-se na Tabela 3.2 abaixo.

Tabela 3.2 – Dados observados no processo

OBS	XFD	XFE	XTD	XTE	YFD	YFE	YTD	YTE
1	0,0	0,0	-0,6	0,4	0,6	-0,7	-1,6	-2,6
2	-1,6	-1,4	2,2	1,9	0,3	-0,3	-1,4	-2,4
3	-0,9	-1,9	1,4	2,4	0,5	-0,6	-1,3	-2,3
4	-0,3	-1,3	0,7	1,7	0,1	-0,2	-1,9	-1,8
5	-0,9	-1,7	1,4	2,3	0,1	-0,2	-1,6	-1,8
6	-1,1	-1,2	1,4	1,7	0,9	-1,0	-0,9	-2,4
7	0,5	0,5	-0,6	-0,2	0,6	-0,7	-1,5	-2,5
8	-0,9	-1,2	0,9	1,7	1,0	-1,1	-1,0	-2,5
9	-1,6	-2,1	2,5	2,5	0,5	-0,6	-0,9	-2,1
10	-2,0	-1,2	2,0	1,4	-0,5	0,5	-2,4	-0,9
11	-0,8	-0,5	0,9	0,7	-0,4	0,5	-1,8	-0,8
12	-2,2	-2,3	2,6	2,6	-0,7	0,7	-2,1	-0,7
13	0,5	-0,5	0,3	1,0	-0,5	0,4	-2,1	-1,1
14	-0,1	-0,1	0,3	0,5	-0,5	0,5	-2,0	-0,4
15	-0,9	-0,1	1,1	0,4	0,3	-0,4	-1,6	-1,7
16	-0,3	-1,1	1,3	1,7	0,7	-0,8	-0,7	-2,6
17	0,5	0,5	-0,5	-0,1	0,1	-0,2	-1,8	-2,3
18	-2,3	-3,7	2,4	4,2	0,7	-0,7	-1,3	-2,4
19	0,7	-0,6	0,8	1,2	0,6	-0,6	-0,7	-2,3
20	-1,7	-1,5	1,8	2,0	0,7	-0,8	-0,9	-2,3
21	-2,9	-1,0	2,4	1,3	0,6	-0,6	-1,2	-1,8
22	-3,9	-3,7	4,1	4,1	0,6	-0,6	-1,0	-1,8
23	0,3	-0,7	0,2	1,3	0,8	-0,8	-0,9	-3,0
24	-3,2	-2,8	3,4	3,1	0,7	-0,6	-0,9	-1,7
25	-3,0	-2,9	3,3	3,2	1,1	-1,0	-0,6	-2,1
26	0,2	-0,8	0,7	1,2	0,4	-0,5	-1,3	-2,0
27	-0,6	-1,2	0,7	1,6	0,8	-0,8	-1,2	-2,3
28	1,7	-0,1	0,4	0,5	0,5	-0,7	-0,4	-3,1
29	1,3	0,4	-0,6	0,1	0,6	-0,5	-1,0	-2,1
30	-0,4	-1,7	1,7	2,2	0,9	-1,0	-0,8	-2,8
31	-2,7	-2,6	2,3	2,9	0,8	-0,8	-1,2	-1,9
32	-2,1	-1,4	2,0	1,7	0,6	-0,7	-1,1	-2,2
33	3,3	2,4	-3,0	-1,9	0,3	-0,3	-1,4	-1,5
34	-1,0	-1,5	1,7	2,0	0,8	-0,8	-0,6	-2,1
35	-4,2	-3,5	4,2	3,8	0,7	-0,7	-1,0	-1,7
36	-1,4	-1,4	1,8	1,8	0,8	-0,8	-1,2	-2,6
37	2,6	1,4	-1,7	-0,8	0,7	-0,7	-1,3	-2,6
38	-1,3	-2,1	1,4	2,6	1,1	-1,2	-1,1	-2,7
39	-4,6	-4,4	4,1	4,8	0,4	-0,5	-1,7	-1,7
40	-2,4	-2,7	2,5	3,0	1,1	-1,1	-0,9	-2,5
41	-2,1	-1,8	1,9	1,9	0,9	-1,0	-1,3	-2,4
42	-5,3	-4,4	5,3	4,8	1,1	-1,0	-0,5	-2,2
43	-2,6	-2,5	3,3	2,9	1,1	-1,1	-0,6	-2,3

3.2 Exploração dos Dados em Busca de Sinais na Fase I

A exploração desses dados seguirá o roteiro sugerido em 2.4.2 e incluirá uma tentativa de interpretação dos sinais que porventura apareçam. Comecemos, então pela análise descritiva das variáveis marginais.

3.2.1 Análise Descritiva das Variáveis Marginais

A Tabela 3.3 mostra algumas estatísticas descritivas das variáveis marginais. Nota-se que as variáveis estão deslocadas do zero. Na verdade o valor nominal não é o valor alvo do processo, servindo apenas como referência para o processo de medição. Portanto, o vetor de médias é um dos parâmetros a serem estimados nesta fase I. Observa-se também que a variabilidade das variáveis da cota X é maior que as da cota Y.

Tabela 3.3 – Estatísticas descritivas das variáveis marginais

Estatística	XFD	XFE	XTD	XTE	YFD	YFE	YTD	YTE
Média	-1,16	-1,40	1,50	1,82	0,52	-0,56	-1,23	-2,07
Mediana	-1,00	-1,40	1,40	1,70	0,60	-0,70	-1,20	-2,20
Desvio padrão	1,81	1,44	1,61	1,42	0,47	0,47	0,47	0,60

A Figura 3.3 mostra os gráficos de tendência e respectivos histogramas para todas as variáveis marginais. Nota-se que as variáveis da cota Y apresentam um comportamento aparentemente assinalável para o conjunto da 10ª à 14ª observações. Já nesses gráficos podemos suspeitar de uma forte estrutura de correlação para o conjunto das variáveis de cada cota. Isso é confirmado analisando-se a Tabela 3.4 em que temos as matrizes de covariância e de correlação respectivamente calculadas usando-se os estimadores $\mathbf{S}_1$ e $\mathbf{S}_3$ vistos no capítulo 2.

