

# ***ESTUDOS DE CAPACIDADE PARA DADOS NÃO-NORMAIS***

***Alberto Wunderler Ramos***

Departamento de Engenharia de Produção - EPUSP

Caixa Postal 61548 São Paulo – SP – Brasil – 05424-970

*awramos@usp.br*

*Abstract:*

*This paper presents a method for calculating capability indexes when data from process is non-normal. It also introduces the Johnson system of curves for data fitting and shows how the indexes need to be adapted for this case. At the end, a complete example is given for a more complete understanding.*

*Key-words: process capability, non-normal data, Johnson system of curves.*

## ***1. INTRODUÇÃO***

Os estudos de capacidade (ou capabilidade, do inglês *capability*) têm por objetivo básico avaliar se um determinado processo consegue gerar produtos que atendam às especificações de engenharia.

Resumidamente, a metodologia consiste em obter uma amostra de produtos fabricados em condições normais de operação, avaliar a estabilidade estatística do processo mediante o emprego de gráficos de controle e, a seguir, determinar a capacidade do processo gerar produtos conformes.

Embora de uso comum no meio industrial, tais estudos são realizados muitas vezes em situações onde os dados não demonstram aderência com a distribuição de probabilidade normal. Alguns softwares estatísticos, tais como Minitab, Statistica e Applied Stats apresentam a opção do emprego da família de curvas de Johnson. Contudo, poucas pessoas fora do meio acadêmico estão familiarizadas com esta

possibilidade e, portanto, este artigo discute e apresenta os seus detalhes. Para maior facilidade de entendimento, um exemplo completo é mostrado.

## 2. ÍNDICES DE CAPACIDADE TRADICIONAIS

A avaliação da capacidade do processo costuma ser realizada mediante o cálculo de índices de capacidade. Existe uma grande variedade destes índices, segundo BOTHE (1997). Entretanto, dois índices são mais frequentemente utilizados:  $C_m$  e  $C_{mk}$ , conforme FINLEY (1992).

O primeiro índice, conhecido como  $C_m$ , é definido como sendo a razão entre a tolerância de engenharia e a dispersão do processo, matematicamente:

$$C_m = \frac{LSE - LIE}{6\sigma} \quad (1)$$

onde:

LSE - limite superior da especificação

LIE - limite inferior da especificação

$\sigma$  - desvio-padrão do processo (população)

Como o desvio-padrão do processo ( $\sigma$ ) é desconhecido, utiliza-se uma estimativa ( $\hat{\sigma}$ ) calculada a partir das amostras obtidas, ou seja

$$\hat{C}_m = \frac{LSE - LIE}{6\hat{\sigma}} \quad (2)$$

O índice  $C_m$  compara a variabilidade total permitida ao produto (ou tolerância de especificação) com a variação do processo (também chamada de tolerância natural). Assim, para o processo ser considerado capaz, o índice  $C_m$  deve ser igual ou maior do que 1, o que equivale a dizer que pelo menos 99,73% dos produtos serão conformes, admitindo-se a distribuição normal válida para a variabilidade dos valores individuais e a média do processo centralizada na especificação.

Intrinsecamente, este índice admite que a média da máquina pode ser facilmente ajustada e, portanto, somente a tolerância de engenharia (que é a distância entre o limite superior e o inferior da especificação) é comparada com a dispersão total. Esta é sempre a melhor condição possível para o estudo, daí o porquê do índice ser habitualmente chamado de capacidade potencial.

O outro índice, conhecido por  $C_{mk}$ , é definido como

$$C_{mk} = \text{mínimo} \{C_{mi}, C_{ms}\} \quad (3)$$

onde  $C_{mi}$  é calculado mediante

$$C_{mi} = \frac{\mu - LIE}{3.\sigma} \quad (4)$$

Como  $\mu$  e  $\sigma$  são geralmente desconhecidos, usam-se as estimativas da média ( $\hat{\mu}$ ) e do desvio-padrão ( $\hat{\sigma}$ ) do processo. Daí segue que um estimador de  $C_{mi}$  é fornecido por

