

Universidade Federal de Santa Catarina
Departamento de Engenharia de Produção e Sistemas
Programa de Pós-Graduação em Engenharia de Produção

**Um modelo para a análise da capacidade de processos
com ênfase na transformação de dados**

Rodrigo Gabriel de Miranda

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

Florianópolis, Fevereiro de 2005

Rodrigo Gabriel de Miranda

**Um modelo para a análise da capacidade de processos
com ênfase na transformação de dados**

Dissertação apresentada à
Universidade Federal de Santa
Catarina para obtenção do título de
Mestre em Engenharia de
Produção

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

Florianópolis, Fevereiro 2005

FICHA CATALOGRÁFICA

AUTOR:

MIRANDA, Rodrigo Gabriel de.

TÍTULO: UM MODELO PARA A ANÁLISE DA CAPACIDADE DE PROCESSOS COM ÊNFASE NA TRANSFORMAÇÃO DE DADOS, 114p

**DISSERTAÇÃO DE MESTRADO
CURSO DE MESTRADO DO PROGRAMA DE PÓS-GRADUAÇÃO EM
ENGENHARIA DE PRODUÇÃO**

ÁREA DE CONCENTRAÇÃO: SISTEMAS DE PRODUÇÃO

ORIENTADOR: ROBERT WAYNE SAMOHYL, *Ph. D*

PALAVRAS CHAVES: ÍNDICES DE CAPACIDADE, NÃO NORMALIDADE, AUTOCORRELAÇÃO, TRANSFORMAÇÃO DE DADOS

RODRIGO GABRIEL DE MIRANDA

**UM MODELO PARA A ANÁLISE DA CAPACIDADE DE
PROCESSOS COM ÊNFASE NA TRANSFORMAÇÃO DE DADOS**

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

Florianópolis, 14 de Fevereiro de 2005.

Prof. Edson Pacheco Paladini, Dr.

Coordenador do Programa de Pós-Graduação em Engenharia de Produção

Prof. Robert Wayne Samohyl, *Ph. D*

Universidade Federal de Santa Catarina

Orientador

BANCA EXAMINADORA

Prof. Pedro Alberto Barbeta, Dr.

Universidade Federal de Santa Catarina

Presidente

Prof. Marcelo Menezes Reis, Dr. Eng.

Universidade Federal de Santa Catarina

Prof. Willy Arno Sommer, Dr.

Universidade Federal de Santa Catarina

DEDICATÓRIA

Dedico este trabalho aos meus familiares, pai, irmãos e em especial a minha mãe.

AGRADECIMENTOS

A realização deste trabalho só foi possível com a ajuda e colaboração das pessoas que venho a agradecer:

ao professor e orientador Robert Wayne Samohyl, pela grande orientação e confiança;

à Universidade Federal de Santa Catarina e o Departamento de Engenharia de Produção e Sistemas, pelas oportunidades oferecidas e condições de estudo;

ao CNPQ, pela bolsa concedida;

aos membros da banca examinadora, pelas sugestões feitas para a melhoria do trabalho;

aos colegas do NNQ, Rubson, Éder, Gueibi, Manoel, Adauto, Andréa, Ary e Custódio pelo apoio e incentivo.

RESUMO

O estudo da capacidade de processos se tornou uma ferramenta fundamental do controle estatístico da qualidade, tendo como principal objetivo verificar se a média e a variabilidade do processo estão de acordo com o alvo e os limites de especificação de projeto. A verificação permite o ajuste do processo de maneira que se reduzam as chances da fabricação de produtos defeituosos. Esta produção de produtos defeituosos é calculada com a utilização de índices desenvolvidos para medir a capacidade de processos, onde os mais utilizados são C_p e C_{pk} . Estes índices são calculados estimando-se a média e o desvio padrão de dados coletados de amostras do processo, portanto não se sabe com certeza o seu valor real. Neste trabalho foram estimados estes índices de capacidade de processo considerando-se o tamanho da amostra e o risco (probabilidade) do índice ser menor ou igual a aquele estabelecido como ideal. Desta forma foi verificado se o processo é realmente capaz, se o tamanho de amostras coletadas é adequado, se são necessárias algumas alterações no processo (ajuste da média ao alvo e redução da variabilidade) e se a capacidade estabelecida como ideal é alcançável. Foram abordados também os casos de não normalidade e autocorrelação nos dados, situações que precisam de um tratamento para que se possam utilizar os índices corretamente. A aplicação da metodologia de análise utilizada foi realizada através de um estudo de caso em uma indústria do setor automotivo.

Palavras-chave: índices de capacidade, não normalidade, autocorrelação, transformação de dados

ABSTRACT

The study of process capability has become a basic tool of statistical quality control, having as its main objective to verify if the process average and variability are in accordance with the target and the specification limits of the project. The verification allows that if the process was adjusted then the possibilities of the manufacture of defective products will be reduced. The percentage of produced defective products is measured with the use of indices to measure the capacity of processes, where the most common are C_p and C_{pk} . These indices are calculated with the estimation of the average and the standard deviation of collected data from samples of the process and therefore its real value is not known with certainty. In this work these process capability indices were estimated considering the sample size and the risk (probability) of the index being less or equal to that established as ideal. In this way it could be verified if the process is really capable in a statistical sense, if the collected sample size is adequate, if some alterations in the process are necessary (adjustment of the average to the target and reduction of the variability) and if the capacity established as ideal is feasible. The presence of non normality and autocorrelation in the data will also be analyzed, situations that require special treatment for the correct use of the indices. The application of methodology proposed here will be tested through a real case study in the automotive industry.

