

**Monitoramento e Controle de
Processo Multivariado**

Roteiro

1. O Problema do Controle da Qualidade Multivariado
2. Descrição de Dados Multivariados
3. Gráfico de Controle T^2 de Hotelling (Subgrupos)
4. Gráfico de Controle T^2 de Hotelling (O. Individuais)
5. Gráfico de Controle EWMA Multivariado
6. Ajustamento de Regressão
7. Gráficos de Controle para Monitoramento da Variabilidade
8. Referências

O Problema do Controle da Qualidade
Multivariado

Processo Multivariado

- Monitoramento e controle univariado:
 - √ Supõe que há apenas uma variável de saída
- Maioria dos cenários de monitoramento e controle de processos envolvem **várias variáveis relacionadas**
- Aplicação de cartas de controle univariadas a cada variável individual
 - √ Não é eficaz e pode levar a conclusões errôneas

- Gráficos de controle multivariados funcionam bem quando o número de variáveis não é muito grande (≤ 10)
- Os gráficos de controle multivariados perdem eficiência em detectar mudança à medida em que o número de variáveis cresce
 - √ Abordagem muito usada: diminuir a dimensão do problema (componentes principais)

- Há situações nas quais é necessário o monitoramento simultâneo ou controle de duas ou mais características de qualidade
- Exemplo: Características de qualidade de mancal
 - √ X_1 : diâmetro interno
 - √ X_2 : diâmetro externo

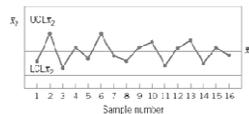
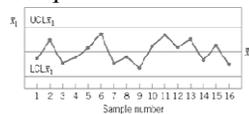
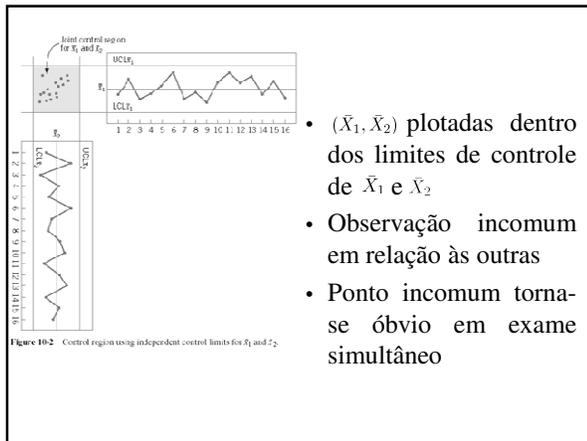


Figure 10-1 Control charts for inner (x_1) and outer (x_2) bearing diameters.



- (\bar{X}_1, \bar{X}_2) plotadas dentro dos limites de controle de \bar{X}_1 e \bar{X}_2
- Observação incomum em relação às outras
- Ponto incomum torna-se óbvio em exame simultâneo

• Suposição: X_1 e X_2 têm distribuições normais independentes – a :

$\alpha_1 = \alpha_2 = 0,0027 P\{\bar{X}_1(\bar{X}_2) \text{ exceder limites } 3\sigma\}$
 $(0,0027)^2 = 0,00000729 P\{\bar{X}_1 \cap \bar{X}_2 \text{ excederem limites } 3\sigma\}$
 $(0,9973)^2 = 0,99460729 P\{\bar{X}_1 \cap \bar{X}_2 \text{ dentro limites } 3\sigma\}$
 $1 - (0,9973)^2 = 0,99460729 P\{\bar{X}_1 \cup \bar{X}_2 \text{ exceder limites}\}$

• Uso de dois gráficos independentes distorce monitoramento simultâneo

√ Erro tipo I não é igual aos níveis anunciados para os gráficos individuais

• Distorção aumenta na medida em que aumenta o número de características de qualidade

• Em geral, se há p características de qualidade independentes e se α é mantido para cada carta

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

– P{erro tipo I} para o procedimento conjunto

$(1 - \alpha)^p$

– P{todas a médias plotadas dentro limites de controle}

- Se p características de qualidade não são independentes não há uma maneira fácil de medir a distorção no procedimento de controle conjunto
- Monitoramento e controle de processo multivariado
 - √ Hotelling, H. (1947) *Multivariate Quality Control*, In: Techniques of statistical Analysis, edited by Eisenhart, Hastay & Wallis, Mc-Graw hill, New York
- Assunto de particular importância nos dias de hoje

Descrição de Dados Multivariados

Normal Multivariada

- Suponha que tenhamos p variáveis X_1, X_2, \dots, X_p
 - √ Vetor de componentes $\mathbf{X}' = [X_1, X_2, \dots, X_p](X')$
 - √ Vetor de médias: $\boldsymbol{\pi}' = [\mu_1, \mu_2, \dots, \mu_p](\boldsymbol{\pi}')$
 - √ Matriz de variâncias e covariâncias

$$\Sigma_X = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1p} \\ & \sigma_{22} & \dots & \sigma_{2p} \\ & & \ddots & \vdots \\ & & & \sigma_{pp} \end{bmatrix}$$
 - √ Variância da variável aleatória X_i : $\text{Var}(X_i) = \sigma_{ii} = \sigma_i^2$
 - √ Covariância entre Variáveis X_i e X_j : $\text{Covar}(X_i, X_j) = \sigma_{ij}$

Função de Densidade de Probabilidade

- Distribuição Normal Univariada: distância quadrática padronizada

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

- Distribuição Normal Multivariada:

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} \exp \left\{ -\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})' \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right\}$$