$$\hat{C}_{mi} = \frac{\hat{\mu} - LIE}{3.\hat{\sigma}} \quad (5)$$

Analogamente, define-se o índice  $C_{ms}$  como sendo

$$C_{ms} = \frac{LSE - \mu}{3.\sigma} \quad (6)$$

e, empregando-se as estimativas da média e desvio-padrão do processo, um estimador de (6) é dado por

$$\hat{C}_{ms} = \frac{LSE - \hat{\mu}}{3.\hat{\sigma}} \quad (7)$$

$C_{mk}$  é recomendado nos casos de especificações unilaterais, onde inexistem LIE ou LSE, ou quando a média do processo não pode ser centralizada na especificação por problema de engenharia ou devido a um elevado custo de alteração ou ajuste.

No índice  $C_{mk}$ , além de avaliar-se a variabilidade total permitida às peças com a tolerância natural de fabricação, verifica-se, também, a posição do processo em relação aos limites (superior e inferior) da especificação. Assim, o valor de  $C_{mk}$  deve ser igual ou superior a 1 para o processo ser considerado capaz. Este índice é, também, conhecido pelo nome de capacidade de máquina.

### **3. ÍNDICES DE CAPACIDADE PARA DADOS NÃO NORMAIS**

Quando os valores individuais não mais apresentam distribuição normal, a interpretação dos índices tradicionais  $C_m$  e  $C_{mk}$  não é mais válida.

Nesta situação, para haver equivalência em termos de área com a distribuição normal (99,73%), pode-se usar os percentis 0,135 ( $P_I$ ) e 99,865 ( $P_S$ ) da outra distribuição, e calcular-se os índices de capacidade através de:

$$\hat{C}_{m^*} = \frac{LSE - LIE}{P_S - P_I} \quad (8)$$

e  $\hat{C}_{mk^*} = \min\{\hat{C}_{mi^*}, \hat{C}_{ms^*}\}$ , com

$$\hat{C}_{mi^*} = \frac{Md - LIE}{Md - P_I} \quad (9)$$

ou ainda

$$\hat{C}_{ms^*} = \frac{LSE - Md}{P_S - Md} \quad (10)$$

onde Md é a mediana (ou percentil 50,000).

Para avaliação destes três últimos índices, existem três alternativas, segundo PYZDEK (1992a, p.420):

- a) Obter uma distribuição de probabilidade que represente satisfatoriamente o fenômeno (uso de distribuições não-normais);
- b) Ajustar uma curva empírica sobre o conjunto de dados;
- c) Trabalhar sem assumir nenhuma distribuição dos dados (modelo não-paramétrico).

No primeiro caso, há realmente uma modelagem do processo, ou seja, com base no conhecimento de seu comportamento (em termos de variação), seleciona-se um modelo probabilístico adequado e, daí em diante, admite-se que este seja válido no futuro. No segundo caso, emprega-se um ajuste empírico de curvas, sem considerar a dinâmica do processo, seja por falta de conhecimento sobre o processo, seja por inviabilidade em obter um modelo probabilístico adequado à sua modelagem. No último caso, nenhuma distribuição de probabilidade é assumida *a priori*, buscando-se contornar as dificuldades das alternativas anteriores.

Obviamente, pelo exposto antes, sempre é preferível utilizar a primeira alternativa, mas, na prática, é muito pouco provável que se tenha um conhecimento

adequado sobre o processo ou produto, que torne possível justificar porque uma certa distribuição de probabilidade é mais adequada para representar o fenômeno estudado do que outra qualquer.

#### **4. USO DE DISTRIBUIÇÕES EMPÍRICAS**

Uma possibilidade para se resolver o problema das distribuições não-normais é aquela em que se ajusta uma distribuição empírica aos dados obtidos do processo. Após o ajuste desta curva, pode-se calcular a porcentagem de itens acima do máximo e abaixo do mínimo da especificação e transformá-los em  $\hat{C}_m$  e  $\hat{C}_{mk}$  equivalentes.

