3
Modelo Matemático

Este trabalho analisa o efeito da imprecisão na estimativa do desvio-padrão do processo sobre o desempenho do gráfico de *S* e sobre os índices de capacidade do processo. A análise deste efeito é feita através da utilização de um modelo matemático, apresentado neste capítulo (3) e utilizado para avaliar o desempenho do gráfico de *S* em função do erro de estimação (*k*) e do deslocamento sofrido pelo desvio-padrão do processo (γ), de acordo com as medidas de desempenho risco α e *Pd* (ver Seção 3.1). As medidas de desempenho calculadas a partir destes (*NMA*, NMA_o, *TES e TMAF*) podem, então, ser obtidas. Distintamente dos trabalhos apresentados na Seção 2.2, neste trabalho não são utilizados *softwares* computacionais de simulação para obtenção das amostras e conseqüentemente, dos valores das medidas de desempenho. Os resultados são obtidos com o auxílio do *software Microsoft* Excel, para diferentes cenários (valores de *k*, γ, α, β ou *Pd*, *m*, *n*, entre outros).

Nas Seções seguintes, (3.2) e (3.3), são apresentadas as distribuições de probabilidade do risco α e *Pd*, e dos índices de capacidade considerados, necessárias para as diversas formas de análise do efeito de estimação dessas medidas de desempenho que serão apresentadas no Capítulo 4. Tal apresentação é relevante porque, como o erro da estimativa é considerado uma variável aleatória e os limites de controle do gráfico de *S* são modelados em função desse erro, também são analisados como tal, possibilitando a apresentação dos resultados em termos da distribuição de probabilidade das medidas de desempenho.

3.1.
Definições Iniciais

Denote-se, em geral, o desvio-padrão do processo por $\sigma = \gamma\sigma_0$, sendo σ_0 o seu desvio-padrão quando sob controle estatístico, e $\gamma \geq 1$ um "fator de aumento", tal que, se $\gamma = 1$, o processo está sob controle; se $\gamma > 1$, houve aumento no desvio-padrão do processo devido a alguma causa especial; e se $0 < \gamma < 1$, houve uma redução na dispersão do processo, situação que não será considerada neste trabalho, já que representa uma situação favorável para a variabilidade do processo. Equivalentemente, pode-se escrever:

$$\gamma = \frac{\sigma}{\sigma_0} \tag{3.1.1}$$

Para analisar o desempenho do gráfico de S em relação a um possível aumento na dispersão do processo por um fator $\gamma > 1$, observa-se o comportamento da medida de desempenho *Pd* (ou do risco β).

Como os riscos α e β são função dos limites de controle do gráfico de S, que por sua vez são influenciados por erros na estimação de σ, o modelo deve considerar, além do fator de aumento na dispersão (γ), a possibilidade de erro na estimação do desvio-padrão do processo.

Sendo $\hat{\sigma}_0$ a estimativa do desvio-padrão do processo em controle (aqui supostamente $\hat{\sigma}_0 = \bar{S}/c_4$), defina-se o fator de erro de estimação k, como:

$$k = \frac{\hat{\sigma}_0}{\sigma_0} \tag{3.1.2}$$

Assim, quando o desvio-padrão do processo for estimado sem erro ($\hat{\sigma}_0 = \sigma_0$), $k = 1$. Quando $0 < k < 1$, $\hat{\sigma}_0 < \sigma_0$, indicando que a dispersão do processo foi subestimada e, quando $k > 1$, $\hat{\sigma}_0 > \sigma_0$, indicando que a dispersão do processo foi superestimada.

3.2. Distribuição do Risco *α* e do poder *Pd* do gráfico de *S* sob erros de estimação

Considerando como sinal o evento "ponto acima do limite superior de controle" (recordando que aqui só estão sendo considerados gráficos de *S* sem limite inferior de controle), a probabilidade deste evento, $P[S > LSC]$, corresponderá, quando o processo estiver sob controle, à probabilidade de alarme falso e, quando o processo estiver fora de controle, ao poder do gráfico de *S*. Assim, esta probabilidade de sinal é o elemento básico para todas as medidas de desempenho que serão consideradas, uma vez que o NMA_0 e o *NMA* correspondem ao seu inverso e o *TMAF* e o *TES* são calculados a partir destes últimos. Em qualquer caso (processo sob controle ou fora de controle), pode-se escrever:

$$\alpha \text{ ou } Pd = P[S > LSC_S] = P\left[\left(\frac{n-1}{\sigma^2}\right)S^2 > \left(\frac{n-1}{\sigma^2}\right)LSC_S^2\right]$$

A igualdade vale porque *S*, LSC_S e $\left(\frac{n-1}{\sigma^2}\right)$ são sempre estritamente positivos. Como $\left(\frac{n-1}{\sigma^2}\right)S^2$ segue uma distribuição qui-quadrado com $n-1$ graus de liberdade, então:

$$\alpha \quad ou \quad Pd = P[S > LSC_S] = P\left[\chi^2_{n-1} > \left(\frac{n-1}{\sigma^2}\right)LSC_S{}^2\right] \tag{3.2.1}$$