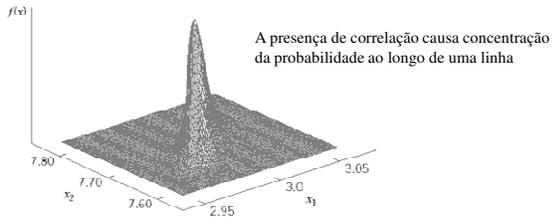
Padronização volume sob superfície

distância generalizada quadrática padronizada

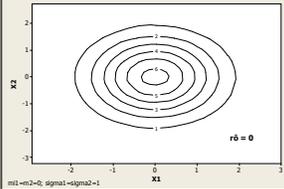
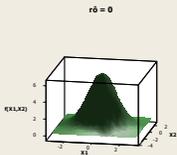
Normal Bivariada

- Função de densidade de probabilidade ($p=2$)

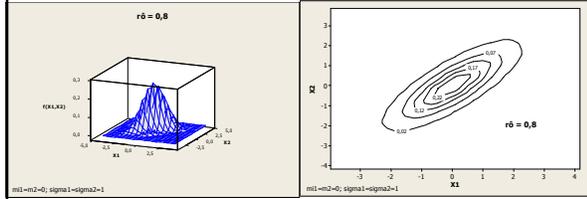
$$f_{X_1, X_2}(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho_{12}^2}} \exp \left\{ -\frac{1}{2(1-\rho_{12}^2)} \left[\left(\frac{x_1 - \mu_1}{\sigma_1} \right)^2 + \left(\frac{x_2 - \mu_2}{\sigma_2} \right)^2 - 2\rho_{12} \left(\frac{x_1 - \mu_1}{\sigma_1} \right) \left(\frac{x_2 - \mu_2}{\sigma_2} \right) \right] \right\}$$



- X_1 e X_2 independentes

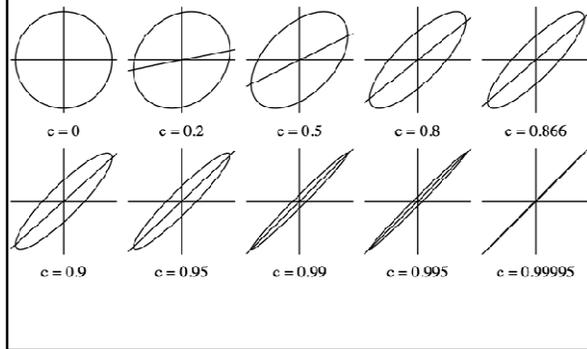


- $\text{Corr}(X_1, X_2) = 0,8$



√ A presença de correlação causa concentração da probabilidade ao longo de uma linha

Efeito Correlação



Vetor de Média Amostral

- Suponha uma amostra aleatória de uma distribuição normal multivariada $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$
- \mathbf{x}_i : i-ésimo vetor amostral
- $\mathbf{x}_i' = [x_{i1}, x_{i2}, \dots, x_{ip}]$
- Vetor de médias amostrais

$$\bar{\mathbf{x}} = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i$$

Matriz de Variâncias e Covariâncias Amostral

$$S = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x})'$$

- Variâncias Amostrais (diagonal de S)

$$S_{jj} = \frac{1}{n-1} \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2 \quad j = 1, 2, \dots, p$$

- Covariâncias amostrais

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

- \bar{x} e S são estimadores não-viciados de μ e Σ , respectivamente

$$E[\bar{X}] = \mu$$

$$E[S] = \Sigma$$

Gráfico de Controle T^2 de Hotelling para Dados Subgrupados

Gráfico de Controle T^2 de Hotelling

- Procedimento de monitoramento e controle do vetor de médias de processo multivariado
 - √ Análogo ao gráfico de \bar{x} de Shewhart;
 - √ Versões para dados subgrupados e observações individuais

Dados Subgrupados

- Suponha duas características de qualidade X_1 e X_2 distribuídas conjuntamente de acordo com uma distribuição normal bivariada
 - μ_1 e μ_2 : valores médios das características de qualidade
 - σ_{11} e σ_{22} : variâncias de X_1 e X_2
 - σ_{12} : covariância entre X_1 e X_2
- Suponha conhecidos σ_{11} , σ_{22} e σ_{12}
- \bar{x}_1 e \bar{x}_2 : médias amostrais das duas características de qualidade
(calculadas a partir de amostra de tamanho n)

- Com essas considerações

$$\chi_0^2 = \frac{n}{\sigma_{11}\sigma_{22} - \sigma_{12}^2} [\sigma_{22}(\bar{x}_1 - \mu_1)^2 + \sigma_{11}(\bar{x}_2 - \mu_2)^2 - 2(\bar{x}_1 - \mu_1)(\bar{x}_2 - \mu_2)\sigma_{12}]$$

$$\sqrt{\chi_0^2} \sim \chi^2_2$$

- Pode ser usada como base de um gráfico de controle para médias do processo μ_1 e μ_2
- Se as médias do processo permanecem no valores μ_1 e μ_2 (processo sob controle) então
 - $\chi_0^2 < LSC = \chi^2_2(\alpha)$
 - $P\{\chi_0^2 > \chi^2_2 \sim\} = \alpha$

- Se pelo menos uma das médias muda para algum novo valor (processo fora de controle), então:
 $P(\chi^2_0 > \chi^2_{2-\alpha})$ aumenta

- $\sigma_{12} = 0$
 X_1 e X_2 independentes
- χ^2_0 define elipse:
centro (μ_1, μ_2)
eixos paralelos a X_1, X_2
- Ponto dentro da elipse
Processo sob controle

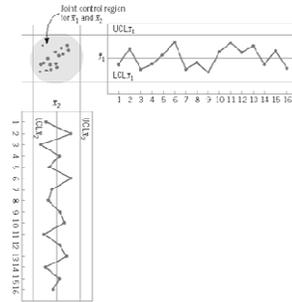
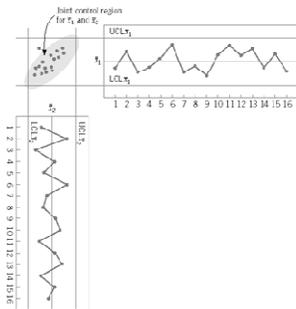


Figure 10-4 A control ellipse for two independent variables.