Um método de realizar este ajuste é o sistema de JOHNSON (1949). Este apresenta algumas vantagens, conforme FARNUM (1997), em relação a outras formas desenvolvidas anteriormente (Gram-Charlie ou Pearson):

- a) o sistema de curvas de Johnson cobre uma vasta gama de diferentes formas de distribuições, maior do que as disponíveis nos outros sistemas sendo, portanto, mais flexível;
- b) o mecanismo para seleção da família de curvas mais adequada aos dados é mais simples que nos demais casos, especialmente considerando-se os aperfeiçoamentos feitos por SLIFKER; SHAPIRO (1980), que substituíram o uso dos momentos amostrais, quantidades imprecisas e com altos vieses, pela utilização de percentis;
- c) este sistema permite que as probabilidades (áreas sobre as curvas ajustadas) sejam calculadas a partir do simples uso de tabelas da distribuição normal padronizada, mediante transformações adequadas das funções;
- d) o sistema de Johnson já se encontra disponível em diversos softwares estatísticos do mercado.

#### **5. CURVAS DE JOHNSON**

O sistema de curvas de Johnson é basicamente constituído por três diferentes famílias de distribuições, geradas a partir de transformações da seguinte forma genérica:

$$z = \gamma + \eta k_i(x; \omega, \kappa) \quad (11)$$

onde  $z$  é a variável normal padronizada (ou reduzida), e  $\eta, \gamma, \omega$  e  $\kappa$  são parâmetros específicos para cada família. Em função do tipo de família selecionada no ajuste, Johnson propôs os seguintes tipos de funções:

- **Família  $S_U$** , ou seja, não limitada (do inglês, unbounded), quando a variável não possui um valor mínimo ou máximo, podendo variar livremente entre  $-\infty$  e  $+\infty$ . Nesta situação, a função proposta é do tipo:

$$k_1(x, \omega, \kappa) = \sinh^{-1}\left(\frac{x - \kappa}{\omega}\right) \quad (12)$$

- **Família  $S_B$** , ou seja, limitada (do inglês, bounded), para aquelas distribuições nas quais a variável assume valores entre  $(\kappa, \kappa + \omega)$ . A função proposta para este caso é do tipo:

$$k_2(x, \omega, \kappa) = \ln\left(\frac{x - \kappa}{\omega + \kappa - x}\right) \quad (13)$$

- **Família  $S_L$** , ou seja, lognormal, cuja função proposta é do tipo:

$$k_3(x, \omega, \kappa) = \ln\left(\frac{x - \kappa}{\omega}\right) \quad (14)$$

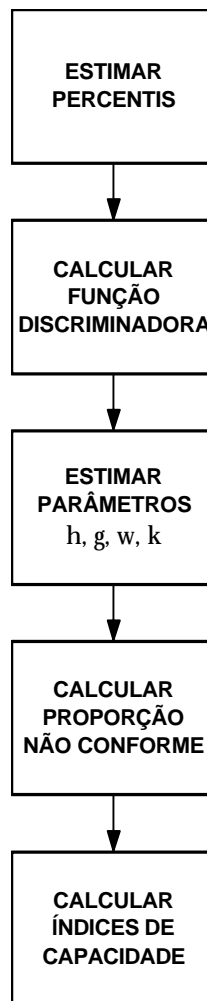
Basicamente, existem 5 grandes etapas na aplicação da metodologia de ajuste de curvas de Johnson, conforme apresentadas na Figura 1.

### 5.1 Estimação dos Percentis da Amostra

A primeira etapa é escolher qual dentre as três famílias de curvas é mais adequada no ajuste aos dados. Isto é feito mediante a obtenção de certos percentis da amostra e do cálculo de uma função discriminadora.

Inicialmente, é preciso escolher um valor arbitrário de  $0 < z < 1$ . Segundo SLIFKER; SHAPIRO (1980), idealmente o valor de  $z$  deve ser próximo a 0,5, já que isto colocará a quantidade  $\pm 3z$  não muito distante na cauda da distribuição selecionada, evitando-se obter estimativas pouco confiáveis. A seguir, determina-se as

probabilidades acumuladas  $\Phi(-3z)$ ,  $\Phi(-z)$ ,  $\Phi(z)$  e  $\Phi(3z)$  associadas a estes quatro números  $(-3z, -z, z$  e  $3z)$ , mediante o uso de uma tabela normal padronizada.