Key-Words: capability index, non normality, autocorrelation, data transformation

LISTA DE FIGURAS

FIGURA 1 – CURVA NORMAL.....	21
FIGURA 2 – DIFERENÇA ENTRE A VARIABILIDADE DE CURTO E LONGO PRAZO.....	36
FIGURA 3 – COMPARAÇÃO DOS LIMITES DE ESPECIFICAÇÃO COM A VARIABILIDADE DO PROCESSO.....	38
FIGURA 4 – RELAÇÃO DA VARIABILIDADE DO PROCESSO EM RELAÇÃO A MÉDIA COM OS LIMITES DE ESPECIFICAÇÃO E A MÉDIA.....	41
FIGURA 5 – PERCENTUAL DE DEFEITUOSOS.....	47
FIGURA 6 – HISTOGRAMA DO GANHO DE AMPLIFICADORES.....	53
FIGURA 7 – GRÁFICO NORMAL DE PROBABILIDADE PARA O GANHO DE AMPLIFICADORES.....	54
FIGURA 8 – GRÁFICO DA FUNÇÃO DE DENSIDADE KERNEL PARA OS DADOS DE AMPLIFICADORES.....	56
FIGURA 9 – HISTOGRAMA, GRÁFICO NORMAL DE PROBABILIDADE E FUNÇÃO DE DENSIDADE KERNEL PARA OS DADOS TRANSFORMADOS DO GANHO DE AMPLIFICADORES.....	67
FIGURA 10 – GRÁFICO DAS OBSERVAÇÕES INDIVIDUAIS DO DIÂMETRO DOS FUROS DE UMA PEÇA.....	73
FIGURA 11 – MODELO DE ANÁLISE DA CAPACIDADE DE PROCESSOS PARA DADOS NORMAIS E INDEPENDENTES (MODELO 1).....	83
FIGURA 12 – MODELO DA ANÁLISE DA CAPACIDADE DE PROCESSOS PARA DADOS NÃO NORMAIS E AUTOCORRELACIONADOS (MODELO 2).....	85
FIGURA 13 – HISTOGRAMA, GRÁFICO NORMAL DE PROBABILIDADE E FDK DOS DADOS DO PROCESSO 1... ..	88
FIGURA 14 – GRÁFICOS DE CONTROLE PARA O PROCESSO 1.....	90
FIGURA 15 – GRÁFICO NORMAL DE PROBABILIDADE E FDK DOS DADOS ORIGINAIS E TRANSFORMADOS DO PROCESSO 2.....	94
FIGURA 16 – GRÁFICOS DE CONTROLE PARA O PROCESSO 2.....	96

LISTA DE TABELAS

TABELA 1 – VANTAGENS E DESVANTAGENS COM O ESTUDO DA CAPACIDADE DE PROCESSOS.	15
TABELA 2 – ASPECTOS TEÓRICOS CONSIDERADOS NO ESTUDO DA CAPACIDADE DE PROCESSOS.	16
TABELA 3 - C_p E A PRODUÇÃO DE NÃO CONFORMES.	38
TABELA 4 – RECOMENDAÇÃO DE VALORES MÍNIMOS PARA OS ÍNDICES C_p E C_{pk}	39
TABELA 5 – RELAÇÃO DE C_p E C_{pk} COM A PRODUÇÃO DE NÃO CONFORMES POR MILHÃO	48
TABELA 6 – GANHO DE AMPLIFICADORES (DB)	52
TABELA 7 – DISTRIBUIÇÃO DE FREQUÊNCIA PARA O GANHO DE AMPLIFICADORES	53
TABELA 8 – COEFICIENTES PARA λ , μ_y E σ_y PARA O TESTE W	60
TABELA 9 – TRANSFORMAÇÕES ASSOCIADAS COM AS FAMÍLIAS DE CURVAS DE JOHNSON	62
TABELA 10 – DIÂMETRO DE EIXOS DE UMA PEÇA	73
TABELA 11 – FUNÇÃO AUTOCORRELAÇÃO E TESTE DE LJUNG-BOX PARA OS DADOS DE DIÂMETROS DO FURO DE UMA PEÇA	74
TABELA 12 – RESULTADOS DO MODELO AUTOREGRESSIVO PARA OS DADOS DE DIÂMETROS DO FURO DE UMA PEÇA.	75
TABELA 13 - FUNÇÃO DE AUTOCORRELAÇÃO E TESTE DE LJUNG-BOX PARA OS RESÍDUOS DO MODELO AUTOREGRESSIVO DOS DADOS DE DIÂMETROS DO FURO DE UMA PEÇA	75
TABELA 14 – TESTE DE AUTOCORRELAÇÃO PARA OS DADOS DO PROCESSO 1	89
TABELA 15 – ESTIMAÇÃO DOS ÍNDICES DE CAPACIDADE DO PROCESSO 1	91
TABELA 16 – TESTE DE AUTOCORRELAÇÃO PARA OS DADOS TRANSFORMADOS DO PROCESSO 2	93
TABELA 17 – ESTIMAÇÃO DOS ÍNDICES DE CAPACIDADE DO PROCESSO 2	97
Tabela 1A - Valores dos coeficientes para construção do gráfico da média.....	105
Tabela 2A - Valores dos coeficientes para construção do gráfico do desvio padrão.....	106
Tabela 3A - Valores dos coeficientes para construção do gráfico da amplitude.....	107
Tabela 4A – Coeficientes a_i para o teste W de Shapiro-Wilk.....	109
Tabela 1B – Valores obtidos aplicando a transformação de Johnson aos dados de ganho de amplificadores do capítulo 3	110
Tabela 1C – Diâmetro interno de uma peça usinada no processo 1 (mm)	111
Tabela 2C – Distância entre faces de uma peça usinada no processo 2 (mm).....	111
Tabela 1D – Resíduos do modelo autoregressivo do processo 1 do capítulo 5.....	112
Tabela 2D – Função de autocorrelação dos resíduos do modelo autoregressivo do processo 1 do capítulo 5.....	112
Tabela 3D – Transformação de Johnson aplicada aos dados do processo 2 capítulo 5.....	113
Tabela 4D - Resíduos do modelo autoregressivo do processo 2 do capítulo 5.....	114

LISTA DE SIGLAS E SÍMBOLOS

<i>ppm</i>	peças por milhão
LC	linha central
<i>LSC</i>	limite superior de controle
<i>LIC</i>	limite inferior de controle
<i>LSE</i>	limite superior de especificação
<i>LIE</i>	limite inferior de especificação
<i>Teste W</i>	teste de normalidade
μ	média populacional
\bar{x}	média amostral
σ	desvio padrão populacional
<i>S</i>	desvio padrão amostral
<i>FDK</i>	Função densidade kernel
H_0	hipótese nula
H_1	hipótese alternativa