- $\sigma_{12} \neq 0$



- ✓ eixos principais não são mais paralelos aos eixos X_1 e X_2
- ✓ Ponto dentro dos limites dos gráficos individuais mas fora da elipse de controle

Desvantagens Associadas às Elipses

- Perde-se a seqüência temporal dos pontos plotados
✓ Pode-se numerar os pontos ou usar símbolos especiais
- Difícil construir a elipse para mais de duas características de qualidade
- Para evitar estas dificuldades
✓ Plota-se Gráfico de controle Qui-quadrado
✓ Valores de χ^2_0 em um gráfico de controle apenas com o limite de controle superior

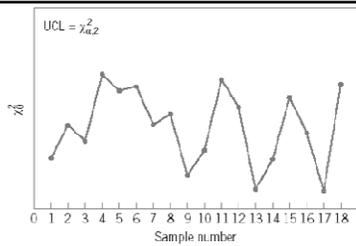


Figure 10-6 A chi-square control chart for $p = 2$ quality characteristics.

- LSC: $\chi^2_2(\alpha)$

Gráfico de Controle de Qui-quadrado – p Variáveis

- Supõe-se que a distribuição de probabilidade conjunta das p características da qualidade seja a distribuição normal p -variada

√ Vetor de médias da amostra $\bar{x}' = [\bar{x}_1, \bar{x}_2, \dots, \bar{x}_p]$

√ Vetor das médias sob controle: $\pi' = [\mu_1, \mu_2, \dots, \mu_p]$

√ Estatística de teste: $\chi_0^2 = n(\bar{x} - \mu)' \Sigma^{-1} (\bar{x} - \mu)$

√ LSC: $\chi^2_p(\alpha)$

Estimação de μ e Σ

- Estimação a partir de amostras preliminares de tamanho n tomadas quando se admite que o processo esteja sob controle

√ Admita m amostras disponíveis

$j=1, 2, \dots, p$ e $k=1, 2, \dots, m$

x_{ijk} : i -ésima observação da j -ésima característica da qualidade na k -ésima amostra

$$\bar{x}_{jk} = \frac{1}{n} \sum_{i=1}^n x_{ijk} \quad S_{jjk} = \frac{1}{n-1} \sum_{i=1}^n (x_{ijk} - \bar{x}_{jk})^2$$

$$S_{jhk} = \frac{1}{n-1} \sum_{i=1}^n (x_{ijk} - \bar{x}_{jk})(x_{ihk} - \bar{x}_{hk}) \quad j \neq k$$

- Médias das estatísticas sobre todas as m amostras

$$\bar{x}_j = \frac{1}{m} \sum_{k=1}^m \bar{x}_{jk}, j = 1, 2, \dots, p. \quad \bar{\mathbf{x}}' = [\bar{x}_1, \bar{x}_2, \dots, \bar{x}_p]$$

$$\bar{S}_{jj} = \frac{1}{m} \sum_{k=1}^m S_{jjk}, j = 1, 2, \dots, p.$$

$$\bar{S}_{jh} = \frac{1}{m} \sum_{k=1}^m S_{jhk}, j \neq h. \quad \mathbf{S} = \begin{bmatrix} \bar{S}_{11} & \bar{S}_{12} & \dots & \bar{S}_{1p} \\ & \bar{S}_{22} & \dots & \bar{S}_{2p} \\ & & \ddots & \vdots \\ & & & \bar{S}_{pp} \end{bmatrix}.$$

- \mathbf{S} é estimador não viciado de Σ quando o processo está sob controle

Gráfico de Controle de T^2

- Suponha:
 - √ \mathbf{S} estima Σ
 - √ $\bar{\mathbf{x}}$ valor sob controle do vetor de médias do processo
- Estatística T^2 de Hotelling: $T^2 = n(\bar{\mathbf{x}} - \bar{\mathbf{x}})' \mathbf{S}^{-1}(\bar{\mathbf{x}} - \bar{\mathbf{x}})$
- Deve-se ter cuidado na seleção dos limites de controle para a estatística T^2 de Hotelling;

Fases dos Gráficos de Controle

- Há duas fases distintas no uso de um gráfico de controle:
 - √ Fase 1: Uso dos gráficos para estabelecimento do controle
 - testar se o processo estava ou não sob controle quando os m subgrupos preliminares forem extraídos e as estatísticas amostrais calculadas
 - obtenção de conjunto de observações sob controle, de maneira a estabelecer os limites de controle da fase 2
 - √ Fase 2: Monitoramento da produção futura

Limites de Controle para o Gráfico T^2

- Fase 1: Estabelecimento do controle

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

- Fase 2: Monitoramento produção futura

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

- Obs:

$$\sqrt{LIC} = 0$$

Amostras Preliminares

- Quando μ e Σ são estimados a partir de um grande número de amostras preliminares costuma-se usar $LSC = \chi^2_p(\alpha)$
- Há situações em que são necessárias muitas amostras preliminares antes que os limites de controle exatos da fase 2 sejam bem aproximados pelos limites qui-quadrado
 - \sqrt{m} sempre superior a 20 amostras preliminares, em geral superior a 50 amostras

Exemplo

- Características da qualidade importantes para serem controladas conjuntamente
 - $\sqrt{X_1}$: Tensão de resistência (psi)
 - $\sqrt{X_2}$: Diâmetro (10^{-2} in)
 - $\sqrt{n = 10; m = 20}$

Interpretação Sinais de Fora de Controle

- Dificuldade na interpretação prática de um sinal fora de controle
 - √ Qual das p variáveis é responsável pelo sinal?
(ou qual subconjunto delas)
- Prática padrão:
 - √ Plotar gráficos \bar{X} univariados para as variáveis individuais
 - Abordagem pode não ser bem sucedida

