UNIVERSIDADE FEDERAL DE SANTA CATARINA

PROGRAMA DE PÓS-GRADUAÇÃO EM

ENGENHARIA DE PRODUÇÃO **e** SISTEMAS

DECOMPOSIÇÃO DA ESTATÍSTICA DO GRÁFICO DE CONTROLE MULTIVARIADO T^2 DE HOTELLING POR MEIO DE UM ALGORITMO COMPUTACIONAL

ANDRÉA CRISTINA KONRATH

Orientador: Robert Wayne Samohyl, *Ph.D.*

Florianópolis, dezembro de 2002.

ANDRÉA CRISTINA KONRATH

DECOMPOSIÇÃO DA ESTATÍSTICA DO GRÁFICO DE CONTROLE MULTIVARIADO T^2 DE HOTELLING POR MEIO DE UM ALGORITMO COMPUTACIONAL

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia de Produção da Universidade Federal de Santa Catarina como requisito parcial para a obtenção do grau de Mestre em Engenharia de Produção.

Orientador: Prof. Robert Wayne Samohyl, *Ph.D.*

Florianópolis, dezembro de 2002.

Andréa Cristina Konrath

DECOMPOSIÇÃO DA ESTATÍSTICA DO GRÁFICO DE CONTROLE MULTIVARIADO T^2 DE HOTELLING POR MEIO DE UM ALGORITMO COMPUTACIONAL

Esta dissertação foi julgada e aprovada para a obtenção do título de Mestre em Engenharia de Produção no Programa de Pós-Graduação em Engenharia de Produção da Universidade Federal de Santa Catarina.

Florianópolis, 13 de dezembro de 2002.

Prof. Edson Pacheco Paladini, Dr.

Coordenador do Programa de Pós-Graduação em Engenharia de Produção

Banca Examinadora

Prof. Robert Wayne Samohyl, *Ph.D.*
Universidade Federal de Santa Catarina.
Orientador

Prof. João Neiva de Figueiredo, *Ph.D.*
Universidade Federal de Santa Catarina.

Prof. Marcelo Menezes Reis, *Dr.*
Universidade Federal de Santa Catarina.

DEDICATÓRIA

Dedico este trabalho ao meu pai, a minha irmã e em especial a minha mãe (*in memorian*).

AGRADECIMENTOS

Para desenvolver este trabalho foi necessário muito empenho, bem como a ajuda e compreensão de inúmeras pessoas que, de forma direta e indireta, participaram de todo o seu processo de elaboração. Assim, manifesto meus sinceros agradecimentos:

ao grande *arquiteto do universo* que me deu forças físicas e emocionais garantindo a execução deste trabalho;

ao professor Robert Wayne Samohyl, alguém que antes de ser um orientador, foi um grande amigo. Saiba que seus incentivos foram fundamentais para o prosseguimento de meu trabalho nos momentos de desânimo;

aos membros da banca examinadora pelas valiosas sugestões apresentadas para o engrandecimento deste trabalho;

à Universidade Federal de Santa Catarina, especialmente ao Departamento de Engenharia de Produção e Sistemas, pela oportunidade de realização do mestrado;

A Éder Daniel Corvalão, pelo auxílio na parte computacional, sugestões e constante colaboração.

A Luciana e Wesley Vieira da Silva, pela força, amizade e companheirismo indispensável.

Muitas são as dificuldades surgidas no decorrer da elaboração de um trabalho de dissertação. Felizmente, existem as pessoas que estão sempre prontas a nos auxiliar, a nos encorajar nos momentos mais difíceis. Gostaria de registrar meus sinceros agradecimentos a estas pessoas que sempre me apoiaram, em particular a Ivanete Masson e Rubson Rocha, Silvia de Almeida e Edson Leal Soares Ramos, Cássia Regina de Lima, Viviane Leite Dias Mattos, Neiva Teresinha Badin, Rogério Orlandeli, Patrícia Tavares, Florence de Araújo, Luciano Antonini, Adauto Scalon.

A todos que colaboraram, direta ou indiretamente, para a elaboração desse trabalho: o meu muito obrigado!

Finalmente, não poderia deixar de agradecer a existência das pessoas mais importantes de minha vida: minha mãe (*in memorian*), meu pai Darcilo e minha irmã Andiara, bem como aos meus tios Lourdes e Mario Tito de Azevedo e Zeno Rech.

SUMÁRIO

LISTA DE FIGURAS ..... *iii*
LISTA DE TABELAS ..... *iv*

1. APRESENTAÇÃO DO TRABALHO **14**

1.1. INTRODUÇÃO 14
1.2. JUSTIFICATIVA E IMPORTÂNCIA 17
1.3. OBJETIVO DO TRABALHO 18
1.4. A METODOLOGIA DE PESQUISA DO TRABALHO 19
1.4.1. TIPO DE ESTUDO 19
1.4.2. MÉTODO DE COLETA E ANÁLISE DE DADOS 19
1.5. ESTRUTURA DO TRABALHO 20
1.6. LIMITAÇÕES DO TRABALHO 20

2. REVISÃO BIBLIOGRÁFICA **22**

2.1. CONTROLE ESTATÍSTICO DE PROCESSOS 22
2.1.1. FUNDAMENTOS DOS GRÁFICOS DE CONTROLE TRADICIONAIS 23
2.1.2. TIPOS DE ERROS EM GRÁFICOS DE CONTROLE 28
2.1.3. IDENTIFICAÇÃO DAS VARIÁVEIS FORA DOS LIMITES DE CONTROLE, POR MEIO DO MÉTODO DE GRÁFICOS UNIVARIADOS 30
2.2. GRÁFICOS DE CONTROLE MULTIVARIADOS 32
2.2.1. GRÁFICO DE CONTROLE QUI-QUADRADO- χ^2 33
2.2.2. GRÁFICOS DE HOTELLING OU GRÁFICOS T^2 36
2.3. CUSTOS DA QUALIDADE 45
2.3.1. CUSTO DE PREVENÇÃO 46
2.3.2. CUSTO DE AVALIAÇÃO 47
2.3.3. CUSTOS DE FALHAS 48
2.3.4. A IMPLICAÇÃO DOS GRÁFICOS DE CONTROLE NOS CUSTOS DA QUALIDADE 49
2.4. CONSIDERAÇÕES FINAIS 49

3. O MÉTODO PROPOSTO **51**

3.1. LINGUAGEM DE PROGRAMAÇÃO 51
3.2. PROCEDIMENTO PARA IDENTIFICAR VARIÁVEIS FORA DOS LIMITES DE CONTROLE EM GRÁFICOS MULTIVARIADOS 52
3.3. IDENTIFICAÇÃO DAS VARIÁVEIS FORA DOS LIMITES DE CONTROLE, ATRAVÉS DOS MÉTODOS DE DECOMPOSIÇÃO DA ESTATÍSTICA T^2 DE HOTELLING 53
3.3.1. O MÉTODO DE DECOMPOSIÇÃO DA ESTATÍSTICA T^2 DE HOTELLING POR MASON ET AL 53
3.3.2. OUTROS MÉTODOS DE DECOMPOSIÇÃO 56
3.4. PASSOS PARA OPERACIONALIZAR O GRÁFICO DE CONTROLE T^2 DE HOTELLING 60
3.4.1. APRESENTAÇÃO DO APLICATIVO EM FLUXOGRAMA 63
3.4.2. DESCRIÇÃO DO FLUXOGRAMA 64
3.4.3. COMO IDENTIFICAR AS VARIÁVEIS QUE ESTÃO INTERFERINDO NA ESTABILIDADE DO PROCESSO 66
3.4.4. APRESENTAÇÃO DO SISTEMA 71
3.5. CONSIDERAÇÕES FINAIS 81

4. CONCLUSÕES **82**

4.1. RECOMENDAÇÕES 83

REFERÊNCIA BIBLIOGRÁFICA 84

ANEXOS 88

ANEXO 1 88

ANEXO 2 89

ANEXO 3 90

LISTA DE FIGURAS

Figura 1- Gráfico de Controle de Shewhart 24
Figura 2: Gráfico de Controle *T^2 de Hotelling* 40
Figura 3: Fluxograma 63
Figura 4: Tela Inicial do Programa 71
Figura 5: Tela para Iniciar a Operacionalização 72
Figura 6: Tela para Digitar os Dados 73
Figura 7: Tela para Processar os Dados 74
Figura 8:Tela para Iniciar a Operacionalização 75
Figura 9: Tela com os Resultados 76
Figura 10: Tela de Apresentação da Decomposição 77
Figura 11: Tela para Acrescentar as Novas Observações no Processo 79
Figura 12: Tela de Apresentação da Decomposição 80
Figura 13: Gráfico de Shewhart para a variável 1 89
Figura 14: Gráfico de Shewhart para a variável 2 90

LISTA DE TABELAS

Tabela 1: Médias, Variâncias e Covariâncias da Força de Tensão e Diâmetro, respectivamente, de uma Fibra Têxtil, (Montgomery, 1996)..... 39

LISTA DE ABREVIATURAS

ARL	*Average Run Length* - Comprimento médio da corrida
CP	Componente Principal
CEP	Controle Estatístico de Processo
CUSUM	Gráfico de Controle para Soma Acumulada
MCUSUM	Gráfico de Controle Multivariado para Soma Acumulada
MEWMA	Gráfico de Controle Multivariado para Média Móvel Ponderada Exponencialmente
GCM	Gráfico de Controle Multivariado
LC	Linha Central
LIC	Limite Inferior de Controle
LSC	Limite Superior de Controle
PDCA	Plan- Do- Check –Action
VBA	Visual Basic for Applications

RESUMO

Este trabalho tem como objetivo apresentar a decomposição da estatística T^2 de Hotelling, através de um algoritmo computacional, para auxiliar na detecção das variáveis significativas no momento em que as características de qualidade do processo monitoradas saem fora dos limites de controle em processos multivariados. O método de decomposição utilizado é de Mason, Tracy &Young (1995), que sugere que a estatística T^2 seja dividida em componentes independentes, cada um refletindo a contribuição de uma variável individual. Os resultados apresentam-se satisfatórios ao objetivo proposto, e a descrição do referido método é apresentada passo a passo, sendo o algoritmo computacional testado com um banco de dados retirado da literatura (Mitra, 1993), bem como a geração de dados aleatoriamente.

ABSTRACT

The objective of this dissertation is to present an application of Hotelling's T^2 statistical decomposition in multivariate process. This was achieved through the development of an algorithm and computer program to help the detection of statistically significant variables at the moment in which the quality characteristics which are being monitored extrapolate control limits. The Mason, Tracy and Young (1995) decomposition method is used, suggesting that the T^2 statistic is divisible into independent components, each one of which reflects the contribution of an individual variable. The method is described in detail and tests of the algorithm using a database from the literature (Mitra, 1993), as well as with data randomly generated are presented. The results were positive.

1. Apresentação do Trabalho

O desenvolvimento deste capítulo tem como finalidade apresentar a problemática do tema abordado, as justificativas, os objetivos a serem alcançados e suas limitações.

1.1. Introdução

A qualidade atualmente desempenha um papel muito importante no mundo industrial e empresarial. Ela tornou-se, nos últimos anos, um tema central nas estratégias de negócios no cenário empresarial cada vez mais competitivo, uma vez que leva a um aumento da produtividade. Além disso, a qualidade passou a ser um fator de decisão básico na escolha de produtos e serviços pelo consumidor, onde produtos podem ser simples parafusos, sofisticados sistemas eletrônicos ou chips complexos; os serviços podem ser, por exemplo, o fornecimento de energia elétrica ou serviços bancários e, por fim, o consumidor pode ser um indivíduo ou uma empresa. Conseqüentemente, a qualidade é um elemento chave no sucesso, no crescimento e na liderança de uma empresa em seus negócios.

De forma suscinta, pode-se afirmar que três fatores são determinantes na busca da competitividade: qualidade, produtividade e custos de operação. Segundo Montgomery (1996) qualidade está relacionada a uma ou mais características que um produto ou serviço deveria apresentar. Garvin (1987), classifica estas características da seguinte forma:

- Desempenho: o produto ou serviço realmente corresponde ao que dele se espera;
- Confiabilidade: com que freqüência o produto ou serviço deixa a desejar;
- Durabilidade: por quanto tempo o produto ou serviço irá durar;
- Manutenabilidade: facilidade para reparar o produto ou serviço;
- Estética: qual a aparência do produto ou serviço;
- Características: o que o produto ou serviço faz, o que ele apresenta como diferencial;

- Qualidade percebida: qual é a reputação da companhia responsável pelo produto ou serviço;
- Conformidade com padrões: o produto ou serviço é exatamente igual ao que foi projetado.

O aperfeiçoamento dinâmico da qualidade torna-se um instrumento para o aumento da produtividade e redução de custos. Neste contexto de melhoria contínua da qualidade surgem alguns fatores que auxiliam na formulação de estratégias de negócio, tais como, (Montgomery 1996):

- O aumento da consciência do consumidor para qualidade e forte orientação para alto desempenho;
- Responsabilidade pelo produto;
- Aumento do custo de trabalho, energia e matéria-prima;
- Competição intensa das empresas;
- Melhoria na produtividade por meio de programas efetivos de engenharia de qualidade.

Segundo Montgomery (1996), houve também a necessidade da avaliação dos custos de qualidade, devido aos seguintes fatores:

- Aumento no custo da qualidade em função do aumento na complexidade dos produtos manufaturados associados a avanços tecnológicos;
- Aumento da consciência dos custos de ciclo de vida, incluindo manutenção, trabalho, partes substituídas e custos de falhas em campo;
- A necessidade de engenheiros e gerentes de qualidade em efetivamente comunicar o custo da qualidade numa linguagem gerencial.

Os custos da qualidade são categorias de custos gerados por meio da atividade de produzir, identificar, evitar ou reparar produtos que não estão dentro dos padrões. Essas categorias de custo estão divididas em: Custos de Prevenção; Custos de Avaliação; Custos de Falhas Internas; Custos de Falhas Externas.

Na gestão da qualidade, o papel crítico dos fornecedores não deve ser esquecido, pois a seleção de fornecedores também deve ser baseada em qualidade, programação de entrega do produto e os custos da qualidade.

A melhoria contínua na qualidade de produtos, processos e serviços, vem sendo auxiliada por técnicas estatísticas, usadas nas indústrias. A técnica de controle estatístico de processo (CEP) proporciona melhores resultados, já que ela faz uso de métodos estatísticos em todos os estágios de desenvolvimento de um projeto e/ou da produção.

A globalização aumenta a competitividade entre todos os setores produtivos e, como conseqüência direta, houve uma rápida evolução tecnológica, na qual a eficácia e custos baixos são exigidos, fazendo com que o CEP, cada vez mais, desempenhe um papel de destaque.

Nas indústrias, o CEP é utilizado para monitorar muitas variáveis que controlam o desempenho do processo, bem como para verificar as múltiplas características de qualidade do produto. O monitoramento é realizado, principalmente, através do uso de Gráficos de Controle Univariados. O número de informações geradas é grande, o que torna difícil sua administração. Para os casos onde é preciso controlar vários fatores do processo e várias características do produto, fez-se necessário o desenvolvimento de gráficos que combinem todas essas características de uma única vez, denominados de gráficos de controle multivariados.

1.2. Justificativa e Importância

A competitividade acirrada induziu as empresas a investirem em parques industriais modernos visando melhoria nos processos e, em função destes investimentos ocorreram mudanças e seus processos passaram por uma avaliação mais específica e criteriosa. O resultado dessas avaliações fez com que a maioria das indústrias passasse a adquirir modernos equipamentos aumentando a complexidade dos processos produtivos e conseqüentemente o controle da qualidade.

A utilização do Controle da Qualidade passou a ser mais constante nas empresas fazendo com que as mesmas investissem em ferramentas de qualidade para que seus processos gerassem resultados mais precisos e mais rápidos. Destacam-se dentre as ferramentas, os gráficos de controle que são de extrema importância para o monitoramento dos processos, bem como para detecção de possíveis problemas dos mesmos.

Os gráficos de controle mais freqüentemente utilizados pelas empresas são os gráficos de controle univariados, pois sempre tiveram uma ampla divulgação na literatura e sua utilização sempre foi mais acessível para os operadores. Já os gráficos multivariados nunca tiveram uma ampla divulgação na literatura, além de serem pouco utilizados nos processos industriais. Os motivos para a pouca utilização são justificados pelos pacotes computacionais existentes que apresentam recursos insuficientes quanto à identificação dos problemas nos processos, além da falta de treinamento dos operadores, que normalmente não são habilitados para identificação das possíveis variáveis de maior relevância no processo que estejam apresentando uma causa especial.

Dentre os gráficos multivariados existentes, o gráfico de controle multivariado T^2 de Hotelling é o mais conhecido na literatura, e sua aplicabilidade é mais recomendada para processos que apresentam várias características de qualidade.Tais características são fortemente correlacionadas e precisam ser monitoradas conjuntamente. Além disso, o gráfico de controle T^2 de Hotelling apresenta mais sensibilidade do que os gráficos univariados, permitindo ao operador detectar mais rapidamente os possíveis problemas existentes no processo e com isso corrigi-los com mais agilidade.

Assim, este estudo torna-se importante, à medida que se propõe a elaborar um algoritmo computacional para a decomposição da estatística T^2 de Hotelling, o que facilita o uso do gráfico multivariado por parte dos operadores e estudiosos no assunto no que diz respeito à detecção de possíveis causas assinaláveis, bem como na operacionalização do manuseio dos resultados.

1.3. Objetivo do Trabalho

Abordar o estudo do Gráfico de Controle Multivariado T^2 Hotelling, decompondo a estatística T^2 por meio de um algoritmo computacional, baseado na abordagem proposta por Mason, Tracy, Young (1995).

1.3.1 Objetivo Principal

Apresentar a decomposição da estatística T^2 através de um algoritmo computacional, que auxilie a detecção das variáveis significativas, podendo assim obter um maior controle sobre os processos multivariados.

1.3.2 Objetivos Específicos

- Desenvolver a teoria detalhada sobre o gráfico de controle multivariado T^2 de Hotelling, bem como os métodos sugeridos por vários autores;
- Revisar a bibliografia existente sobre o gráfico de controle multivariado, mais especificamente o gráfico de controle multivariado T^2 Hotelling;
- Desenvolver o algoritmo computacional que apresente a decomposição da estatística T^2 de Hotelling, proposta por Mason, Tracy, Young (1995);
- Apresentar o algoritmo computacional.

1.4. A Metodologia de Pesquisa do Trabalho

1.4.1. Tipo de Estudo

O presente estudo foi baseado numa pesquisa aplicada quantitativa, que segundo Menezes e Silva (2001), tem como objetivo gerar conhecimentos para aplicação prática dirigida à solução de problemas específicos com o uso de recursos e técnicas estatísticas.

1.4.2. Método de Coleta e Análise de Dados

Quanto ao método de coleta dos dados será utilizado um banco de dados disponível na literatura sobre o assunto abordado, bem como serão gerados dados aleatórios para trabalhar com mais variáveis (características da qualidade). Assim, os dados serão analisados (testados) pelo algoritmo computacional desenvolvido para o propósito deste estudo.

Vale salientar, que a utilização do gráfico de controle multivariado T^2 de Hotelling compõe uma das etapas envolvidas no método de gerenciamento de melhorias, comumente conhecido como ciclo do PDCA, cuja denominação é: Planejar (Plan), Executar (Do), Verificar (Check) e Agir (Act), (Britz, Emerling, Hare, Hoerl, Janis, Shade, 1999).

A etapa de planejar consiste definir o que se quer planejar, o que será feito, estabelecer metas, definir os métodos e atingir as metas propostas. A fase de execução consiste em tomar iniciativa, educar, treinar, implementar, executar o planejado conforme as metas e métodos definidos. A fase de verificar consiste em verificar os resultados que se obteve, acompanhar continuamente os trabalhos e ao mesmo tempo analisar se o resultado correspondeu ao que foi planejado. Por fim, a etapa de agir é fazer correções de rotinas se for necessário, ou seja, ações corretivas ou de melhoria, caso tenha sido constatada na fase anterior.

