

# Implementação Computacional em *S-Plus* dos Testes Estatísticos de Hotelling e de Hayter e Tsui e suas Aplicações em Controle de Qualidade

Fernanda Karine Ruiz Colenghi (UFMG) fernandak@ufmg.br

Sueli Aparecida Mingoti (UFMG) sueli@est.ufmg.br

## Resumo

*Com as exigências de um mercado cada vez mais competitivo, as empresas necessitam de certificados de controle de qualidade do processo de produção ou serviços. Nesse contexto, o controle de qualidade multivariado aparece como uma ferramenta extremamente importante para avaliar várias características que afetam diretamente e simultaneamente a qualidade final de determinado produto. Devido a esses fatores vemos a importância do estudo das técnicas estatísticas multivariadas para controle de processos. Neste trabalho apresentamos o conceito de algumas das técnicas estatísticas multivariadas e mostramos exemplos com a implementação computacional usando a linguagem de programação do software estatístico S-Plus.*

*Palavras-chave: Controle Multivariado de Processos; Testes Estatísticos Multivariados; S-Plus.*

## 1. Introdução

No processo de produção da maioria das indústrias existentes, é comum utilizar-se técnicas de controle de qualidade para verificar-se a qualidade final de um determinado produto bem como do processo de produção em questão. A situação mais comum é aquela em que o processo é monitorado por várias características que afetam diretamente a qualidade final do produto (controle simultâneo de características de qualidade).

No caso multivariado, a aplicação de gráficos de controle pode ser feita analisando-se cada característica separadamente através de gráficos de controle de Shewhart (Montgomery, 2004), no entanto esta análise não é eficaz e pode levar a conclusões errôneas. Portanto, estudaremos métodos multivariados, que consideram as variáveis em conjunto.

Dentro da área de testes de hipóteses multivariados encontra-se o teste  $T^2$  de Hotelling (1947) para comparação de vetores de médias populacionais. Deste modo, podem ser encontrados vários artigos na literatura propondo o uso deste teste para testar a hipótese de controle do processo considerando-se todas as características medidas simultaneamente. A partir do teste  $T^2$  de Hotelling constrói-se o elipsóide de confiança, usando a distribuição conjunta das  $p$  características de qualidade, que permite verificar se o processo está ou não sob controle considerando-se todas as características simultaneamente. No entanto, a construção de regiões de confiança em termos gráficos torna-se impraticável quando  $p \geq 4$  variáveis. Neste caso, o mais comum é transformar-se a informação multivariada para uma estatística unidimensional (ou bidimensional) que possa ser disponibilizada em termos gráficos. Uma das alternativas é o gráfico univariado  $T^2$  de Hotelling. Outra é a sumarização da informação multivariada em combinações lineares (Johnson & Wichern, 2002), chamadas de componentes principais (Mingoti, 2001). Uma crítica em relação a este

procedimento é que nem sempre as componentes têm uma interpretação que faz sentido para o usuário que está supervisionando o processo. No entanto, pode constituir numa técnica importante para síntese de variáveis criando-se então, índices globais de qualidade do processo (Jackson, 1991).

O uso de gráficos de controle multivariados é recente e algumas referências interessantes de aplicação são: Mason, Chou et. al. (2001), Mason, Tracy et. al. (1997) e Nomikos & MacGregor (1995) que faz uma espécie de sumário sobre o uso da estatística  $T^2$ .

Uma possível crítica ao uso do teste  $T^2$  Hotelling para avaliar o processo vem do fato de que no momento em que a hipótese nula é rejeitada torna-se necessário identificar as características de qualidade responsáveis pela sua rejeição, o que muitas vezes é feito através de gráficos de Shewhart (Montgomery, 2004) para cada variável isoladamente, corrigindo-se ou não, os níveis de significância dos testes feitos separadamente para a média de cada característica. Testes de comparações múltiplas de Bonferroni (Johnson & Wichern, 2002), por exemplo, podem ser usados, no entanto esses não levam em consideração a correlação entre variáveis respostas. Algumas alternativas ao teste  $T^2$  de Hotelling estão publicadas na literatura, dentre elas citam-se Hayter & Tsui (1994), Tiku & Singh (1982), Tiku & Balakrishnan (1988), Mudholkar & Subbaih (1980), Mudholkar & Srivastava (1999) e mais recentemente os testes *stepwise* robustos de Mudholkar & Srivastava (2000) sendo este último mais poderoso que o teste  $T^2$  de Hotelling inclusive para amostras de tamanho pequeno é mais robusto em relação a desvios da normalidade multivariada.

Alguns destes testes multivariados não estão implementados em *softwares* estatísticos. Mesmo aqueles que estão disponíveis, não estão no formato padrão da área de controle estatístico de qualidade. Neste artigo apresentamos a implementação computacional em *S-Plus* (Venables & Ripley, 2003; Reis, 1997) de duas técnicas: a  $T^2$  de Hotelling e a de Hayter & Tsui.