SUMÁRIO

1	INTRODUÇÃO	14
1.1	JUSTIFICATIVA E IMPORTÂNCIA	16
1.2	OBJETIVOS DO TRABALHO	17
1.2.1	<i>Objetivo Geral</i>	17
1.2.2	<i>Objetivos específicos</i>	17
1.3	DELIMITAÇÕES DO TRABALHO	18
1.4	ESTRUTURA DO TRABALHO	18
2	REVISÃO TEÓRICA	20
2.1	ESTATÍSTICA EM CONTROLE ESTATÍSTICO DE QUALIDADE	20
2.1.1	<i>Distribuição de probabilidade</i>	20
2.1.1.1	<i>Distribuição Normal</i>	20
2.1.1.2	<i>Distribuição Chi-Quadrado (χ^2)</i>	22
2.1.1.3	<i>Distribuição Chi (χ)</i>	23
2.1.2	<i>Amostragem</i>	24
2.1.3	<i>Teste de hipótese</i>	24
2.1.4	<i>Intervalo de confiança</i>	27
2.2	GRÁFICOS DE CONTROLE	27
2.2.1	<i>Gráfico de amplitude móvel (MR) e medidas individuais ($\bar{I}\bar{X}$)</i>	29
2.2.2	<i>Gráfico da média (\bar{X}) e amplitude (R)</i>	30
2.2.3	<i>Gráfico da média (\bar{X}) e desvio padrão (S)</i>	32
2.3	ÍNDICE DE CAPACIDADE DE PROCESSOS	33
2.3.1	<i>Histórico dos índices</i>	33
2.3.2	<i>Condições necessárias para a utilização dos índices de capacidade</i>	34
2.3.3	<i>Estimação da média ($\hat{\mu}$) e desvio padrão ($\hat{\sigma}$)</i>	35
2.3.4	<i>Índices de Capacidade da Primeira Geração ($C_p, C_{pk}, k, C_{pl}, C_{pu}$)</i>	37
2.3.5	<i>Segunda Geração (C_{pm})</i>	44
2.3.6	<i>Terceira Geração de índices</i>	46
2.3.7	<i>Relação entre os índices de capacidade e a produção de não conformes</i>	47
3	ANÁLISE DA CAPACIDADE NA PRESENÇA DE DADOS NÃO NORMAIS E AUTOCORRELACIONADOS	50
3.1	CAUSAS PARA NÃO NORMALIDADE	50
3.2	TESTE DE NORMALIDADE	51
3.2.1	<i>Método gráfico (histograma, gráfico normal de probabilidade e função de densidade kernel)</i> 51	
3.2.2	<i>Método estatístico: o teste W de Shapiro-Wilk</i>	56
3.3	TRANSFORMAÇÃO DE JOHNSON	61
3.4	UTILIZAÇÃO DOS ÍNDICES DE CAPACIDADE PARA DADOS NÃO NORMAIS	67
3.5	ESTIMAÇÃO DOS ÍNDICES DE CAPACIDADE NA PRESENÇA DE DADOS AUTOCORRELACIONADOS	70
3.5.1	<i>Teste estatístico para a autocorrelação</i>	71
3.5.2	<i>Solução para o problema de autocorrelação</i>	74
4	MODELOS DE ANÁLISE DA CAPACIDADE DE PROCESSOS	78
4.1	MODELO DA ANÁLISE DA CAPACIDADE DE PROCESSOS PARA DADOS NORMAIS E INDEPENDENTES (MODELO 1)	78
4.2	MODELO DA ANÁLISE DA CAPACIDADE DE PROCESSOS PARA DADOS NÃO NORMAIS E AUTOCORRELACIONADOS (MODELO 2)	84
5	APLICAÇÃO DO MODELO EM UMA FÁBRICA DO SETOR AUTOMOTIVO	86
5.1	PROCESSO 1 – CARACTERÍSTICA AVALIADA: DIÂMETRO INTERNO	87
5.2	PROCESSO 2 – CARACTERÍSTICA AVALIADA: DISTÂNCIA ENTRE FACES	92

6 CONCLUSÕES	98
REFERÊNCIA BIBLIOGRÁFICA	101
APÊNDICE A – CÁLCULO DOS COEFICIENTES a_i E DO W DO TESTE W DE SHAPIRO- WILK PARA O EXEMPLO DO CAPÍTULO 3.....	108
A1. CÁLCULO DOS COEFICIENTES a_i	108
A2. CÁLCULO DE W	109
APÊNDICE B – TABELA COM OS VALORES DA TRANSFORMAÇÃO DE JOHNSON DO EXEMPLO DO CAPÍTULO 3.	110
APÊNDICE C – DADOS DE MEDIDAS DE PEÇAS DOS DOIS PROCESSOS DE USINAGEM DO CAPÍTULO 5.	111
APÊNDICE D – RESÍDUOS DOS MODELOS DOS ESTUDOS DE CASO DOS PROCESSOS (1 E 2) APRESENTADOS NO CAPÍTULO 5.....	112
D1. PROCESSO 1	112
D2. PROCESSO 2	113

1 INTRODUÇÃO

O estudo de capacidade de processos é uma das ferramentas existentes do controle estatístico de qualidade, e desde o início da década de 80 vem sendo utilizado nas indústrias do Japão, um dos países que são referência quando se fala em qualidade. Com o sucesso obtido por essas empresas japonesas, mais empresas de outras partes do mundo também começaram a se utilizar desta ferramenta da qualidade. No setor automotivo, por exemplo, pioneiro na utilização de índices de capacidade de processos, é exigido em contrato, entre cliente e fornecedor, um valor mínimo de índice para garantir a qualidade dos produtos (ausência de defeitos).

A qualidade dos produtos produzidos, conhecida como qualidade de conformação, pode ser atingida quando os processos que a produzem são capazes, isto é, produzem as peças de acordo com o que foi especificado em projeto. As especificações de projeto são os valores alvos de características de produtos que se querem produzir mais as tolerâncias admitidas em torno destes alvos. Esta tolerância em torno de um alvo é definida por que existe variabilidade em processos produtivos, não sendo possível produzir todas as peças com exatamente as mesmas medidas. Como existe esta variabilidade, o estudo da capacidade estima qual a proporção de produtos produzidos está dentro dos limites de especificação.

Quando a produção de defeituosos ocorre acima do admitido, o processo é considerado incapaz, basicamente por dois motivos: a variabilidade do processo é muito grande em relação aos limites de especificação e a média do processo não está centrada no alvo. Tendo conhecimento destas informações, pode-se então tomar medidas corretivas no processo, reduzindo a produção de defeituosos, com conseqüentes reduções de custos.