1.5. Estrutura do trabalho

O trabalho proposto está estruturado em 4 capítulos, os quais estão descritos a seguir:

no capítulo 1, apresenta-se o tema abordado, as justificativas para a escolha do mesmo, os objetivos a serem alcançados e as limitações destes;

no capítulo 2, desenvolve-se um embasamento teórico sobre os assuntos pertinentes ao tema principal da dissertação. Através desta fundamentação teórica procura-se apresentar de forma clara e detalhada as pesquisas realizadas até o momento na área de controle estatístico de processo multivariado;

no capítulo 3, apresenta-se a Linguagem de Programação a ser utilizada para o desenvolvimento do algoritmo computacional, a decomposição estatística T^2 de Hotelling por meio do método proposto por Mason, Tracy, Young (1995). Em seguida apresentam-se as etapas do desenvolvimento do algoritmo e os resultados;

no capítulo 4, são apresentadas as conclusões e as proposições de possíveis extensões a este trabalho.

1.6. Limitações do Trabalho

As limitações encontradas ao longo do trabalho são apresentadas a seguir:

- O banco de dados trabalhado no algoritmo computacional não foi um banco de dados real;
- A distribuição da probabilidade dos dados gerados foi a normal;
- A escassez de material bibliográfico referente ao processo de decomposição da estatística T^2 de Hotelling impediu a utilização de outros métodos existentes no presente trabalho.

- A ausência de um estudo de caso não proporcionou análises complementares que pudessem aprimorar a utilização do algoritmo computacional desenvolvido.

2. Revisão Bibliográfica

Neste capítulo, é feita uma breve introdução sobre o Controle Estatístico de Processo. Apresenta-se uma revisão dos gráficos de controle tradicionais e dos gráficos de controle multivariados, mais especificamente o gráfico de controle de T^2 Hotelling, bem como os tipos de erros em gráficos de controle. São apresentados também os principais custos da qualidade e suas categorias.

2.1. Controle Estatístico de Processos

A maioria das indústrias tem como o objetivo comum, a busca da manutenção e melhoria contínua da qualidade dos seus produtos no atual mercado competitivo. Além disso, qualidade passou a ser um fator de decisão básica na escolha de produtos e serviços pelo consumidor.

Conseqüentemente, a qualidade é o elemento chave do sucesso, crescimento e liderança de uma empresa ou negócio. De uma forma sucinta pode-se afirmar que três fatores são determinantes na competitividade: qualidade, produtividade e custos de operação, influenciando no custo final do produto e em seu preço de venda. Desta forma, para que haja uma melhoria na qualidade de produtos, processos e serviços, as técnicas de controle estatístico de processo têm sido amplamente usadas.

O controle estatístico de processo (*CEP*) trata do uso de métodos estatísticos em todos os estágios considerados críticos ou de muita importância no desenvolvimento de um projeto e/ou da produção. Segundo Montgomery (1996) o principal objetivo do *CEP* é monitorar o desempenho de processos ao longo do tempo e detectar eventos incomuns que influenciem nas propriedades determinantes da qualidade do produto final. Uma vez encontradas as causas especiais responsáveis pelo evento incomum, melhorias no processo e na qualidade dos produtos, podem ser sugeridas. Na implementação destas melhorias utiliza-se o gráfico de controle, introduzido por

Walter A. Shewhart (1924), um físico dos Laboratórios da Bell Telephone, na década de 20, com o objetivo de controlar a variabilidade dos processos.

Shewhart classificou a variabilidade como sendo provocada por causas comuns (pequenas variações aleatórias inerentes ao processo) e por causas especiais ou assinaláveis (presença de eventos anormais no processo que prejudicam a qualidade do produto). De acordo com Hradesky (1989), a principal vantagem proporcionada pelo *CEP* é a realimentação de informações sobre o próprio processo em tempo real. Tal realimentação deve ser contínua e estatisticamente fundamentada. Assim, nas subseções a seguir estabelece-se uma revisão de *CEP* e dos gráficos de controle tradicionais e do gráfico de controle multivariados de Hotelling.

2.1.1. Fundamentos dos Gráficos de Controle Tradicionais

Os gráficos de controle são registros dos dados mensurados em um ponto crítico do processo, construídos num sistema de coordenadas cartesianas. Assim, o eixo das ordenadas é representado pelas mensurações feitas de uma determinada característica do produto e o eixo das abscissas representado pelos subgrupos. A figura 1 apresenta o gráfico de controle de *Shewhart,* o qual apresenta três linhas paralelas ao eixo da abscissa, denominadas de linha central (*LC)* que é comumente definida a partir do valor médio, o limite superior de controle (*LSC*) e o limite inferior de controle (*LIC*). No processo de construção desses limites comumente são usados em geral $\pm$ 3 desvios padrões, que pode ser justificado pelos bons resultados obtidos na prática, com o nível de confiança estabelecido de 99,73% na análise dos dados.

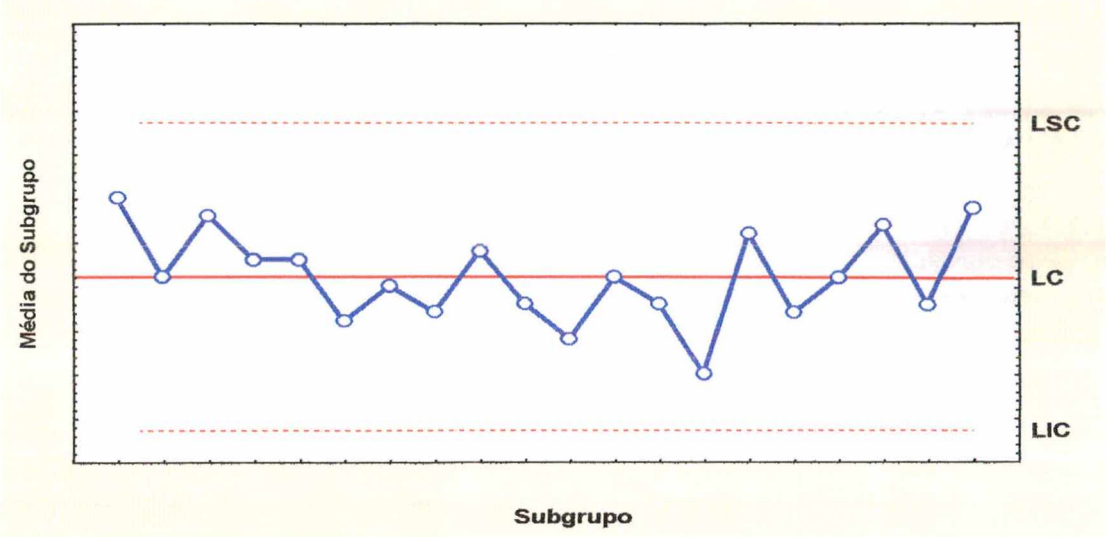


Figura 1- Gráfico de Controle de Shewhart

Os gráficos de controle possibilitam análise das amostras coletadas numa etapa do processo como observações individuais ou com a formação de subgrupos.

Conforme Siqueira (1997), um subgrupo pode ser definido como uma divisão racional da amostra coletada, de modo a facilitar a identificação de uma possível causa especial que seja assinalada no gráfico de controle. Porém, o tamanho dessa divisão depende do tipo de gráfico que estiver sendo usado, volume de produção, custo de inspeção e importância da informação obtida, ou periodicidade da coleta da amostra.

Os gráficos de controle devem ser implementados para monitorar as variáveis que influenciam a qualidade dos itens produzidos. Dessa forma, uma determinada variável de qualidade ou característica de qualidade é monitorada por meio de sucessivas amostras espaçadas no tempo.

As medições obtidas em cada amostra são representadas em ordem cronológica num gráfico e comparadas com os limites de controle obtidos da amostra. Caso alguma medição ultrapasse os limites de controle, o referido processo é considerado fora dos limites de controle estatístico. Se uma seqüência de medições apresentar um comportamento sistemático não aleatório, mesmo que nenhuma medição ultrapasse os limites de controle do processo, o referido processo também poderá

ser considerado fora dos limites de controle. Esta verificação/investigação de causas especiais são conhecidas na literatura como *Run Tests* (ou testes seqüenciais).

Vale ressaltar que em uma primeira fase de implantação do *CEP*, deve-se monitorar uma amostra que sirva de referência para as etapas posteriores. Isso significa que, após verificar as condições da matéria-prima, equipamentos, treinamento de pessoal, coleta-se uma amostra e verifica-se, se esta apresenta um comportamento sob controle. Caso o processo já esteja sob controle, pode-se passar para a segunda fase, na qual são realizadas outras coletas de amostras para proceder-se o monitoramento contínuo, caso apresente um estado fora dos limites de controle, deve-se investigar o processo até que a amostra do processo se normalize.

Assim, os gráficos de controle podem monitorar uma variável de qualidade como o número de produtos conformes ou não conformes (número de defeituosos). Esses tipos de gráficos são chamados de gráficos de controle por atributos. Os principais gráficos de controle por atributos são o gráfico *p*, que monitora a fração de produtos não conforme, o gráfico *np*, que monitora o número de produtos não conforme, a carta *c*, que monitora o número de não conformidades por unidade do produto e o gráfico *u*, que monitora o número médio de não conformidades por unidade do produto.

Outros gráficos de controle monitoram as características de qualidade que são expressas numa escala contínua de valores, como: temperatura e pressão. Esses gráficos são denominados de gráficos de controle para variáveis.

Na prática, os gráficos de controle para variáveis mais utilizados são os gráficos da média e amplitude ($\overline{X}$ *e R*) (Montgomery,1996). As médias amostrais ou valor alvo de uma variável é monitorado através do gráfico $\overline{X}$. Já o gráfico *R* é utilizado para monitorar, a partir de amplitudes amostrais, a variabilidade do processo. Outros tipos de gráficos de controle são utilizados para monitorar a média do processo, como: gráficos de controle para a soma acumulada (*CUSUM- Cumulativa Sum Control Chart*) e o gráfico de controle para as médias móveis ponderadas exponencialmente (*Gráfico EWMA- Exponentially Weighted Moving Average Control Chart).*

Segundo Montgomery (1996) pode-se monitorar a variabilidade do processo através do gráfico *S* que utiliza o desvio padrão no lugar da amplitude amostral. A escolha

do gráfico cuja medida é o desvio padrão ao invés do gráfico da amplitude *R* que só é justificado nos casos em que o tamanho do subgrupo for superior a 10, já que neste caso o gráfico *S* apresenta melhores resultados.

As equações para o cálculo dos limites de controle para os gráficos $\overline{X}$ e *S* podem ser determinadas da seguinte forma:

$$\overline{S} = \frac{\sum_{i=1}^{g} S_i}{g} \qquad \text{e} \qquad \overline{\overline{X}} = \frac{\sum_{i=1}^{g} \overline{X}_i}{g}. \tag{2.1}$$

Para os cálculos dos respectivos valores centrais dos gráficos $\overline{X}$ e *S*, temos que:

S_i : desvio padrão do *i-ésimo* subgrupo;

$\overline{S}$ *:* médias dos desvios padrões dos subgrupos;

$\overline{\overline{X}}$ *:* média das médias dos subgrupos;

$\overline{X}_i$: média do *i-ésimo* subgrupo e

g: número de subgrupos.

Para os limites de controle superior e inferior dos Gráficos $\overline{X}$ e S, respectivamente, considere

$$LSC_{\overline{X}} = \overline{\overline{X}} + A_3\overline{S} \qquad \text{e} \qquad LIC_{\overline{X}} = \overline{\overline{X}} - A_3\overline{S} \tag{2.2}$$

$$LSC_S = B_4\overline{S} \qquad \text{e} \qquad LIC_S = B_3\overline{S}\,. \tag{2.3}$$

Onde:

LSC: limite superior de controle;

LIC: limite inferior de controle,

A_3, B_3, e B_4: são fatores retirados a partir de tabelas próprias, encontrados em praticamente todos os livros textos da área, em função do tamanho dos subgrupos.

Já para os gráficos $\overline{X}$ e *R*, os limites, de controle são obtidos utilizando, respectivamente, as seguintes fórmulas:

$$LSC_{\overline{X}} = \overline{\overline{X}} + A_2\overline{R} \qquad \text{e} \qquad LIC_{\overline{X}} = \overline{\overline{X}} - A_2\overline{R} \tag{2.4}$$

$$LSC_R = D_3 \times \overline{R} \qquad \text{e} \qquad LIC_R = D_4 \times \overline{R}\,. \tag{2.5}$$

Assim, as expressões acima são estabelecidas para a construção dos gráficos de controle por variáveis desenvolvidos por Shewhart, servindo como base para o desenvolvimento de outros gráficos de controle.

Vale salientar, que esses tipos de gráficos devem ser aplicados a um conjunto de dados que possuam uma distribuição normal e que não possuam autocorrelação serial; pois a violação desses pressupostos pode levar a interpretações erradas e, conseqüentemente, a conclusões distorcidas sobre o processo. Além disso, esses gráficos ora apresentados são indicados para detectar variações no processo acima de dois desvios-padrão (2σ), não sendo tão eficaz na detecção de desvios/variações pequenos, ou seja, abaixo de dois desvios-padrão (2σ). Este valor de dois desvios-padrão com um valor mínimo de sensibilidade do gráfico de controle de Shewhart é recomendado na maioria dos livros e artigos pertinentes ao assunto, como Montgomery (1996), Hawnkins e Owell (1997), entre outros.

2.1.2. Tipos de Erros em Gráficos de Controle

Existem dois tipos de erros possíveis de ocorrer no monitoramento de processos por meio dos gráficos de controle. Quando um gráfico de controle indica um sinal, face a não ocorrência de mudança no processo, tal sinal é caracterizado pela ocorrência de um alarme falso.

O erro cometido na tomada de decisão neste caso indica-se por erro tipo I e a probabilidade dele ocorrer representa-se por α. Se o processo é considerado sob controle, a presença de causas determinísticas, dize-se que ocorre o erro tipo II e a probabilidade dele correr representa-se por β. A escolha dos limites de controle influencia diretamente as probabilidades de ocorrência dos erros tipo I e tipo II. Quanto mais distantes da linha central estiverem os limites de controle, menor será a probabilidade do erro tipo I, aumentando, todavia, a probabilidade do erro tipo II. Em contrapartida, aproximar os limites de controle da linha central provoca uma situação inversa.

Outro aspecto indispensável ao planejamento de um gráfico de controle (*GC*) é a determinação do tamanho da amostra e da freqüência de amostragem. Para tanto, deve ser considerado o custo de amostragem, as perdas pela produção de produtos de baixa qualidade e o custo para identificar as causas especiais. Além disso, devem ser levados em consideração aspectos estatísticos, como a minimização de alarmes falsos. Isto é, a ocorrência de uma causa especial quando não houve alteração no processo e, ainda, maximizar a capacidade de detecção de mudanças reais no processo.

Assim, o tamanho da amostra e a freqüência de amostragem devem ser definidos tendo em vista (*i*) o tamanho da mudança no processo a ser detectada pelo gráfico de controle e (*ii*) a rapidez com que se deseja detectar esta mudança. Desta forma, pequenas mudanças no processo podem ser rapidamente detectadas utilizando-se grandes amostras tomadas freqüentemente, embora esta estratégia possa ser economicamente inviável (Michel & Fogliatto, 2000).

Uma outra maneira de definir o tamanho da amostra e a freqüência de amostragem é através do comprimento médio da corrida (*ARL- Average Run Length*) do gráfico de controle. O *ARL* é o número médio de pontos que devem ser plotados antes que um ponto indique uma causa especial no gráfico de controle. O *ARL* pode ser usado também para avaliar o desempenho do gráfico de controle. Para um gráfico de controle de Shewhart, o *ARL* pode ser calculado como (Montgomery, 1996):

$$ARL = \frac{1}{P\,(ponto\ fora\ dos\ limites\ de\ controle)} \tag{2.6}$$

Considera-se para um *ARL* sob controle:

$$ARL_0 = \frac{1}{\alpha}. \tag{2.7}$$

E para um *ARL* fora de controle:

$$ARL_1 = \frac{1}{1-\beta}. \tag{2.8}$$

Os limites de controle utilizados no gráfico de controle de Shewhart são dados na equação (2.4). Assim, α= 0,0027, ou seja, existe uma probabilidade de 0,27% do gráfico de controle acusar incorretamente uma mudança na média do processo. O tamanho médio da corrida para este processo, na ausência de causas especiais, é dado por $ARL_0 = 1/\alpha = 1/0{,}0027 \approx 370$ amostras. Isto significa que, estando o processo sob controle, espera-se retirar, em média, 370 amostras sucessivas antes que o gráfico gere um alarme falso.

Quando a média do processo desviar-se do valor nominal em um desvio padrão $ARL_1 = 1/(1-\beta) = 1/0{,}0228 \approx 44$ amostras, onde *(1-β)* é a probabilidade de que o gráfico detecte um desvio de 1σ na média do processo, isto significa que serão necessárias 44 amostras sucessivas para que o gráfico sinalize uma mudança na média do processo.

Uma prática adequada, bastante utilizada na indústria, é a escolha entre pequenas amostras tomadas freqüentemente ou grandes amostras tomadas em intervalos de tempo maiores (Michel & Fogliatto, 2000).

2.1.3. Identificação Das Variáveis Fora dos Limites de Controle, Por Meio do Método de Gráficos Univariados

Alt (1985) sugere o uso do Gráfico Multivariado de Hotelling, simultaneamente a utilização de p gráficos univariados de Shewhart no controle do processo. A idéia é utilizar o gráfico multivariado para detectar mudanças no processo, e assim que for gerado um sinal fora dos limites de controle no gráfico multivariado, pode-se verificar através dos gráficos univariados qual grupo de variáveis pode ter sido responsável pelo sinal. Portanto, para minimizar o erro tipo I é necessário o uso dos limites de controle de *Bonferroni*. Tais limites são obtidos substituindo $\alpha/2$ por $\alpha/2p$ em cada um dos p gráficos univariados. Desta forma, ampliam-se os limites de controle nestes gráficos buscando-se obter uma probabilidade total de alarme falso α *(erro tipo I*, arbitrado no gráfico de Hotelling).

Woodall e Ncube (1985), Healy (1987) e Hawkins (1991) sugerem os procedimentos de controle que utilizam múltiplos gráficos CUSUM univariados (abreviado *MCU*) para detectar mudanças no vetor μ_0 de médias do processo. Sendo que os procedimentos *MCU* utilizam gráficos CUSUM univariados projetados para detectar a possível presença de uma causa especial da variação conhecida.

Pignatiello e Runger (1990) também sugerem a utilização de gráficos univariados para auxiliar na interpretação dos sinais fora dos limites de controle, gerados pelo gráfico *MC*1 (procedimento com melhor desempenho do *ARL*).

Mason *et al* (1997) apresentam restrições ao uso dos gráficos univariados para diagnosticar um sinal fora dos limites de controle num processo multivariado. Assim,

é difícil determinar a probabilidade total α de alarme falso, associado ao uso simultâneo de *p* gráficos univariados.

Existem situações que é preciso controlar mais que uma característica de qualidade. Isso pode levar a conclusões erradas do tipo: a probabilidade que ambas as variáveis excedam simultaneamente seus limites de controle quando o processo está sob controle é (0,0027) × (0,0027) que é igual a 0,0000073 que é muito menor que 0,0027, ou ainda, que a probabilidade de que ambas apareçam sob controle simultaneamente quando os processos estão sob controle é (0,9973) ×(0,9973) que é igual a 0,9946 e é diferente de 0,9973. Assim, essas distorções aumentam à medida que aumenta o número de variáveis a serem controladas.

O procedimento sugerido por Alt (1985) é o uso dos limites de *Bonferroni* para minimizar a probabilidade α. Se existem muitas variáveis, tais limites serão demasiadamente alargados nos *p* gráficos univariados, dificultando assim a detecção de um efeito no deslocamento da média de alguma variável no processo. Portanto, quando o gráfico multivariado detectar um desvio no processo, não significa necessariamente que apenas a média de uma ou mais variáveis tenha mudado, mas sim que a atual estrutura de correlação pode não estar de acordo com a estrutura de correlação das variáveis no processo de controle. Devido ao fato que a correlação entre as variáveis não é levada em consideração, pois não usam a matriz de covariância Σ_0 do processo, os gráficos univariados tornam-se pouco sensíveis para detectar mudanças desta natureza.