## 2. Gráfico $T^2$ de Hotelling

O gráfico de controle de  $T^2$  de Hotelling, chamado de gráfico qui-quadrado, é usado para monitoramento do vetor de médias do processo. Ele é um análogo direto do gráfico de  $\bar{X}$  de Shewhart univariado. Mostramos a seguir, duas versões do gráfico  $T^2$  de Hotelling: uma para dados em subgrupos e outra para observações individuais.

### 2.1. Dados em Subgrupos

Suponha que a distribuição de probabilidade conjunta das  $p$  características da qualidade ( $X_1, X_2, \dots, X_p$ ) seja a distribuição normal  $p$ -variada. Sejam  $\mu_1, \mu_2, \dots, \mu_p$  os valores médios das características de qualidade e  $\Sigma$  a matriz de covariâncias da distribuição dessas características. A partir de uma amostra aleatória do processo de tamanho  $n$  observa-se o vetor  $X = [X_1 \ X_2 \ \dots \ X_p]'$  que é medido em cada um dos  $n$  elementos amostrais. Quando  $n > 1$  o procedimento de amostragem é chamado de dados em subgrupos.

O procedimento requer o cálculo da média amostral para cada uma das  $p$  características de qualidade a partir de  $m$  amostras de tamanho  $n$  do processo. Esse conjunto de médias das características da qualidade é representado pelo vetor de médias amostral  $\bar{X} = [\bar{X}_1 \ \bar{X}_2 \ \dots \ \bar{X}_p]'$ . A estatística de teste plotada no gráfico de controle qui-quadrado para cada amostra  $t, t=1,2,\dots,m$ , quando  $\mu$  e  $\Sigma$  são conhecidos é a  $T_t^2$  dada por:

$$T_i^2 = n(\bar{X} - \mu)' \Sigma^{-1} (\bar{X} - \mu) \quad (1)$$

onde  $\mu$  e  $\Sigma$  são os parâmetros do processo sob controle. O limite superior de controle é dado por  $LSC = \chi_{1-\alpha, p}^2$ , onde  $\chi_{1-\alpha, p}^2$  é o valor obtido na distribuição qui-quadrado com  $p$  graus de liberdade correspondente a probabilidade acumulada de  $(1-\alpha)$ ,  $0 < \alpha < 1$ , ou seja,  $P[\chi_p^2 \leq \chi_{(1-\alpha), p}^2] = 1 - \alpha$ .

Na prática, em geral, é necessário estimar  $\mu$  e  $\Sigma$  a partir de  $m$  amostras preliminares de tamanho  $n$  do processo, obtidas quando se admite que o processo está sob controle. Nesse caso, as médias e variâncias amostrais para cada amostra são calculadas para cada variável  $X_i$  como usual, isto é,

$$\bar{X}_{jk} = \frac{1}{n} \sum_{i=1}^n \bar{X}_{ijk} \quad \begin{cases} j = 1, 2, \dots, p \\ k = 1, 2, \dots, m \end{cases} \quad (2)$$

$$S_{jjk} = S_{jk}^2 = \frac{1}{n-1} \sum_{i=1}^n (X_{ijk} - \bar{X}_{jk})^2 \quad \begin{cases} j = 1, 2, \dots, p \\ k = 1, 2, \dots, m \end{cases} \quad (3)$$

onde  $X_{ijk}$  é a  $i$ .ª observação da  $j$ .ª característica da qualidade na  $k$ .ª amostra e  $\bar{X}_{jk}$  é a média das  $n$  observações na  $j$ .ª característica da  $k$ .ª amostra. A covariância amostral entre as características da qualidade  $j$  e  $h$  na  $k$ .ª amostra é dada por:

$$S_{jhk} = \frac{1}{n-1} \sum_{i=1}^n (X_{ijk} - \bar{X}_{jk})(X_{ihk} - \bar{X}_{hk}) \quad \begin{cases} k = 1, 2, \dots, m \\ j \neq h \end{cases} \quad (4)$$

Tomam-se, então, as médias das estatísticas  $\bar{X}_{jk}$ ,  $S_{jk}^2$ ,  $S_{jhk}$  sobre todas  $m$  amostras, obtendo-se:

$$\bar{\bar{X}}_j = \frac{1}{m} \sum_{k=1}^m \bar{X}_{jk} \quad j = 1, 2, \dots, p \quad (5)$$

$$\bar{S}_{jj} = \frac{1}{m} \sum_{k=1}^m S_{jk}^2 \quad j = 1, 2, \dots, p \quad (6)$$

$$\bar{S}_{jh} = \frac{1}{m} \sum_{k=1}^m S_{jhk} \quad j \neq h \quad (7)$$

Os  $\{\bar{\bar{X}}_j\}$  são elementos do vetor  $\bar{\bar{X}}$ , e a matriz  $S$  de dimensão  $p \times p$  é a média das  $m$  matrizes de covariâncias amostrais é construída como:

$$S = \begin{bmatrix} \bar{S}_{11} & \bar{S}_{12} & \bar{S}_{13} & \cdots & \bar{S}_{1p} \\ & \bar{S}_{22} & \bar{S}_{23} & \cdots & \bar{S}_{2p} \\ & & \bar{S}_{33} & & \vdots \\ & & & \ddots & \bar{S}_{pp} \end{bmatrix} \quad (8)$$

A matriz  $S$  é um estimador não-viciado de  $\Sigma$ , quando o processo está sob controle (Johnson & Wichern, 2002).