Soluções

- Uso de gráficos com limites de controle do tipo Bonferroni (Alt, 1985)
 - √ Substituir $z_{\alpha/2}$ por $z_{\alpha/2p}$
 - √ Substitui o número de alarmes falsos associados ao uso de muitos gráficos univariados simultâneos
- Procedimento para intervalo de confiança simultâneos exatos (Hayter & Tsui, 1994)
 - √ Pode ser usado também em situações em que a hipótese de normalidade não é válida

- Gráficos de controle baseados nas p componentes principais (combinações lineares das variáveis originais) (Jackson, 1980)
 - √ Desvantagem: componentes principais em geral não fornecem uma interpretação clara da situação em relação às variáveis originais
 - √ Vantagem: eficaz em casos em que as componentes principais têm uma interpretação clara

- Decomposição da estatística T^2 em componentes que refletem a contribuição de cada variável (Runger, Alt & Montgomery, 1996)

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

d_i : indicador da contribuição relativa da i -ésima variável para a estatística global

T^2 : valor da estatística corrente

$T_{(i)}^2$: valor da estatística para todas as variáveis do processo exceto a i -ésima

- √ Quando um sinal de fora de controle é gerado, calculam-se os valores de $d_i, i=1,2,\dots,p$, concentrando a atenção nas variáveis para os quais os d_i são relativamente grandes

- Exemplo – (Runger, Alt & Montgomery, 1996)

√ $p = 3$ e Σ conhecida

√ Variáveis são escalonadas

$$y_{ij} = \frac{x_{ij} - \mu_j}{\sqrt{(m-1)\sigma_{22}}}$$

√ S está na forma de matriz de correlação

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

√ Valor sob controle da média do processo: $\pi' = [0, 0, 0]$

- S é conhecida \rightarrow LSC = $\chi^2_3(0,005)=12,84$

Observation Vector y'	Control Chart	$d_i = T^2 - T_{(i)}^2$		
	Statistic $T_{(i)}^2 (= d_i^2)$	d_1	d_2	d_3
(2, 0, 0)	27.14	27.14	6.09	6.09
(1, 1, -1)	26.79	6.79	6.79	25.73
(1, -1, 0)	20.00	14.74	14.74	0
(0.5, 0.5, 1)	15.00	3.69	3.68	14.74

√ Sugestão: Corte aproximado para a magnitude de um d_i individual $\chi^2_1(\alpha)$. Escolhendo $\alpha = 0,01$

$$\chi^2_1(0,01)=6,63$$

- Qualquer d_i que exceder este valor seria considerado grande colaborador

√ Decomposição T^2 dá diretriz clara sobre quais variáveis mudaram no vetor de observações

Gráfico de Controle T^2 de Hotelling para
Observações Individuais

Observações Individuais

- Em alguns ambientes industriais o tamanho do subgrupo é $n = 1$
 - √ Frequentemente, nas indústrias químicas e de processamento
 - √ Estas indústrias têm, em geral, múltiplas características de qualidade a serem monitoradas

- Suponha disponíveis m amostras, cada uma de tamanho $n = 1$ e que seja p o número de características de qualidade observadas em cada uma delas

- Sejam:

√ Vetor de médias amostral: $\bar{\mathbf{x}}' = [\bar{x}_1, \bar{x}_2, \dots, \bar{x}_p]$

√ Matriz de covariâncias amostral:

$$\mathbf{S} = \frac{1}{n-1} \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})' \quad \mathbf{S} = \begin{bmatrix} S_{11} & S_{12} & \cdots & S_{1p} \\ & S_{22} & \cdots & S_{2p} \\ & & \ddots & \vdots \\ & & & S_{pp} \end{bmatrix}$$

- Estatística T^2 de Hotelling:

$$T^2 = (\mathbf{x} - \bar{\mathbf{x}})' \mathbf{S}^{-1} (\mathbf{x} - \bar{\mathbf{x}})$$

- Limites de Controle da fase 2 – Monitoramento:

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

- Quando o número m de amostras preliminares é grande ($m > 100$), pode-se usar limites aproximados

$$LSC = \frac{p(m-1)}{m-m} F_{p, m-p}(\alpha) \quad \text{Razoável para } m > 100$$

$$LSC = \chi_p^2(\alpha) \quad \text{Apropriado apenas se } \Sigma \text{ é conhecida}$$

- Limites da Fase 1 – Estabelecimento Controle

√ Se $n=1$, limites da fase 1 devem se basear em distribuição Beta (Tracy, Young & Mason, 1992)

$$LSC = \frac{(m-1)^2}{m} \beta_{p/2, (m-p-1)/2}(\alpha)$$

√ Para os limites da fase, aproximações com base em distribuições F ou qui-quadrado serão, provavelmente, imprecisas

Estimação de Σ

- Problema significativo no caso de observações individuais

√ Sullivan & Woodall (1996): comparação estimadores

- Estimação usual:

√ Combinação de todas as m observações

$$S_1 = \frac{1}{m-1} \sum_{i=1}^m (\mathbf{x}_i - \bar{\mathbf{x}}) (\mathbf{x}_i - \bar{\mathbf{x}})'$$

√ Sensível a valores discrepantes ou observações fora de controle na amostra original de m observações

- Uso da diferença entre pares sucessivos de observações (Holmes & Mergen, 1993)

√ Vetor das i -ésimas diferenças:

$$v'_i = x_{i+1} - x_i, \quad i = 1, 2, \dots, m-1$$

√ Matriz desses vetores:

$$V = \begin{bmatrix} v'_1 \\ v'_2 \\ \vdots \\ v'_{m-1} \end{bmatrix}$$

√ Estimador para Σ : $S_2 = \frac{1}{2} \frac{V'V}{(m-1)}$

Exemplo

- Sullivan & Woodall (1996) usando os dados de Holmes e Mergen (1993) e a estatística T^2