Para limites de três-sigma, LSC_S corresponde a $\left(c_4 + 3\sqrt{1-c_4^2}\right)\hat{\sigma}_0$. Substituindo-se a expressão do LSC_S e as equações (3.1.1) e (3.1.2) na equação (3.2.1), chega-se após simplificação, a:

$$\alpha \quad ou \quad Pd = P[S > LSC_S] = P\left[\chi^2_{n-1} > \left(\frac{n-1}{\gamma^2}\right)\left(k^2\right)\left(c_4 + 3\sqrt{1-c_4^2}\right)^2\right] \tag{3.2.2}$$

De acordo com a expressão acima, para valores dados de n (que determina c_4), γ e k, é possível obter a probabilidade de alarme falso α, quando $\sigma = \sigma_0$ (ou $\gamma = 1$), e o poder do gráfico de controle de S, quando $\sigma > \sigma_0$ (ou $\gamma > 1$).

Como, para o gráfico de S, o desvio-padrão do processo é estimado a partir do estimador não-viesado $\overline{S}/c_4$, e ainda como, pelo Teorema Central do Limite e pelas expressões (2.1.1) e (2.1.3) para a média e o desvio-padrão de S, se o número de amostras m for suficientemente grande e as amostras tiverem sido retiradas de um processo em controle ($\sigma = \sigma_0$), a distribuição de $\overline{S} = \frac{1}{m}\sum_{j=1}^{m} S_j$ é bem aproximada por $N\left(c_4\sigma_0; \frac{\left(1-c_4^2\right)\sigma_0^2}{m}\right)$, então as seguintes distribuições para $\hat{\sigma}_0$ e k se tornam válidas:

$$\hat{\sigma}_0 = \frac{\overline{S}}{c_4} \sim N\left[\sigma_0; \frac{\left(1-c_4^2\right)\sigma_0^2}{c_4^2 m}\right] \tag{3.2.3}$$

$$k = \frac{\hat{\sigma}_0}{\sigma_0} \sim N\left(1; \frac{\left(1-c_4^2\right)}{c_4^2 m}\right) \tag{3.2.4}$$

Tem-se assim que k é uma variável aleatória contínua com distribuição normal, com média igual a 1 (pois o estimador é não-viesado) e variância que é função apenas de n e m. Assim sendo, o risco α e o Pd do gráfico de S [equação (3.2.2)] dependem dos valores de n, m e γ.

Além disso, é fácil ver que Pd e α são funções monotonicamente decrescentes de k (pois, quanto maior k, maior o valor de LSC_S e menor a probabilidade de sinal); portanto, para qualquer valor k específico, sendo a e z os valores de α e Pd, respectivamente, que a ele correspondem por (3.2.2), valem as seguintes relações:

$$P(Pd \le z) = P(K > k) = 1 - P(K \le k) \tag{3.2.5}$$

$$P(\alpha > a) = P(K \leq k) \tag{3.2.6}$$

Assim, usando a distribuição normal acumulada com os parâmetros dados em (3.2.4) é possível, variando k, determinar a densidade de probabilidade acumulada para k, $F(k) = P(K \leq k)$, parametrizada por n e m, e daí, por (3.2.5) e (3.2.6) podem-se obter as distribuições acumuladas de α e *Pd* (para qualquer fator de aumento γ na dispersão do processo).

Neste trabalho, as distribuições das medidas de desempenho foram levantadas com o auxílio do *Microsoft Excel*, variando k de 0,50 a 1,50, espaçados de 0,05 em 0,05 unidades. Isso porque para valores de k menores que 0,50, a densidade de probabilidade acumulada de k sofre pouca alteração, se aproximando de zero; e para valores de k maiores que 1,50, a densidade de probabilidade acumulada de k também sofre pouca alteração, se aproximando de 100%.

O modelo desenvolvido até aqui fornece essas distribuições para o caso do gráfico de S com limites de controle de três-sigma, pois utiliza a equação (3.2.2). No caso de limites de probabilidade, estipulados para fornecer uma probabilidade de alarme falso α especificada, foi visto que o LSC_S é dado por $LSC_S = \sqrt{\chi^2_{n-1,\alpha}\hat{\sigma}^2_0 / (n-1)}$. Substituindo-se esta expressão do LSC_S e as equações (3.1.1) e (3.1.2) na equação (3.2.1), chega-se após simplificação, à seguinte fórmula para o risco α ou *Pd* do gráfico de S com limites de probabilidade:

$$\alpha \text{ ou } Pd = P[S > LSC_S] = P\left[\chi^2_{n-1} > \left(\frac{k^2}{\gamma^2}\right)\chi^2_{n-1,\alpha}\right] \tag{3.2.7}$$

Com $\gamma = 1$, a equação (3.2.7) fornece o risco α do gráfico de S com limites de probabilidade; com $\gamma > 1$, fornece o poder do gráfico de S com limites de probabilidade de sinalizar aumentos na dispersão do processo por um fator γ.