2.2. Gráficos de Controle Multivariados

Segundo Mason *et al* (1995), a utilização dos gráficos de controle multivariados é para monitorar processos industriais que estão se tornando cada vez mais difundidos. Esse monitoramento deve-se ao avanço da atual tecnologia de manipulação dos dados. Desta forma, computadores realizam em tempo real várias mensurações sobre as variáveis do processo e da qualidade final do produto, disponibilizando-se inúmeras informações sobre o processo. Em conseqüência desse avanço, tornou-se necessário desenvolver ferramentas mais robustas, que absorvam uma grande capacidade de informações de dados no monitoramento dos processos.

Os gráficos de controle multivariados (*GCM*) são utilizados no monitoramento conjunto de duas ou mais variáveis do processo ou produto. Nestes gráficos os pontos *p*-dimensionais (ou seja, são valores das *p* variáveis aleatórias ou estatísticas de interesse delas derivadas) são representados uni-dimensionalmente e *plotados* em gráficos similares aos gráficos de Shewhart, simplificando a tarefa de controle simultâneo de variáveis.

Em situações onde exista correlação significativa entre as variáveis a serem monitoradas os *GCM* são particularmente recomendados; pois, em processos multivariados, as variáveis interferem e sofrem interferências entre si, possuindo assim uma forte correlação. No caso em que as variáveis não sejam fortemente correlacionadas, e seja aplicado o *GCM,* há possibilidade de surgirem alarmes falsos no processo investigado, já que o operador pode constatar problemas em uma determinada variável que não esteja necessariamente interferindo no processo; ou também pode considerar que o processo está sob controle sem que o mesmo esteja, em função desta estar trabalhando com variáveis que não interferem no mesmo, mas que influenciam na média e desvio-padrão. Uma forma de identificar a correlação existente entre as variáveis é através da aplicação do teste de Fischer em correlações amostrais, em que se pode determinar a significância das mesmas; (Freund & Simon, 1997, p. 538).

Biblioteca Universitária
UFSC
0-360-749-0

Assim, os gráficos *Qui-quadrado* e de T^2 Hotelling, descritos na seqüência, são gráficos de controle do tipo Shewhart; no que diz respeito à detecção de grandes variações no processo (acima de 2σ), bem como na verificação do comportamento do processo através da média.

2.2.1. Gráfico de Controle Qui-quadrado- χ^2

Suponha que exista apenas uma variável de interesse, designada por *X*, a ser monitorada em um determinado processo. Amostras periódicas independentes, compostas por *n* observações da variável $(x_1 ..., x_n)$ são obtidas do processo. Sendo $\overline{X}$ a média amostral das *n* observações. A variável *X* segue uma distribuição Normal, com média $\mu = \mu_0$ e desvio padrão $\sigma = \sigma_0$ conhecidos e obtidos do processo sob controle estatístico; ou seja, $X \sim N(\mu_0, \sigma_0)$, média e desvio padrão, então o valor

$$Z = \frac{(\overline{X} - \mu_0)\sqrt{n}}{\sigma_0}, \qquad (2.9)$$

é uma variável que segue uma distribuição normal com média 0 e desvio-padrão unitário, ou seja, $Z \sim N(0,1)$. As médias amostrais são monitoradas através da variável *Z*. O valor de *Z* representa a distância padronizada entre uma média amostral $\overline{X}$ e a média do processo μ_0. Caso o processo esteja sob controle, espera-se apenas pequenos desvios aleatórios em relação à média do processo. Logo, se diz que apenas causas comuns de variação estão presentes e que o processo está sob controle estatístico.

A partir da expressão (2.9) tem-se o valor *Z* correspondente à média da *k-ésima* amostra retirada do processo (*com k= 1,2...,n).* Caso o processo seja considerado sob controle estatístico espera-se que o número de amostras seja suficientemente

grande até que um ponto ultrapasse os limites de controle. Ao utilizam-se a variável *Z*, os limites de controle são expressos da seguinte forma:

$$LSC = L = Z_{\alpha/2},$$
$$e \qquad (2.10)$$
$$LIC = -L = -Z_{\alpha/2}.$$

onde $Z_{\alpha/2}$ é o percentil da distribuição Normal, o qual é obtido da probabilidade α e estabelecido para o intervalo de confiança, ou seja (1-α).
Utiliza-se $\mu_0 \pm 3\sigma_0/\sqrt{n}$ com ($Z \in$ *[-3; +3] ou seja* | *L*| *=3)* como intervalo de tolerância para as médias amostrais obtidas de um processo sob controle estatístico (Montgomery, 1996). Assim, $Z_{\alpha/2}$=3 como *Z~N (0,1)*, a probabilidade de alarme falso é 0,27%, ou seja, (1-α).

Caso a distância padronizada entre uma média amostral $\overline{X}$ e a média histórica do processo μ_0 for superior a *L* ou inferior *–L*, conclui-se que existe um indicativo que a média do processo sofreu uma alteração ($\mu \neq \mu_0$). Portanto, existem indícios de que o processo possa estar fora de controle estatístico, sendo necessário uma investigação para encontrar as possíveis causas desta variação. Tal procedimento de controle, criado por Walter A. Shewhart é conhecido na literatura como Gráfico de Controle de Shewhart.

Assim, χ_1^2, segue uma distribuição do *Qui-quadrado* com um grau de liberdade e representa a distância padronizada entre a média $\overline{X}$ obtida em cada amostra e a média do processo μ_0. Elevando a expressão (2.9) ao quadrado, ela passa a ser expressa na seguinte forma.

$$\chi_1^2 = \frac{(\overline{X} - \mu_0)^2 n}{\sigma_0^2} \qquad \text{ou} \qquad \chi_1^2 = n(\overline{X} - \mu_0)(\sigma_0^2)^{-1}(\overline{X} - \mu_0). \qquad (2.11)$$

Logo, pode-se monitorar as médias amostrais $\bar{X}$ através do valor χ^2 obtido na expressão (2.11).Portanto quando *L= 3* então

$$LSC= 9 \qquad \text{e} \qquad LIC=0 \tag{2.12}$$

Observa-se a partir da equação (2.11) que o valor mínimo da estatística é zero *(LIC=0)*, pois a mesma representa a medida da distância padronizada. Quando a distância quadrada padronizada entre a média amostral $\bar{X}$ e a média do processo μ_0 ultrapassar 9 $[(\pm 3)^2]$ indica que o processo está fora dos limites de controle estatístico.

Supõe-se que existam *p* variáveis de interesse sendo monitoradas em um determinado processo e que estas variáveis estejam correlacionadas. Seja $\mathbf{X} = [X_1, X_2 \ldots, X_p]$ um vetor contendo *p* variáveis aleatórias, tal que $X_i, i = 1, \ldots, p$, seja a *i-ésima* variável de interesse, e $x_{i1}, \ldots, x_{in}$ suas *n* observações. Para cada amostra periódica independente de *n* observações dessas *p* variáveis obtém-se o vetor de médias $\bar{\mathbf{x}} = \left[\bar{x}_1, \bar{x}_2, \ldots \bar{x}_p\right]'$. Supondo que as *p* variáveis sigam uma distribuição Normal *p*-variada, com vetor de médias $\mu=\mu_0$ e matriz de covariâncias $\Sigma=\Sigma_0$ conhecidos e dados por:

$$\boldsymbol{\mu}_0 = \begin{bmatrix} \mu_{01} \\ \mu_{02} \\ \cdots \\ \mu_{0p} \end{bmatrix} \qquad e \qquad \Sigma_0 = \begin{bmatrix} \sigma^2_{0,1} & \sigma_{0,12} & \cdots & \sigma_{0,1p} \\ \sigma_{0,21} & \sigma^2_{0,2} & \cdots & \sigma_{0,2p} \\ \cdots & \cdots & \cdots & \cdots \\ \sigma_{0,p1} & \sigma_{0,p2} & \cdots & \sigma^2_{0,p} \end{bmatrix}$$

onde μ_{0i} e $\sigma^2_{0,i}$ para *i, j* = 1, ...p, correspondem à média e à variância da *i-ésima* variável, e $\sigma_{0,ij}$, representa a covariância entre variáveis dependentes *i* e *j*. Logo, diz-se que $X \sim N(\mu_0, \Sigma_0)$.

Supondo que o vetor de médias μ_0 e a matriz de covariância Σ_0 caracterizem um processo em estado de controle estatístico, a expressão (2.11) pode ser

generalizada para o caso onde *p* variáveis são amostradas simultaneamente (Lowry & Montgomery, 1995). Tem-se então a seguinte estatística:

$$\chi^2_p = n\,(\overline{X} - \mu_0)'\,\Sigma_0^{-1}\,(\overline{X} - \mu_0)\,, \qquad (2.13)$$

na qual, χ^2_p segue uma distribuição do *Qui-quadrado* com *p* graus de liberdade e representa a distância padronizada *p-dimensional*, entre um vetor de médias $\overline{X}$ e o vetor da média do processo μ_0. A raiz quadrada χ_p dessa quantidade é conhecida como "*distância de Mahalanobis*"; Rencher, (1993). O valor de χ^2_p não deve ultrapassar certos limites de controle para que o processo gerador dos dados seja considerado sob controle estatístico. Esses limites são obtidos da seguinte forma

$$LSC = \chi^2_{p,\alpha} \qquad \text{e} \qquad LIC=0 \qquad (2.14)$$

onde $\chi^2_{p,\alpha}$, representa o percentil da distribuição do *Qui-quadrado* com *p* graus de liberdade, obtido através da probabilidade α de alarme falso adotado.

2.2.2. Gráficos de Hotelling ou Gráficos T^2

O gráfico T^2 foi desenvolvido por Hotelling (1947), o pioneiro na pesquisa sobre os gráficos de controle multivariado. Ele utilizou-se da abordagem multivariada de controle em dados contendo informações sobre localizações de bombardeios na segunda guerra mundial. O desenvolvimento teórico proposto por Hotelling é descrito a seguir.

Na seção anterior, analisou-se um processo no qual o vetor de médias μ_0 e a matriz de covariância Σ_0 das variáveis envolvidas eram conhecidos. Na prática μ_0 e Σ_0 são freqüentemente estimados através de amostras preliminares recolhidas do processo, quando este está sob controle estatístico.

Considerando as estimativas dos parâmetros, a equação será expressa na seguinte forma:

$$T^2 = n(\overline{\mathbf{X}} - \overline{\overline{\mathbf{X}}})'\overline{S}^{-1}(\overline{\mathbf{X}} - \overline{\overline{\mathbf{X}}}). \tag{2.15}$$

onde $\overline{\overline{\mathbf{X}}}$ e $\overline{\mathbf{S}}$ representam, respectivamente, as estimativas para o vetor de médias e matriz de covariâncias do processo, sendo obtidos na seguinte forma:

$$\overline{\overline{\mathbf{X}}} = \frac{1}{c}\sum_{k=1}^{c}\overline{\mathbf{X}}_k. \tag{2.16}$$

e

$$\overline{\mathbf{S}} = \frac{1}{c}\sum_{k=1}^{c}\mathbf{S}_k. \tag{2.17}$$

com

$$\mathbf{S}_k = \frac{1}{n-1}\sum_{i=1}^{n}(\mathbf{X}_i - \overline{\mathbf{X}}_k)(\mathbf{X}_i - \overline{\mathbf{X}}_k). \tag{2.18}$$

onde *c* representa o número de amostras preliminares de tamanho *n* retiradas do processo e $\mathbf{X}_i$ é a *i-ésima* observação da *p*-variada referente à *k-ésima* amostra.

A expressão estatística (2.15) é utilizada como base para o gráfico de controle T^2 de Hotelling (Lowry & Montgomery, 1995). Ryan (1989) derivou a distribuição dessa

estatística, em ambas as fases. Na primeira fase, os limites de controle para a estatística T^2 são expressas como segue:

$$LSC = \frac{p(c-1)(n-1)}{cn-c-p+1} F_{\alpha,p,cn-c-p+1} \tag{2.19}$$

Onde $F_{\alpha,p,cn-c-p+1}$ representa o percentil da distribuição *F* com p e (*cn-c-p+1*) graus de liberdade, obtido a partir da probabilidade α de alarme falso adotada.

Na segunda fase de controle os novos limites são estabelecidos apenas para monitorar as observações futuras, utilizando os limites de controle mostrados na equação (2.20).

$$LSC = \frac{p(m+1)(n-1)}{mn-m-p+1} F_{\alpha,p,mn-m-p+1}, \tag{2.20}$$

$$LIC = 0$$

Onde:

p: é o número de variáveis (característica de qualidade);

n: é o tamanho de subgrupo;

m: é o número de observações da amostra que representa o processo sob controle, filtradas a partir das *c* amostras. Nesta fase se *m* < *c*, ou seja, *c-m* amostras foram descartadas por não caracterizarem o padrão das variáveis no processo, recalcula-se $\overline{\overline{\mathbf{X}}}$ e $\overline{\mathbf{S}}$ para monitoramento das amostras futuras substituindo c por m nas expressões (2.16) e (2.17).

Para ilustrar a operacionalização dos *GCM*, considere duas variáveis de interesse (força de tensão e diâmetro) mensuradas em um mesmo processo de produção de Fibra Têxtil, (Montgomery, 1996). Foram coletadas *m*=20 amostras do processo, em condições desejáveis de operação com a formação de subgrupo de tamanho *n* = 10. Assim, conforme a Figura 2, as estatísticas T^2 de Hotelling foram calculadas a partir da equação (2.15).

Tabela 1: Médias, Variâncias e Covariâncias da Força de Tensão e Diâmetro de uma Fibra Têxtil, (Montgomery, 1996, p. 368).

Amostra	Médias Amostrais		Variâncias e Covariâncias			Hotelling
K	$\overline{X}_{1k}$	$\overline{X}_{2k}$	S^2_{1k}	S^2_{2k}	S^2_{12k}	T^2
1	115.25	1.04	1.25	0.87	0.80	2.16
2	115.91	1.06	1.26	0.85	0.81	2.14
3	115.05	1.09	1.30	0.90	0.82	6.77
4	116.21	1.05	1.02	0.85	0.81	8.29
5	115.90	1.07	1.16	0.73	0.80	1.89
6	115.55	1.06	1.01	0.80	0.76	0.03
7	114.98	1.05	1.25	0.78	0.75	7.54
8	115.25	1.10	1.40	0.83	0.80	3.01
9	116.15	1.09	1.19	0.87	0.83	5.92
10	115.92	1.05	1.17	0.86	0.95	2.41
11	115.75	0.99	1.45	0.79	0.78	1.13
12	114.90	1.06	1.24	0.82	0.81	9.96
13	116.01	1.05	1.26	0.55	0.72	3.86
14	115.83	1.07	1.17	0.76	0.75	1.11
15	115.29	1.11	1.23	0.89	0.82	2.56
16	115.63	1.04	1.24	0.91	0.83	0.70
17	115.47	1.03	1.20	0.95	0.70	0.19
18	115.58	1.05	1.18	0.83	0.79	0.00
19	115.72	1.06	1.31	0.89	0.76	0.35
20	115.40	10.4	1.29	0.85	0.68	0.62

Logo, para o gráfico de controle T^2 de Hotelling, através dos limites de controle calculados a partir da equação (2.20), encontrou-se o valor de $LSC = 21{,}93$, com $\alpha = 0{,}001$, conforme apresentado na Figura 2. Os pontos no gráfico caracterizam um processo estável.

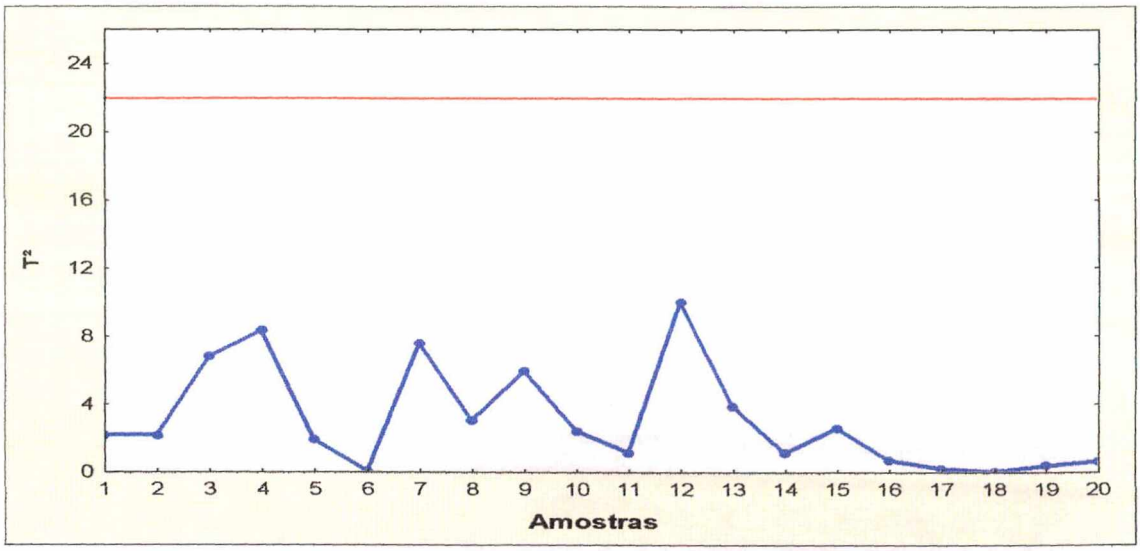


Figura 2: Gráfico de Controle T^2 *de Hotelling*

Para os gráficos de controle univariados, a utilização de amostras preliminares com $m \geq 20$ ou 25 existe uma distinção entre os limites de controle na primeira e segunda fase para o gráfico de médias que é desnecessária. Já a seleção dos limites de controle para os gráficos de controle multivariados deve ser cuidadosa, recomenda-se a construção dos gráficos de controle em ambas as fases.

Lowry & Montgomery (1995) apresentaram tabelas indicando o número mínimo recomendado de *m* amostras preliminares para tamanhos de n =3, 5 e 10 e para p = 2, 3, 4, 5, 10 e 20 variáveis, para que os limites de controle da segunda fase sejam bem aproximados, através dos limites de controle obtidos, via aproximação da distribuição *Qui-quadrado*. Os valores recomendados para *m* foram sempre maiores que 50. Logo, quando o número de variáveis *p* aumenta, maior o valor de *m* necessário para uma boa aproximação dos limites de controle da segunda fase.

Em algumas situações, o processo é monitorado utilizando amostras de tamanho unitário ($n = 1$). Nestes casos, não é possível a seleção de subgrupos homogêneos de tamanho grande. Portanto, não faz sentido agrupar uma seqüência de observações consecutivas das variáveis, pois verifica-se, a cada instante, se as variáveis estão seguindo as trajetórias especificadas para um determinado processo que deseja-se construir um gráfico de controle multivariado para *p* variáveis, a partir

de *m* amostras preliminares, de tamanho *n* = 1, retirados do processo sob controle estatístico. Neste caso, a estatística T^2 de Hotelling da equação (2.15) pode ser expressa na seguinte forma:

$$T^2 = (\mathbf{x} - \bar{\mathbf{x}})\mathbf{S}^{-1}(\mathbf{x} - \bar{\mathbf{x}}). \tag{2.21}$$

onde $\bar{\mathbf{x}}$ e $\boldsymbol{S}$ representam respectivamente as estimativas para o vetor das médias e a matriz de covariância do processo.

Na primeira fase, quando selecionam-se *c* amostras preliminares que representam o processo sob controle, Tracy *et al* (1992) mostraram que os limites de controle para a estatística T^2 é dado por:

$$LSC = \frac{(c-1)^2}{c}\beta(\alpha; p/2, (c-p-1)/2)$$
$$LIC = 0. \tag{2.22}$$

Onde $\beta(\alpha; p/2, (c-p-1)/2)$ representa o percentil de uma distribuição *beta* com *p/2* e *c-p-1/2* graus de liberdade, obtido a partir da probabilidade α de alarme falso adotado. Valores na distribuição *Beta* podem ser aproximados pela distribuição *F*. Assim, pode-se obter limites de controle na expressão (2.21) da seguinte forma (Lowry & Montgomery,1995):

$$LSC = \frac{(c-1)^2}{c} \times \frac{(p/(c-p-1)F(\alpha; p, (c-p-1)}{1 + (p/(c-p-1))F(\alpha; p, c-p-1)},$$
$$LIC = 0. \tag{2.23}$$

A partir de observações individuais nos gráficos construídos, utilizam-se *m* amostras preliminares na primeira fase para o monitoramento de *p* variáveis, o limite de controle durante a segunda fase é definido pela equação (2.24); Ryan (1989).

$$LSC = \frac{p(m+1)(m-1)}{m^2 - mp} F_{\alpha,p,m-p,} \cdot \qquad (2.24)$$

onde, $F_{\alpha,p,\ m\text{-}p}$ significa o percentual da distribuição *F* com *p* e *m-p* graus de liberdade e (1-α) denota o intervalo de confiança representado pelo *GCM*. O limite de controle inferior é zero, por definição.