√ Cálculo usando os estimadores S_1 e S_2

√ $m = 56$

√ Variáveis:

- % das classificações como grande (G), média (M) e pequena (P)

√ Usadas apenas as duas primeiras componentes (somam 100%)

Table 10-2 Example from Sullivan and Woodall (1995) Using the Data from Holmes and Mergen (1993) and the T^2 Statistics Using Estimators S_1 and S_2

i	$L=x_{i1}$	$M=x_{i2}$	$S=x_{i3}$	T^2_{i1}	T^2_{i2}	i	$L=x_{i1}$	$M=x_{i2}$	$S=x_{i3}$	T^2_{i1}	T^2_{i2}
1	5.4	93.6	1.0	4.496	6.429	29	7.4	83.6	9.0	1.594	3.261
2	3.2	92.6	4.2	1.739	4.227	30	6.8	84.8	8.4	0.912	1.743
3	5.2	91.7	3.1	1.460	2.200	31	6.3	87.1	6.6	0.110	0.266
4	3.5	85.9	9.6	4.933	7.643	32	6.1	87.2	6.7	0.077	0.166
5	2.9	90.4	6.7	2.690	5.565	33	6.6	87.3	6.1	0.255	0.564
6	4.6	92.1	3.3	1.272	2.258	34	6.2	84.8	9.0	1.358	2.069
7	4.4	91.5	4.1	0.797	1.676	35	6.5	87.4	6.1	0.203	0.448
8	5.0	99.3	4.7	0.337	0.045	36	6.0	86.8	7.2	0.193	0.317

$$V = \begin{bmatrix} -2,20 & -1,00 \\ 2,00 & -0,90 \\ \vdots & \vdots \\ 0,90 & 1,00 \end{bmatrix}_{55 \times 2}$$

- Estimação usual de Σ :

$$S_1 = \frac{1}{m-1} \sum_{i=1}^m (x_i - \bar{x})(x_i - \bar{x})' = \begin{bmatrix} 3,770 & -5,495 \\ -5,495 & 13,53 \end{bmatrix}.$$

$$S_2 = \frac{1}{2(m-1)} V'V = \begin{bmatrix} 1,562 & -2,093 \\ -2,093 & 6,721 \end{bmatrix}.$$

- Estatísticas T^2 :

$$T_{11}^2 = [5,4 - 5,68; 93,6 - 88,22] \begin{bmatrix} 3,770 & -5,495 \\ -5,495 & 13,53 \end{bmatrix}^{-1} \begin{bmatrix} -0,28 \\ 5,38 \end{bmatrix} = 1,594$$

$$T_{21}^2 = [5,4 - 5,68; 93,6 - 88,22] \begin{bmatrix} 1,562 & -2,093 \\ -2,093 & 6,721 \end{bmatrix}^{-1} \begin{bmatrix} -0,28 \\ 5,38 \end{bmatrix} = 3,261$$

- Limites de controle exatos por simulação (Sullivan & Woodall, 1996)

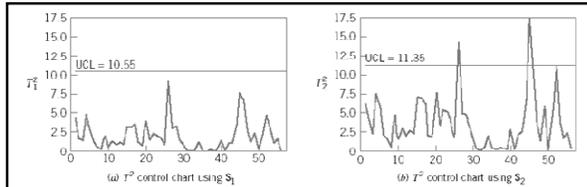


Figure 10-8 T^2 control charts for the data in Table 10-2.

- √ Se consideradas amostras 1 a 24: $\bar{x}'_{1-24} = [4,23; 90,8]$
- √ Se consideradas amostras 25 a 56: $\bar{x}'_{25-56} = [6,77; 86,3]$
- √ Matrizes de covariâncias “dentro” não são significativamente diferentes
- √ Há uma aparente mudança do vetor de médias em seguida à amostra 24:
 - Corretamente detectado em S_2

Gráfico de Controle *EWMA* Multivariado

EWMA Multivariado

- Gráficos χ^2 e T^2 são gráficos de controle do tipo Shewhart
 - √ Usam informação apenas da amostra corrente e são, conseqüentemente insensíveis a mudanças pequenas e moderadas no vetor de médias

- Lowry et al. (1992) desenvolveram versão multivariada do gráfico de controle EWMA:

$$z_i = \lambda x_i + (1 - \lambda)z_{i-1}, \text{ com } 0 < \lambda \leq 1 \text{ e } z_0 = \mathbf{0}$$

√ Quantidade plotada no gráfico: $T_i^2 = z_i' \Sigma_{z_i}^{-1} z_i$

√ Matriz de Covariâncias: $\Sigma_{z_i} = \frac{\lambda}{2 - \lambda} [1 - (1 - \lambda)^{2i}] \Sigma$

√ $\lambda=1$; EWMA multivariado é equivalente ao gráfico de controle T^2 (ou qui-quadrado)

- Análise completa do NMA de EWMA multivariado usando abordagem por cadeia de Markov

√ Limite H escolhido para dar $NMA_0 = 200$

√ Tamanho da mudança (parâmetro de não-centralidade)

$$\delta = (\mu' \Sigma^{-1} \mu)^{1/2}$$

- valores grandes de $\delta \rightarrow$ mudanças maiores na média
- $\delta = 0$ (estado sob controle)