**Figura 1** – Metodologia para ajuste de curvas de Johnson.

## 5.2 Cálculo da Função Discriminadora

Prosseguindo-se com o método, as probabilidades acumuladas são empregadas para se encontrar os percentis da amostra, respectivamente,  $x_{-3z}$ ,  $x_{-z}$ ,  $x_z$  e  $x_{3z}$ , o que costuma requerer interpolações no seu cálculo. Calcula-se, então, a função discriminadora:

$$\frac{mn}{p^2} \quad (15)$$

onde



$$m = x_{3z} - x_z \quad (16)$$

$$n = x_{-z} - x_{-3z} \quad (17)$$

$$p = x_z - x_{-z} \quad (18)$$

Se  $mn/p^2 > 1$ , a família  $S_U$  deve ser a escolhida. Se  $mn/p^2 < 1$ ,  $S_B$  é a melhor opção e, se  $mn/p^2 = 1$ , então  $S_L$  deve ser preferida.

### 5.3 Estimação dos Parâmetros

Uma vez selecionada a família de curvas mais adequada ao conjunto de dados existente, a etapa seguinte consiste em estimar os parâmetros desta.

a) **Família  $S_U$** : nesta família os quatro parâmetros são estimados através de

$$\eta = 2z \left\{ \cosh^{-1} \left[ \frac{1}{2} \left( \frac{m}{p} + \frac{n}{p} \right) \right] \right\}^{-1} \quad (19)$$

$$\gamma = \eta \sinh^{-1} \left\{ \left( \frac{n}{p} - \frac{m}{p} \right) \left[ 2 \left( \frac{m}{p} \frac{n}{p} - 1 \right)^{1/2} \right]^{-1} \right\} \quad (20)$$

$$\omega = 2p \left( \frac{m}{p} \frac{n}{p} - 1 \right)^{1/2} \left[ \left( \frac{m}{p} + \frac{n}{p} - 2 \right) \left( \frac{m}{p} + \frac{n}{p} + 2 \right) \right]^{-1/2} \quad (21)$$

$$\kappa = \frac{x_z + x_{-z}}{2} + p \left( \frac{n}{p} - \frac{m}{p} \right) \left[ 2 \left( \frac{m}{p} + \frac{n}{p} - 2 \right) \right]^{-1} \quad (22)$$

b) **Família  $S_B$** : nesta família os parâmetros são estimados mediante as fórmulas

$$\eta = z \left\{ \cosh^{-1} \left( \frac{1}{2} \left[ \left( 1 + \frac{p}{m} \right) \left( 1 + \frac{p}{n} \right) \right]^{1/2} \right) \right\}^{-1} \quad (23)$$

$$\gamma = \eta \sinh^{-1} \left\{ \left( \frac{p}{n} - \frac{p}{m} \right) \left[ \left( 1 + \frac{p}{m} \right) \left( 1 + \frac{p}{n} \right) - 4 \right]^{1/2} \left[ 2 \left( \frac{p}{m} \frac{p}{n} - 1 \right) \right]^{-1} \right\} \quad (24)$$

$$\omega = p \left\{ \left[ \left( 1 + \frac{p}{m} \right) \left( 1 + \frac{p}{n} \right) - 2 \right]^2 - 4 \right\}^{1/2} \left( \frac{p}{m} \frac{p}{n} - 1 \right)^{-1} \quad (25)$$

$$\kappa = \frac{x_z + x_{-z}}{2} - \frac{\lambda}{2} + p \left( \frac{p}{n} - \frac{p}{m} \right) \left[ 2 \left( \frac{p}{m} \frac{p}{n} - 1 \right) \right]^{-1} \quad (26)$$

c) *Família  $S_L$* : nesta última, as estimativas dos parâmetros são obtidas através de

$$\eta = \frac{2z}{\ln \left( \frac{m}{p} \right)} \quad (27)$$

$$\gamma^* = \eta \ln \left( \frac{\frac{m}{p} - 1}{p \left( \frac{m}{p} \right)^{1/2}} \right) \quad (28)$$

$$\kappa = \frac{x_z + x_{-z}}{2} - \frac{p}{2} \left( \frac{\frac{m}{p} + 1}{\frac{m}{p} - 1} \right) \quad (29)$$