Quando se utilizam grandes quantidades de amostras preliminares na primeira fase, (m > 100) o limite de controle da segunda fase pode ser aproximado pela seguinte equação (Jackson, 1985).

$$LCS = \frac{p(m-1)}{m-p} F_{\alpha,p,m-p,} \cdot \qquad (2.25)$$

Quando as amostras possuem tamanho unitário, os limites de controle são calculados através da equação (2.25) ou aproximado através da distribuição χ^2. A comparação do limite de controle exato na segunda fase (dada pela equação (2.24)) para amostras de *n* = 1, com o limite de controle aproximado pela distribuição *Qui-quadrado,* para *p*=2, 3, 4, 5, 10 e 20 variáveis foi apresentada por Lowry & Montgomery (1995). À medida que *p* aumenta, maior é o valor de *m,* necessário para manter o erro relativo entre o limite de controle exato e a aproximação χ^2. Para *p* =3, por exemplo, para manter o erro relativo 0,1 é necessário um número m = 150.

Para a construção de gráficos de Controle Multivariados com observações individuais (*n*=1) em ambas as fases, foram apresentados cálculos para o limite de controle por Sullivan & Woodall (1996). Eles ainda apresentaram diferentes estimadores da matriz de covariância (Σ), e analisaram sua utilização nos gráficos de controle quanto à sensibilidade às mudanças no vetor das médias do processo. Dois tipos de alterações no vetor de médias do processo foram investigadas: mudanças bruscas (*step shift*) ou mudanças graduais (*ramp shift*). Observações adjacentes tendem a ser obtidas do mesmo vetor de médias do processo. Se houver uma mudança brusca no vetor das médias, somente um par de observações sucessivas será obtido a partir de diferentes vetores das médias. Se houver uma mudança gradativa, em contrapartida, sucessivos pares serão obtidos a partir de diferentes

vetores de médias. Tais diferenças, porém, serão pequenas se comparadas as duas observações aleatoriamente escolhidas, (Sullivan & Woodall, 1996).

Holmes & Mergen (1993), propuseram um estimador para Σ calculado a partir da diferença entre os pares de observações sucessivas. Eles consideraram V_i (com i = 1,..., *m-1*) como a diferença entre os pares de observações sucessivas dado pela equação (2.26) e **V** a matriz formada pelas diferenças V_i:

$$V_i = X_{i+1} - X_i . \tag{2.26}$$

Assim, o estimador proposto para Σ é dado pela equação (2.27) a seguir:

$$\overline{S} = \frac{1}{2}\frac{\boldsymbol{V'V}}{(m-1)} . \tag{2.27}$$

Onde:

$$\boldsymbol{V} = \begin{bmatrix} \mathbf{v}_1 \\ \mathbf{v}_2 \\ \cdots \\ \mathbf{v}_{m-1} \end{bmatrix}$$

onde $\overline{\mathbf{S}}$ representa a matriz de covariância dessas diferenças. Usando $\overline{\mathbf{S}}$ em lugar de **S** na estatística de controle de T^2 de Hotelling na expressão (2.15), obtém-se um gráfico de controle mais sensível a possíveis mudanças no vetor de médias do processo.

A determinação do tamanho da amostra e da freqüência de amostragem é um aspecto indispensável ao planejamento do gráfico de controle multivariado. Tais parâmetros devem ser selecionados de modo a tornar o *GCM* mais sensível à presença de causas especiais. Para tanto, agrupam-se os dados multivariados do processo em subgrupos racionais. Esses subgrupos devem ser selecionados, de acordo com as características do processo, a fim de maximizar as diferenças dentro dos subgrupos, sempre que causas especiais estiverem presentes (Sullivam e Woodall, 1996). Para obter-se esses grupos racionais, existem diferentes

estratégias, conforme as características do processo produtivo (ver Montgomery (1996, p. 367)). Alguns métodos são recomendados por autores que tratam o caso de observações multivariadas individuais, como, por exemplo, Jackson (1985) Tracy *et al.* (1992) Lowry & Montgomery (1995) e Wierda (1994).

Além do gráfico de controle T^2 de Hotelling, existem outros tipos de gráficos de controle que trabalham com processos multivariados, a saber: gráfico de controle multivariado para a soma acumulada (*gráfico de controle MCUSUM*) e o gráfico de controle multivariado para a média móvel ponderada exponencialmente (*gráfico de controle MEWMA).* Estes gráficos de controle multivariado possuem a característica de detectar pequenas mudanças no processo (abaixo de 2σ), ao contrário do gráfico de controle multivariado T^2 Hotelling que detecta grandes variações; maiores detalhes ((Lowry & Montgomery, 1995), (Lowry *et al*, 1992), (Montgomery, 1996)).

2.3. Custos da Qualidade

Na atual situação de competitividade em mercados nacionais e internacionais, a qualidade tem uma influência decisiva sobre o preço de venda de qualquer produto, além do cumprimento de prazos de entrega. Para que a empresa possa garantir a qualidade sob seus serviços é preciso um gerenciamento da qualidade durante todo o processo produtivo. Neste sentido, é necessário investir em metodologias que minimizam os custos da má-qualidade. Esses investimentos trarão benefícios futuros como: redução de custos devido à menor incidência de serviço com uma qualidade insatisfatória, redução de retrabalho dentre outros.

Tagushi *et al.* (1990) ensina que a qualidade de um produto é identificada através de algumas características que reduzem a perda total para o consumidor. Em outras palavras: qualidade é a crítica científica da má-qualidade. O consumidor tendo comprado um produto mais barato, porém de qualidade inferior, terá maiores perdas com quebras e defeitos, e ainda que este esteja na garantia, as perdas continuarão, decorrentes da não-utilização. Desta forma, tendo investido pouco em qualidade, o consumidor terá alta perda devida à má-qualidade e o resultado líquido será uma perda total desvantajosa.

A empresa que administra mal seus recursos, desperdiçando materiais, depreciando seus equipamentos, não aproveitando do potencial de sua mão de obra, sub-utilizando a informação disponível, conseqüentemente terá perdas significativas. Novamente, vê-se que um baixo investimento em qualidade acarreta em altas perdas em função da má-qualidade, que resulta em perdas finais insatisfatórias, Frota (1998).

Em contrapartida, uma empresa que trabalha com alta qualidade, utiliza racionalmente seus recursos e conseqüentemente suas perdas totais serão baixas. O empresário que contrata pessoal capacitado e a compra de equipamentos de marca confiável terá alguns custos adicionais, para o funcionamento da empresa será dinâmico e suas perdas minimizadas e sua receita maximizada. Mesmo que seu produto não seja "o-mais-barato", terá vantagens econômicas para o cliente em

função das suas características aprimoradas pela qualidade, ou seja, seu custo benefício é vantajoso. A função da qualidade de uma empresa abrange todos os setores e suas atividades de forma direta e indireta, com a finalidade de melhorar a qualidade do produto final mantendo a confiabilidade do produto.

A qualidade, enquanto conjunto de atividades, custa dinheiro, ainda que tais custos fiquem espalhados pelos vários departamentos da empresa. Os custos operacionais da qualidade, no entanto, não representam apenas uma soma monetária, mas alguns elementos que possuem ação positiva ou negativa que geram custos. Tudo o que é negativo gera custo, e na maioria das vezes, custos maiores do que é gasto com ações positivas. Os custos operacionais da qualidade podem ser definidos como tudo o que é despendido na melhoria continua da qualidade.

A partir dessa definição inicial, pode-se dividir os custos operacionais da qualidade em categorias como os custos que aparecem na falha da qualidade e podem ser divididos em: custos de falhas internas e custos de falhas externas; de acordo com a situação do produto no instante da falha, se a ocorrência for dentro da empresa é caracterizado como uma falha interna e caracterizado como falha externa se a falha ocorrer após ter sido vendido aos clientes. Os custos da qualidade podem ser divididos em custos de prevenção, que englobam tudo o que é feito para prevenir defeitos, bem como os custos de avaliação, que reúnem os esforços despendidos para remover do processo, ações e materiais que são as causas dos produtos defeituosos.

2.3.1. Custo de Prevenção

Para Moreira (1996) e Slack *et al* (1997), o custo de prevenção representa os gastos de evitar o trabalho com defeito, falhas ou quebras, ou seja, são custos incorridos na prevenção de problemas. Crosby (1994) considera o custo de prevenção como todas as atividades empreendidas para prevenir defeitos de *design*, de mão-de-obra, de aquisição de materiais já com defeitos e na criação e desenvolvimento dos produtos.

O custo de prevenção incorre antes que o produto seja feito ou o serviço prestado. Assim, garante-se que o produto entregue ao cliente apresentará a qualidade desejada e atendendo as perspectivas dos adquirentes.

Slack *et al* (1997) incluem nas atividades relacionadas ao custo de prevenção a identificação de problemas e a correção do processo; o desenvolvimento de *design* de produtos, serviços e processos para reduzir os problemas que ocasionam má qualidade; e o treinamento para o pessoal executar seu trabalho da melhor maneira possível.

2.3.2. Custo de Avaliação

Segundo Moreira (p. 48, 1996) os custos de avaliação são aqueles incorridos para verificar se os produtos estão dentro dos limites de especificação. São os custos que incorrem durante as inspeções, testes e avaliações planejadas com a finalidade de se determinar as conformidades dos produtos com as suas especificações, conforme afirma Crosby (1994).

O custo da avaliação, segundo Slack *et al* (p. 659, 1997), está associado ao "controle da qualidade, que visa checar se ocorreram problemas ou erros durante o processo produtivo e após a criação do produto".

Pode-se exemplificar o custo de avaliação como os custos de auditar procedimentos, tais como: inspecionar produtos e processos, vistoriar os materiais enviados pelos fornecedores, condução de pesquisa junto aos clientes, entre outros.

Desta forma, considera-se o custo da avaliação os gastos que a empresa discorre em decorrência da correta determinação de condições dos produtos. Isso ocorre depois que o bem já foi feito, mas antes de ser liberado e enviado ao consumidor.

2.3.3. Custos de Falhas

Os custos de falhas estão associados a fatores que, segundo se verificou, não estão de acordo com os requisitos, no desempenho, na avaliação, e no aspecto dos negócios Crosby (p. 135, 1994). Esses custos representam a diferença nos requisitos previamente determinados aos produtos.

Essa diferença nos requisitos pode ser em decorrência de falhas nas atividades ou nos processos detectados em operações internas, ou seja, antes que o produto chegue ao cliente. Quando ocorre nesta etapa são considerados custos de falhas internas.

Neste grupo englobam as despesas advindas de unidades viram refugos, retrabalhos e retestes de itens defeituosos. Além de gastos com tempo de produção perdido em razão do erro. Em decorrência desses fatos, os produtos perdem valor quando são vendidos, pois possuem uma qualidade inferior aos demais e, conseqüentemente, preços inferiores. (Moreira, 1996)

Por outro lado, pode ocorrer falhas que são detectadas somente fora das operações da empresa, quando o produto já se encontra junto ao consumidor. Assim, nestes casos, são consideradas falhas externas.

Neste grupo podem estar incluídos os custos com garantias, com devoluções, com reclamações, entre outros. Além disso, ocorre a perda de confiança do cliente. Em decorrência destas falhas, futuros negócios podem ser afetados, conforme afirmam Slack *et al* (1997).

Moreira (1996) acrescenta que quando ocorre a perda de futuros negócios em decorrência de falhas externas, geram custo de difícil mensuração. Desta forma ,quando os custos da qualidade focalizam além de requisitos, da conformidade dos requisitos, falhas dos produtos, o impacto que a empresa pode provocar ao meio ambiente, e identifica os gastos associados à sua preservação, inicia-se o desenvolvimento do custo ambiental.

2.3.4. A implicação dos Gráficos de Controle nos Custos da Qualidade

Os gráficos de controle, conforme mencionado anteriormente, possuem a vantagem de monitorar sistematicamente as características da qualidade do processo, procurando detectar possíveis alterações. Assim, a utilização dos gráficos de controle pode proporcionar a redução dos custos de avaliação à medida que o monitoramento sistemático do processo ajuda a mantê-lo de acordo com as especificações do projeto. Em adição ao monitoramento sistemático, já que ocorre uma melhora na prevenção, detectando rapidamente possíveis problemas existentes no processo. Os gráficos de controle agilizam também na busca das causas desses problemas, evitando com isso o retorno do problema e indicando desta forma medidas de Prevenção.

2.4. Considerações Finais

Foi apresentado neste capitulo uma breve revisão sobre Controle Estatístico de Processo, bem como uma abordagem a respeito dos gráficos de controle tradicionais e dos gráficos de controle multivariados, mais especificamente o gráfico de controle de T^2 Hotelling. Além disso, foram apresentados os custos da qualidade e suas divisões, a saber: custos de prevenção, custos de avaliação e os custos de falha.

Conforme descrito nas seções anteriores, pode-se considerar que os gráficos de controle multivariados são de suma importância nos processos industriais, podendo ser destacado dentre estes, o gráfico de controle Hotelling, por trabalhar com várias variáveis simultaneamente, já que nas indústrias os problemas apresentados no processo são influenciados por um conjunto de fatores, normalmente provocados por mais de uma característica. Desta forma, a utilização do gráfico de controle T^2 de Hotelling pode reduzir os custos de falhas internas bem como de avaliação, à medida que proporciona aos engenheiros monitorar várias características da

qualidade simultaneamente e investigar com mais rapidez as causas dos problemas encontrados no processo, minimizando assim as perdas e, conseqüentemente, maximizando o lucro.

No capítulo a seguir será apresentado o método proposto no presente trabalho, bem como a linguagem de programação a ser utilizada no desenvolvimento do algoritmo computacional para a decomposição da estatística T^2 de Hotelling.

3. O Método Proposto

Neste capítulo propõe-se um estudo referente os procedimentos da identificação das variáveis fora dos limites de controle em gráficos multivariados, por meio do método de decomposição da estatística T^2 de Hotelling. Na elaboração do algoritmo será utilizada a abordagem sugerida por Mason *et al* (1995).

3.1. Linguagem de Programação

No mundo dos negócios, informação é dinheiro. A capacidade de converter rapidamente dados brutos em informações úteis leva a melhores decisões, e conseqüentemente, a maiores lucros.

Como as demais áreas de tecnologia, na década de oitenta o controle de qualidade experimentou um tremendo crescimento em aplicações computacionais. Porém, não se sabe exatamente quando os computadores foram utilizados pela primeira vez para estes propósitos. O *Journal of Quality Technology*, em 1969, mostra um programa computacional para análise de dados no controle da qualidade.

Com a chegada dos *softwares* de análise geral de dados, implementados em forma de planilhas, houve um grande avanço na análise de informações gerais das empresas, embora estes softwares fossem limitados, ou seja, não previam funções mais elaboradas, específicas para a análise de dados.

As planilhas estão por toda parte e fazem parte do nosso dia a dia, como por exemplo: empresas, universidades, comércio, indústria, governo, instituições e imprensa. Assim, as planilhas de cálculo transformaram-se em uma plataforma de desenvolvimento graças à linguagem de programação *VBA*. Este fato possibilitou o aprimoramento de aplicações mais sofisticadas, facilitando a interface amigável com o usuário.

Assim, a linguagem de programação *Visual Basic* originou novos produtos, como o VbScript que é uma linguagem de programação utilizada em aplicações *WEB*. Existem alguns seguimentos específicos de aplicação usando parte dos recursos da linguagem *VB*, como o VBA (Visual Basic for *Applications)* e *VBScript*.

No presente trabalho, será utilizado a planilha de cálculo com *VBA*, desenvolvido pela Microsoft Corporation®. Trata-se de uma ferramenta que apresenta recursos necessários para o desenvolvimento do algoritmo computacional, proposto neste trabalho. Além de serem de fácil utilização e amplamente difundidos na área empresarial e industrial, é de baixo custo.

3.2. Procedimento para Identificar Variáveis Fora dos Limites de Controle em Gráficos Multivariados

Segundo Jackson (1991), um procedimento de controle multivariado deve satisfazer alguns requisitos, como: (*i*) responder o mais rápido possível se o processo está sob controle ou fora de controle; (*ii*) estabelecer com precisão a probabilidade de erro do tipo I; (*iii*) verificar a estrutura de correlação existente entre as variáveis envolvidas; (*iv*) oferecer uma rápida identificação sob as causas especiais incidentes sobre o processo.

O item (*iv*) é o de mais difícil obtenção. Quando um gráfico multivariado detecta a possível presença de causas especiais de variação no processo, não possui uma informação clara sobre as causas que atuaram sobre as variáveis, o que leva o processo ao estado fora dos limites de controle. Tal dificuldade de interpretação do ponto fora dos limites de controle num gráfico multivariado tem sido algo de muita discussão por vários autores. Por estas dificuldades, vários métodos foram propostos, podendo ser divididos em dois grupos. O primeiro seria o procedimento de diagnóstico via gráficos univariados, visto no capítulo anterior, e o segundo procedimento de diagnóstico via decomposição da estatística T^2 de Hotelling.

3.3. Identificação das Variáveis Fora dos Limites de Controle, Através dos Métodos de Decomposição da Estatística T^2 de Hotelling

3.3.1. O método de decomposição da estatística T^2 de Hotelling por Mason *et al*

Vários autores têm sugerido técnicas para diagnosticar sinais fora dos limites de controle em gráficos multivariados baseadas na decomposição da estatística T^2 de Hotelling. A idéia é dividir T^2 em duas ou mais partes, onde cada parte representa um subgrupo de *q (< p)* variáveis.

Na segunda fase, pode-se monitorar o vetor *p*-dimensional μ_0 de médias do processo, a partir de *m* amostras preliminares unitárias (*n* = 1) recolhidas do processo sobre controle estatístico testado na primeira fase, conforme mencionada no capítulo anterior. Assim, quando uma nova amostra é coletada, uma observação $\mathbf{x}_k'$ *(com k=m+1)* é obtida e, o devido valor T_k^2 de Hotelling é calculado através da expressão (2.21).

Para um dado valor de α, se $T_k^2 > [p(m+1)(m-1)/m(m-p)]\, F_{\alpha,p,m-p}$, se o limite superior de controle calculado pela expressão (2.24) gerar um ponto fora dos limites de controle é necessário verificar as variáveis responsáveis por esse ponto fora dos limites de controle, através da decomposição da estatística T^2. Contudo, existem várias maneiras de efetuá-la que estão sugeridas na literatura e algumas delas são apresentadas a seguir.

Mason *et al.* (1995) sugerem que a estatística T^2 seja dividida em componentes independentes, cada um refletindo a contribuição de uma variável individual. Utilizando um resultado originalmente obtido por Rencher (1993), tais autores, apresentam, inicialmente, a seguinte decomposição.

$$T^2 = T^2_{p-1} + T^2_{p1,2..p-1} \cdot \qquad (3.1)$$

Onde:

$T^2_{p-1} = (\mathbf{X}_k^{(p-1)} - \bar{\mathbf{X}}^{(p-1)})' \mathbf{S}^{-1(p-1)} (\mathbf{X}_k^{(p-1)} - \bar{\mathbf{X}}^{(p-1)})$, é a estatística T^2 de Hotelling em relação às *p-1* primeiras variáveis, representando a distância padronizada em *p-1* dimensões, entre o *k-ésimo* vetor de observações $\mathbf{X}_k^{(p-1)}$ *(K=1,2...)* e o vetor $\bar{\mathbf{X}}^{(p-1)}$ das médias estimadas do processo e $\mathbf{S}^{(p-1)}$ representando a matriz de covariâncias estimada do processo, ambos ($\bar{\mathbf{X}}^{(p-1)}$ e $\mathbf{S}^{(p-1)}$) obtidos a partir de *m* amostras preliminares;

$$T^2_{p.1.2...p-1} = \frac{\mathbf{X}_{p,k} - \bar{\mathbf{X}}_{p.1,....p-1}}{\mathbf{S}_{p.1,...p-1}}, \qquad \text{é o } p\text{-ésimo componente do vetor } X_k,$$

ajustado pelas estimativas da média $\bar{X}_{p.1...p-1}$ e o desvio padrão $S_{p.1...p-1}$ da distribuição condicional de X_p dado $X_1, X_2, ...X_{p-1}$.