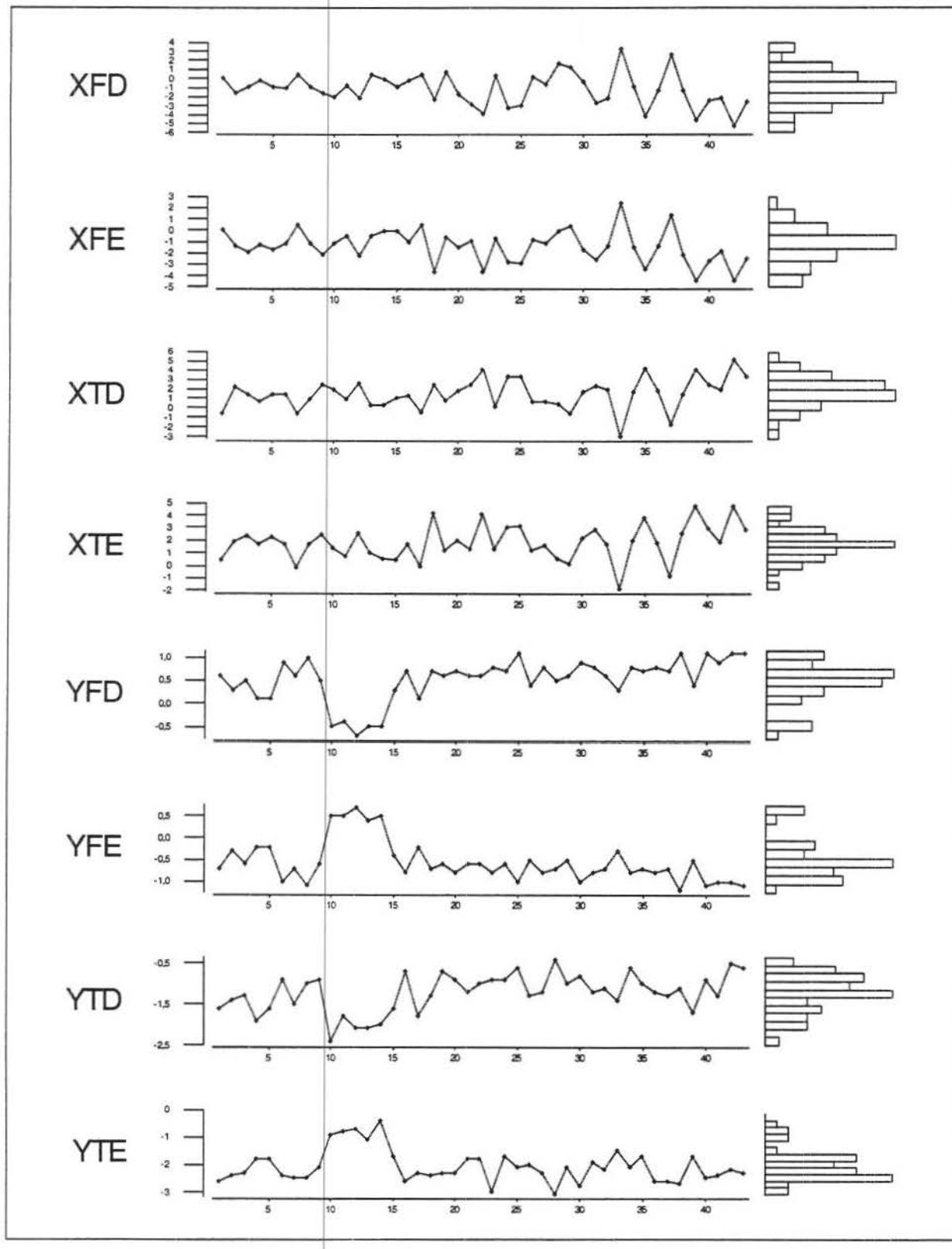

Figura 3.3 – Gráficos de tendência e respectivos histogramas para as variáveis marginais

Tabela 3.4 – Matrizes de covariância e de correlação estimadas por $\mathbf{S}_1$ e $\mathbf{S}_3$

Matriz de Covariância estimada por $\mathbf{S}_1$

XFD	3,2649							
XFE	2,3833	2,0771						
XTD	-2,7880	-2,1895	2,5974					
XTE	-2,2893	-2,0406	2,1262	2,0181				
YFD	-0,1918	-0,2030	0,1848	0,2120	0,2175			
YFE	0,1549	0,1823	-0,1549	-0,1928	-0,2162	0,2201		
YTD	-0,1227	-0,1768	0,2159	0,1919	0,1830	-0,1794	0,2243	
YTE	-0,1723	-0,0096	0,0898	-0,0177	-0,2202	0,2345	-0,1892	0,3598
	XFD	XFE	XTD	XTE	YFD	YFE	YTD	YTE

Matriz de Correlação estimada a partir de $\mathbf{S}_1$

XFD	1,0000							
XFE	0,9152	1,0000						
XTD	-0,9574	-0,9427	1,0000					
XTE	-0,8919	-0,9967	0,9287	1,0000				
YFD	-0,2275	-0,3020	0,2459	0,3199	1,0000			
YFE	0,1827	0,2697	-0,2049	-0,2893	-0,9880	1,0000		
YTD	-0,1433	-0,2590	0,2828	0,2851	0,8283	-0,8076	1,0000	
YTE	-0,1590	-0,0111	0,0929	-0,0207	-0,7873	0,8334	-0,6660	1,0000
	XFD	XFE	XTD	XTE	YFD	YFE	YTD	YTE

Matriz de Covariância estimada por $\mathbf{S}_3$

XFD	2,7595							
XFE	2,1056	1,9082						
XTD	-2,3845	-1,9762	2,2956					
XTE	-2,0449	-1,8954	1,9443	1,8939				
YFD	-0,0758	-0,1036	0,0826	0,1106	0,0811			
YFE	0,0543	0,0964	-0,0751	-0,1032	-0,0799	0,0840		
YTD	-0,0530	-0,1373	0,1625	0,1543	0,0786	-0,0795	0,1414	
YTE	-0,2018	-0,0670	0,0911	0,0507	-0,0765	0,0914	-0,0950	0,1875
	XFD	XFE	XTD	XTE	YFD	YFE	YTD	YTE

Matriz de Correlação estimada a partir de $\mathbf{S}_3$

XFD	1,0000							
XFE	0,9176	1,0000						
XTD	-0,9474	-0,9442	1,0000					
XTE	-0,8945	-0,9970	0,9325	1,0000				
YFD	-0,1603	-0,2633	0,1915	0,2823	1,0000			
YFE	0,1127	0,2408	-0,1710	-0,2587	-0,9677	1,0000		
YTD	-0,0848	-0,2642	0,2852	0,2981	0,7338	-0,7294	1,0000	
YTE	-0,2805	-0,1120	0,1388	0,0851	-0,6209	0,7283	-0,5834	1,0000
	XFD	XFE	XTD	XTE	YFD	YFE	YTD	YTE