#### 5.4 Cálculo da Proporção Não Conforme

A partir das estimativas dos parâmetros, obtidas através das fórmulas 19 a 29 anteriores e, também, mediante o uso das transformações para cada família de curvas para convertê-las em distribuição normal, é possível encontrar-se as áreas procuradas, conforme indicado na tabela abaixo.

**Tabela 1** – Transformações para famílias  $S_U$ ,  $S_B$  e  $S_L$ .

<i>Família</i>	<i>Transformação</i>	
$S_U$	$z = \gamma + \eta \operatorname{senh}^{-1} \left( \frac{x - \kappa}{\omega} \right)$	(30)
$S_B$	$z = \gamma + \eta \ln \left( \frac{x - \kappa}{\omega + \kappa - x} \right)$	(31)
$S_L$	$z = \gamma^* + \eta \ln(x - \kappa)$	(32)

### 5.5 Cálculo de Índices de Capacidade

Conforme exposto anteriormente, quando a distribuição não mais é normal, a forma de calcular os índices de capacidade necessita ser modificada, ou seja, empregam-se os índices de capacidade modificados de (8) a (10).

## 6. EXEMPLO DE APLICAÇÃO

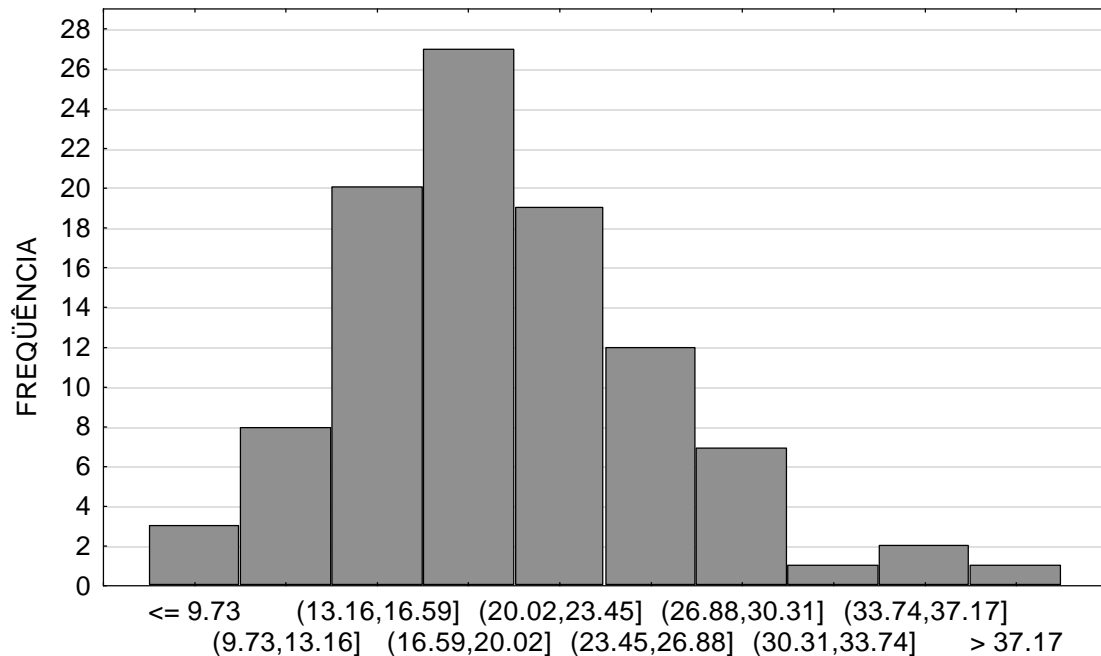
Os dados da Tabela 2 abaixo foram retirados de FARNUM (1997, p.332) e já estão apresentados ordenados para facilitar a sua manipulação.