Para facilitar esta estimação da proporção de defeituosos produzidos foram desenvolvidos índices de capacidade, que relacionam o seu valor com a produção de defeituosos. Os primeiros índices desenvolvidos foram C_p e C_{pk} , sendo os mais utilizados na indústria e constituem a base de estudo deste trabalho. A vantagem da utilização destes índices é que eles são adimensionais, facilitando a comparação de processos produtivos, independente do que se esteja produzindo.

Segundo um estudo realizado por [Deleryd \(1999a\)](#) em 97 empresas da Suécia as principais vantagens e desvantagens apontadas com a análise da capacidade de processos são descritos na [Tabela 1](#).

Tabela 1 – Vantagens e desvantagens com o estudo da capacidade de processos.

Vantagens	Desvantagens
Ter conhecimento sobre o processo	Consumo de recursos
Tomar decisões baseadas em fatos	Dificuldade teórica
Identificar prioridades para realização de melhorias	Difícil de motivar o trabalho em equipe
Melhorar a qualidade dos produtos	Demora em obter resultados
Foco nas melhorias	Dificuldade de análise
Aumento da motivação para o trabalho em equipe	Muito tempo e esforço gasto com treinamento

Fonte: Deleryd (1999a)

A segunda desvantagem indicada na pesquisa (dificuldade teórica) é o foco deste trabalho. Esta dificuldade teórica foi tema de outro artigo publicado com este mesmo estudo [Deleryd \(1998\)](#), onde foram apontados quais são os principais aspectos teóricos que as empresas têm que trabalhar no dia-dia, e qual o percentual delas que sabiam trabalhar com cada um deles.

Tabela 2 – Aspectos teóricos considerados no estudo da capacidade de processos

Aspecto teórico	Percentual de empresas que sabiam lidar com o aspecto teórico
Calibração dos instrumentos de medição	33%
Estabelecer controle estatístico antes de conduzir o estudo	25%
Coleta de dados consistente	50%
Utilizar vários índices de capacidade simultaneamente	70%
Usar limites de confiança e testes de hipótese	10%
Checar suposições sobre a distribuição dos dados e lidar com assimetria	10%

Fonte: Deleryd (1998)

Dentre os aspectos citados, os que serão abordados neste trabalho com maior ênfase são os dois últimos da [Tabela 2](#), os que menos as empresas sabiam trabalhar. A utilização de limites de confiança e teste de hipótese e a não normalidade dos dados. Ainda outro tema que será estudado e não foi citado na pesquisa, é a autocorrelação nos dados.

1.1 Justificativa e importância

O estudo da capacidade de processos se mostrou ao longo das décadas de 80 e 90 uma ferramenta fundamental do controle estatístico de processos (CEP). Muitas empresas adotaram a sua prática após o sucesso obtido pelo setor automotivo, pioneiro neste tipo de estudo e na utilização de índices de capacidade de processo. A montadora Ford, por exemplo, exige que 904 de seus 1188 fornecedores comprovem a qualidade de seus produtos via estudo de capacidade. São 4569 de 10735 processos de seus fornecedores monitorados e sua capacidade avaliada (Kotz e Lovelace, 1998). São centenas, às vezes milhares de componentes que precisam ser montados e todos devem seguir as especificações de projeto para que o produto final funcione como planejado. Outro grande exemplo do sucesso obtido por este tipo de estudo, muito citado na

literatura, é o programa de qualidade da Motorola, que admite apenas 3,4 defeitos por milhão em seus produtos fabricados.

Mesmo com o sucesso obtido por estas empresas, que realizaram a análise da capacidade de processos e utilizaram os índices desenvolvidos para este fim, existem ainda muitas críticas e descrença ao uso da estatística e índices de capacidade para avaliar a capacidade de processos produtivos. Estas críticas surgem principalmente pela utilização incorreta dos índices, que necessitam do conhecimento de suas propriedades estatísticas para avaliação mais correta da real capacidade. Neste trabalho pretende-se apresentar várias destas propriedades, pois se considera que esta é uma ferramenta de qualidade valiosa. A oportunidade de se reduzir a fabricação de produtos defeituosos justifica o empenho em se mostrar a teoria existente até o momento para o uso da estatística e índices de capacidade.

1.2 Objetivos do trabalho

Os objetivos do trabalho podem ser divididos em dois grupos, o objetivo geral e os específicos.

1.2.1 Objetivo Geral

Criar um modelo e propor soluções quando necessário para avaliar a capacidade de processos, utilizando a estimação de índices de capacidade.

1.2.2 Objetivos específicos

- Fazer uma revisão sobre a literatura referente a índices de capacidade de processos e como utilizá-los na presença de não normalidade e autocorrelação.
- Aplicar o modelo desenvolvido em um estudo de caso em uma empresa do setor automotivo considerando os seguintes aspectos:

- Verificar a normalidade dos dados coletados no processo e quando esta não ocorrer utilizar a transformação de dados;
- Verificar a independência dos dados coletados no processo e quando esta não ocorrer utilizar modelos autorregressivos;
- Calcular o tamanho de amostra necessária para um certo intervalo de confiança do índice estimado.

1.3 Delimitações do trabalho

As delimitações do trabalho são destacadas a seguir:

- Apenas foram estudados os casos de distribuição contínua de dados, não sendo considerado os casos para dados discretos.
- Apenas os casos de distribuição univariada foram abordados.
- Falta de teoria para o cálculo de limites de confiança quando os dados não seguem a distribuição normal e se utiliza a transformação de Johnson.
- Não inclusão dos erros de medição na estimação dos índices de capacidade.
- A falta de um modelo de custos para a decisão, quando várias soluções proposta.

1.4 Estrutura do trabalho

O trabalho proposto está estruturado em 6 capítulos, que são descritos a seguir:

- capítulo 1 – apresenta o tema abordado, as justificativas e importância do tema escolhido, os objetivos a serem alcançados e as limitações;
- capítulo 2 – apresenta a revisão teórica referente a conceitos de estatística utilizados no controle estatístico de qualidade que compreendam os gráficos de controle e

índices de capacidade. A ênfase da revisão deste capítulo são os índices de capacidade (como se estimar, intervalos de confiança, teste de hipótese).

- capítulo 3 – apresenta a revisão teórica e as técnicas para que se possa utilizar os índices de capacidade quando os dados não são normais e possuem autocorrelação.
- capítulo 4 – apresenta um modelo para avaliação da capacidade de processos;
- capítulo 5 – apresenta um estudo de caso de capacidade de processo, em uma fábrica do setor automotivo, onde foram aplicados os conceitos vistos na revisão teórica dos capítulos 3 e 4 e o modelo desenvolvido no capítulo 4;
- capítulo 6 – apresenta as conclusões do trabalho e recomendações para trabalhos futuros.