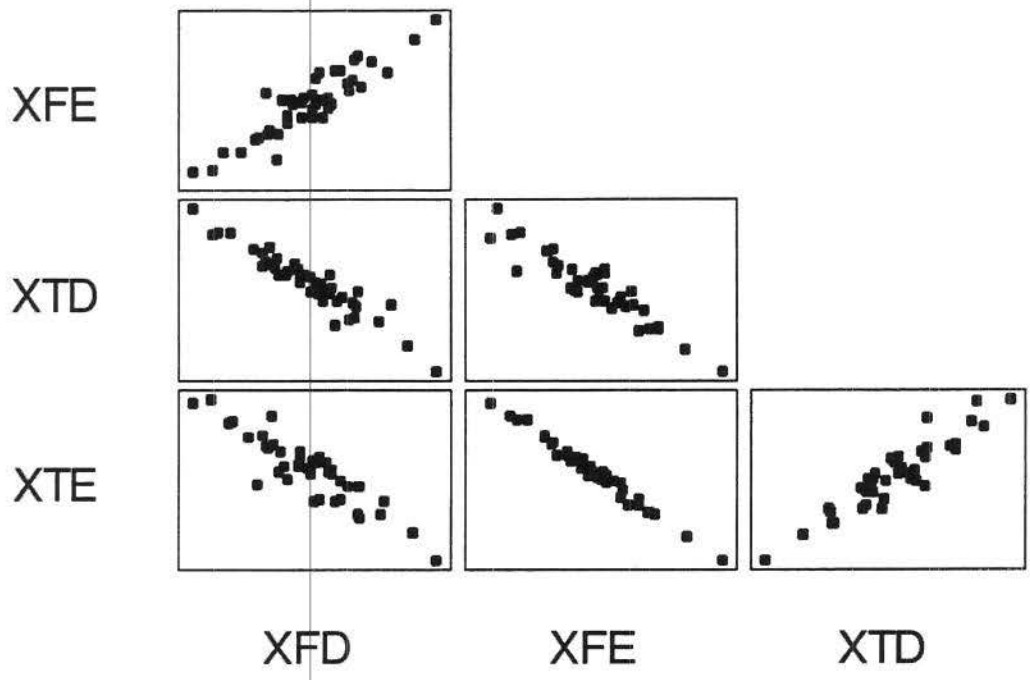

Figura 3.4 – Gráficos de dispersão para as variáveis marginais da direção X

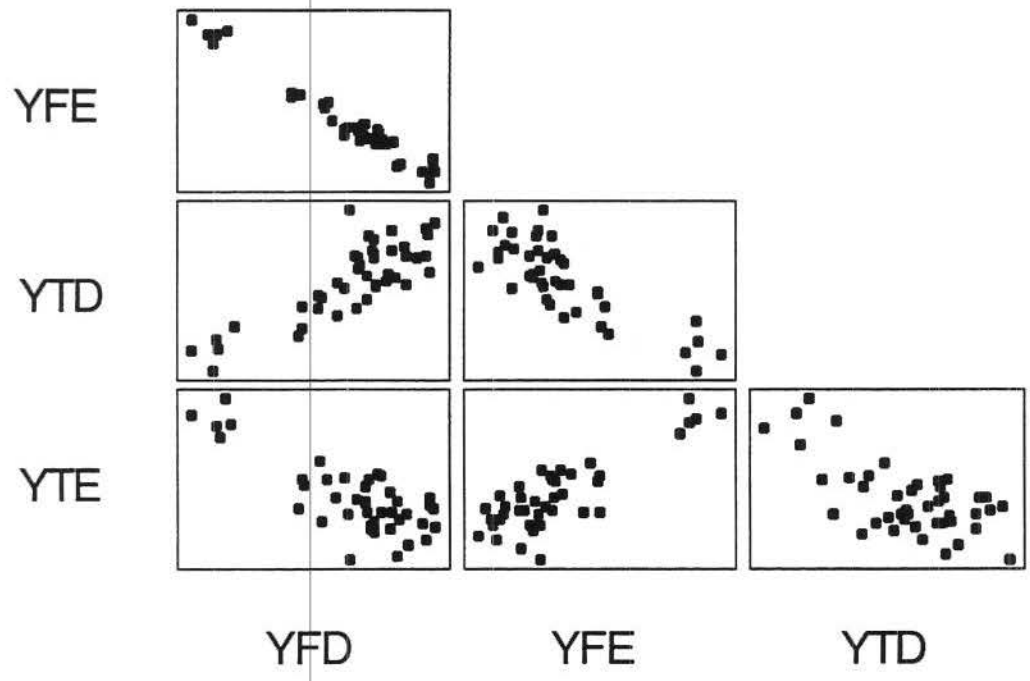

Figura 3.5 – Gráficos de dispersão para as variáveis marginais da direção Y

Podemos notar que os conjuntos de variáveis de cada cota X e Y não têm correlação entre si, assim podemos visualizar a estrutura de correlação em separado para cada grupo através dos gráficos de dispersão das Figuras 3.4 e 3.5. Vemos, também que a estimativa da matriz de covariância através de $\mathbf{S}_1$ é ligeiramente inflacionada com relação a $\mathbf{S}_3$.

Podemos imaginar como essa estrutura pôde ser obtida se visualizarmos as variações relativas dos pontos A, B C e D mais freqüentes que possivelmente a provocaram. As quatro situações representadas na Figura 3.6 mostram isso. Os pontos A e B movem-se juntos na direção X e no sentido contrário de C e D ao passo que na coordenada Y, A e C movem-se juntos e em sentido inverso de B e D. Essas variações têm maior amplitude na direção transversal ao veículo como vimos pela diferença de variabilidade entre as cotas X e Y.