**Tabela 2** – Dados para exemplo de aplicação

6,3	6,8	9,3	10,4	11,1	11,6	12,2	12,5	12,5	12,6
12,9	13,2	13,2	13,3	13,3	13,5	13,5	13,9	14,0	14,4
14,8	14,8	15,2	15,4	15,7	15,8	15,9	16,2	16,3	16,5
16,5	16,7	16,9	17,0	17,1	17,7	17,8	17,9	18,0	18,1
18,1	18,1	18,1	18,1	18,1	18,4	18,4	18,7	18,7	18,8
19,1	19,3	19,3	19,5	19,6	19,7	19,7	19,9	20,2	20,3
20,6	20,6	20,7	20,8	21,4	21,5	21,9	22,0	22,0	22,1
22,3	22,6	22,7	22,9	23,0	23,3	23,3	23,5	24,0	24,2
24,7	25,0	25,1	25,5	25,5	25,7	25,9	26,0	26,1	29,3
29,4	29,6	29,6	29,8	29,9	29,9	31,4	34,0	34,9	40,6

Uma análise preliminar destes dados, feita através do histograma apresentado na Figura 2, revela que a distribuição normal não é adequada à representação da

variabilidade dos dados. De fato, o teste de Anderson-Darling aplicado a estes dados fornece  $A^2 = 0,688$ , que corresponde a uma probabilidade de 0,070 (ou 7%) de que esta estatística seja superior ao seu valor crítico.



**Figura 2** – Histograma dos dados da Tabela 2

### 6.1 Seleção da Família de Curvas

Vamos selecionar o valor de  $z = 0,524$ , similarmente ao proposto por SLIFKER; SHAPIRO (1980, p.240). Este valor mostra-se extremamente adequado, uma vez que:

$$\Phi(-3z) = \Phi(-1,572) = 0,058$$

$$\Phi(-z) = \Phi(-0,524) = 0,300$$

$$\Phi(3z) = \Phi(0,524) = 0,700$$

$$\Phi(z) = \Phi(1,572) = 0,942$$

Visando reduzir o viés na estimação dos percentis, pode-se empregar a seguinte fórmula (COSTA NETO, 1977, p. 135):

$$i = n \cdot \Phi(.) + \frac{1}{2}$$

onde  $n$  é o tamanho da amostra disponível de dados e  $\Phi(.)$  é a probabilidade normal acumulada. Conseqüentemente obtém-se, conforme a Tabela 3

**Tabela 3** – Percentis para os dados

F	i	x
0,058	$100 \cdot (0,058) + \frac{1}{2} = 6,3$	11,78
0,300	$100 \cdot (0,300) + \frac{1}{2} = 30,5$	16,50
0,700	$100 \cdot (0,700) + \frac{1}{2} = 70,5$	22,20
0,942	$100 \cdot (0,942) + \frac{1}{2} = 94,7$	29,87

Para calcular a função discriminadora é necessário antes se obter:

$$m = x_{3z} - x_z = 29,87 - 22,20 = 7,67$$

$$n = x_{-z} - x_{-3z} = 16,50 - 1,78 = 4,72$$

$$p = x_z - x_{-z} = 22,20 - 16,50 = 5,70$$

e, portanto,  $mn/p^2 = 1,114$  o que implica que a família a ser selecionada é a  $S_U$ .

### 6.2 Estimação dos Parâmetros da Família $S_U$

Empregando-se as fórmulas (19) a (22) vistas anteriormente, resulta em  $\eta = 2,519$ ,  $\gamma = -1,767$ ,  $\omega = 10,826$ ,  $\kappa = 10,942$ .