2 REVISÃO TEÓRICA

Neste capítulo será apresentada a revisão teórica referente a conceitos sobre estatística, que são utilizados no controle estatístico de qualidade, [seção 2.1](#). Na [seção 2.2](#) uma revisão sobre gráficos de controle e na [seção 2.3](#) índices de capacidade de processo. Esta última seção é o foco central deste capítulo, sendo o seu conteúdo abordado com mais profundidade.

2.1 Estatística em controle estatístico de qualidade

Nesta seção será apresentada uma revisão de conceitos estatísticos utilizados em controle estatístico de qualidade. Na [seção 2.1.1](#) serão apresentadas algumas distribuições de probabilidades contínuas, na [seção 2.1.2](#) o processo de amostragem, na [seção 2.1.3](#) teste de hipótese e na [seção 2.1.4](#) intervalos de confiança. Uma grande fonte de informações onde se podem encontrar os conceitos que serão apresentados aqui é o livro de [Montgomery e Runger \(2003\)](#).

2.1.1 Distribuição de probabilidade

As distribuições de probabilidade apresentadas aqui são utilizadas na estimação de parâmetros e seus intervalos de confiança utilizados no controle estatístico de processo. Tanto na parte de gráficos de controle quanto em índices de capacidade.

2.1.1.1 Distribuição Normal

É o modelo de distribuição de probabilidade mais importante, pois a maior parte da teoria de controle estatístico de qualidade foi desenvolvida considerando que os dados seguem esta distribuição ou podem ser aproximadas pelo teorema central do limite. Uma notação utilizada para definir a distribuição é $x \sim N(\mu, \sigma^2)$, onde significa

que a variável x aleatória é normalmente distribuída com média μ e variância σ^2 . Os estimadores amostrais naturais para a média μ , variância σ^2 e desvio padrão σ são respectivamente:

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n} \quad (2.1)$$

$$S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1} \quad (2.2)$$

$$S = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}} \quad (2.3)$$

A função de densidade da distribuição normal é descrita pela equação (2.4) e sua forma pode ser vista na [Figura 1](#), é simétrica em relação à média.

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad -\infty < x < \infty \quad (2.4)$$

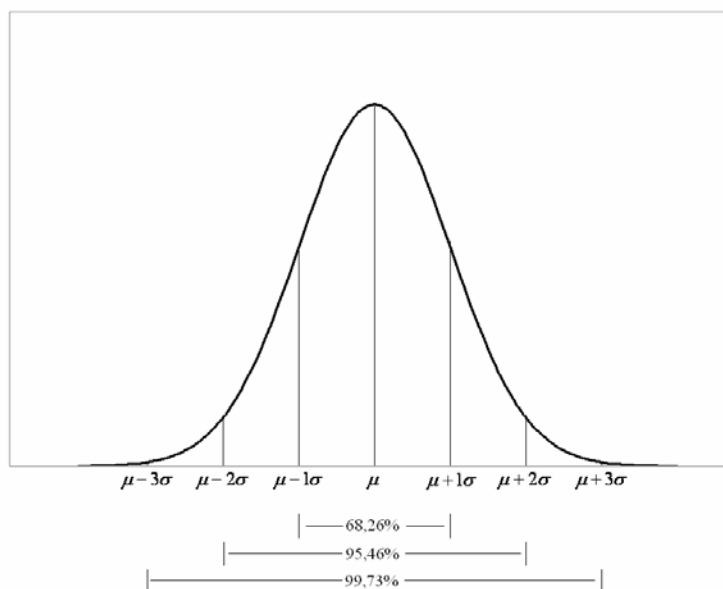


Figura 1 – Curva Normal

Observa-se que 68,26 % da população estão contidos entre os limites de $\mu \pm 1\sigma$, 95,46% entre $\mu \pm 2\sigma$ e 99,73% entre $\mu \pm 3\sigma$. Existe uma relação direta entre o número

de desvios padrão em relação à média e a probabilidade de ocorrência de um determinado valor. Por exemplo, a probabilidade de um valor ocorrer acima de $\mu + 2\sigma$ é de 2,27% e $\mu + 3\sigma$ é de apenas 0,135%.

Uma outra maneira de se representar a distribuição normal é através de sua forma padronizada, em que se faz uma mudança de variável.

$$z = \frac{x - \mu}{\sigma} \quad (2.5)$$

Pode-se transformar qualquer variável aleatória $x \sim N(\mu, \sigma^2)$ para $z \sim N(0,1)$, onde z possui média zero e variância igual a um. A função de distribuição cumulativa da distribuição normal padronizada recebe a notação de $\Phi(\cdot)$ e sua tabela pode ser encontrada em qualquer livro de estatística. Como exemplo da utilização da forma padronizada, a probabilidade de x ser menor ou igual a a é

$$P\{x \leq a\} = P\left\{z \leq \frac{a - \mu}{\sigma}\right\} \equiv \Phi\left(\frac{a - \mu}{\sigma}\right)$$

No estudo da capacidade de processos este cálculo é utilizado para se definir qual o percentual da população que está além dos limites de especificação.

2.1.1.2 Distribuição Chi-Quadrado (χ^2)

Neste trabalho este modelo de distribuição é utilizado para o cálculo de intervalo de confiança da variância da distribuição normal e no teste estatístico de autocorrelação (equação (3.37) da página 71). Se x_1, x_2, \dots, x_n forem variáveis aleatórias independentes e normalmente distribuídas com média zero e variância um, então a variável

aleatória $y = x_1^2 + x_2^2 + \dots + x_n^2$ é distribuída como uma distribuição χ^2 com n graus de liberdade.

A função de densidade é

$$f(y) = \frac{1}{2^{n/2} \Gamma\left(\frac{n}{2}\right)} y^{(n/2)-1} e^{-y/2} \quad (2.6)$$

onde $\Gamma(\alpha)$ é a função gama definida pela equação (2.7) e $\Gamma(\alpha) = (\alpha - 1)!$ para todos os inteiros positivos α .

$$\Gamma(\alpha) = \int_0^{\infty} y^{\alpha-1} e^{-y} dy \quad (2.7)$$

Sua média e variância são

$$\mu = n \quad (2.8)$$

$$\sigma^2 = 2n \quad (2.9)$$

Os pontos percentuais da distribuição χ^2 podem ser encontrados em livros de estatística.