Em ambos os casos (risco α ou *Pd*), o termo após a desigualdade no segundo membro da equação (3.2.2) para limites de três-sigma e na equação

(3.2.7) para limites de probabilidade pode ser substituído por $\chi^2_{n-1,\alpha_{real}}$, mas calculado distintamente. No caso de limites três-sigma, $\chi^2_{n-1,\alpha_{real}}$ corresponde à expressão $\left(n-1/\gamma^2\right)\left(k^2\right)\left[c_4+3\sqrt{\left(1-c_4^2\right)}\right]^2$ e varia de acordo os valores de n, k e γ especificados; e no caso de limites de probabilidade, $\chi^2_{n-1,\alpha_{real}}$ é determinado a partir da expressão $\left(k^2/\gamma^2\right)\chi^2_{n-1,\alpha}$ e assume diferentes valores de acordo com k, γ e o valor especificado para α. Tanto para limites de três-sigma quanto para limites de probabilidade, $\chi^2_{n-1,\alpha_{real}}$ representa o valor real qui-quadrado, que corresponde ao limite superior de controle do gráfico de S, dado que existe um erro na estimação e/ou deslocamento de σ_0; e difere do valor "nominal", neste trabalho referido como $\chi^2_{n-1,\alpha_{nom}}$, desejado ou especificado.

3.3. Distribuição dos índices de capacidade *Cp* e *Cpk* sob erros de estimação

Como mencionado anteriormente, a análise de capacidade de processos através dos índices de capacidade também é baseada em estimativas dos parâmetros dos processos e por isso, afetada pela imprecisão dessas estimativas.

O índice *Cp* é calculado através da equação (2.1.22). Admitindo-se que o desvio-padrão σ_0 necessário para o cálculo de *Cp* seja desconhecido e necessite ser estimado e ainda, que este seja estimado com erro ($\hat{\sigma}_0 \neq \sigma_0$), este erro será evidentemente propagado para a estimativa de *Cp*. Definindo Cp_a como o índice de capacidade aparente calculado quando o desvio-padrão do processo é estimado com erro e considerando a definição do fator de erro *k* [equação (3.1.2)], então:

$$Cp_a = \frac{LSE - LIE}{6\hat{\sigma}_0} = \frac{LSE - LIE}{6k\sigma_0} = \frac{1}{k}Cp \tag{3.3.1}$$

Assim, o fator de erro na estimativa de *Cp* é o inverso do fator de erro na estimativa de σ_0. Quando o desvio-padrão do processo for estimado sem erro ($k = 1$), Cp_a corresponderá ao *Cp* verdadeiro. Quando a dispersão do processo for subestimada ($k < 1$), a capacidade do processo terá sido superestimada ($Cp_{aparente} > Cp$) e vice-versa.

Assumindo que a média μ do processo seja estimada sem erro, a mesma relação vale entre o erro de estimação de σ_0 e o erro de estimação de *Cpk*. Definindo Cpk_a como o índice *Cpk* aparente, basta ver a partir da equação (2.1.23) que

$$Cpk_a = Min\left[\frac{LSE - \mu}{3\hat{\sigma}_0}; \frac{\mu - LIE}{3\hat{\sigma}_0}\right] = Min\left[\frac{LSE - \mu}{3k\sigma_0}; \frac{\mu - LIE}{3k\sigma_0}\right] = \frac{1}{k}Cpk \tag{3.3.2}$$

Assim, o fator de erro na estimação da capacidade do processo é o mesmo tanto para a capacidade potencial (indicada por *Cp*) como para a capacidade atual (indicada por *Cpk*). Para esses índices de capacidade são apresentados três diferentes análises, possivelmente, de maior interesse em termos práticos. A saber: (a) obtenção da distribuição acumulada do fator de erro $\frac{1}{k} = \frac{Cp_a}{Cp} = \frac{Cpk_a}{Cpk}$ obtida a partir da distribuição acumulada do inverso do fator de erro, que é o próprio *k* (fator de erro na estimação de σ_0); (b) obtenção de intervalos de confiança (IC) para o inverso do fator de erro (*k*), ou seja, IC´s para *Cp* e *Cpk*; e (c) determinação do valor de *m* que, para um valor de *n* dado, garante que $\left(1-\varepsilon/100\right)Cp_a \le Cp \le \left(1+\varepsilon/100\right)Cp_a$ ou $\left(1-\varepsilon/100\right)Cpk_a \le Cpk \le \left(1+\varepsilon/100\right)Cpk_a$, para uma semi-largura $\varepsilon\%$ do intervalo de confiança 100%*c*, ambos especificados.

Neste trabalho, para a análise da capacidade do processo, a distribuição do fator 1/*k* foi obtida a partir da distribuição de *k*, com o auxílio do *Microsoft Excel*, variando *k* de 0,70 a 1,30, espaçados de 0,05 em 0,05 unidades. Isso porque para valores de *k* menores que 0,70, a densidade de probabilidade acumulada de 1/*k* sofre pouca alteração, se aproximando de 100%; e para valores de *k* maiores que 1,30, a densidade de probabilidade acumulada de *k* também sofre pouca alteração, se aproximando de zero.