Suponha que  $S$  seja usada para estimar  $\Sigma$ , e que o vetor  $\bar{\bar{X}}$  seja considerado o valor sob controle do vetor médio do processo. Neste caso, a estatística que deverá ser calculada para cada amostra  $t$ ,  $t=1,2,\dots, m$ , e plotada no gráfico de controle é dada por:

$$T_t^2 = n(\bar{X} - \bar{\bar{X}})' S^{-1}(\bar{X} - \bar{\bar{X}}) \quad (9)$$

Alt (1985) observou que, nas aplicações de controle de qualidade multivariado, deve-se ter cuidado na seleção dos limites de controle para a estatística  $T^2$  de Hotelling, dependendo de como o gráfico esteja sendo usado. Há duas fases distintas no uso de um gráfico de controle. Fase 1 é o uso dos gráficos para o estabelecimento do limite de controle, isto é, testar se o processo estava, ou não, sob controle quando os  $m$  subgrupos preliminares foram extraídos e as estatísticas amostrais  $\bar{X}$  e  $S$ , calculadas. Essa fase só se aplica no caso em que  $\mu$  e  $\Sigma$  são desconhecidos, uma vez que a amostragem preliminar não seria necessária para o estabelecimento dos limites de controle quando os parâmetros de interesse já são conhecidos. O objetivo, na fase 1, é a obtenção de um conjunto de observações sob controle, de modo que os limites de controle possam ser estabelecidos para a fase 2, que é o monitoramento da produção futura. Isto é, às vezes, chamada análise retrospectiva.

Os limites de controle para o gráfico de controle  $T^2$  na fase 1 são dados por:

$$\text{LSC} = \frac{p(m-1)(n-1)}{mn-m-p+1} F_{\alpha,p,mn-m-p+1} \quad \text{e} \quad \text{LIC} = 0 \quad (10)$$

Na fase 2, quando o gráfico  $T^2$  é usado para monitorar a produção futura, os limites de controle são os seguintes:

$$\text{LSC} = \frac{p(m+1)(n-1)}{mn-m-p+1} F_{\alpha,p,mn-m-p+1} \quad \text{e} \quad \text{LIC} = 0 \quad (11)$$

Quando  $\mu$  e  $\Sigma$  são estimados a partir de um grande número de amostras preliminares, costuma-se usar  $\text{LSC} = \chi_{1-\alpha,p}^2$  como limite superior de controle em ambas fases. A análise retrospectiva das amostras preliminares para testar o controle estatístico e estabelecer os limites de controle também ocorre no ambiente gráfico de controle univariado. Para o gráfico de controle univariado  $\bar{X}$ , se usarmos  $m \geq 20$  ou 25 amostras preliminares, a distinção entre os limites da fase 1 e da fase 2 é, em geral, desnecessária, porque esses limites praticamente coincidirão.

## 2.2. Observações individuais

Em alguns ambientes industriais como em indústrias químicas de processamento (Montgomery, 2004), o tamanho do subgrupo é  $n=1$  e há várias características da qualidade a serem monitoradas. Suponha que estejam disponíveis  $m$  amostras independentes do processo, cada uma tamanho  $n=1$ , e que  $p$  seja o número de características da qualidade observadas em cada uma delas. Sejam  $\bar{X}$  e  $S$ , respectivamente, o vetor de médias amostral e a matriz de covariâncias amostral calculadas com essas  $m$  observações. A estatística  $T^2$  de Hotelling que

deverá ser calculada para cada amostra  $t$ ,  $t=1,2,\dots, m$ , e plotada no gráfico de controle é dada por:

$$T_t^2 = (X - \bar{X})' S^{-1} (X - \bar{X}) \quad (12)$$

Sullivan e Woodall (1995) fornecem discussões e análises de vários estimadores da matriz de covariâncias, no entanto neste artigo usamos a matriz de covariâncias amostral  $S$  definida por:

$$S_{pp} = \frac{1}{m-1} \sum_{i=1}^m (X_i - \bar{X})(X_i - \bar{X})' \quad (13)$$

Tracy, Mason e Young (1992) enfatizam que, se  $n=1$ , os limites da fase 1 devem basear-se em uma distribuição Beta. Da mesma forma que foi visto para dados em subgrupos, a análise da fase 1 só é realizada quando  $\mu$  e  $\Sigma$  são desconhecidos. Nesse caso, os limites de controle são:

$$\text{LSC} = \frac{(m-1)^2}{m} \beta_{\alpha, p/2, (m-p-1)/2} \quad \text{e} \quad \text{LIC} = 0 \quad (14)$$

onde  $\beta_{\alpha, p/2, (m-p-1)/2}$  é a ordenada correspondente ao percentil de ordem  $(1-\alpha)$  da distribuição Beta com  $p/2$  e  $(m-p-1)/2$ . Segundo Tracy, Mason e Young (1992) as aproximações pela  $F$  e qui-quadrado na fase 1, quando  $n=1$ , serão, provavelmente, imprecisas para determinar os limites de controle.