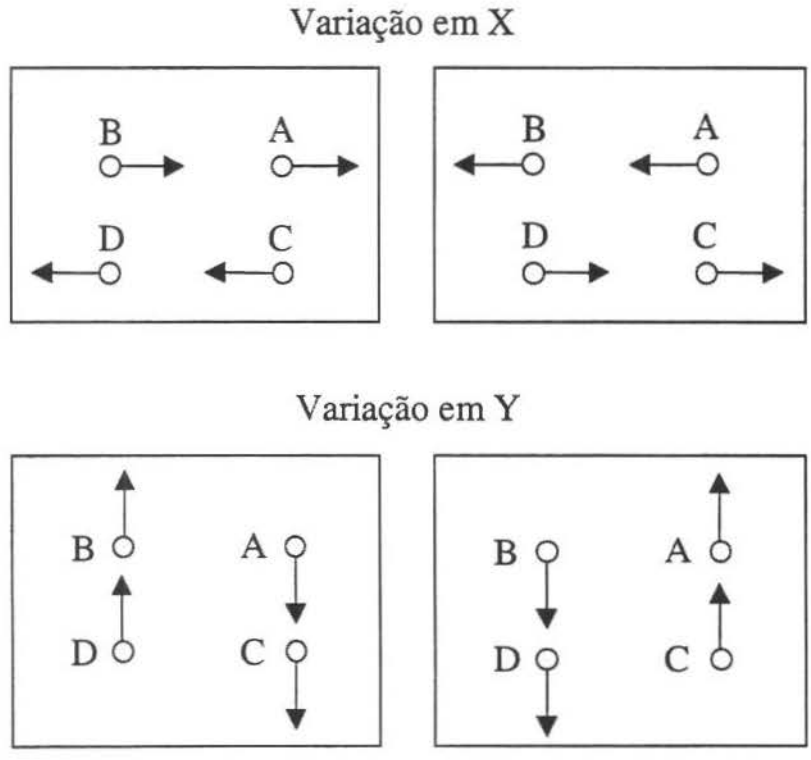

Figura 3.6 – Variações relativos que possivelmente provocaram a estrutura de correlação

3.2.2 *Aplicação dos Gráficos de Controle Multivariados*

Com os estimadores $\mathbf{S}_1$ e $\mathbf{S}_3$ calculados, construímos os gráficos multivariados estudados no item 2.2, o que é mostrado na Figura 3.7. Vemos que o gráfico com o estimador $\mathbf{S}_1$ acusa somente a observação 28 como causa assinalável e o gráfico feito a partir do estimador $\mathbf{S}_3$ vai além, assinalando as observações 10 a 14, 28 e 33. Além disso, poderíamos suspeitar das observações 18, 21 e 41 por estarem muito próximas ao limite superior de controle. O grupo de observações de 10 a 14 provavelmente estão assinaladas devido ao degrau observado nas variáveis marginais na direção Y. Veja, com isso, a importância de utilizar o estimador de diferenças sucessivas $\mathbf{S}_3$ nesse caso para captar bem esse sinal.

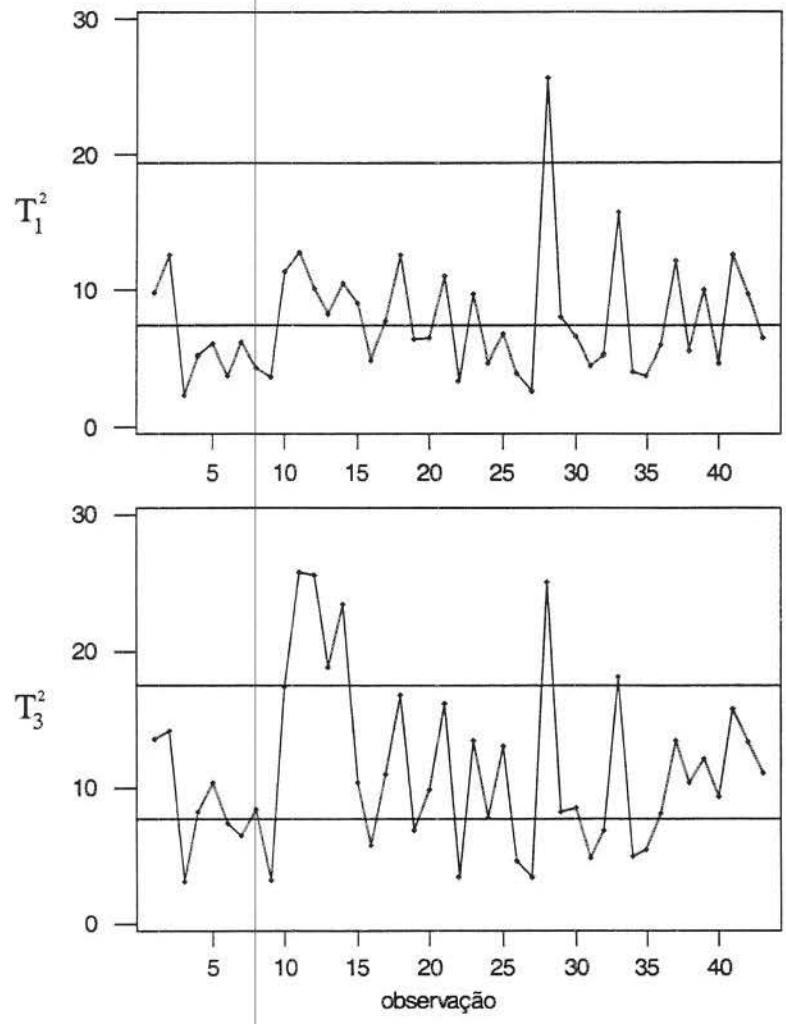

Figura 3.7 – Gráfico de controle multivariado construído a partir do estimadores $\mathbf{S}_1$ e $\mathbf{S}_3$

3.2.3 *Aplicação de Componentes Principais*

A Análise de Componentes Principais foi aplicada aos dados com as variáveis deslocadas para o zero e o resultado é resumido na Tabela 3.5. Vemos que a componente 1 carrega 87% da variação total dos dados e representa a variabilidade do conjunto de variáveis na direção X, por ser um contraste da média das variáveis de X à frente contra a média das da traseira do veículo. A componente 2, com 8% da variação dos dados, traz informação do conjunto de variáveis na direção Y, pois trata-se do contraste entre as médias das variáveis dessa direção da direita contra a esquerda do veículo. Esforços de entendimento do significado das outras componentes serão feitos para diagnóstico de causas especiais.