Table 10-3 Average Run Lengths (zero state) for the MEWMA Control Chart [from Prabhu and Runger (1997)]

p	δ	λ							
		0.05	0.10	0.20	0.30	0.40	0.50	0.60	0.80
2		$H = 7.35$							
	0.0	199.93	199.98	199.91	199.82	199.83	200.16	200.04	200.20
	0.5	26.61	28.07	35.17	44.00	53.82	64.07	74.50	95.88
	1.0	11.23	10.15	10.20	11.36	13.26	15.88	19.24	28.65
	1.5	7.14	6.11	5.49	5.48	5.78	6.36	7.25	10.28
	2.0	5.28	4.42	3.78	3.56	3.53	3.62	3.84	4.79
3.0	3.56	2.93	2.42	2.20	2.05	1.95	1.90	1.91	
4		$H = 11.22$							
	0.0	199.84	200.12	199.94	199.91	199.96	200.05	199.99	200.05
	0.5	32.29	35.11	46.30	59.28	72.43	85.28	97.56	120.27
	1.0	13.48	12.17	12.67	14.81	18.12	22.54	28.06	42.58

- ✓ Para um dado δ , os NMAs geralmente tendem a crescer na medida que λ cresce
 - (exceto para valores muito grandes de δ)
- ✓ Para $\lambda = 1$ (gráfico T^2/χ^2) o EWMA multivariado é mais sensível a mudanças menores

Gráficos EWMA – Planejamento

- Gráficos EWMA multivariados ótimos:
 - ✓ Várias mudanças (δ) e valores-alvo sob controle NMA_0 de 500 ou 1.000
 - ✓ $NMA_{\text{mínimo}}$: valor mínimo alcançado pelo NMA_1 para valor especificado de λ

Table 10-3 Average Run Lengths (zero state) for the MEWMA Control Chart [from Prabhu and Runger (1997)]

p	δ	λ							
		0.05	0.10	0.20	0.30	0.40	0.50	0.60	0.80
2		$H = 7.35$							
	0.0	199.93	199.98	199.91	199.82	199.83	200.16	200.04	200.20
	0.5	26.61	28.07	35.17	44.00	53.82	64.07	74.50	95.88
	1.0	11.23	10.15	10.20	11.36	13.26	15.88	19.24	28.65
	1.5	7.14	6.11	5.49	5.48	5.78	6.36	7.25	10.28
	2.0	5.28	4.42	3.78	3.56	3.53	3.62	3.84	4.79
3.0	3.56	2.93	2.42	2.20	2.05	1.95	1.90	1.91	
4		$H = 11.22$							
	0.0	199.84	200.12	199.94	199.91	199.96	200.05	199.99	200.05
	0.5	32.29	35.11	46.30	59.28	72.43	85.28	97.56	120.27
	1.0	13.48	12.17	12.67	14.81	18.12	22.54	28.06	42.58

Table 10-4 Optimal MEWMA Control Charts

δ	ARL ₀ =	$p = 4$		$p = 10$		$p = 20$	
		500	1000	500	1000	500	1000
0.5	λ	0.04	0.03	0.03	0.025	0.03	0.025
	H	13.37	14.68	22.69	24.70	37.09	39.63
	ARL _{min}	42.22	49.86	55.94	66.15	70.20	83.77
1.0	λ	0.105	0.09	0.085	0.075	0.075	0.065
	H	15.26	16.79	25.42	27.38	40.09	42.47
	ARL _{min}	14.60	16.52	19.29	21.74	24.51	27.65
1.5	λ	0.18	0.18	0.16	0.14	0.14	0.12
	H	16.03	17.71	26.58	28.46	41.54	43.80
	ARL _{min}	7.65	8.50	10.01	11.07	12.70	14.01
2.0	λ	0.28	0.26	0.24	0.22	0.20	0.18
	H	16.40	18.06	27.11	29.07	43.15	44.45
	ARL _{min}	4.82	5.30	6.25	6.84	7.88	8.60
3.0	λ	0.52	0.46	0.42	0.40	0.36	0.34
	H	16.84	18.37	27.55	29.45	42.80	45.08
	ARL _{min}	2.55	2.77	3.24	3.50	4.04	4.35

Note: ARL₀ and ARL_{min} are zero-state average run lengths.

Exemplo

- Suponha $p=6$ e matriz de covariâncias dada por:

$$\Sigma = \begin{bmatrix} 1 & 0,7 & 0,9 & 0,3 & 0,2 & 0,5 \\ 0,7 & 1 & 0,8 & 0,1 & 0,4 & 0,2 \\ 0,9 & 0,8 & 1 & 0,1 & 0,2 & 0,1 \\ 0,3 & 0,1 & 0,1 & 1 & 0,2 & 0,1 \\ 0,2 & 0,4 & 0,2 & 0,2 & 1 & 0,1 \\ 0,5 & 0,2 & 0,1 & 0,1 & 0,1 & 1 \end{bmatrix}$$

√ Interesse: mudança no processo de $\mu' = \mathbf{0}$ para $\mu' = [1,1,1,1,1,1]$

– essencialmente mudança de um σ para cima em todas as p variáveis

$\delta = (\mu' \Sigma^{-1} \mu)^{1/2} = 1,75$ Tab. 10.3 \rightarrow NMA₀ = 200 (sob controle)
NMA₁ entre 4,88 e 7,32

$\lambda = 0,2$ e $H = 17,51$

Gráfico de Controle T^2 de Hotelling – Considerações Gerais

Considerações Gerais

- Gráfico de Controle T^2 de Hotelling (χ^2):

√ Idéia geral:

– Testar a hipótese de que o vetor de médias de uma distribuição normal multivariada é igual a uma constante, contra a hipótese alternativa de que o vetor de médias não é igual àquela constante

√ É uma estatística de teste ótima para essa hipótese

√ Não é necessariamente um procedimento ótimo de gráfico de controle para detectar mudanças na média

√ O EWMA multivariado pode ser planejado para ter uma capacidade de detecção mais rápida (menores valores de NMA_1)

√ T^2 de Hotelling não é ótimo para mudanças mais estruturadas na média, tais como mudanças em apenas algumas das variáveis do processo

√ T^2 de Hotelling e qualquer método que use a estrutura de forma quadrática (tal como EWMA multivariado) serão sensíveis a mudanças na variância bem como na média