Os limites de controle da fase 2 para o processo são:

$$\text{LSC} = \frac{p(m+1)(m-1)}{m^2 - mp} F_{\alpha, p, m-p} \quad \text{e} \quad \text{LIC} = 0 \quad (15)$$

### 2.3 Interpretação dos Sinais Fora de Controle

Quando se observa na fase 2 algum valor numérico de  $T^2$  fora do limite de controle, é de interesse saber qual ou quais das  $p$  variáveis é responsável por esse sinal. A abordagem que iremos usar foi sugerida por Runger, Alt e Montgomery (1996b.). Eles sugerem a decomposição da estatística  $T^2$  em componentes que refletem a contribuição de cada variável individual. Se  $T^2$  é o valor corrente da estatística para a respectiva amostra, e  $T_{(i)}^2$  é o valor da estatística para todas as variáveis do processo exceto a  $i$ -ésima, tem-se que:

$$d_i = T^2 - T_{(i)}^2 \quad (16)$$

é um indicador da contribuição relativa da  $i$ .ª variável para a estatística global. Deve-se ter atenção nas variáveis que geram  $d_i$ 's relativamente grandes. Além disso, eles sugerem que um corte aproximado para a magnitude de um  $d_i$  individual seja  $\chi_{1-\alpha, p}^2$ , isto é, a ordenada correspondente ao percentil de ordem  $(1-\alpha)$  da distribuição qui-quadrado com 1 grau de liberdade. Nesse caso, pode-se testar a significância de  $d_i$  usando um valor de nível de significância  $\alpha$  fixo para o teste ou através do cálculo da probabilidade de significância.

Para um melhor entendimento desse procedimento, ilustraremos com o exemplo de Runger, Alt e Montgomery (1996a.). Há três características de qualidade, a matriz de covariâncias é conhecida e a matriz de correlação é como se segue:

$$\Sigma = \begin{bmatrix} 1 & 0,9 & 0,9 \\ 0,9 & 1 & 0,9 \\ 0,9 & 0,9 & 1 \end{bmatrix}$$

A média do processo é  $\mu' = [0, 0, 0]$ .

A Tabela 1, mostra os valores de  $d_i = T^2 - T_{(i)}^2$  calculados para cada característica de quatro vetores amostrais. Como  $\Sigma$  é conhecida, o limite superior de controle para o gráfico é igual a  $\chi_{0,01;1}^2 = 6,36$ . Qualquer  $d_i$  que exceder esse valor é uma indicação de que a variável  $i$ , da correspondente amostra é responsável pelo processo estar fora de controle. A decomposição de  $T^2$  fornece uma diretriz sobre quais variáveis afetam mais acentuadamente o valor de  $T^2$ . Pela Tabela 1, observa-se que a variável 1 na amostra 1; as variáveis 1, 2 e 3 na amostra 2; 1 e 2 na amostra 3 e 3 na amostra 4, são as responsáveis pelo sinal de fora de controle do processo.

Vetor de amostras $X'$	Valor de $T_0^2$ observado	$d_1$	$d_2$	$d_3$	p-valores		
					Variável 1	Variável 2	Variável 3
(2,0,0)	27,14	27,14	6,09	6,09	0,0000	0,0136	0,0136
(1,1,-1)	26,79	6,79	6,79	25,73	0,0092	0,0092	0,0000
(1,-1,0)	20,00	14,74	14,74	0	0,0001	0,0001	1,0000
(0,5;0,5;1)	15,00	3,69	3,68	14,74	0,0547	0,0551	0,0001

Tabela 1 - Representação da variável  $T^2$  decomposta para cada variável

### 3. Método de Hayter & Tsui

Como uma alternativa para o teste de Hotelling, Hayter & Tsui (1994) propuseram um método que além de descobrir quais variáveis estão fora de controle, fornece também informações sobre a possível alteração ocorrida nos valores de média do processo através da construção de intervalos de confiança para a média verdadeira de cada uma das  $p$ -variáveis. Além disso, ao contrário de metodologias como a de Bonferroni, os intervalos são construídos de modo que a constante de referência que determina a abertura dos intervalos é a mesma para todas as variáveis, sendo que a obtenção dessa constante leva em consideração a estrutura de correlação existente entre as variáveis.

Seja  $X = [X_1, X_2, \dots, X_p]'$  o vetor contendo as  $p$  características de qualidade de interesse, com distribuição Normal  $p$ -variada e com vetor de médias e matriz de covariâncias definidos na seção anterior. Considera-se portanto que  $X \sim N_p(\mu_0, \Sigma_{p \times p})$ . De acordo com Hayter & Tsui, para cada variável  $X_i$  os limites de confiança de  $(1-\alpha)100\%$ , são dados pela equação:

$$P \left[ \left| \frac{X_i - \mu_i^0}{\sigma_i} \right| \leq C_{R,\alpha}, \forall i = 1, 2, \dots, p \right] = 1 - \alpha \quad (17)$$

o que significa dizer que a probabilidade de que o intervalo  $[X_i - \sigma_i C_{R,\alpha}; X_i + \sigma_i C_{R,\alpha}]$  contenha o verdadeiro valor de  $\mu_i^0$  para todo  $i, i=1,2,\dots,p$ , é igual a  $(1-\alpha)$ .