2.1.1.3 Distribuição Chi (χ)

Na estimação dos índices de capacidade é necessário que se estime o desvio padrão σ das amostras do processo. O intervalo de confiança para o desvio padrão segue este modelo de distribuição. A função de densidade é

$$f(x) = \frac{2^{1-\frac{n}{2}} x^{n-1} e^{-\frac{x^2}{2}}}{\Gamma\left(\frac{n}{2}\right)} \quad (2.10)$$

Sua média e variância são

$$\mu = \frac{\sqrt{2}\Gamma\left(\frac{1}{2}(n+1)\right)}{\Gamma\left(\frac{n}{2}\right)} \quad (2.11)$$

$$\sigma^2 = \frac{2\left[\Gamma\left(\frac{n}{2}\right)\Gamma\left(1+\frac{n}{2}\right) - \Gamma^2\left(\frac{1}{2}(n+1)\right)\right]}{\Gamma^2\left(\frac{n}{2}\right)} \quad (2.12)$$

2.1.2 Amostragem

O objetivo da amostragem é a redução de custos para a determinação de algum parâmetro que se deseja saber. Em processos produtivos quer se saber a tendência central e a variabilidade do processo. Poderiam se medir todas as peças para descobrir o valor destes parâmetros, mas com algumas amostras é possível atingir este objetivo com certo risco de se errar. Este risco de errar (que depende do tamanho da amostra) também tem um custo associado e pode ser calculado. Deve-se então encontrar um equilíbrio entre estes custos para determinar qual o tamanho da amostra necessária.

No estudo da capacidade de processos as amostras devem representar todas as fontes de variabilidade do processo, como o *set-up*, mudanças ambientais, mudança de operadores, possíveis ciclos e tendência que podem afetar na estimação do desvio padrão. O tamanho de amostra mínimo recomendado para o estudo da capacidade é de 100 amostras pelo menos [Kotz e Lovelace \(1998\)](#).

2.1.3 Teste de hipótese

O teste de hipótese é realizado quando se quer saber se o valor de algum parâmetro de alguma população é igual ou diferente de um valor pré-estabelecido. Como a

população é modelada por uma distribuição de probabilidade, testam-se os parâmetros desta distribuição. O teste é feito considerando-se duas hipóteses, a nula H_0 e uma alternativa H_1 . Segundo [Montgomery \(1996, p.96\)](#), geralmente existem três maneiras de se especificar o valor de um parâmetro para o teste de hipótese:

1. O valor resulta de um estudo passado. Este procedimento é utilizado em gráficos de controle, onde se estimam os valores da média e desvio padrão (ou amplitude) e faz-se o teste para saber se estes valores se modificaram ao longo do tempo.
2. O valor resulta de alguma teoria ou modelo.
3. O valor escolhido é resultado de um contrato com especificações de projeto entre o cliente e o fabricante. Este procedimento é muito utilizado na escolha do valor dos índices de capacidade de processo no setor automotivo, onde se definem o valor dos índices em contrato.

Antes de se definir se a hipótese nula não pode ou pode ser rejeitada é necessária à definição de alguns conceitos:

- Região crítica: são os valores de probabilidades das caudas da distribuição que levam a rejeição da hipótese nula.
- Valor crítico: são os valores que delimitam até aonde o parâmetro estimado pode estar contido antes que se rejeite a hipótese nula.
- Região de aceitação: é a região em que não se rejeita a hipótese nula. Se o valor estimado estiver dentro do intervalo formado pelos valores críticos então não se rejeita a hipótese nula.

Existem dois erros que podem ser cometidos quando se testa uma hipótese. O primeiro é rejeitar a hipótese nula quando ela é verdadeira, e o segundo é falhar em

rejeitar a hipótese nula quando ela é falsa. As probabilidades destes dois tipos de erro podem ser definidas como mostradas abaixo:

$$\alpha = P\{\text{erro tipo I}\} = P\{\text{rejeitar } H_0 \mid H_0 \text{ é verdadeiro}\}$$

$$\beta = P\{\text{erro tipo II}\} = P\{\text{falhar em rejeitar } H_0 \mid H_0 \text{ é falso}\}$$

Outra definição utilizada também é o poder do teste, que é a probabilidade de se rejeitar a hipótese nula corretamente.

$$\text{Poder do teste} = 1 - \beta = P\{\text{rejeitar } H_0 \mid H_0 \text{ é falso}\}$$

O teste pode ser definido como bilateral ou unilateral (superior e inferior). No primeiro caso, por exemplo, se a média μ_0 de um processo fosse testada as duas hipóteses seriam:

$$H_0 : \mu = \mu_0$$

$$H_1 : \mu \neq \mu_0$$

Nota-se neste caso que o valor da hipótese alternativa pode ser tanto μ_0 maior ou menor que μ por isto o teste é bilateral. A região crítica é separada em duas partes, cauda superior e inferior da distribuição de probabilidade. No unilateral superior e inferior os testes seriam respectivamente:

$$H_0 : \mu = \mu_0 \quad \text{e} \quad H_0 : \mu = \mu_0$$

$$H_1 : \mu > \mu_0 \quad \text{e} \quad H_1 : \mu < \mu_0$$

No unilateral superior a região crítica esta na cauda superior e no unilateral inferior na cauda inferior da distribuição de probabilidade.