√ Foram desenvolvidos métodos para monitorar processos que não dependem da estatística de teste T^2 de Hotelling

Ajustamento de Regressão

Ajustamento de Regressão

- Desenvolvido por Hawkins (1991)
- Procedimento:
 - √ Plotar gráficos de controle univariado para os resíduos de cada variável, quando aquela variável é regredida sobre todas as outras
 - √ Bastante aplicável a medidas individuais

- Aplicação muito importante:
 - √ Situações em que o processo tem uma hierarquia distinta de variáveis tais como um conjunto de variáveis de entrada do processo (digamos os x 's) e um conjunto de variáveis de saída (digamos os y 's)
 - Ex.:

Exemplo

- Processo em cascata
 - √ 40 observações, com 9 variáveis de entrada e 2 de saída
 - √ Demonstração do ajustamento utilizará apenas uma das variáveis de saída (y_1)

Table 10-5 Cascade Process Data

Observation	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	y_1	Residuals	y_2
1	12.79	0.15	91	56	1.54	7.20	1.75	5.09	1.11	951.5	0.81498	87
2	14.97	0.1	90	49	1.54	7.14	1.71	5.91	1.109	952.2	-0.31685	88
3	15.43	0.07	90	41	1.47	7.33	1.64	5.92	1.104	952.3	-0.28369	86
4	14.95	0.12	89	43	1.54	7.21	1.93	5.71	1.103	951.8	-0.45924	89
5	16.17	0.1	83	42	1.67	7.23	1.86	5.63	1.103	952.3	-0.36512	86
6	17.25	0.07	84	54	1.49	7.15	1.68	5.8	1.099	952.2	-0.22592	91
7	16.57	0.12	89	61	1.64	7.23	1.82	5.88	1.096	950.2	-0.55431	90
8	19.31	0.08	99	60	1.46	7.74	1.69	6.13	1.092	950.5	-0.18874	10
9	18.75	0.04	99	52	1.89	7.57	2.02	6.27	1.084	950.6	-0.15245	10
10	16.99	0.09	98	57	1.66	7.51	1.82	6.38	1.086	949.8	-0.33580	10

- Gráfico de controle para observações individuais e gráfico de controle de amplitude móvel
 - √ 40 observações de y_1

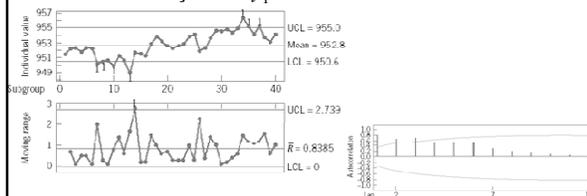


Figure 10-9 Individuals and moving range control charts for y_1 from Table 10-5. Figure 10-11 Sample autocorrelation function for y_1 from Table 10-5.

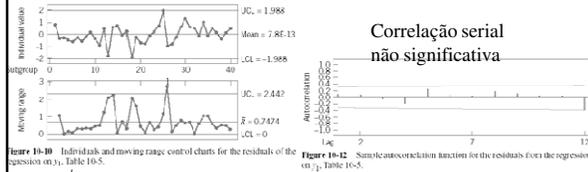
- √ Há 7 pontos fora de controle para as observações individuais
- √ Autocorrelação significativa entre saídas (y_1)

- Ajuste de regressão de mínimos quadrados:

$$\hat{y}_1 = 826 + 0,474x_1 + 1,41x_2 - 0,117x_3 - 0,0824x_4 - 2,93x_5 - 1,30x_6 + 2,18x_7 + 2,98x_8 + 113x_9$$

✓ Resíduos: $y_i - \hat{y}_i$

- Gráfico de observações individuais para resíduos



Correlação serial
não significativa

✓ Apenas um ponto fora de controle (carta AM)

✓ Impressão geral sobre estabilidade processo diferente

- Outra boa característica:
 ✓ Se é incluído no modelo o conjunto apropriado de variáveis, os resíduos do modelo serão tipicamente não-correlacionados, mesmo que a variável de interesse original (y_i) tenha exibido correlação
- procedimento de ajustamento de regressão tem muitas aplicações possíveis nas indústrias químicas e de processamento, onde há, frequentemente:
 ✓ processos em cascata com várias entradas e apenas poucas saídas
 ✓ muitas das variáveis são altamente correlacionadas

Gráficos de Controle para Monitoramento da Variabilidade

Monitoramento Multivariado de Variabilidade

- Introdução ao problema e apresentação de dois procedimentos úteis (Alt, 1985)
- 1º Procedimento
 - √ Extensão direta do gráfico de controle univariado S^2
 - √ É equivalente a testes de significância repetidos da hipótese de que a matriz de covariâncias é igual a uma particular matriz de constantes Σ .

- Estatística plotada no gráfico para a i -ésima amostra:

$$W_i = -pn + pn \ln(n) - n \ln \left(\frac{|A_i|}{|\Sigma|} \right) + \text{tr}(\Sigma^{-1} A_i)$$

- √ S_i : matriz de covariâncias amostral para a amostra i
- √ tr : operador traço (soma dos elementos da diagonal principal)
- Limite Superior de Controle

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

- 2º Procedimento:

- √ Baseia-se na variância amostral generalizada $|S|$
 - Determinante da matriz de covariâncias amostral
 - Medida amplamente usada para dispersão multivariada

- √ Montgomery & Wadsworth (1972)
 - Usaram aproximação normal assintótica

- √ Outro método:
 - usar média e variância de $|S|$
 - maior parte da distribuição de probabilidade está contida no intervalo

$$E(|S|) \pm 3\sqrt{\text{Var}(|S|)}$$

- Pode-se mostrar que:

$$E(|S|) = b_1|\Sigma|, \text{ com } b_1 = \frac{1}{n-1} \prod_{i=1}^p (n-i)$$

e

$$\text{Var}(|S|) = b_2|\Sigma|, \text{ com}$$

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

- Parâmetros do gráfico de controle para |S|:

$$\text{LSC} = |\Sigma|(b_1 + 3\sqrt{b_2})$$

$$\text{LC} = b_1|\Sigma|$$

$$\text{LIC} = |\Sigma|(b_1 - 3\sqrt{b_2})$$

√ Usualmente, Σ será estimada por uma matriz de covariâncias S , baseada na análise de amostras preliminares

- Substituir $|\Sigma|$ por $|S|$, pois $|S|/b_1$ é não viciada para Σ .