### 6.3 Cálculo das Probabilidades

Admitindo-se que os limites (inferior e superior) de especificação sejam respectivamente 5 e 40, vem que:

$$z_I = \gamma + \eta \operatorname{senh}^{-1} \left( \frac{\text{LIE} - \kappa}{\omega} \right) = -1,767 + 2,519 \cdot \operatorname{senh}^{-1} \left( \frac{5 - 10,942}{10,826} \right) = -3,09$$

e

$$z_S = \gamma + \eta \operatorname{senh}^{-1} \left( \frac{\text{LSE} - \kappa}{\omega} \right) = -1,767 + 2,519 \cdot \operatorname{senh}^{-1} \left( \frac{40 - 10,942}{10,826} \right) = 2,55$$

Desta forma, a porcentagem estimada de itens não-conformes será:

$$P(X < \text{LIE}) = P(z < z_I) = P(z < -3,09) = 0,0010$$

e

$$P(X > \text{LSE}) = P(z > z_S) = P(z > 2,55) = 0,0054$$

#### 6.4 Cálculo dos Índices de Capacidade

A partir dos parâmetros estimados anteriormente, pode-se calcular os índices de capacidade, definidos de (8) a (10). Contudo, antes é necessário encontrar-se os percentis 0,135 e 99,865 ( $P_I$  e  $P_S$ ):

$$P_I = \kappa - \omega \operatorname{senh} \left( \frac{\gamma + 3}{\eta} \right) = 5,429$$

e

$$P_S = \kappa + \omega \operatorname{senh} \left( \frac{\gamma - 3}{\eta} \right) = 46,044$$

Portanto:

$$\hat{C}_m^* = \frac{\text{LSE} - \text{LIE}}{P_S - P_I} = \frac{40 - 5}{46,044 - 5,429} = 0,862$$

e

$$\hat{C}_{mi}^* = \frac{\text{Md} - \text{LIE}}{\text{Md} - P_I} = \frac{18,950 - 5}{18,950 - 5,429} = 1,032$$

e

$$\hat{C}_{ms}^* = \frac{LSE - Md}{P_s - Md} = \frac{40 - 18,950}{46,044 - 18,950} = 0,777$$

resultando em

$$\hat{C}_{mk}^* = \min \{ \hat{C}_{mi}^*, \hat{C}_{ms}^* \} = 0,777$$

## 7. CONCLUSÕES

A complexidade para lidar com os dados onde não se pode assumir a distribuição normal é muito maior. Isto faz com que muitos estudos de capacidade sejam feitos de forma equivocada, sem consideração das modificações pertinentes nos índices ou, então, que fiquem relegados a um segundo plano, em função da dificuldade intrínseca existente nestas situações.

Os softwares estatísticos hoje existentes para o tratamento dos dados provenientes de um processo estável, porém não com distribuição normal, em geral têm-se concentrado mais na forma de cálculo de índices  $\hat{C}_m^*$  e  $\hat{C}_{mk}^*$  equivalente àqueles para distribuições normais.

A família das curvas de Johnson tem sido freqüentemente empregada nestas situações e fornece resultados que, manualmente seriam de difícil realização em virtude da quantidade de cálculos envolvida.

Com a evolução dos softwares, espera-se que novas alternativas estejam disponíveis àqueles que são responsáveis pela monitoração e melhoria de processos de manufatura ou transacionais, tal como é o caso das empresas que adotam a metodologia Seis Sigma.

## 8. BIBLIOGRAFIA

BOTHE, D.R. *Measuring process capability*. New York, McGraw-Hill, 1997.

COSTA NETO, P. L. O. *Estatística*. São Paulo, Edgard Blücher, 1977.

FARNUM, N. R. Using Johnson curves to describe non-normal process data. *Quality Engineering*, Monticello, v. 9, n.2, p. 39-36, 1997

FINLEY, J.C. What is capability? or What is Cp and Cpk? In: ASQC QUALITY CONGRESS TRANSACTIONS, 46<sup>th</sup>, Nashville, 1992. *Proceedings*. Milwaukee, ASQC, p.186-92, 1992.

JOHNSON, N. L. Systems of frequency curves generated by methods of translation. *Biometrika*, London, v.36, p.149-76, 1949.

PYZDEK, T. Process capability analysis using personal computers. *Quality Engineering*, Monticello, v.4, n.3, p. 419-40, 1992.

SLIFKER, J.F.; SHAPIRO, S.S. The Johnson system: selection and parameter estimation. *Technometrics*, Alexandria, v.22, n.2, p.239-46, 1980.