2.1.4 Intervalo de confiança

Quando se deseja estimar parâmetros de uma população retirando amostras desta, não se tem certeza sobre o verdadeiro valor do parâmetro da população. Então se calcula um intervalo de valores na qual o verdadeiro valor do parâmetro está contido com certa probabilidade. Existem três intervalos de confiança $100(1-\alpha)\%$ possíveis de serem construídos:

- O bilateral, onde se encontra um limite superior S e inferior I para o parâmetro. Este é utilizado nos limites do gráfico de controle (seção 2.2).

$$P\{S \leq \text{parâmetro} \leq I\} = 1 - \alpha \quad (2.13)$$

- O unilateral superior, onde se encontra um limite superior S para o parâmetro.

$$P\{\text{parâmetro} \leq S\} = 1 - \alpha \quad (2.14)$$

- O unilateral inferior, onde se encontra um limite inferior I para o parâmetro. Este é utilizado na estimação de índices de capacidade, quando se quer saber qual é o menor valor estimado do índice para que o processo seja considerado capaz (número máximo de defeituosos admitidos).

$$P\{I \leq \text{parâmetro}\} = 1 - \alpha \quad (2.15)$$

2.2 Gráficos de controle

Nesta seção será apresentada uma breve revisão de gráficos de controle, um texto mais detalhado pode ser encontrado em [Montgomery \(1996\)](#) no capítulo 5. O gráfico é uma forma visual de se testar a estabilidade do processo, tanto da média

quanto da variabilidade, foi desenvolvido por Shewhart na década de 20 e continua sendo utilizado até os dias de hoje. É a ferramenta mais importante do controle estatístico de processos. O gráfico é formado por uma linha central (LC), limite superior (LSC) e inferior de controle (LIC). Quando uma observação ocorre além dos limites de controle e se encontra uma explicação para isto, se diz que processo não está sob o controle estatístico ou é instável. Esta instabilidade se deve ao que foi denominado como causas especiais. Estas causas são responsáveis por variações (tanto na média quanto na variabilidade) não aleatórias do processo.

Quando a variação é aleatória se diz que o processo está sob efeito de causas comuns. Este efeito é resultado do somatório de pequenas causas que não são identificáveis e são inerentes ao processo. Quando o processo está apenas sob o efeito de causas comuns (todas as observações estão dentro dos limites de controle e são distribuídas aleatoriamente), se diz que está sob o controle estatístico ou é estável. A grande vantagem de se monitorar e atingir a estabilidade é que se sabe o que vai acontecer no futuro, isto é, que as observações vão ocorrer dentro de uma faixa de valores com uma certa probabilidade.

Estes gráficos nada mais são do que se fazer testes de hipótese sobre a estabilidade do processo. Quando ocorre um ponto fora dos limites de controle, e existe uma causa, está se rejeitando a hipótese nula de estabilidade. Nos testes definidos na [seção 2.1.3](#) existem dois erros cometidos, o erro tipo um e dois. Em gráficos de controle quando se comete o erro tipo um, significa dizer que o processo está fora do controle, quando na verdade não está. Este erro é definido pela escolha dos limites de controle. Já quando o erro tipo dois é cometido, está se dizendo que o processo está sob controle,

quando na verdade não está. Este erro depende da escolha dos limites de controle e do tamanho das amostras, quanto maior menor a chance de se errar.

Estas amostras podem ser coletadas individualmente ou em subgrupos (m), quanto maior o tamanho do subgrupo menor é o erro tipo dois cometido. A escolha do tamanho das amostras vai depender das características de cada processo e com custos associados ao processo de amostragem. Em processos químicos, por exemplo, o tamanho das amostras é igual a um, onde se produz uma batelada e é retirada uma amostra. Na produção de bens com alto valor agregado e quando o teste é destrutivo, o tamanho das amostras tende a ser de tamanho um também. Já em processos de produção de parafusos, com baixo valor agregado e fácil medição, o tamanho das amostras pode ser maior que um.

Para estes diversos tamanhos de subgrupos existem três opções de gráficos (gráficos Shewhart) que podem ser utilizados e serão mostrados na próxima seção. Dentre estas três opções estão: o gráfico da amplitude móvel e medidas individuais, média e amplitude, média e desvio padrão. Para a construção destes gráficos recomenda-se utilizar pelo menos 25 subgrupos.

2.2.1 Gráfico de amplitude móvel (MR) e medidas individuais ($\bar{I}\bar{X}$)

Quando as amostras retiradas do processo são de tamanho igual a um ($n = 1$) os gráficos a serem utilizados são: amplitude móvel e medidas individuais. O primeiro controla a variabilidade do processo e consiste em se medir a amplitude móvel de duas observações consecutivas. O segundo controla a média das observações individuais. Os limites de controle respectivamente são:

a) Gráfico de amplitude móvel.

$$\begin{aligned} LSC &= D_4 \overline{MR} \\ LC &= \overline{MR} \\ LIC &= 0 \end{aligned} \quad (2.16)$$

$$MR_i = |x_i - x_{i-1}| \quad (2.17)$$

O valor de D_4 para $n = 2$ é tabelado e pode ser encontrado na tabela 3A do anexo A.

b) Gráfico de medidas individuais.

$$\begin{aligned} LSC &= \bar{x} + 3 \frac{\overline{MR}}{d_2} \\ LC &= \bar{x} \\ LIC &= \bar{x} - 3 \frac{\overline{MR}}{d_2} \end{aligned} \quad (2.18)$$

O valor de d_2 para $n = 2$ é tabelado e pode ser encontrado na tabela 3A do anexo A.

2.2.2 Gráfico da média (\bar{X}) e amplitude (R)

Quando o tamanho das amostras (n) retiradas do processo são maiores que 1 e menores que 6 recomenda-se utilizar estes dois gráficos. A vantagem do gráfico da amplitude é a facilidade com que se realizam os cálculos necessários (pois são utilizados apenas o maior e o menor valor de cada amostra para a estimação da variabilidade, equação (2.21)), podendo facilmente ser utilizado no chão de fábrica. A amplitude de cada amostra R_i e a amplitude média \bar{R} são definidas nas equações abaixo.

$$R_i = x_{\max} - x_{\min} \quad i = 1, 2, \dots, m \quad (2.19)$$

$$\bar{R} = \frac{R_1 + R_2 + \dots + R_m}{m} \quad (2.20)$$

$$\sigma = \frac{\bar{R}}{d_2} \quad (2.21)$$

O valor de d_2 é tabelado (tabela 3A do anexo A) e depende do tamanho da amostra n .