Tabela 3.5 – Resumo dos resultados da Análise de Componentes Principais com os dados

	$\hat{\lambda}_1$	$\hat{\lambda}_2$	$\hat{\lambda}_3$	$\hat{\lambda}_4$	$\hat{\lambda}_5$	$\hat{\lambda}_6$	$\hat{\lambda}_7$	$\hat{\lambda}_8$
Autovalor	9,5316	0,9096	0,3180	0,1422	0,0613	0,0115	0,0033	0,0017
proporção	0,8680	0,0830	0,0290	0,0130	0,0060	0,0010	0,0000	0,0000
				Coeficientes				
Variável	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
XFD	-0,569	0,231	0,578	0,433	-0,042	-0,313	0,026	0,036
XFE	-0,458	-0,102	-0,457	0,135	-0,084	0,046	-0,737	-0,018
XTD	0,513	-0,056	-0,205	0,667	-0,300	-0,390	-0,053	0,041
XTE	0,446	0,156	0,552	-0,096	0,098	0,122	-0,661	-0,021
YFD	0,043	0,435	-0,214	-0,152	0,342	-0,333	-0,059	0,713
YFE	-0,038	-0,450	0,194	0,180	-0,213	0,452	0,034	0,690
YTD	0,039	0,408	-0,149	0,509	0,410	0,602	0,099	-0,096
YTE	0,013	-0,592	0,061	0,161	0,749	-0,234	-0,033	-0,058

Gráficos de controle univariados para valores individuais para os escores das componentes (utilizando o estimador de amplitudes móveis para a variabilidade) são mostrados na Figura 3.8. Para cada gráfico o erro tipo I utilizado foi de 0,27% (correspondendo a 3 sigmas) porém, sabemos pela discussão em 1.2.6, que esse erro não é preservado. Portanto, aparecem como suspeitos as observações 10 a 14 (CP2), 18 e 21 (CP3), 28 (CP4), 28 e 41 (CP7). Os pontos 10 a 14 assinalados na componente 2 reafirmam a tese de que houve um degrau nas variáveis da direção Y nesse período.

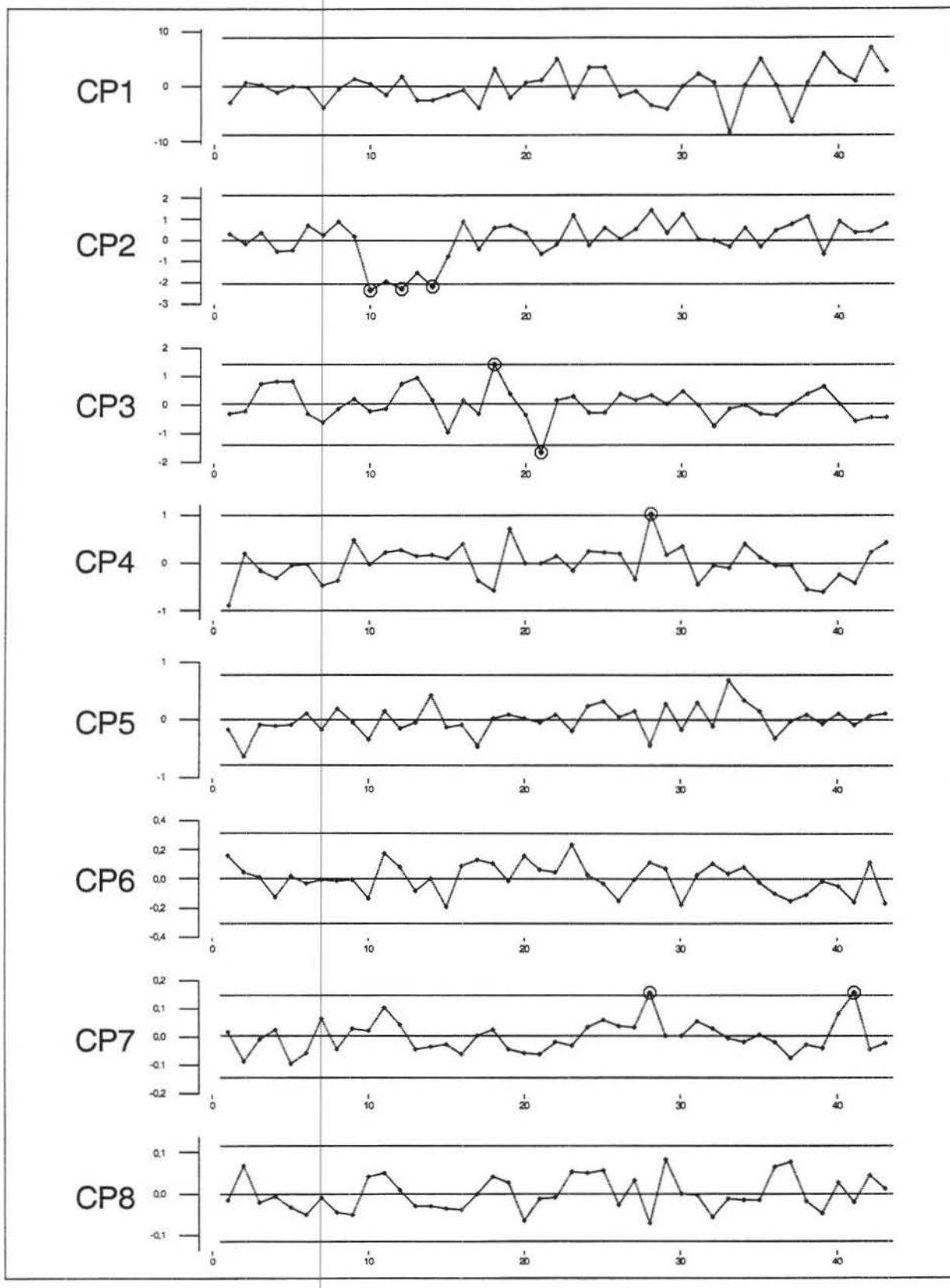

Figura 3.8 – Gráficos de controle univariados para os escores das Componentes Principais

A estatística D^2 foi calculada no intuito de captar pontos fora de controle nas primeiras componentes. O gráfico dos quantis ajustando-se uma distribuição Gama mostrado na Figura 3.9 aponta a observação 33 como suspeita, porém sem grandes evidências. Já a estatística U^2 calculada a partir da terceira componente, aponta nesse mesmo tipo de gráfico (Figura 3.10) a observação 28 como aberrante.