Tais estimativas são obtidas através de um modelo de regressão linear múltipla, usando *m* amostras preliminares, de X_p em função de $X_1, X_2...X_{p-1}$. Assim, o escore $T^2_{p.1,2...p-1}$ representa a distância padronizada unidimensional entre a observação $X_{p,k}$ e a média esperada dessa variável, dados os valores $\mathbf{X}_k^{(p-1)}$ obtidos nas *p-1* demais variáveis. De uma nova observação $\mathbf{x}_k'$ *(com k=m+1,...)* pode-se utilizar a estatística $T^2_{p1,2...p-1}$ para verificar se a relação entre a variável *p* e as demais *p-1* variáveis não se alterou, considerando a estrutura padrão de correlação entre elas.

Portanto, a equação (3.1) pode-se decompor em *p* partes independentes, obtendo-se o seguinte resultado:

$$T^2 = T^2_1 + T^2_{2.1} + T^2_{3.1,2} + T^2_{4.1,2,3} + \ldots + T^2_{i+1,....i} \cdot \qquad (3.2)$$

Mason, *et al* (1995) demonstraram que a estatística $T^2_{i+1,1..i}$, multiplicada por uma constante segue uma distribuição *F*, com graus de liberdade *1* e *(m-1)*. A significância α de qualquer um dos termos apresentados em (3.2) pode ser verificada comparando os valores obtidos com os seguintes limites:

$$LSC = \frac{(m+1)}{m} F_{\alpha(1,m-1)}.$$

$$LIC = 0. \qquad (3.3)$$

Desta forma, a estatística T^2 pode ser decomposta de diferentes maneiras. Se o valor T^2 apresentou um resultado significativo ($T^2 > LSC$), pode-se, por exemplo, verificar a relação de cada variável com as *p-1* variáveis restantes, obtendo-se os seguintes escores:

$$T^2_{1.2,\ldots p}, T^2_{2.1,3..,p}, T_{3.1,.2,4\ldots p}, T^2_{p,1\ldots,p-1}. \qquad (3.4)$$

onde a significância de cada termo pode ser verificada comparando estes valores com os limites de controle na expressão (3.3). Portanto, deve-se adotar uma estratégia de decomposição considerando as características do processo em questão, e a determinação das razões que levaram esse processo a sair de controle sejam facilmente identificadas. Segundo Mason *et al* (1995) os principais procedimentos propostos na literatura são casos especiais do método de partição apresentado acima. Maiores detalhes sobre a decomposição proposta por Mason *et al* serão apresentados na seção 3.4.

3.3.2. Outros Métodos de Decomposição

Hawkins (1993) também apresenta três diferentes estratégias para a decomposição de T^2. Duas dessas, destacam-se e relacionam-se com as decomposições propostas por Mason *et al.* (1995) nas equações (3.2) e (3.4). A primeira estratégia baseia-se na seguinte decomposição:

$$T^2 = \sum_{i=1}^{p} (X_i - \bar{X}_i) Z_i \ . \qquad (3.5)$$

onde Z_i é o resíduo padronizado obtido através da regressão linear múltipla da *i-ésima* variável sobre *p-1* demais variáveis no vetor **X** (X_i em função de $X_1, X_2, X_3, \ldots X_{i-1}, X_{i+1}, \ldots X_p$) obtido a partir de *m* amostras preliminares. T^2 é expressa como uma soma de produtos entre Z_i e o correspondente desvio de X_i em relação a sua média $\bar{X}_i$. A uma nova observação $\mathbf{x}_k$ (*com k=m+1,....),* Hawkins (1993) sugere uma estratégia de diagnóstico utilizando *p* gráficos de controle, uma para cada $Z_{i,k}$.

Se $T_k^2 > LSC$ e, se para alguma variável *i*, o valor de $Z_{i,k}$ ultrapassar certos limites de controle pode-se concluir que a relação desta variável com as demais *p-1* não está em harmonia com a estrutura padrão de correlação que existe entre elas. Mason *et al* (1995) mostram que a estatística Z_i pode ser escrita como:

$$Z_i = \frac{1}{S_{i.1,\ldots i-1,i+1,\ldots p}} T^2_{i.1,\ldots i-1,i+1,\ldots p} \ . \qquad (3.6)$$

Assim, Z_i, na expressão acima, pode ser obtido através de um produto entre a estatística T^2, utilizando a decomposição da equação (3.2) e o inverso do desvio-padrão condicional da *i-ésima* variável, dados $X_1 \ldots X_{i-1}, X_{i+1}, \ldots X_p$.

Já a segunda estratégia proposta por Hawkins (1993) baseia-se na seguinte decomposição:

$$T^2 = \sum_{i=1}^{p} Y_i^2 \quad . \tag{3.7}$$

onde Y_i, é o resíduo padronizado, obtido através da regressão múltipla da *i-ésima* variável em função das *i-1* variáveis anteriores no vetor $\boldsymbol{X}$ (X_i em função $X_1, X_2, \ldots Y_{i-1}$) a partir de *m* amostras preliminares. Uma estratégia sugerida por Hawkins (1993) para diagnosticar seria a utilização de *p* gráficos de controle, uma para cada Y_i. Logo, para uma nova observação $\mathbf{x}_k$ (com *k= m+1,...*) se $T_k^2 > LSC$ e, se alguma variável *i*, o valor $Y_{i,k}$, ultrapassar certos limites, logo se pode concluir que a relação desta variável com as primeiras *i-1* variáveis não está condizente com a estrutura padrão de correlação entre elas.

Wade & Woodall (1993), Lowry & Montgomery (1995) sugerem um procedimento de diagnóstico denominado *cause-selecting chart* (gráfico de controle para detecção de causas). Tais autores defendem a idéia da construção de um gráfico de controle para monitorar uma variável X_i, do vetor **X** fazendo o ajuste das observações via regressão. Tal estratégia é semelhante às propostas por Hawkins (1993).

A diferença que existe é que o modelo de regressão é proposto para monitorar a média de X_i considerando o efeito de alguma covariável preditora, ou seja, não faz parte do vetor **X**. Os autores Wade & Woodall (1993) fazem uma revisão propondo alguns ajustes..

Wierda, apud Mason *et al* (1997) apresenta uma estratégia para a decomposição de T^2 denominada *Step-down*. Tal procedimento está baseado numa ordenação *a priori* das médias *p* variáveis. Conforme a ordenação, o procedimento particiona o vetor de médias em *q (<p)* subvetores e define sub-hipóteses, isto é, cada sub-hipótese testa a possibilidade de um determinado sub-vetor de médias não ter sofrido alterações. Portanto, o gráfico de controle é associado a cada uma das *q* sub-hipóteses. A estatística de controle utilizada no *i-ésimo* gráfico é

$$F_i = \frac{T_i^2 - T_{i-1}^2}{1 + T_{i-1}^2/(m-1)} \quad . \tag{3.8}$$

onde T_i^2 e T_{i-1}^2, representam a estatística T^2 em relação às i e *i-1* primeiras variáveis do sub-vetor escolhido.

Utilizando a equação (3.2), pode-se escrever:

$$F_i = \frac{1}{1+T_{i-1}^2/(m-1)} T_{i.1,\ldots p}^2 . \qquad (3.9)$$

Pode-se verificar que o procedimento acima é semelhante à segunda estratégia proposta por Hawkins (1993).

Murphy (1987) apresenta as *p* variáveis particionadas em dois grupos, onde um deles representa as variáveis suspeitas de estarem causando o ponto fora dos limites de controle. Tal estatística de controle é dada pela equação:

$$D_i = T^2 - T_i^2 . \qquad (3.10)$$

onde T_i^2 representa a estatística T^2 considerando as *i* primeiras variáveis do sub-grupo suspeito. Neste método não existe uma ordenação a *priori* das médias das *p* variáveis, ao contrário do método proposto por Mason *et al* (1997). Utilizando as equações (3.1) e (3.2), Mason *et al* (1995) mostraram a estatística D_i expressa da seguinte forma:

$$D_i = \sum_{j=i+1}^{p} T_{j.1,2\ldots j-1}^2 . \qquad (3.11)$$

Doganaksoy *et al* (1991), sugere uma estratégia de decomposição denominada ranqueamento univariado. Esta estratégia consiste em particionar a estatística T^2 em *p* termos incondicionais; ou seja; o valor da estatística de Hotelling é obtido através de cada variável individualmente. Assim, temos *p* gráficos univariados de Hotelling. Tal estratégia é semelhante ao procedimento de Alt (1985) que representa um caso particular em relação aos critérios de decomposição sugeridos por Mason *et al* (1995).

Jackson (1980), a exemplo de Alt (1985), propõe o uso do gráfico de Hotelling, para verificar mudanças no processo, juntamente com os gráficos univariados, para diagnosticar o problema. A estratégia de Jackson (1980) se diferencia do procedimento de Alt (1985), pois o mesmo utiliza os gráficos univariados para monitorar *p* Componentes Principais ao contrário das *p* variáveis originais. As projeções das variáveis originais de controle nas direções de máxima variabilidade dos dados são representados pelos componentes principais (CP's). A estatística T^2 pode ser escrita em função dos CP's, segundo Jackson.

Logo, a equação (2.21) pode ser vista da através da seguinte expressão:

$$T^2 = \sum_{i=1}^{p} \frac{C_i^2}{\lambda_i}. \qquad (3.12)$$

onde C_i, para i= 1, 2,...p, representa o *i- ésima* componente principal (*CP*) e λ_i a sua variância.

Portanto, alguns *CP's* por serem independentes, refletem um conjunto específico de eventos com causas especiais de variação no processo. Desta forma, se existir uma causa especial, um dos *CP's* deve apresentar um desvio em relação a seu padrão. Como cada *CP* é uma combinação linear das *p* variáveis originais, pode-se determinar as variáveis que mais contribuem para sua formação, assim o grupo de variáveis responsáveis pelo ponto fora de controle plotado pelo gráfico de Hotelling pode ser encontrado.

Inicialmente, a idéia de diagnosticar o processo através dos *CP's* foi criticado, pois muitas vezes não possuem a interpretação física. Miller *et al* (1993) sugerem um método para determinar o efeito exercido por cada variável em um *CP*, aprimorando a idéia original de Jackson (1991) que propõe o uso de gráficos de controle univariados para *CP*, utilizando os limites de controle de *Bonferroni*. Tais limites resolvem o problema do erro tipo I, apresentado no procedimento de Alt (1985), pois há independência entre componentes principais.

Na próxima seção será elaborado um procedimento computacional que demonstrará os passos para a decomposição da estatística T^2 de Hotelling.

3.4. Passos para Operacionalizar o Gráfico de Controle T^2 de Hotelling

A distância T^2 Hotelling é uma constante multiplicada por uma forma quadrática. Para um grupo de *m* variáveis e *p* observações, esta forma quadrática pode ser obtida através dos seguintes passos:

- ✓ Inicia-se a operacionalização pela matriz de dados X_{mxp} (m x p observações);
- ✓ Define-se o vetor alvo X,
- ✓ Calcula-se a matriz de covariância **S** e sua inversa $\mathbf{S}^{-1}$, conforme a expressão (2.15) vista no Capítulo 2.
- ✓ Calcula-se a matriz de desvios $(\mathrm{X}-\overline{\mathrm{X}})$ e sua transposta $(\mathrm{X}-\overline{\mathrm{X}})'$.

Assim, a forma quadrática, para o caso de observações com subgrupos, pode ser calculada através da expressão (2.15), ou seja, o valor de cada T^2 para cada *m* amostra é calculado utilizando a expressão (2.15) e é comparada com o limite de controle superior da expressão (2.20). Se o valor de T^2 para a *j-ésima* amostra, ou seja, T_j^2 for acima do limite de controle superior, temos um ponto fora dos limites de controle, que deverá ser investigado.

Para calcular o valor de T_j^2, o vetor de médias das amostras é dado por

$$\overline{\mathrm{X}}_{\mathrm{j}} = \begin{bmatrix} \overline{\mathrm{X}}_{1\mathrm{j}} \\ \overline{\mathrm{X}}_{2\mathrm{j}} \\ \cdots \\ \overline{\mathrm{X}}_{\mathrm{pj}} \end{bmatrix}, \quad \mathrm{j}=1,2,\ldots\mathrm{m}$$

onde $\overline{\mathrm{X}}_{\mathrm{ij}}$ representa a média da amostra da *i-ésima* característica para a *j-ésima* amostra, ou seja:

$$\overline{X}_{ij} = \frac{\sum_{k=1}^{n} X_{ijk}}{n} \qquad i = 1,2,...,p \quad e \quad j = 1,2,...m \tag{3.13}$$

Onde:

X_{ijk} representa o valor da *k-ésima* observação da *i-ésima* característica da *j-ésima* amostra.

A variância da amostra para a *i-ésima* característica na *j-ésima* amostra é dada por:

$$S_{ij}^2 = \frac{1}{n-1}\sum(\mathbf{X}_{ijk} - \overline{\mathbf{X}}_{ij})^2 \qquad i = 1,2,...,p \quad e \quad j = 1,2,...,m \tag{3.14}$$

A covariância da característica *i* e a característica *h* na *j-ésima* amostra é calculado da seguinte forma

$$S_{ihj} = \frac{1}{n-1}\sum_{k=1}^{n}(\mathbf{X}_{ijk} - \overline{\mathbf{X}}_{ij})(\mathbf{X}_{hjk} - \overline{\mathbf{X}}_{hj}) \qquad i = 1,2,...,m \quad e \quad i \neq h \tag{3.15}$$

O vetor $\overline{\overline{\mathbf{X}}}$ é a média de cada característica para *m* amostra, ou seja :

$$\overline{\overline{\mathbf{X}}}_i = \frac{\sum_{j=1}^{m} \overline{\mathbf{X}}_{ij}}{m} \qquad i = 1,2,...,p. \tag{3.16}$$

A matriz de variância e covariância ***S*** da expressão (2.15) é estimado a partir das médias das *m* amostras, apresentadas nas expressões seguintes:

$$S_i^2 = \frac{\sum_{j=1}^{m} S_{ij}^2}{m} \qquad i = 1,2,...,p. \tag{3.17}$$

$$S_{ih} = \frac{\sum_{j=1}^{m} S_{ihj}}{m} \qquad i \neq h \cdot \tag{3.18}$$

Finalmente, o vetor $\overline{\overline{\mathbf{X}}}$ é estimado utilizando os elementos $\left\{\overline{\overline{X}}_i\right\}$ e a matriz **S** é estimado a seguir

$$\mathbf{S} = \begin{bmatrix} S_1^2 & S_{12} & \cdots & S_{1p} \\ & S_2^2 & \cdots & S_{2p} \\ & & & \vdots \\ & & & S_p^2 \end{bmatrix}. \tag{3.19}$$

A utilização da equação (2.15), requer a inversão da matriz acima.

3.4.1. Apresentação do Aplicativo em Fluxograma

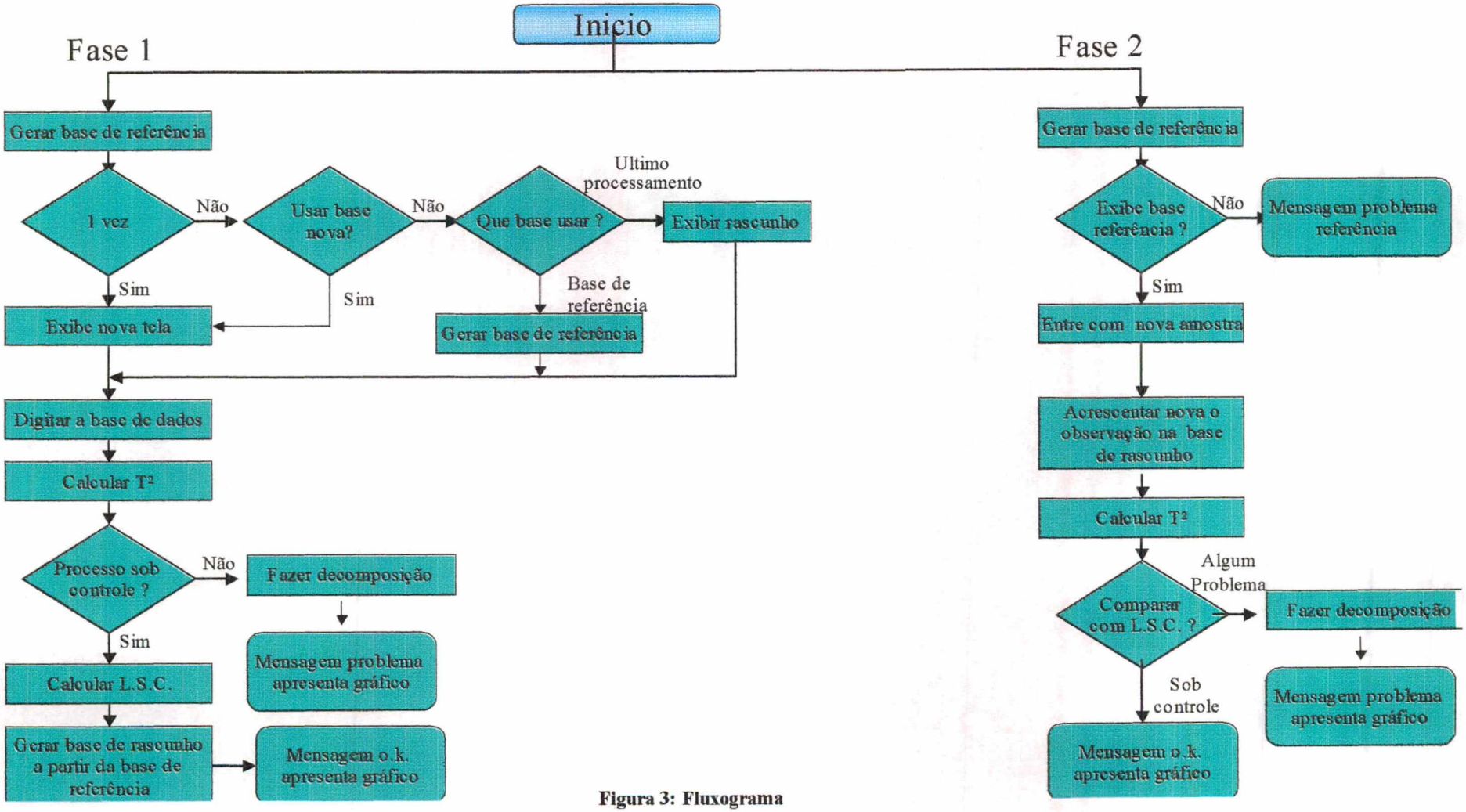


Figura 3: Fluxograma

3.4.2. Descrição do Fluxograma

1. Inicialmente é exibida a Figura 4, referente ao inicio do aplicativo (planilha INSTRUÇÕES). O usuário pode selecionar entre Gerar base de referência ou Monitoramento.

2. Gerar base de referência:

2.1 Quando for a primeira vez que o usuário estiver utilizando o aplicativo, ele deverá selecionar entre uma base nova ou dados do exemplo (vide Anexo 1). Se já existir uma base de referência o usuário poderá selecionar entre rascunho, referência, nova e dados de exemplo.

2.2 Na planilha DADOS o usuário informará os dados que irá calcular T^2

2.3 Exibe-se uma tela na qual são informados o intervalo de dados, quantidade de características, tamanho do subgrupo, e nível de significância;

2.4 Em seguida será calculado: a média; variância; covariância, a estatística T^2 *de Hotelling,* e o LCS (limite de controle superior).

2.5 Será criada uma nova planilha com o gráfico e valores do processamento, se algum dos valores da estatística T^2 *de Hotelling* estiverem acima do LCS, este será destacado em vermelho e aparecerá um botão para processar a decomposição dos valores fora dos limites de controle.