A escolha do valor crítico  $C_{R,\alpha}$  depende da matriz de correlação  $P_{p \times p}$ . Na prática, é necessário estimar  $P_{p \times p}$  através da matriz de correlação amostral  $R_{p \times p}$  do vetor aleatório  $X$ . Logo a estrutura de correlação de  $X$  afeta todos os intervalos simultaneamente e deste modo, o processo é considerado como fora de controle quando observamos que:

$$M = \max \left\{ \left| \frac{X_i - \mu_i^0}{\sigma_i} \right|, i = 1, 2, \dots, p \right\} > C_{R,\alpha} \quad (18)$$

Para construir o gráfico de controle plota-se a estatística  $M$  para cada observação amostral multivariada e usa-se o valor crítico  $C_{R,\alpha}$  como limite superior de controle. O gráfico individual de controle para cada uma das  $p$  variáveis pode ser mantido com  $[\mu_i^0 \pm \sigma_i C_{R,\alpha}]$  como os limites de controle.

Hayter e Tsui sugerem o uso de simulação para encontrar  $C_{R,\alpha}$ . Primeiramente é necessário gerar um grande número  $N=100000$  de vetores de uma distribuição normal multivariada com vetor de médias zero e matriz de covariâncias  $R_{p \times p}$ , isto é  $Z^i, i = 1, \dots, N$ . Em seguida calcula-se as  $M$  estatísticas para cada um dos vetores. Depois calcula-se  $M^i = \max_{1 \leq j \leq p} |Z_j^i|$  para  $i=1,2,\dots,N$ . Finalmente encontra-se a ordenada correspondente ao  $(1-\alpha)$  é-simo percentil da amostra  $\{M^1, \dots, M^N\}$  e usa-se este como estimativa para o ponto crítico  $C_{R,\alpha}$ .

A estatística de Hayter e Tsui foi utilizada também por Mingoti e Glória (2003) na construção de índices de capacidade multivariados. Nesse artigo, os autores mostram que para o encontro da constante  $C_{R,\alpha}$  basta utilizar uma amostra de  $N=10000$  observações da distribuição normal p-variada padronizada e não  $N=100000$  como sugerido no artigo de Hayter e Tsui (1994). Neste artigo será adotado  $N=10000$  na simulação para encontrar  $C_{R,\alpha}$ .

Na fase 1, quando o processo estiver fora de controle as amostras responsáveis pelo sinal devem ser retiradas e o valor de  $C_{R,\alpha}$  deve ser recalculado. Na fase 2, quando o processo estiver fora de controle deve-se construir o intervalo de confiança  $[\mu_i - \sigma_i C_{R,\alpha}; \mu_i + \sigma_i C_{R,\alpha}]$  para cada uma das  $p$  variáveis. Com esses limites verifica-se quais observações estão fora dos limites. A partir delas verifica-se quais variáveis são as responsáveis pelo sinal de fora de controle do processo. O procedimento adotado neste artigo trabalha tanto com observações individuais, como com dados em subgrupos.

#### 4. Implementação Computacional do Programa no S-Plus

Um programa computacional chamado de *MQCI (Multivariate Quality Control Implementation)*, foi desenvolvido para implementação em *S-Plus*, versão 2000, das metodologias de  $T^2$  e de Hayter & Tsui para dados em subgrupos e dados individuais. A Figura 1 mostra os procedimentos que o programa utiliza como base de seu algoritmo.