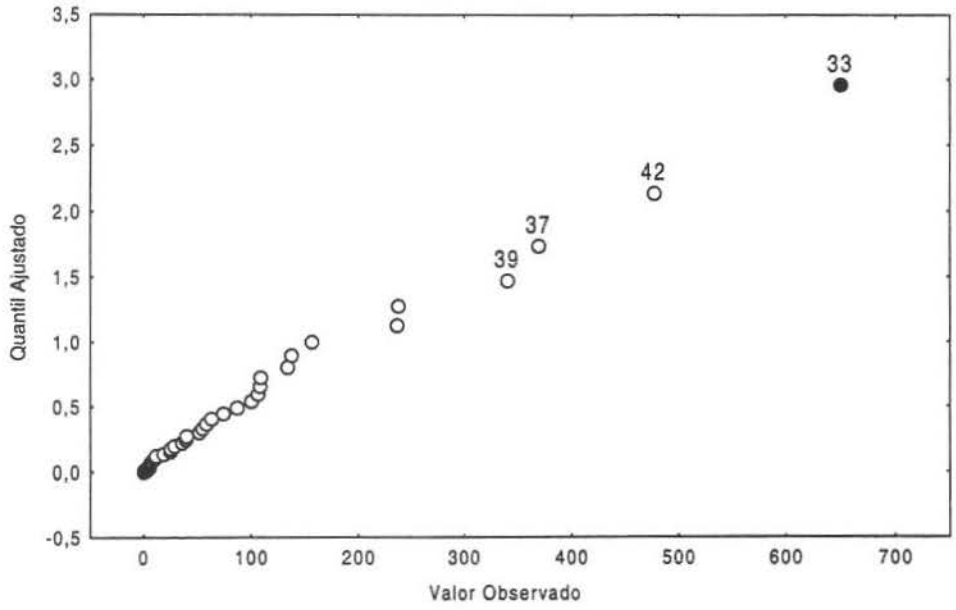

Figura 3.10 – Q-Q Plot (distribuição Gama) para a estatística D^2

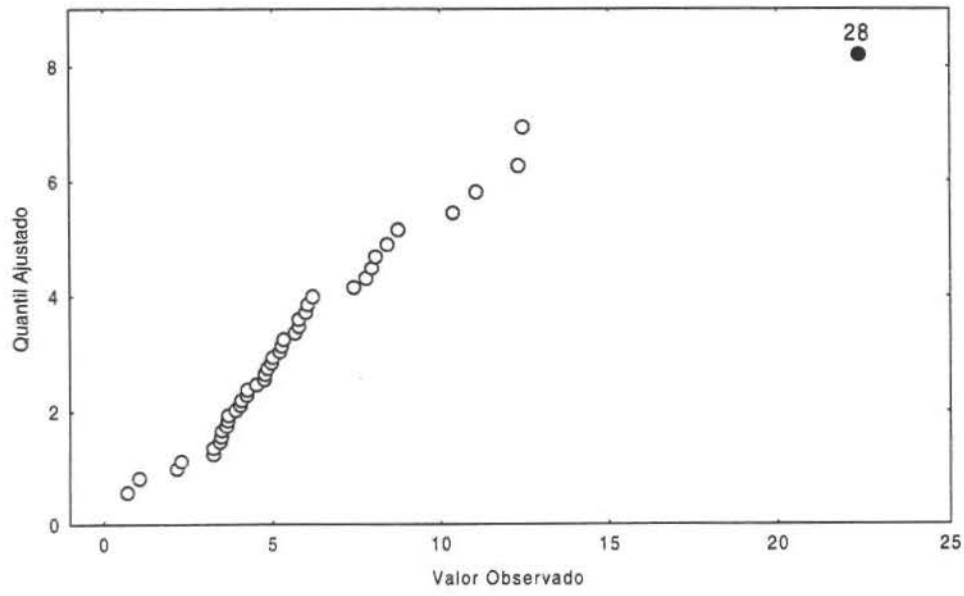

Figura 3.11 – Q-Q Plot (distribuição Gama) para a estatística U^2

Diante dessas informações, decidimos fazer uma incisão nos dados retirando apenas as observações 10 a 14, 28 e 33 numa primeira etapa de limpeza dos dados para em seguida procurar por sinais nos dados remanescentes.

3.2.4 Diagnóstico dos Sinais

Sobre as observações 10 a 14 e 33 já dissemos que o motivo básico porque foram assinaladas foi devido a um deslocamento de média a favor da estrutura de correlação. Porém, a observação 28 não parece ter essa explicação. Calculamos, assim, a estatística C_i^2 para as observação 28 e também para a 41, essa última como apoio para comparação de diagnóstico pois ambas pontuam alto em CP7.

O gráfico de barras da Figura 3.11 mostra os resultados. Note que para a observação 28, as variáveis que mais contribuem para a causa especial são XFE, XTE, YFD e YTD, portanto são distorções nas estruturas de correlação dessas variáveis em ambas as direções. Veja que a observação 28 pontua alto em duas das componentes principais, CP4 e CP7 e sem o auxílio da estatística C_i^2 ficaria difícil o seu diagnóstico.

Comparativamente temos a observação 41 que pontua alto somente em CP7 e as variáveis que provocam isso são XFE e XTE conforme os coeficientes para essa componente. Isso vai ao encontro do resultado para a estatística C_i^2, portanto nesse caso a análise da componente seria suficiente para o diagnóstico.

Infelizmente não há o histórico de fabricação para podermos associar esses diagnósticos aos eventos ocorridos nos períodos assinalados.