2.6 Caso o usuário esteja satisfeito com os valores calculados poderá gerar a base de referência, ou se for o caso, poderá efetuar o processo de decomposição.

2.7 Se optar pela geração da base de referência, será calculado o LCS para a segunda fase (processo em controle) e geradas as bases de referência e a base de rascunho, que são as bases para o futuro monitoramento.

3. Monitoramento

3.1 Para realizar a etapa de "Monitoramento" é necessário que o usuário tenha a sua "base de referência":

3.2 Quando existir a base de referência, o usuário digitará uma nova observação que será acrescentada no processo;

3.3 Tomam-se os valores que foram utilizados na base de referência e calcula-se apenas a estatística da observação que foi acrescentada no processo;

3.4 Será criada uma nova planilha com o gráfico e valores do processamento, se algum dos valores da estatística T^2 *de Hotelling* estiverem acima do *LCS*, este será destacado em vermelho e aparecerá um botão para processar a decomposição dos valores fora dos limites de controle.

3.4.3. Como Identificar as Variáveis que Estão Interferindo na Estabilidade do Processo

Conforme descrito na seção 3.3.2, existem diversos métodos que auxiliam na identificação da(s) variável(eis) que interfere(m) na estabilidade do processo. A escolha de um método em relação a outro é uma decisão que leva em consideração a experiência do pesquisador, bem como a facilidade de operacionalização do mesmo, seja através de um algoritmo computacional, ou através de cálculos manuais.

Além disso, os outros métodos existentes na literatura sobre a decomposição da estatística T^2 de Hotelling, não possuem explicações suficientes para a completa compreensão dos métodos sugeridos. A escassez de material bibliográfico também contribuiu para o não aprofundamento nos outros métodos de decomposição já descritos brevemente no capitulo 3.

Assim sendo, a escolha do método de decomposição proposto por Mason, Tracy, Young (1995) para o presente trabalho levou em consideração as explicações descritas acima. Logo, as etapas realizadas no desenvolvimento do algoritmo computacional podem ser vistas a seguir:

1- Comparar a estatística T^2 global com o valor da estatística F. Ao comparar o valor da decomposição das variáveis com a estatística F deve-se observar o teste de hipótese para saber se a variável deve ou não ser investigada, ou seja, se $T^2 > F$, rejeita-se H_0, caso contrário, se $F > T^2$ não se pode rejeitar H_0 ou seja, quando $T^2 > F$.
2- Para calcular T^2 utiliza-se a expressão de decomposição (3.20) e compara-se as p-1 variáveis restantes com a estatística F para verificar a significância de cada uma.

$$T^2 = T_1^2 + T_{2.1}^2 + T_{3.1,2}^2 + T_{4.1,2,3}^2 + \ldots + T_{i+1,1\ldots i}^2$$

$$T^2 = T_1^2 + \sum_{j=1}^{p-1} T_{j+1.1,\ldots j} \quad . \tag{3.20}$$

Onde:

T^2: é o valor da estatística T^2 global de uma determinada observação

T_1^2: é a estatística da primeira variável, que é o quadrado da variável reduzida para a estatística *t* univariada da variável inicial;

$T_{2.1}^2$: é a estatística da segunda variável combinada com a primeira variável;

$T_{3.1,2}^2$: é a estatística da terceira variável, que combinada com a primeira e segunda variável.

3-Para se chegar ao valor do T^2 global é necessário calcular o valor de T_1^2 e todas as possíveis combinações com as *p* variáveis restantes. Assim o valor de T_1^2 pode ser calculado da seguinte expressão:

$$T_1^2 = \frac{(\mathbf{X}_{i1} - \overline{\mathbf{X}}_1)^2}{\mathbf{S}_1^2} \quad . \tag{3.21}$$

Onde:

$\mathbf{X}_{i1}$: é o valor da *i-ésima* nova observação que foi retirada do processo;

$\overline{\mathbf{X}}_1$: é o vetor de médias da primeira variável (*p*);

$\mathbf{S}_1^2$: é o valor da variância da primeira variável.

Conhecendo-se a expressão para efetuar o cálculo do valor de T_1^2 pode-se calcular as demais componentes da decomposição, conforme a fórmula a seguir:

$$T_{p.1\ldots,p-1} = \frac{\mathbf{X}_{ip} - \overline{\mathbf{X}}_{p.1,\ldots p-1}}{\mathbf{S}_{p.1,\ldots p-1}} \quad . \tag{3.22}$$

Onde: $\overline{X}_{p.1}\ldots, p_1 = \overline{X}_p + b_p'(X_i^{(p-1)} - \overline{X}^{(p-1)})$,

$\overline{\mathbf{X}}_p$ = é a média amostral das *m* observações das *p* variáveis;

$\mathbf{X}_i^{(p-1)}$ = é o vetor das medidas individuais no período *i*, excluindo as variáveis *p*;

$\overline{\mathbf{X}}^{(p-1)}$ = é o vetor das médias amostrais das *m* observações para as primeiras *(p-1)* variáveis;

$$\mathbf{b}_p' = \mathbf{S}_{XX}^{-1}\,\mathbf{S}_{xX} \text{ para } \mathbf{S} = \begin{bmatrix} \mathbf{S}_{XX} & \mathbf{S}_{xX} \\ \mathbf{S}'_{xX} & \mathbf{S}_{x^2} \end{bmatrix},$$

Onde:

$\mathbf{S}_{XX}^{-1}$: é a matriz inversa da variância, referente a primeira variável;

$\mathbf{S}_{xX}$: é o valor da covariância da primeira e segunda variável.

Em seguida calcula-se a variância:

$$S_{2.1}^2 = S_x^2 - S'_{xX} S_{XX}^{-1} S_{xX} \tag{3.23}$$

Onde:

$\mathbf{S}_x^2$: é a matriz de variância da segunda variável

$\mathbf{S}'_{xX}$: é a matriz transposta da covariância da primeira e segunda variável.

Logo, o valor da segunda observação $\mathbf{X}_i$ que foi retirada do processo é subtraído da média $\overline{\mathbf{X}}_p$ e é dividida pela variância de $\mathbf{S}_{2.1}^2$, retornando-se assim, para a mesma expressão dada em (3.21), com a variável $T_{2.1}^2$, calculada através da seguinte expressão:

$$T_{2.1}^2 = \frac{(\mathbf{X}_{i1} - \overline{\mathbf{X}}_{2.1})^2}{\mathbf{S}_{12}^2} \,. \tag{3.24}$$

Para calcular o valor de $T^2_{3.1,2}$, a partir das equações (3.23) e (3.24) , deve-se observar atentamente a representação da matriz para o calculo de $\mathbf{b}_p' = \mathbf{S}_{XX}^{-1}\,\mathbf{S}_{xX}$ para $\mathbf{S} = \begin{bmatrix} \mathbf{S}_{XX} & \mathbf{S}_{xX} \\ \mathbf{S}'_{xX} & \mathbf{S}_{x^2} \end{bmatrix}$, para mais de duas variáveis, é a seguinte:

$$\mathbf{S} = \begin{bmatrix} x_{11} & x_{21} & \vdots & x_{13} \\ x_{21} & x_{22} & \vdots & x_{23} \\ \cdots & \cdots & \vdots & \cdots \\ x_{31} & x_{32} & \vdots & x_{33} \end{bmatrix}. \qquad (3.25)$$

Onde

$S_{XX}^{-1} = \begin{bmatrix} x_{11} & x_{21} \\ x_{21} & x_{22} \end{bmatrix}$ e

$S_{xX} = [x_{31} \quad x_{32}] = S'_{xX}$, ou seja a transposta de S_{xX}

$S_{X^2} = [x_{33}]$

Esse passo deverá ser desenvolvido sucessivamente conforme o número de variáveis que estejam envolvidas no processo e todas as suas possíveis combinações, ou seja:

$$T^2 = T_1^2 + T_{2.1}^2 + T_{3.1,2}^2 + T_{4.1,2,3}^2 + \ldots + T_{i+1,1,\ldots,i}^2, \qquad (3.26)$$

$$\text{logo } T^2 = T_1^2 + \sum_{j=1}^{p-1} T_{j+1.1,\ldots,j}\,.$$

4- Outra maneira de encontrar as variáveis que estão interferindo no processo, seria calcular as estatísticas T^2 das p variáveis sem as combinações, ou seja $T_1^2, T_2^2, T_3^2 \ldots T_p^2$, e compará-las com a estatística F, para verificar a significância das mesmas.

5- A utilização do segundo passo, descrito acima, muitas vezes apresentará vários T^2 condicionais significantes, dificultando a interpretação do pesquisador. Assim, uma maneira rápida de identificar qual das variáveis está interferindo no processo, seria calcular o T^2 global de uma dada observação, e subtraí-lo do valor de T^2 de cada variável, por exemplo:

$$T^2 - T_1^2, T^2 - T_2^2, \ldots, T^2 - T_p^2 \ . \tag{3.27}$$

Desta forma, se qualquer destas diferenças forem significantes, seguindo uma distribuição F, elas indicam que os valores condicionais restantes devem ser calculados e examinados, seguindo todas as indicações dadas a partir da segunda etapa.

3.4.4. *Apresentação do Sistema*

Nessa seção, apresenta-se o Tutorial do Sistema " *DEC_Hotelling"*, desenvolvido na linguagem de programação Visual Basic com a planilha de cálculo. O sistema decompõe as variáveis do gráfico de controle T^2 *de Hotelling,* que interferem na estabilidade do processo.

Ao executar o aplicativo *"DEC_Hotelling"* disponibiliza-se a tela inicial contendo um resumo da estatística T^2 *de Hotelling,* e sua utilidade. O sistema possui dois processos executados pela página inicial através dos botões "Gerar Base de Referência" e "Monitoramento". O botão "Gerar Base de Referência" possibilita ao usuário acessar uma janela com varias opções, para gerar a base de referência. O botão de " Monitoramento" serve para monitoramento de futuras amostras, quando o processo se apresentar estável, a partir da "Gerar Base de Referência".

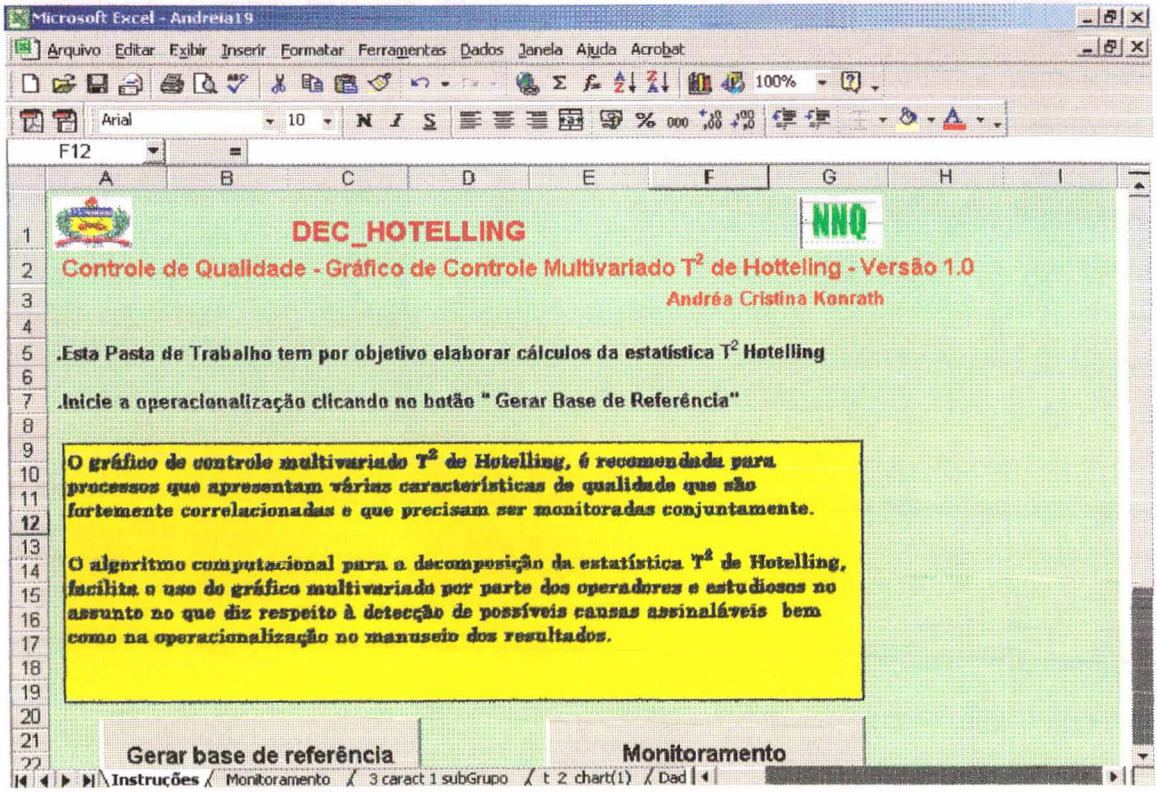


Figura 4: Tela Inicial do Programa

A Figura 5, evidencia a tela a qual o usuário, ao clicar no botão "Gerar base de referência" aparecerá uma janela na qual ele poderá optar quanto a utilização dos dados. Esta tela disponibilizará várias opções para a operacionalização do sistema de acordo com o objetivo do usuário, como: usar dados da base de rascunho (que já constam no sistema são acrescidos das novas observações monitoradas) dados da base de referência (dados utilizados na geração da base de referência atual), uma base nova de dados; ou usar a base de dados do exemplo.

Uma vez escolhido os dados para análise, clicando-se no botão "Processar" e apresenta-se a tela de preenchimento dos dados (figura 6).

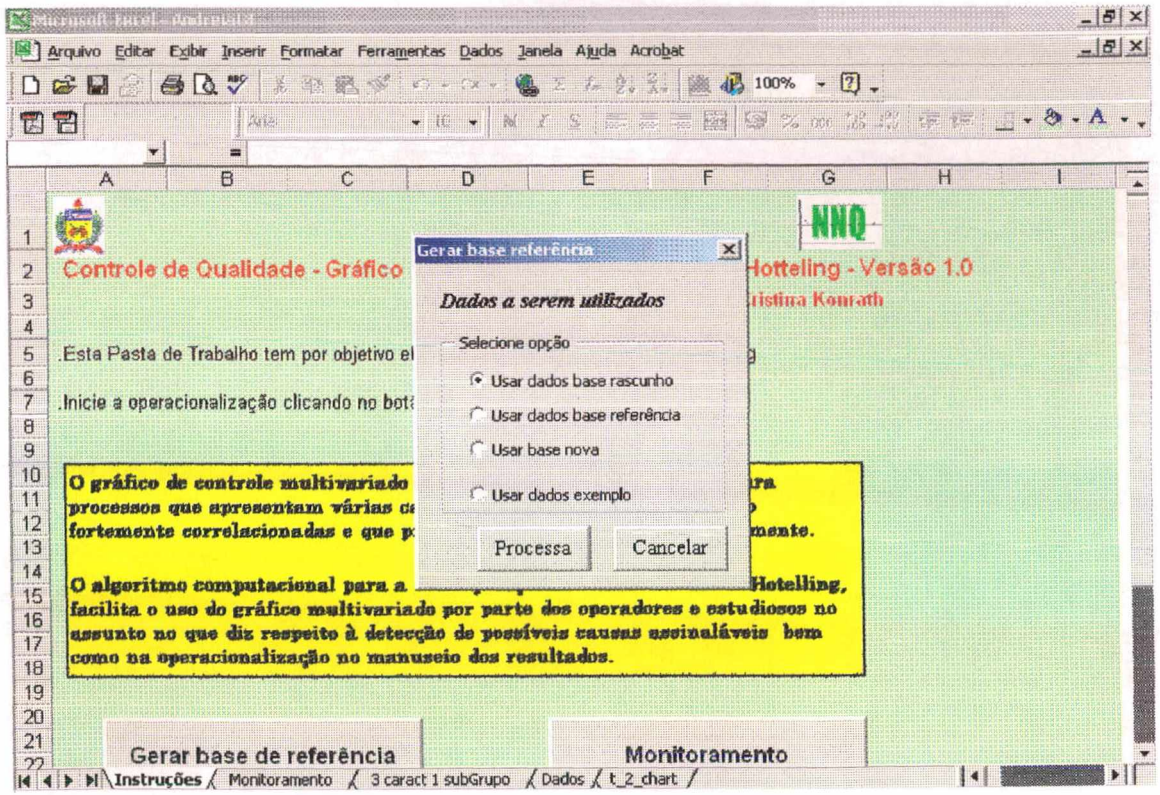


Figura 5: Tela para Iniciar a Operacionalização

Nessa tela, apresentada na Figura 6, será introduzida a base de dados. As observações devem ser acrescentadas seqüencialmente nas linhas. Enquanto que as variáveis (características da qualidade) e subgrupos, nas colunas.

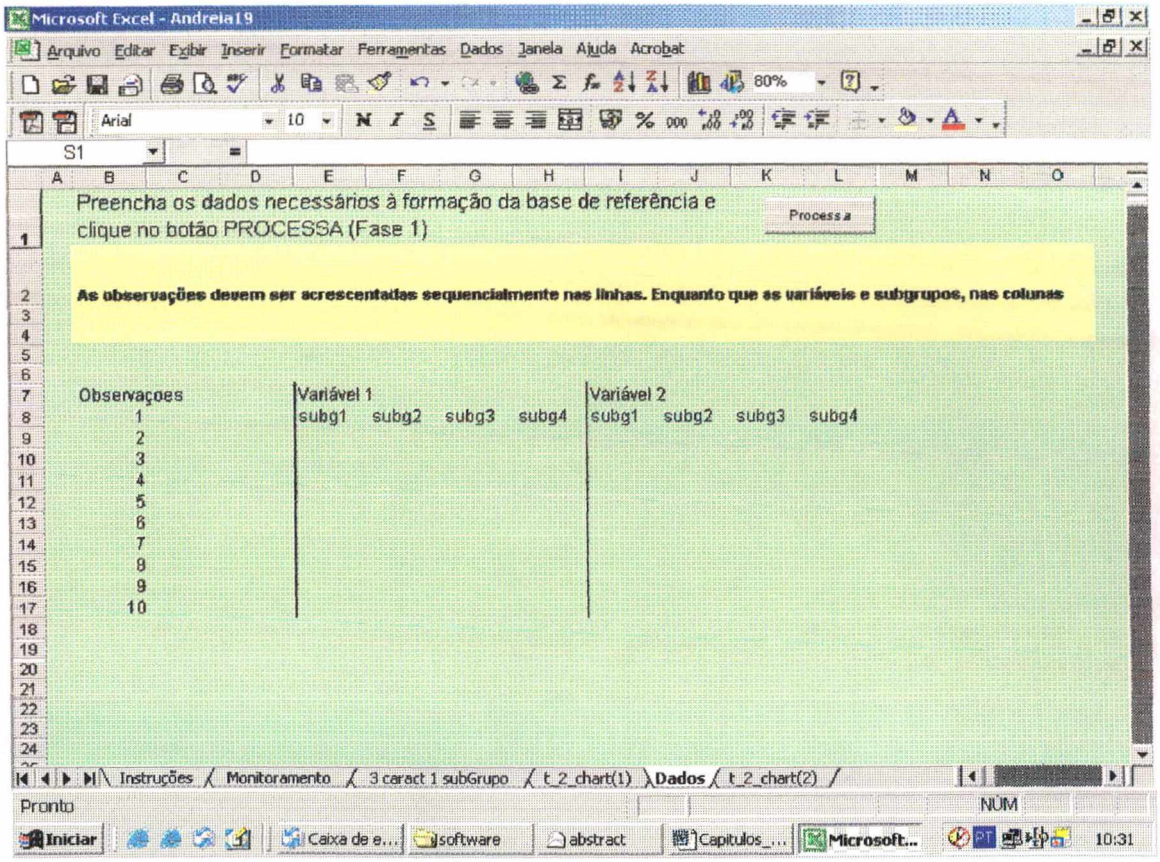
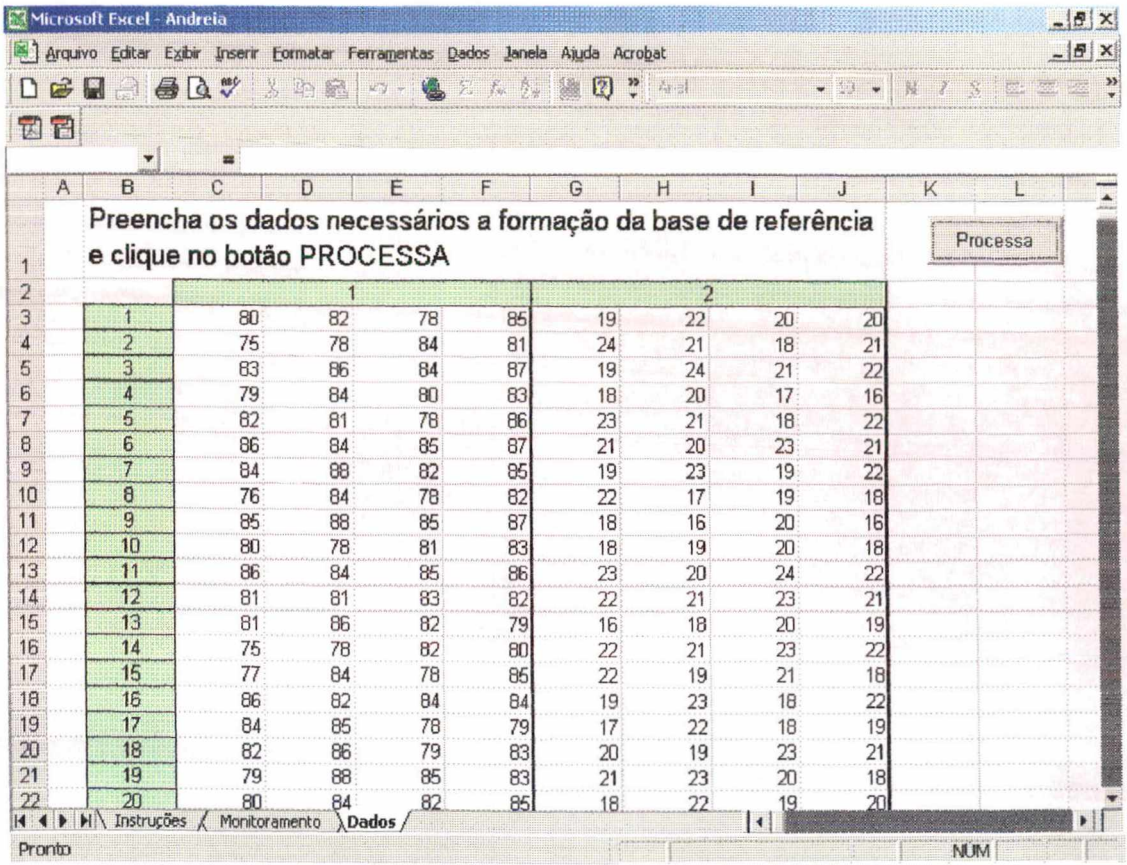