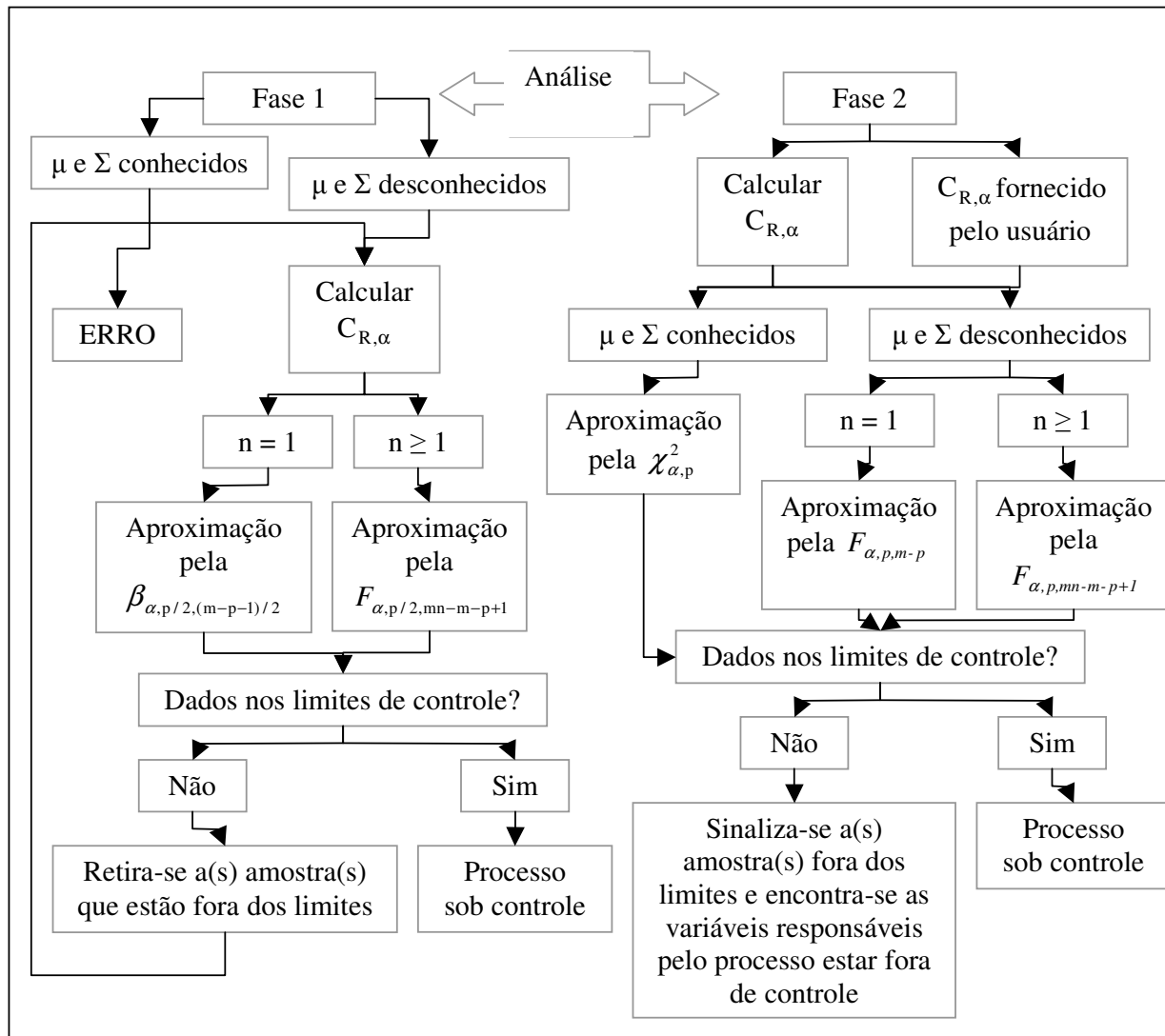


Figura 1 – Fluxograma dos procedimentos realizados pelo programa *MQCI*.

De uma forma geral, o programa é de fácil utilização. No ambiente *S-Plus*, após carregar os parâmetros populacionais de cada processo na sua memória interna do *software*, o programa pode ser chamado no *prompt* de comando através da seguinte instrução:

```
controle(m,dados,i,m,n,p,alfa,fase,valores,dadmedia,dadcov,estcralfa,cr)
```

Deve-se colocar nessa ordem:

1. O arquivo de extensão *.sdd* com os dados a serem analisados (*dados*);
2. A coluna de identificação dos subgrupos (*i*), indicando de qual subgrupo provém a observação;
3. O número de subgrupos (*m*);
4. O tamanho dos subgrupos (*n*);
5. O número de variáveis (*p*);
6. O nível de significância para os testes (*alfa*);
7. A indicação de qual fase, 1 ou 2, o usuário deseja que o teste seja feito (*fase*);



8. A indicação se os parâmetros populacionais, vetor de médias e matriz de covariâncias são estimados (valores=1, dadmedia=0, dadcov=0) ou se são fornecidos pelo usuário (valores=2), tal que os parâmetros dadmedia e dadcov recebem o nome dos arquivos de extensão *.sdd*.
9. A indicação se o usuário deseja estimar  $C_{R,\alpha}$  com os dados (estcralfa= 1, cr=0)
10. Se desejar usar o valor fornecido pelo usuário, colocar estcralfa=2 e colocar o valor fornecido (cr)

Como a metodologia adota a distribuição normal  $p$ -variada, a matriz de covariâncias deverá ser inversível. Se isso não ocorrer, o programa indicará uma mensagem de erro, uma vez que não será possível fazer o cálculo da estatística  $T^2$ . Para melhor ilustrar o uso do *MQCI*, apresentamos a seguir um exemplo.

### 5. Exemplo Usando o Programa *MQCI*

Dados históricos de  $m=50$  peças de aviões coletados pela Companhia Pratt & Whitney foram apresentados em Niverthi & Dey (2000). Observou-se  $p=10$  características diferentes de um componente do motor, sendo essas características medidas com grande precisão. Deseja-se verificar se o processo de fabricação das peças está sob controle estatístico na fase 2, pois deseja-se monitorar a produção futura. O vetor de médias e a matriz de covariâncias foram são estimados com os dados amostrais, assim como a constante  $C_{R,\alpha}$ . Dessa forma a chamada para o programa *MQCI* fica:

```
controlelem(aviao,1,50,1,10,0.05,2,1,0,0,1,0)
```