Exemplo 10.2 – Fibra Textil

- Variáveis: Tensão de resistência e diâmetro fibra

Ex. 10-4 Data for Example 10-1

Sample Number	(a) Means		(b) Variances and Covariances			(c) Control Chart Statistics	
	Tensile Strength (\bar{y}_{1j})	Diameter (\bar{y}_{2j})	s_{11}^2	s_{22}^2	s_{12}	\bar{y}_j^*	$ S_j $
1	115.25	1.34	1.25	0.57	0.30	2.36	0.45
2	115.51	1.36	1.26	0.55	0.31	2.11	0.41
3	115.05	1.39	1.30	0.50	0.32	0.77	0.39
4	116.21	1.35	1.02	0.55	0.31	8.29	0.21
5	115.00	1.37	1.16	0.73	0.30	1.89	0.21
6	115.25	1.36	1.01	0.80	0.76	0.03	0.25
7	114.98	1.35	1.22	0.78	0.75	7.54	0.41
8	115.25	1.40	1.46	0.43	0.30	3.41	0.52
9	116.15	1.39	1.19	0.57	0.33	5.52	0.35
10	115.02	1.35	1.17	0.86	0.65	2.01	0.19
11	115.75	0.90	1.45	0.70	0.78	1.13	0.51
12	114.90	1.36	1.28	0.52	0.31	9.56	0.36
13	115.01	1.35	1.26	0.55	0.72	3.86	0.17
14	115.85	1.37	1.17	0.76	0.79	1.11	0.33
15	115.29	1.11	1.23	0.89	0.32	2.56	0.42
16	115.63	1.34	1.24	0.91	0.53	0.85	0.44
17	115.47	1.33	1.26	0.35	0.70	0.39	0.65
18	115.28	1.35	1.13	0.53	0.79	0.80	0.36
19	115.72	1.36	1.31	0.89	0.76	0.55	0.59
20	115.40	1.34	1.20	0.85	0.68	0.62	0.63

verages $\bar{y}_1 = 115.89$ $\bar{y}_2 = 1.36$ $\bar{y}_1^* = 1.21$ $\bar{y}_2^* = 0.83$ $r_{12} = 0.79$

- Estimaco Σ com base nas 20 amostras preliminares:

$$S = \begin{bmatrix} 1,23 & 0,79 \\ 0,79 & 0,83 \end{bmatrix} \Rightarrow |S| = 0,3968$$

- Cculo constantes b_1 e b_2 :

$$b_1 = \frac{1}{n-1} \prod_{i=1}^p (n-i)$$

$$= \frac{1}{10-1} (10-1)(10-2) = \frac{(8)(9)}{81}$$

$$b_1 = 0,8889$$

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

$$= \frac{1}{(10-1)^{2 \times 2}} (10-1)(10-2) [(10-1+2)(10-2+2) - (10-1)(10-2)]$$

$$= \frac{(9)(8)[(11)(10) - (9)(8)]}{6561} = 0,4170$$

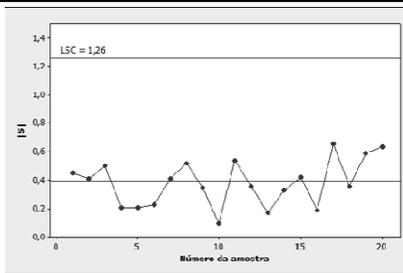
- Substituindo $|S|$ por $|S|/b_1$: $\frac{|S|}{b_1} = \frac{0,3968}{0,8889} = 0,4464$

- Limites de Controle

$$LSC = \frac{|S|}{b_1} (b_1 + 3\sqrt{b_2}) = 0,4464[0,8889 + 3\sqrt{0,4170}] = 1,26$$

$$LC = |S| = 0,3968$$

$$LIC = \frac{|S|}{b_1} (b_1 - 3\sqrt{b_2}) = 0,4464[0,8889 - 3\sqrt{0,4170}] = -0,47 = 0$$



- Nenhum ponto excede limites de controle \rightarrow variabilidade do processo sob controle

Comentários

- Embora a variância amostral generalizada seja uma medida de dispersão multivariada amplamente usada, lembre-se de que é uma representação escalar relativamente simplista de um problema multivariado complexo

- É fácil se enganar se olharmos apenas para $|\mathbf{S}|$

√ Exemplo:

$$S_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$S_2 = \begin{bmatrix} 2,32 & 0,40 \\ 0,40 & 0,50 \end{bmatrix}$$

$$S_3 = \begin{bmatrix} 1,68 & -0,40 \\ -0,40 & 0,50 \end{bmatrix}$$

√ $|\mathbf{S}_1| = |\mathbf{S}_2| = |\mathbf{S}_3| = 1$, mas as três matrizes transmitem informação consideravelmente diferente sobre a variabilidade do processo e sobre a correlação entre as variáveis

- Provavelmente, é boa idéia usar gráficos de controle univariados para a variabilidade em conjunção com o gráfico de controle para $|\mathbf{S}|$

Referências

Bibliografia Recomendada

- Montgomery, D. C. (LTC)
Introdução ao Controle Estatístico de Qualidade
- Mitag, H.-J. e Rinne, H. (Chapman & Hall)
Statistical Methods of Quality Assurance