Figura 6: Tela para Digitar os Dados

A Figura 7 apresenta um exemplo contendo um banco de dados retirado do livro de Mitra (1993). Nota-se que no exemplo abaixo, tem-se duas características de qualidade; tamanho de subgrupo igual a quatro e o tamanho amostra igual a vinte.



		1				2			
1	1	80	82	78	85	19	22	20	20
2	2	75	78	84	81	24	21	18	21
3	3	83	86	84	87	19	24	21	22
4	4	79	84	80	83	18	20	17	16
5	5	82	81	78	86	23	21	18	22
6	6	86	84	85	87	21	20	23	21
7	7	84	88	82	85	19	23	19	22
8	8	76	84	78	82	22	17	19	18
9	9	85	88	85	87	18	16	20	16
10	10	80	78	81	83	18	19	20	18
11	11	86	84	85	86	23	20	24	22
12	12	81	81	83	82	22	21	23	21
13	13	81	86	82	79	16	18	20	19
14	14	75	78	82	80	22	21	23	22
15	15	77	84	78	85	22	19	21	18
16	16	86	82	84	84	19	23	18	22
17	17	84	85	78	79	17	22	18	19
18	18	82	86	79	83	20	19	23	21
19	19	79	88	85	83	21	23	20	18
20	20	80	84	82	85	18	22	19	20

Figura 7: Tela para Processar os Dados

Ao clicar no botão "Processa", será apresentada uma janela (Figura 8), na qual o usuário irá selecionar o intervalo onde se encontram os dados; o número de características envolvidas e o tamanho do subgrupo. O usuário também poderá escolher o nível de significância para calcular os limites de controle.

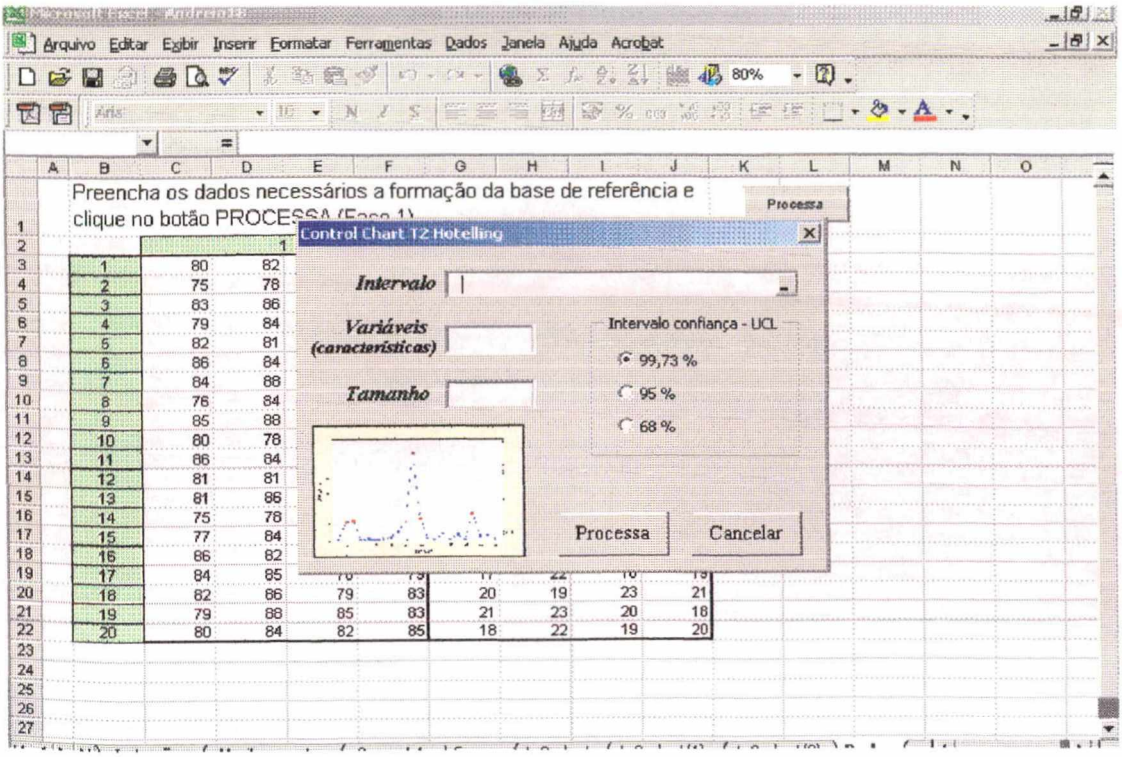
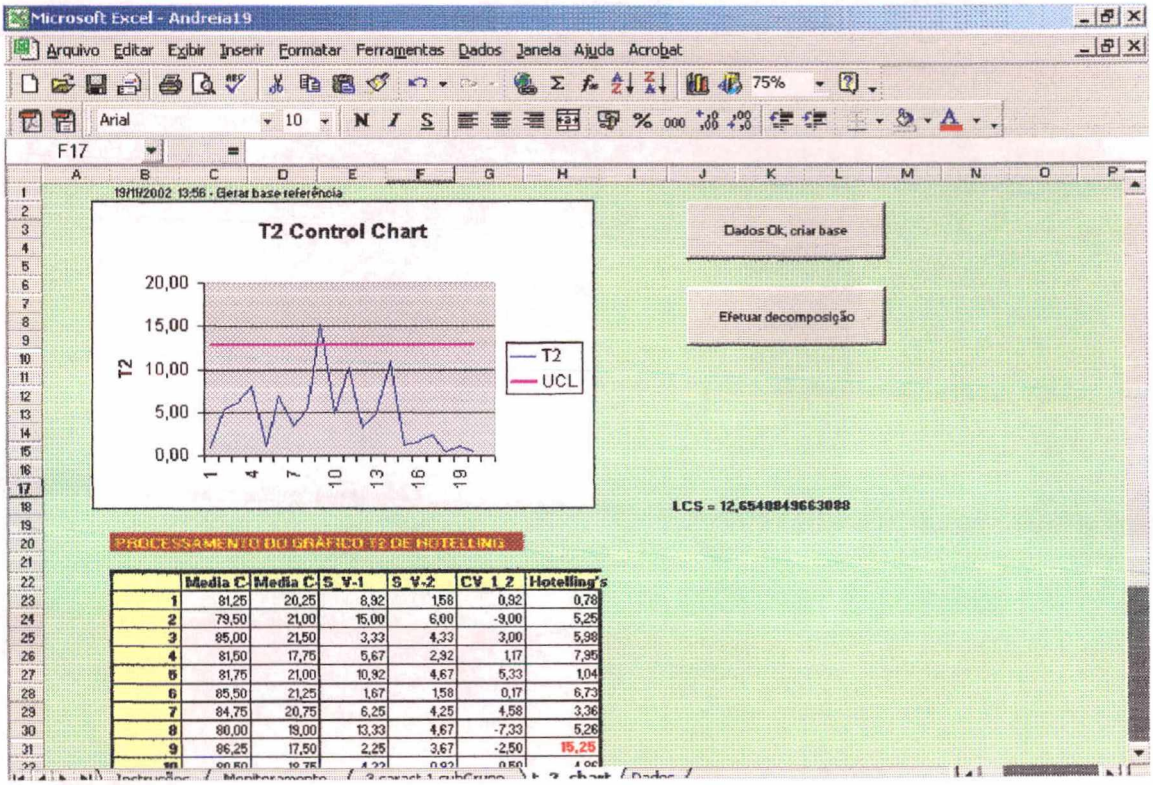


Figura 8:Tela para Iniciar a Operacionalização

Após executar o botão "Processa" (Figura 8), será apresentado o cálculo das médias, variância, covariância e a estatística do gráfico de controle multivariado T^2 *de Hotelling,* e seu respectivo gráfico, juntamente com um botão que possibilita ao usuário gerar a base de referência a partir dos dados processados (Figura 9). Caso o gráfico de controle multivariado T^2 *de Hotelling* acusar algum ponto fora do limite de controle será possível fazer a decomposição.

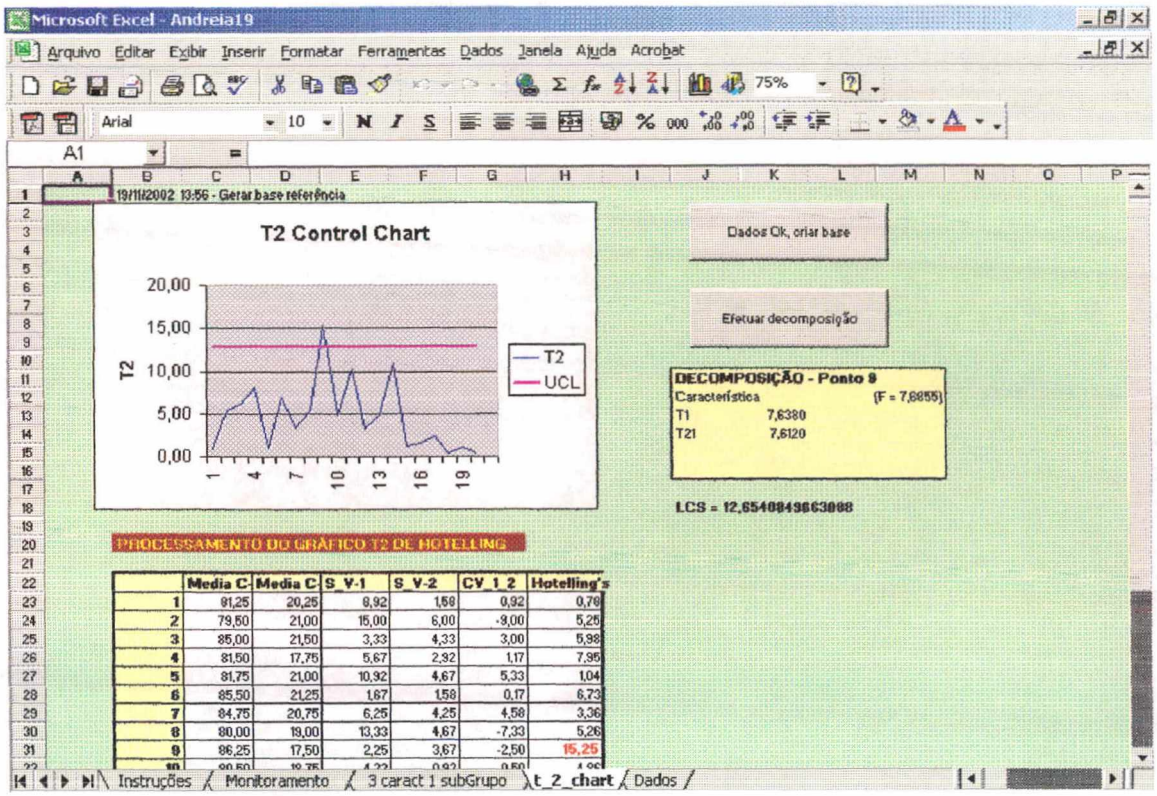
Havendo a necessidade de efetuar a decomposição, deve-se clicar no botão "Efetuar decomposição".



	Media C-1	Media C-2	S_V-1	S_V-2	CV_1_2	Hotelling's
1	81,25	20,25	8,92	1,58	0,92	0,78
2	79,50	21,00	15,00	6,00	-9,00	5,25
3	85,00	21,50	3,33	4,33	3,00	5,98
4	81,50	17,75	5,67	2,92	1,17	7,95
5	81,75	21,00	10,92	4,67	5,33	1,04
6	85,50	21,25	1,67	1,58	0,17	6,73
7	84,75	20,75	6,25	4,25	4,58	3,36
8	80,00	19,00	13,33	4,67	-7,33	5,26
9	86,25	17,50	2,25	3,67	-2,50	15,25

Figura 9: Tela com os Resultados

Conforme visto na Figura 9, a observação número 9, está fora do limite superior de controle e portanto, ela deve ser investigada. A Figura 10 apresenta a decomposição da estatística T^2 *de Hotelling,* após o usuário ter clicado no botão "Efetuar decomposição", disponibiliza-se um quadro com os valores da contribuição de cada característica e o valor da estatística F[1].



	Media C-1	Media C-2	S_V-1	S_V-2	CV_1_2	Hotelling's
1	81,25	20,25	8,92	1,58	0,92	0,70
2	79,50	21,00	15,00	6,00	-9,00	5,25
3	85,00	21,50	3,33	4,33	3,00	5,98
4	81,50	17,75	5,67	2,92	1,17	7,95
5	81,75	21,00	10,92	4,67	5,33	1,04
6	85,50	21,25	1,67	1,58	0,17	6,73
7	84,75	20,75	6,25	4,25	4,58	3,36
8	80,00	19,00	13,33	4,67	-7,33	5,26
9	86,25	17,50	2,25	3,67	-2,50	15,25

Figura 10: Tela de Apresentação da Decomposição

Ao efetuar a decomposição das duas características de qualidade e comparar os valores com a estatística F, utiliza-se o seguinte teste de hipótese :

H_0: todas as variáveis (características de qualidade) possuem, em média, uma variabilidade uniforme;

H_1: todas as variáveis (características de qualidade) não possuem, em média, uma variabilidade uniforme.

---

[1] Caso a variável for significativa, ou seja, $T^2 > F$, a mesma será apresentada com o símbolo de "asterisco", alertando o operador.

Assim sendo, a variável utilizada no presente banco de dados não precisa ser investigada, já que as variabilidades apresentadas foram consideradas inerentes ao processo (causas comuns), conforme foi também ratificado pelo teste de hipótese.

Vale salientar que a importância da decomposição torna-se ainda maior à medida que os dados são testados no gráfico de controle de Shewhart (ver anexo 2 e 3), já que no caso do exemplo apresentado no pacote computacional nenhum ponto apresenta-se fora dos limites de controle, podendo assim, levar o operador a não investigar o processo para possíveis descobertas e correções.

Quando o processo estiver estável, o usuário deverá clicar no botão "Gerar base de referência", os dados serão armazenados numa base de referência e também numa base de rascunho, bem como será calculado o limite de controle para a fase 2.

O sistema retornará automaticamente para a planilha inicial. Ao clicar no botão "Monitoramento", o usuário poderá inserir novas amostras para controlar esta nova observação.

Na Figura 11 são apresentados os dados da base de rascunho e o usuário deverá acrescentar uma nova observação para monitoramento.

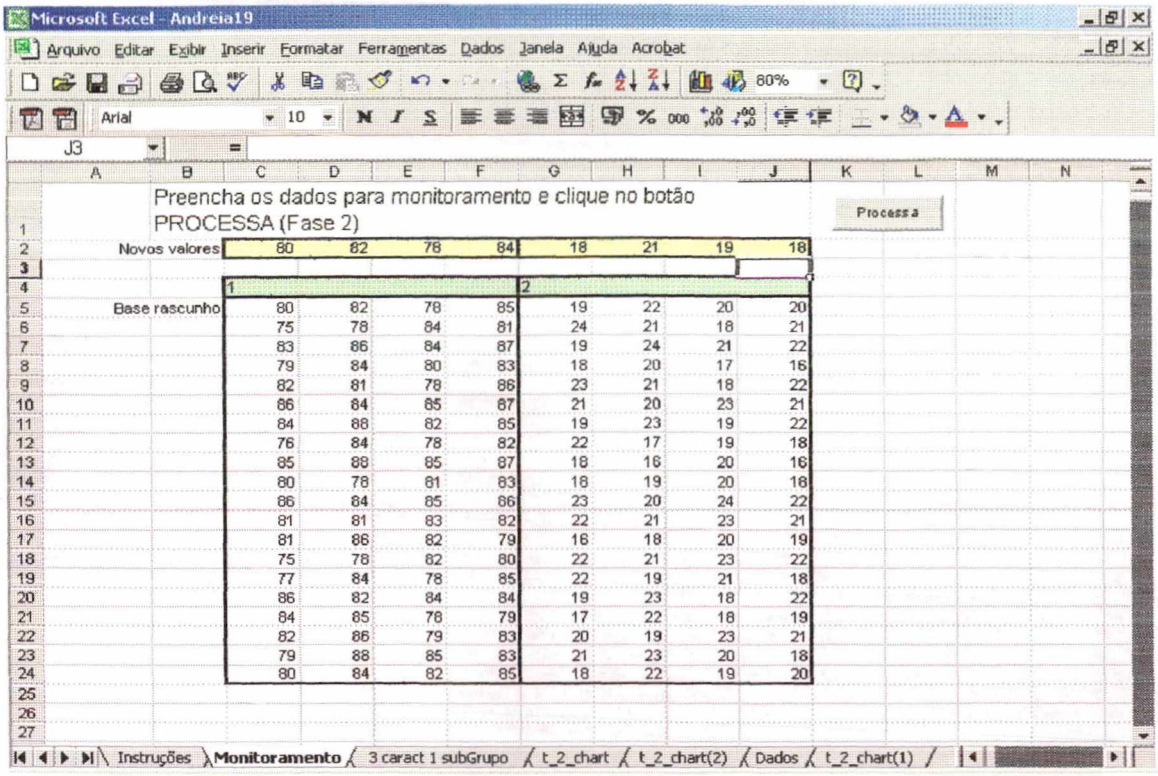
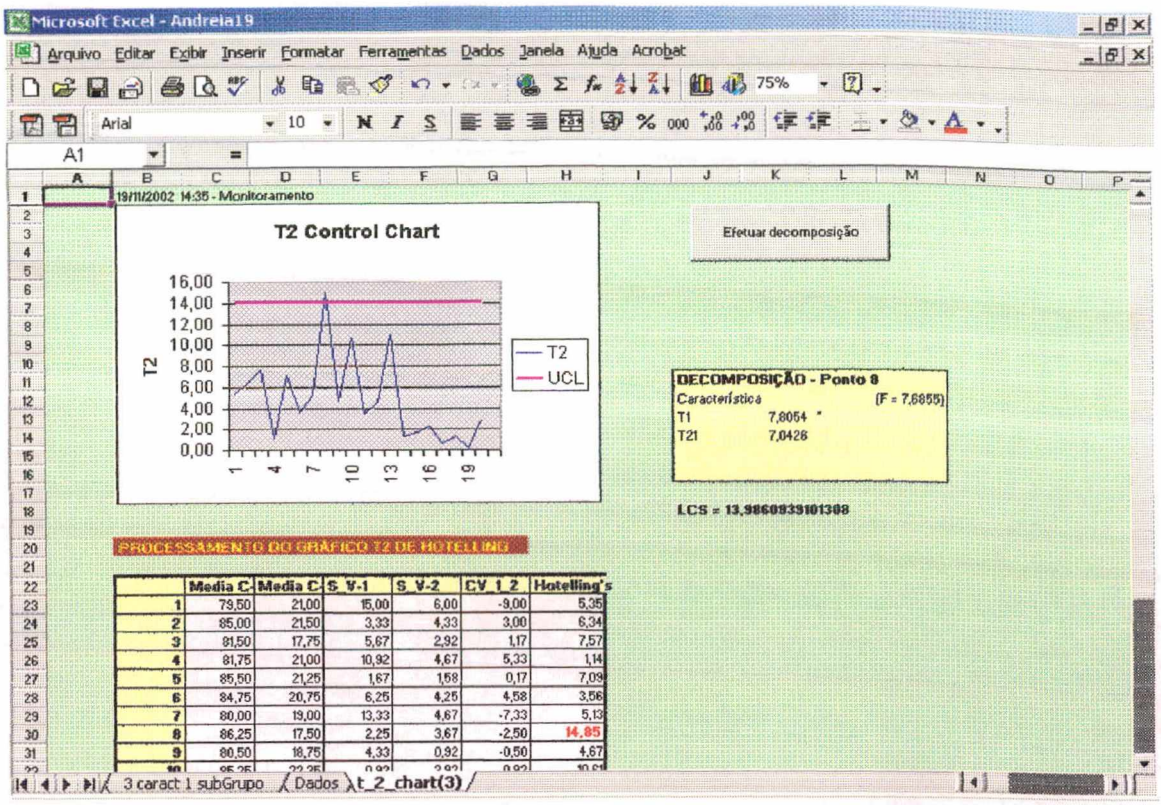


Figura 11: Tela para Acrescentar as Novas Observações no Processo

Após a inserção da nova amostra, clicando-se no botão "Processa" esta nova observação é armazenada na base de rascunho e são efetuados todos os cálculos da estatística T^2 *de Hotelling,* conforme procedimento empregado na primeira fase.