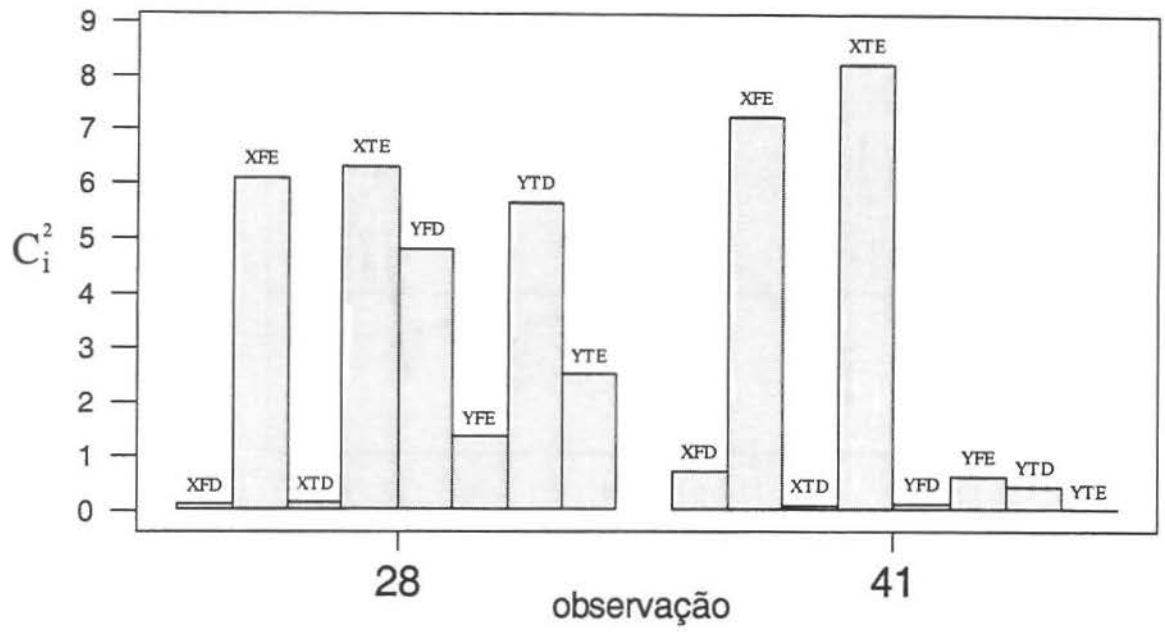

Figura 3.11 – Valores das contribuições marginais para as observações 28 e 41

3.2.5 Análise dos Dados Remanescentes

Realizamos a procura de causas assinaláveis com os dados remanescentes como nas análises anteriores e foi detectado ainda uma observação aberrante, a de número 37, veja o gráficos de T^2 para ambos os estimadores de sigma nas Figuras 3.12 e 3.13.

Após a retirada dessa observação, restando 35 observações, temos um conjunto limpo de dados para estabelecer as estimativas dos parâmetros e passar à fase II do controle estatístico do processo. Além do entendimento do modo de variação dos dados e de informações de possíveis causas assinaláveis que podem aparecer no processo.

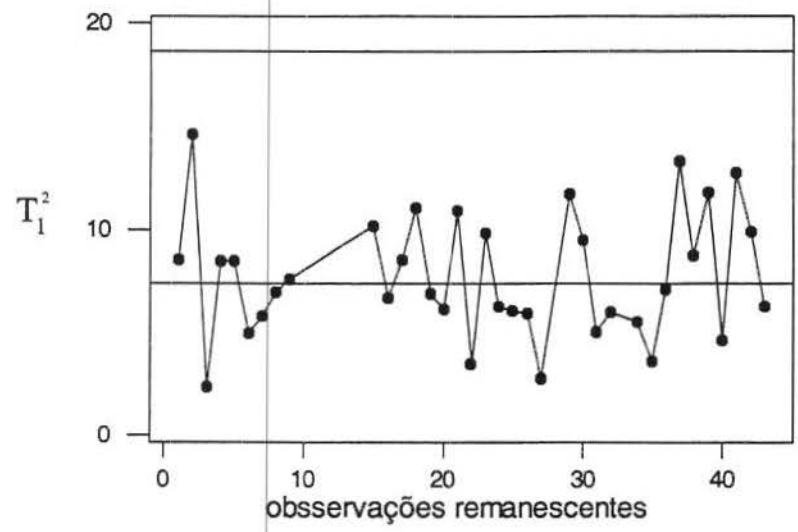

Figura 3.12 – Gráfico de T^2 construído a partir de $\mathbf{S}_1$ com as observações remanescentes

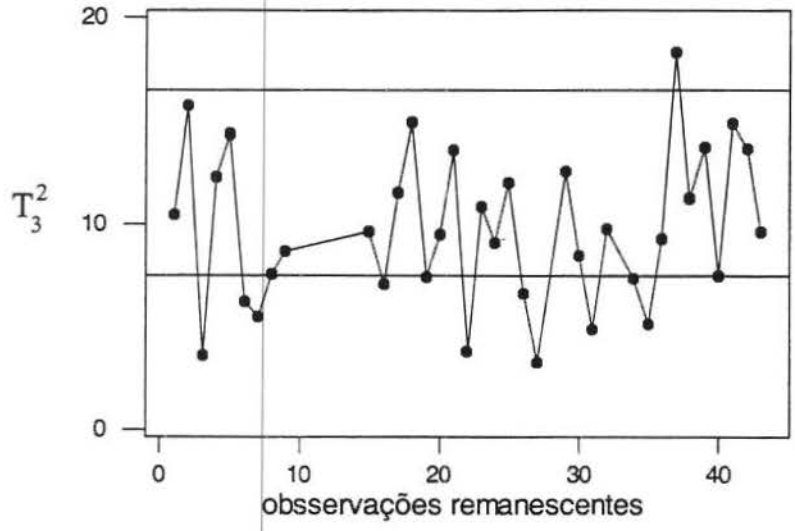

Figura 3.13 – Gráfico de T^2 construído a partir de $\mathbf{S}_3$ com as observações remanescentes

Considerações Finais

Discutimos, neste trabalho, os problemas relacionados com os gráficos de controle multivariados para observações individuais na fase I. Concluímos que o gráfico T^2 de Hotelling pode ser aplicado nesse caso, sendo uma ferramenta eficaz na detecção de causas especiais desde que usemos estimadores da matriz de covariância mais robustos, como aquele correspondente ao de amplitudes móveis do caso univariado visto no Capítulo 2. Aumenta-se esse poder de detecção se usarmos as informações dadas pelos escores das componentes principais cujas informações servirão, também, como base para o diagnóstico do porquê essas causas especiais ocorreram. Um complemento simples e útil para ajudar nesse diagnóstico é calcular e interpretar as contribuições das marginais ao sinal, como visto no item 2.5. Todos esses procedimentos mostraram-se bastante fáceis de serem implantados no MINITAB, não requerendo programação elaborada.

Trabalhamos essas técnicas compondo uma estratégia de investigação voltada para grupos de trabalhos com conhecimento introdutório de Estatística e para ser aproveitada nos esforços de melhoria. O Capítulo 3 mostrou como essa estratégia é útil para obter uma compreensão da variação de um processo ao ser aplicada num conjunto real de dados.