E a saída é como a apresentada na Figura 2.

```
O vetor de medias e de desvios amostrais:
c1 6.395132 0.000278
c2 0.597060 0.001151
c3 8.297904 0.001158
c4 7.894170 0.000489
c5 22.049168 0.000303
c6 1.854400 0.000342
c7 7.300242 0.000570
c8 6.393156 0.000869
c9 3.046770 0.001901
c10 23.679186 0.000376

Valores de T2:
[1]14.375173 8.410156 3.746982 3.180537 17.731231 10.549631 9.216774 1.878478 7.253897 16.285869
[11]13.518123 10.478595 18.368323 3.263727 15.116299 15.003099 12.397441 7.664800 7.100254 26.890423
[21]16.773686 11.210012 10.243943 3.957178 8.795765 4.397813 5.625022 14.124616 1.843576 6.590485
[31] 5.716509 9.300553 14.408364 10.958746 6.912074 4.581471 9.034929 3.231915 3.083766 6.603396
[41]10.442798 13.209359 10.475923 2.385135 11.752258 11.805913 14.236605 11.601765 14.079513 10.187100
```

Figura 2 – Resultados do *MQCI* para dados da companhia aérea.

Pelas Figuras 3 e 4, observa-se que a amostra 20 apresentou problema. Portanto é necessário identificar qual(is) variável(is) é(são) responsáveis por esse sinal. Pela Figura 4 tem-se que  $\chi_{0,95;1}^2 = 3,84$  é o limite superior de controle. Como  $d_4 = 17,77$  nota-se que a quarta variável é a responsável pela amostra 20 estar fora de controle.

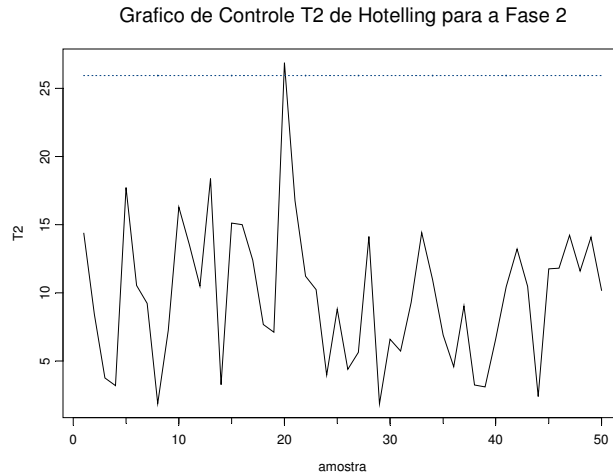


Figura 3 – Gráfico de controle para dados da companhia aérea.

```

O numero de amostras que apresentaram problemas e: 1
A matriz dos di's de tal forma que as colunas mostram os di's das amostras
problematicas e as linhas mostram os di's das diferentes variveis.
      [,1]
[1,]  0.12663397
[2,]  1.41454995
[3,]  0.05824681
[4,] 17.77294493
[5,]  0.42179647
[6,]  0.05863551
[7,]  1.41759787
[8,]  0.06925157
[9,]  0.03390790
[10,] 3.28455067
Limite de corte para os di's com 0.05 de significancia e:  3.84145882069413
A matriz com os p-valores desses di's e:
      [,1]
[1,] 0.72194790481
[2,] 0.23430254138
[3,] 0.80928869518
[4,] 0.00002488975
[5,] 0.51604199801
[6,] 0.80866571839
[7,] 0.23379918681
[8,] 0.79242962432
[9,] 0.85390285975
[10,] 0.06993476839

```

Figura 4 – Resultados do *MQCI* para dados da companhia aérea.

Usando a metodologia de Hayter & Tsui vemos os resultados apresentados na Figura 5. Observa-se que as variáveis 10, 8 e 4 são as responsáveis pelo processo estar fora de controle. Logo, nota-se que a metodologia proposta por Runger, Alt e Montgomery (1996b.) não detecta as variáveis 10 e 8 como possíveis responsáveis pelo processo estar fora de controle. A observação 23,6805 da 10ª variável está fora dos limites (23,6781307; 23,6802413) construídos usando a metodologia de Hayter & Tsui. A Figura 5 mostra as outras observações que também estão fora dos limites de controle. Embora, por exemplo, para a 10ª variável o valor observado esteja fora dos limites apenas a partir da quarta casa decimal, é uma diferença importante de ser detectada em vista da grande precisão usada na medição da variável.

```

Valor de cralfa: [1] 2.799934
Existe pelo menos uma observacao de forma padronizada que apresenta limites maiores
que o cralfa.
Matriz com os limites inferior e superior de controle de Hayter e Tsui para todas
variaveis:
      [,1]      [,2]
c1  6.3943514  6.3959126
c2  0.5938357  0.6002843
c3  8.2946599  8.3011481
c4  7.8928004  7.8955396
c5  22.0483187  22.0500173
c6  1.8534400  1.8553600
c7  7.2986439  7.3018401
c8  6.3907209  6.3955911
c9  3.0414472  3.0520928
c10 23.6781307  23.6802413

As observacoes fora dos limites de controle em Hayter e Tsui sao:
Variavel e valor observado
10      23.6805
8        6.39
4        7.8927
4        7.892

```

Figura 5 – Resultados do *MQCI* para dados da companhia aérea.