	Media C	Media C	S_V-1	S_V-2	CV_1_2	Hotelling's
1	79,50	21,00	15,00	6,00	-9,00	5,35
2	85,00	21,50	3,33	4,33	3,00	6,34
3	81,50	17,75	5,67	2,92	1,17	7,57
4	81,75	21,00	10,92	4,67	5,33	1,14
5	85,50	21,25	1,67	1,58	0,17	7,09
6	84,75	20,75	6,25	4,25	4,58	3,56
7	80,00	19,00	13,33	4,67	-7,33	5,13
8	86,25	17,50	2,25	3,67	-2,50	14,85
9	80,50	18,75	4,33	0,92	-0,50	4,67

Figura 12: Tela de Apresentação da Decomposição

Através da Figura 12, pode-se observar que na decomposição realizada pelo software, a variável 1 apresenta o valor da estatística T^2 *de Hotelling* maior do que o valor de referência de F; ou seja, $T_1^2 = 7,8054 > F=7,6855$. Assim, o resultado indica que a variável 1 deve ser investigada para que se possa encontrar as possíveis causas assinaláveis dessa variabilidade.

3.5. Considerações Finais

Neste capitulo, foram comentadas as facilidades de implementação da planilha de calculo em conjunto com o *VBA*. Foram apresentados alguns procedimentos para identificar variáveis fora dos limites de controle em gráficos multivariados, bem como suas aplicações através da abordagem proposta por Mason, Tracy, Young (1995). Através desta abordagem, desenvolveu-se o algoritmo computacional, o qual realiza o cálculo da estatística T^2 de Hotelling, efetuando a decomposição das variáveis que estejam interferindo no processo que foi o objetivo deste trabalho. Ao comparar o valor da decomposição das variáveis com a estatística *F*, deve-se observar o teste de hipótese para saber se a variável deve ou não ser investigada, ou seja, se $T^2 > F$ rejeita-se H_0, caso contrario se $F > T^2$ não se pode rejeitar H_0.

Portanto o referido trabalho apresenta resultados satisfatórios, alcançando o objetivo proposto. No próximo capítulo são apresentadas as conclusões e as recomendações para trabalhos futuros.

4. CONCLUSÕES

A constatação da importância da utilização do controle de qualidade por meio de métodos estatísticos em um ambiente altamente competitivo encorajou a aplicação da técnica de decomposição da estatística do gráfico de controle multivariado T^2 *de Hotelling* por meio de um algoritmo computacional .

Neste sentido, o objetivo geral desta dissertação foi plenamente atingido. Ao propor um modelo para a decomposição da estatística T^2 de Hotelling, por meio de algoritmo computacional, em linguagem VBA (*Visual Basic Application*) verificou-se a possibilidade de maior agilidade na detecção das variáveis que são significativas, no momento em que as características de qualidade monitoradas do processo saem fora dos limites de controle em processos multivariados. Como foi demonstrado na seção 3.4.4, um programa de melhoria de qualidade bem dimensionado, o qual possibilita a redução de custos e o incremento dos resultados financeiros.

Os objetivos específicos também foram atingidos:

- Buscou-se por meio da fundamentação teórica destacar a importância de conceitos referentes aos gráficos de controle multivariados, mais especificamente o gráfico de controle T^2 de Hotelling, além de abordar o procedimento utilizado para a identificação de variáveis responsáveis por pontos fora dos limites de controle e seus diversos métodos de decomposição.

- Destacou-se também um estudo teórico dos gráficos de controle tradicionais.

- Desenvolveu-se um algoritmo computacional, o qual apresenta a decomposição estatística T^2 de Hotelling que possibilita a detecção das variáveis significativas que saem fora dos limites de controle em processos multivariados.

- Apresentou-se o algoritmo computacional descrevendo cada etapa para sua operacionalização para o monitoramento das variáveis de interesse, mantendo assim, um controle das características da qualidade do processo.

4.1. RECOMENDAÇÕES

Em continuidade ao desenvolvimento deste estudo, sugere-se:

- para trabalhos futuros que seja feito um estudo mais aprofundado sobre os outros métodos de decomposição da estatística T^2 de Hotelling, comparando-os com o utilizado no presente trabalho a fim de verificar qual apresenta melhor performance.

- inserir no pacote do algoritmo computacional a decomposição da estatística dos gráficos de controle somas acumuladas (MCUSUM).

- ampliar o estudo buscando alternativas para a melhoria dos processos industriais a partir do aprofundamento dos estudos de processos que utilizam gráficos de controle multivariados.

- realizar estudos de casos em empresas aplicando o algoritmo desenvolvido.

REFERÊNCIA BIBLIOGRÁFICA

ALT, F.B. Multivariatre Quality Control, in **The Encyclopedia of Statistical Sciences**. New York: John Wiley & Sons, Inc., 1985.

BRITZ, G. C., EMERLING, D. W., HARE, L.B., HOERL, R. W., JANIS, S. J., SHADE J. E. **Improving Performance Through Statistical Thinking** ASQ Statistics Division**.** Milwaukee: Printed in the United States of America, 1999.

CROSBY, P. B. **Qualidade é Investimento.** Rio de Janeiro. Ed. José Olympio. 1994.

DOGANAKSOY, N.FALTIN, F.W.& TUCKER, W.T. Identification of Out of Control Quality Characteristics in a Multivariate Manufacturing Environment **Communications in Statistics- Theory and Methods**, v.20 p.1775-2790,1991.

FREUND, J.E. & SIMON, G .**A. Modern and Elementary Statistics**. 9ª Ed. Revisada. New Jersey, Prentice Hall, Inc., 1997.

FROTA, A. Custos de Qualidade: Disponível em : http://www.calidad.org/articles/jan98/3jan98.htm. Acessado em 10 de setembro de 2002.

GARVIN, D. A. Competing in the Eight Dimensions of Quality. Harvard Business **Review**, Sept-Oct 1987.

HAWKINS, D.M. OLWELL, D.H. **Cumulative Sum Charts and Charting for Quality Improvement.** New York: Springer-Verlang, 1997.

HAWKINS, D.M. Multivariate Quality Control Based no Regression-Adjusted Variables. **Technometrics**, v.33, nº 1, p. 61-75,1991.

HAWKINS, D.M. Regression Adjustment for Variables in Multivariate Quality Control. **Journal of Quality Technology**, v.25, nº 3, p.170-182,1993

HEALY, J.D. A Note on multivariate CUSUM Procedures. **Technometrics** v.29, nº 4, p. 409-412, 1987.

HOLMES, D. S. & MERGEN, A. E. Improving the Performance of the T^2 Control Chart. **Quality Engineering**, v.5, p. 619-625, 1993.

HOTELLING, H. Multivariate Quality Control- illustrated by the air testing of sample bombsights, in **Techniques of Statistical Analysis**. McGraw Hill, New York, p. 111-184, 1947.

HRADESKY, J. L. **Aperfeiçoamento da Qualidade e da Produtividade: Guia Prático para Implementação do Controle Estatístico de Processo-CEP** *Tradução:* Maria Claudia de Oliveira Santos. São Paulo: McGraw-Hill, p.151, 1989

JACKSON, J.E. Principal Component and factor Analysis: Part I- Principal Components,. **Journal of Quality Technology**, v.12, nº 4, p. 201-213, 1980.

JACKSON, J.E. Multivariate Quality Control. **Communications in Statistics- Theory and Methods**, v. 14 p. 657-688,1985.

JACKSON, J.E. **A User's Guide to Principal Components**. John Wiley &sons, New York, 1991.

LOWRY, C.A..& MONTGOMERY, D.C. A Revew of Multivariate Control Charts. **IIE Transactions**, v.27, p.800-810, 1995.

LOWRY, C.A..et al. A Multivariate Exponentially Weighted Moving Average Control Chart. **Technometrics**, v.34, nº 1, p.46-53, 1992.

MASON, R. L. CHAMP, C. W. TRACY, N.D. WIERDA, S.J. & YOUNG, J.C. Assessment of Multivariate Process Control Techniques. **Journal of Quality Technology,** v.29, nº 2, p.140-143, 1997.

MASON, R. L., TRACY, N.D & YOUNG, J.C. Decomposition of T^2 for Multivariate Control Chart Interpretion. **Journal of Quality Technology**, v.27, nº 2, p.99-108, 1995.

MICHEL, R. & FLOGLIATTO, F.S. Projeto Econômico de Cartas Adaptativas para Monitoramento de Processos. Artigo submetido à **revista Produto & Produção** para publicação, 2000.

MILLER, P. SWANSON, R.E. & HECKLER, C.E. Contribution Plots: The Missing Link in Multivariate Quality Control. *Unpublished Manuscript to be Submitted to* **Journal of Quality Technology,** 1993.

MITRA, A. **Fundamentals of Quality Control and Improvement**. 2º Ed. Prentice Hall, New Jersey, 1993, p. 309-316.

MONTGOMERY, D.C. **Introduction to Statistical Quality Control.** 3º Ed. John Wiley & Sons, New York, 1996.

MOREIRA, D. A. **Dimensões do Desempenho em Manufatura e Serviços**. São Paulo: Pioneira, 1996.

MURPHY, B.J. Selecting Out of Control, Variables With the T^2 Multivariate Quality Control Procedure. **The Statistician**, v.36, p.571-583, 1987.

PIGNATELLO, J.J.Jr & RUNGER, G.C. Comparisons of Multivariate CUSUM charts. **Journal of Quality Technology**, v.22, nº 3, p. 173-186, 1990.

RENCHER, A.C. The Contribuition of Individual Variables to Hotelling's T^2, Wilk's Λ. And R^2:. **Biometrics**, v.49, p.479-489, 1993.

RYAN, T. P. **Statistical Methods for Quality Improvement**. 1ª Ed. John Wiley & Sons, Inc. New York, 1989.

SLACK, N. CHAMBERS S. , HARLAND, C. **Administração da Produção**. São Paulo. Atlas. 1997.

SIQUEIRA, L. G. P. **Controle Estatístico de Processos**. Equipe Grifo. São Paulo: Pioneira, 1997, Série Qualidade Brasil. Livro nº 9, p. 12-13.

SULLIVAN, J.H. & WOODALL, W. H. Comparison of Multivariate Control Charts for Individual Observations. **Journal of Quality Technology**, v.28, nº 4, p.398-408, 1996.

TAGUCHI, G; ELSAYED, A; HSIANG, Thomas C. **Engenharia da Qualidade em Sistemas de Produção.** São Paulo: McGraw Hill 1990.

TRACY, N.D. YOUNG, J.C. & MASON, R. L. Multivariate Control Charts for Individual Observations. **Journal of Quality Technology**, v.24, nº 2, p.88-95, 1992.

WADE, M.R. & WOODALL, W.H. A Revew and Analisys of Cause-Selecting Control Charts. **Journal of Quality Technology**, v.25, nº 3, p.161-169, 1993.

WIERDA, S.J. Multivariate Statisitical Process Control-Recent Results and Directions for Future Research. **Statistica Neerlandica**, v.48 p.147-168,1994.

WOODALL, W.H. & NCUBE, M.M. Multivariate CUSUM Quality Control Procedures. **Technometrics**, v. 27, nº 3, p.285-292, 1985.

ANEXOS

ANEXO 1

Table 7-14- Bivariate Data For the Fabric-Production Example (Mitra, 1993, p.314)

Sample	Single-Strand Break Factor - i=1				Weight of Textile Fibers i= 2 hands/pound			
1	80	82	78	85	19	22	20	20
2	75	78	84	81	24	21	18	21
3	83	86	84	87	19	24	21	22
4	79	84	80	83	18	20	17	16
5	82	81	78	86	23	21	18	22
6	86	84	85	87	21	20	23	21
7	84	88	82	85	19	23	19	22
8	76	84	78	82	22	17	19	18
9	85	88	85	87	18	16	20	16
10	80	78	81	83	18	19	20	18
11	86	84	85	86	23	20	24	22
12	81	81	83	82	22	21	23	21
13	81	86	82	79	16	18	20	19
14	75	78	82	80	22	21	23	22
15	77	84	78	85	22	19	21	18
16	86	82	84	84	19	23	18	22
17	84	85	78	79	17	22	18	19
18	82	86	79	83	20	19	23	21
19	79	88	85	83	21	23	20	18
20	80	84	82	85	18	22	19	20

ANEXO 2

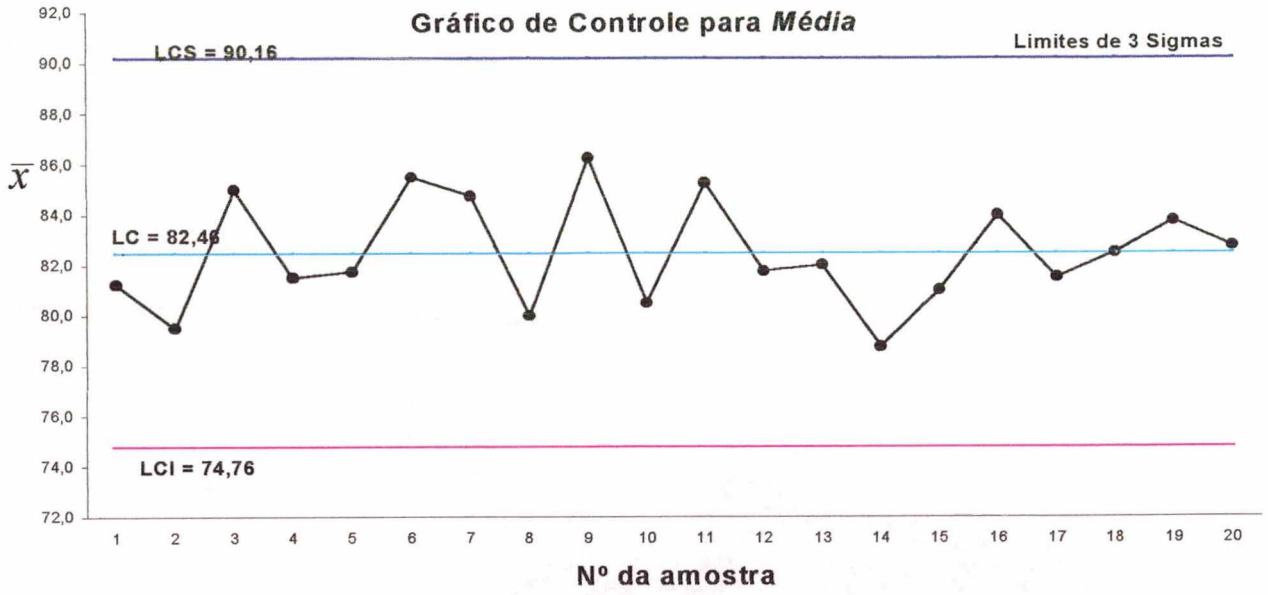


Figura 13: Gráfico de Shewhart para a variável 1

ANEXO 3

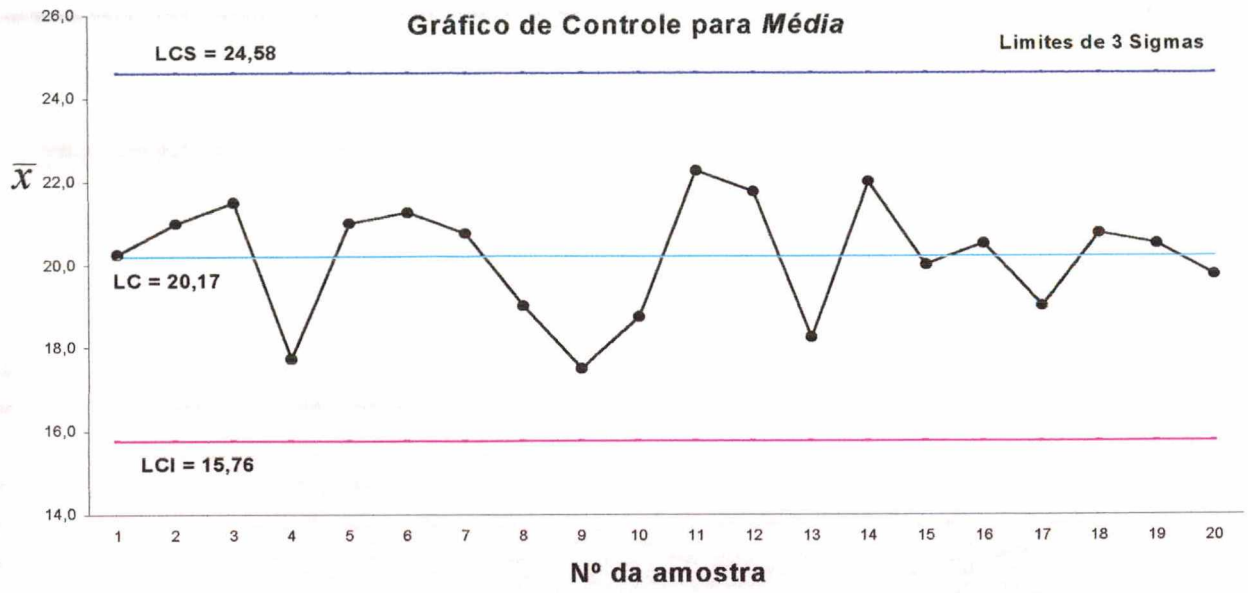


Figura 14: Gráfico de Shewhart para a variável 2