O assunto "Controle Multivariado da Qualidade" tem tido muito interesse na literatura devido ao aumento crescente da informatização dos processos. Esse interesse concentra-se, carecendo de bastantes estudos complementares, nas seguintes áreas:

i) Interpretação de causas especiais. Muitos pesquisadores, como dissemos no Capítulo 2, estão trabalhando para desenvolver métodos para isso. A maior desvantagem da maioria dos gráficos multivariados é que eles não trazem diretamente a informação que um operador necessita para atuar no processo.

ii) Observações auto-correlacionadas. É natural e freqüente o aparecimento de auto-correlação em processos, principalmente nas indústrias químicas. Nesse caso é devido, por exemplo, aos tanques de armazenagem intermediários, aos próprios volumes dos equipamentos e, aos reciclos que são muito comuns

nesses processos. A abordagem multivariada para controle estatístico, nesse caso, pode ser muito complicada e há poucas publicações na área.

iii) Gráficos de detecção rápida. Os gráficos equivalentes ao CUSUM e EWMA do caso univariado foram já todos desenvolvidos com várias abordagens para o caso multivariado. A maioria, porém, está fortemente ligada à suposição de multinormalidade. Há ainda algum trabalho nessa área para torná-los mais acessíveis em termos de implementação.

iv) Abordagens não-paramétricas. Alguns trabalhos têm aparecido sugerindo uma abordagem não-paramétrica do assunto. Liu (1995) trabalhou com o conceito de profundidade de dados propondo um método sem suposições sobre os dados que parece promissor.

v) Integração com o controle automático. Esforços de integração do controle estatístico e automático de processos têm sido foco de alguns pesquisadores. Box & Luceño (1997) dão uma boa introdução no assunto. A abordagem multivariada ainda está praticamente inexplorada.

vi) Grandes massas de dados. É uma área que necessita com urgência de ferramentas, já que muitos usuários têm disponível uma avançada tecnologia de coleta e armazenamento de dados, sem as devidas ferramentas para transformá-los em informação adequada para a ação. O controle multivariado de processos pode integrar-se a técnicas como, por exemplo, de redes neurais.

Consideramos este trabalho bem como essas áreas de desenvolvimento de grande potencial de aplicação do conhecimento acadêmico na prática, e incentivamos os esforços que venham a ser feitos para tornar as metodologias desenvolvidas acessíveis aos grupos em empresas e instituições que buscam excelência em seus produtos e serviços que é, sem dúvida, um passo importante para o desenvolvimento do país.

Bibliografia

[1] Alt, F. B. (1985), "Multivariate quality control", in: S. Kotz e N. L. Johnson (eds.), *Encyclopedia of statistical sciences*, Vol. 6, Wiley, NY, 110-122.

[2] Anderson, T. W. (1959), "An introduction to multivariate statistical analysis", John Wiley & Sons, NY

[3] API (1998), "The Improvement Handbook – Model, Methods, and Tools for Improvement", Associates in Process Improvement, Austin, Tx

[4] Atkinson, A. C. and Mulira, H. M. (1993), "The stalactite plot for the detection of multivariate outliers", *Statistics and Computing* 3, pp. 27-35.

[5] Box, G. and Luceño, A. (1997), "Statistical control by monitoring and feedback adjustment", John Wiley & Sons, NY

[6] Deming, W. E. (1990), "Qualidade: a revolução da administração", Marques Saraiva, RJ

[7] Fuchs, C. and Benjamini, Y. (1994), "Multivariate profile charts for statistical process control", *Technometrics* 36 (2), pp. 182-195.

[8] Gnanadesikan, R. and Kettenring, J. R. (1972), "Robust estimates, residuals, and outlier detection with multiresponse data", *Biometrics* 28, pp. 81-124.

[9] Jackson, J. E. (1980), "Principal components and factor analysis: part I – principal components", *Journal of Quality Tecnology*, 12 (4), 201-213

[10] Jackson, J. E. (1985). "Multivariate Quality Control". *Communications in Statistcs – Theory and Methods* 14 (11), pp. 2657-2688.

[11] Jackson, J. E. and Mudholkar, G. S. (1979), "Control procedures for residuals associated with principal component analysis", *Tecnometrics* 21 (3), pp. 341-349.

[12] Johnson, R. A. and Wichern, D. W. (1998), "Applied multivariate statistical analysis", Prentice Hall, Upper Saddle River, New Jersey.

[13] Kourti, T. and MacGregor, J. F. (1996), "Multivariate SPC methods for process and product monitoring", *Journal of Quality Technology* 28 (4), pp. 409-428.

[14] Langley, G. J. et alli (1996), "The improvement guide: a practical approach to enhancing organizational performance", Jossey-Bass Publishers, San Francisco.

[15] Liu, R. Y. (1995), "Control charts for multivariate processes", *Journal of The American Statistical Association* 90, pp. 1380-1387.

[16] Lowry, C. A. and Montgomery (1995), "A review of multivariate control charts", *IIE Transactions* 27, pp. 800-810.

[17] Marriot, F. H. C. (1974), "The Interpretation of Multiple Observations", Academic Press, London.

[18] Mason, R. L., Champ, C. W., Tracy, J. C. (1995), "Assessment of multivariate processo control techniques", *Journal of Quality Technology* 29 (2), pp. 140-143.

[19] Montgomery, D. C. (1997), "Introduction to statistical quality control", 3nd ed. John Wiley & Sons, New York, NY

[20] Murphy, B. J. (1987), "Selecting out of control variables with the T^2 multivariate quality control procedures", *The statistician*, 36, pp. 571-583.

[21] Runger, G. C. (1996), "Projections and the U^2 multivariate control chart", *Journal of Quality Technology* 28 (3), pp. 313-319.

[22] Runger, G. C., Alt, F. B. and Montgomery, D. C. (1996), "Contributors to a multivariate statistical process control chart signal", *Communications in Statistcs – Theory and Methods* 25 (10), pp. 2203-2213.

[23] Tracy, N. D., Young, J. C. and Mason R. L. (1995a). "Decomposition of T^2 for Multivariate Control Chart Interpretation", *Journal of Quality Technology* 27 (2), pp. 99-108.

[24] Tracy, N. D.; Young, J. C. and Mason R. L. (1995b), "A Bivariate Control Chart for Paired Measurements", *Journal of Quality Technology* 27 (4), pp. 370-376.

[25] Tracy, N. D.; Young, J. C. and Mason R. L. (1997), "A practical approach for interpreting multivariate T^2 control charts signals", *Journal of Quality Technology* 29 (4), pp. 396-406.

[26] Tracy, N. D.; Young, J. C.; e Mason, R. L. (1992), "Multivariate Control Charts for Individual Observations", *Journal of Quality Tecnology*, 24, 88-95

[27] Wierda, S. J. (1994), "Multivariate statistica process control – recent results and directions for future research". *Statistica Neerlandica*, 48, 147-168

[28] Wilks, S. S. (1963), "Matematical statistics" John Wiley & Sons, Wiley, NY

[29] Woodall, W. H. and Sullivan, J. H. (1996), "A comparison of multivariate control charts for individual observation", *Journal of Quality Technology* 28 (4), pp. 398-408.