## 6. Considerações Finais

Com o desenvolvimento desse trabalho pode-se entender melhor a metodologia de controle de qualidade multivariado e suas aplicações. A implementação do programa *MQCI* no *S-Plus* é de grande valia para analisar dados de controle multivariado, uma vez que alguns dos testes não estão implementados nos *softwares* estatísticos mais utilizados ou não estão no formato padrão da área de controle estatístico de qualidade. Neste trabalho vimos a implementação computacional do controle de  $T^2$  de Hotelling com a decomposição da estatística  $T^2$  e a metodologia proposta por Hayter & Tsui (1994) sendo que essa última não está ainda implementada em nenhum *software* estatístico . Brevemente o programa estará disponível para *download* na internet. Pretende-se num trabalho futuro implementar computacionalmente outras técnicas estatísticas multivariadas para controle de qualidade que ainda não estão disponíveis.

## Referências

- ALT, F. B. Multivariate Quality Control, *Encyclopedia of Statistical Sciences*, Vol. 6, editado por N.L. Johnson e S. Kotz, Wiley, New York, 1985.
- HAYTER, A. J.; TSUI, K. L. Identification and Quantification in Multivariate Quality Control Problems. *Journal of Quality Technology*, Atlanta, v. 26, n. 3, p. 197-208, julho. 1994.
- HOTELLING, H. Multivariate quality control. *Techniques of Statistical Analysis*, editado por Eisenhart, Hastay, e Wallis, MacGraw-Hill, New York, 1947.
- JACKSON, J. E. *A user's guide to principal components*. New York: John Wiley, 1991.
- JOHNSON, R. A.; WICHERN, W. D. *Applied Multivariate Statistical Analysis*. 4ª ed. New Jersey, 2002.
- MASON, R. L.; CHOU, Y.M.; YOUNG, J. C. Applying Hotelling's  $T^2$  statistic to batch processes, *Journal of Quality Technology*, 33, 4, 466-479, 2001.

- MASON, R. L.; TRACY, N.D.; YOUNG, J. C. A practical approach for interpreting multivariate  $T^2$  chart signals, *Journal of Quality Technology*, 29, 99-108,1997.
- MINGOTI, S. A., GLÓRIA, F. A. A. Uma proposta de modificação do índice de capacidade multivariado de Chen. In *Anais do XXIII ENEGEP*, Ouro Preto, Minas Gerais, 2003 (em cd-rom).
- MINGOTI, S. A. Pareto charts and principal components analysis combined. *Revista Produto & Produção*, 5,1.21-29, 2001.
- MONTGOMERY, D. C. *Introdução ao Controle Estatístico da Qualidade*.4ª ed. Rio de Janeiro, 2004.
- MUDHOLKAR, G. S.; SRIVASTAVA, D. K. A class of robust stepwise alternatives to Hotelling's  $T^2$  tests. *Journal of Applied Statistics*, New York, v. 27, n. 5, p. 599-619, 2000.
- MUDHOLKAR,G.S.; SRIVASTAVA, D. K. Robust analogs of Hotelling two sample  $T^2$ .*Communications in Statistics-Theory and Methods*, 1999.
- MUDHOLKAR, G. S.; SUBBAIAH, P. Testing a significance of a mean vector- a possible alternative do Hotelling's  $T^2$ , *Annals of the Institute of Statistical Mathematics*, 32, Part A, 43-52,1980.
- NIVERTI, M.; DEY, D. K. Multivariate Process Capability a Bayesian Perspective, *Communications in Statistics-Theory and Methods*, 2000.
- NOMIKOS, P.; MACGREGOR, J. F. Multivariate SPC charts for monitoring batch processes, *Technometrics*, 37, 41-59, 1995.
- REIS, E. A. *Noções Básicas de S-Plus for Windows*, 1997.
- RUNGER, G. C.; ALT, F. B.; MONTGOMERY, D. C. Controlling Multiple Stream Processes with Principal Components, *Communications in Statistics-Theory and Methods*, 1996a.
- RUNGER, G. C.; ALT, F. B.; MONTGOMERY, D. C. Contributors to a Multivariate Statistical Process Control Signal, *Communications in Statistics-Theory and Methods*, 1996b.
- SULLIVAN, J. H.; WOODALL, W. H. A Comparison of Multivariate Quality Control Charts for Individual Observations, *Journal of Quality Technology*, Vol. 27, 1995.
- TIKU, M. L.; BALAKRISHNAN ,N, Robust Hotelling-type  $T^2$  statistics based on modified maximum likelihood estimators, *Communications in Statistics-Theory and Methods*, 17,6, 1789-1810, 1988.
- TIKU, M. L.; SINGH, M. Robust statistics for testing mean vectors of multivariate distributions, *Communications in Statistics-Theory and Methods*, 11,9, 227-234, 1982.
- TRACY, N. D.; MASON, R. L.; YOUNG, J. C. Multivariate Control Charts for Individual Observations, *Journal of Quality Technology*, Vol. 24, 1992.
- VENABLES, W. N.; RIPLEY, B. D. *Modern Applied Statistics with S*, 4ª ed. New York, 2003.