Os limites de controle para os dois gráficos são:

a) Gráfico da amplitude.

$$\begin{aligned} LSC &= D_4 \bar{R} \\ LC &= \bar{R} \\ LSC &= D_3 \bar{R} \end{aligned} \quad (2.22)$$

Os valores de D_3 e D_4 são tabelados (tabela 3A do anexo A)

b) Gráfico da média

$$\begin{aligned} LSC &= \bar{\bar{X}} + A_2 \bar{R} \\ LC &= \bar{\bar{X}} \\ LIC &= \bar{\bar{X}} - A_2 \bar{R} \end{aligned} \quad (2.23)$$

Onde

$$\bar{X}_i = \frac{X_1 + X_2 + \dots + X_n}{n} \quad (2.24)$$

e

$$\bar{\bar{X}} = \frac{\bar{X}_1 + \bar{X}_2 + \dots + \bar{X}_m}{m} \quad (2.25)$$

O valor de A_2 é tabelado (tabela 1A do anexo A), mas seu valor também pode ser encontrado por

$$A_2 = \frac{3}{d_2} \quad (2.26)$$

2.2.3 Gráfico da média (\bar{X}) e desvio padrão (S)

Quando o tamanho das amostras (n) retiradas do processo são maiores que 7 recomendam-se utilizar estes dois gráficos. Para estes tamanho de amostras o gráfico da amplitude tem uma perda de eficiência considerável para medir a variabilidade do processo. Então se utiliza o gráfico do desvio padrão. O desvio padrão de cada amostra S_i e o desvio médio \bar{S} são definidos nas equações abaixo.

$$S_i = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}} \quad (2.27)$$

$$\bar{S} = \frac{1}{m} \sum_{i=1}^m S_i \quad (2.28)$$

a) Gráfico do desvio padrão.

$$\begin{aligned} LSC &= B_4 \bar{S} \\ LC &= \bar{S} \\ LSC &= B_3 \bar{S} \end{aligned} \quad (2.29)$$

Os valores de B_3 e B_4 são tabelados (tabela 2A do anexo A)

b) Gráfico da média.

$$\begin{aligned}
 LSC &= \bar{\bar{x}} + A_3\bar{S} \\
 LC &= \bar{\bar{x}} \\
 LIC &= \bar{\bar{x}} - A_3\bar{S}
 \end{aligned}
 \tag{2.30}$$

Os valores de A_3 são tabelados (tabela 2A do anexo A)

2.3 Índice de capacidade de processos

Nesta seção será apresentada a revisão teórica referente a índices de capacidade de processos. O histórico do desenvolvimento dos índices, sua estimação e os seus intervalos de confiança.

2.3.1 Histórico dos índices

O primeiro estudo publicado foi feito por [Juran et al. \(1974\)](#), onde se considerou um relacionamento entre a variabilidade do processo e as especificações do cliente. Foi criada uma razão de capacidade, onde se compara a variabilidade do processo com os limites de especificação (tolerância admitida).

$$\text{Razão de capacidade} = \frac{\text{variação de } 6\sigma}{\text{Limites de tolerância}}
 \tag{2.31}$$

Após este primeiro estudo surgiram os primeiros índices de capacidade no Japão, que tiveram seu uso nas indústrias apenas no início da década de 80. Originalmente foram cinco os índices criados: C_p , C_{pk} , k , C_{pu} e C_{pl} . Estes são classificados como pertencentes à primeira geração, e ainda são os mais utilizados nas indústrias. A segunda geração foi desenvolvida por Taguchi, quando ele criou o conceito de função de perda em 1985, onde se têm uma penalidade por estar longe do alvo de especificação ([HSIANG e TAGUCHI, 1985](#)). Esta perda foi definida através de

uma função quadrática e foi incorporada ao índice C_{pm} (BOYLES, 1991). Então quando a média está fora do alvo se tem uma penalização.

A terceira geração é considerada como índices que foram desenvolvidos após C_{pm} , principalmente na década de 90. Surgiram uma quantidade enorme, cada um deles desenvolvido para lidar com casos específicos de dados, como para dados não-normais ou autocorrelacionados por exemplo. Poucos destes são utilizados, por sua difícil interpretação e pela dificuldade em se mudar de índices tradicionais para novos. Geralmente procura-se fazer o tratamento dos dados e aplicar para as condições que foram estabelecidas para os índices da primeira geração.

2.3.2 Condições necessárias para a utilização dos índices de capacidade

Para que se possam utilizar os índices de capacidade da primeira geração corretamente algumas condições devem ser satisfeitas Kotz e Lovelace (1998, p. 35):

- a) O processo deve ser estável. Caso esta condição não seja cumprida o índice estimado só representa uma fotografia daquele momento do processo, não se podendo tirar conclusões sobre a capacidade no futuro.
- b) Os dados devem seguir a distribuição normal. Caso esta condição não seja cumprida existirá viés na estimativa dos índices, já que estes são baseados nas estimativas da média e desvio padrão de uma distribuição normal. Como utilizar os índices na condição de não normalidade será visto no [capítulo 3](#).
- c) Os dados devem ser independentes. Caso esta condição não seja cumprida existirá a subestimação do desvio padrão, causando a superestimação do índice de capacidade. O processo parece ser melhor do que na realidade é.

2.3.3 Estimação da média ($\hat{\mu}$) e desvio padrão ($\hat{\sigma}$)

Para se utilizar os índices de capacidade que serão vistos no decorrer deste capítulo, são necessárias as estimações da média μ e desvio padrão σ , baseados nas amostras retiradas do processo. A média pode ser estimada da maneira clássica, como na equação (2.32).

$$\hat{\mu} = \bar{X} = \frac{\sum_{i=1}^n X_i}{n} \quad (2.32)$$

Já o desvio padrão pode ser estimado de duas maneiras

$$\hat{\sigma} = \frac{\bar{R}}{d_2} \quad (2.33)$$

ou

$$S = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}} \quad (2.34)$$

onde \bar{R} é definido pela equação (2.20).

Mas afinal qual a diferença entre as duas equações mostradas acima? A equação (2.33) é utilizada no cálculo do desvio padrão em gráficos de controle da amplitude e estima a variabilidade dentro de cada amostra. É conhecida também como variabilidade de curto prazo ou instantânea. Esta estimativa é utilizada principalmente pelos práticos, pela facilidade que se tem em calcular e por ter sido implementada na origem dos gráficos de controle.

Já os estatísticos utilizam a equação (2.34) como estimativa do desvio padrão do processo, onde se está estimando a variabilidade de longo prazo, pois é considerada a

variabilidade de todas as observações individuais e não apenas a variabilidade dentro de cada subgrupo. Outra explicação para a utilização de S por parte dos estatísticos é o fato de ser tratável estatisticamente, enquanto que $\hat{\sigma}$ é considerado intratável (KOTZ e LOVELACE,1998).

A diferença entre S e $\hat{\sigma}$ pode ser representada na Figura 2, onde S é o desvio padrão de todas as observações e R_1 , R_2 e R_3 é a amplitude de cada subgrupo. A estimação de $\hat{\sigma}$ é obtida pelo valor médio das amplitudes das amostras.

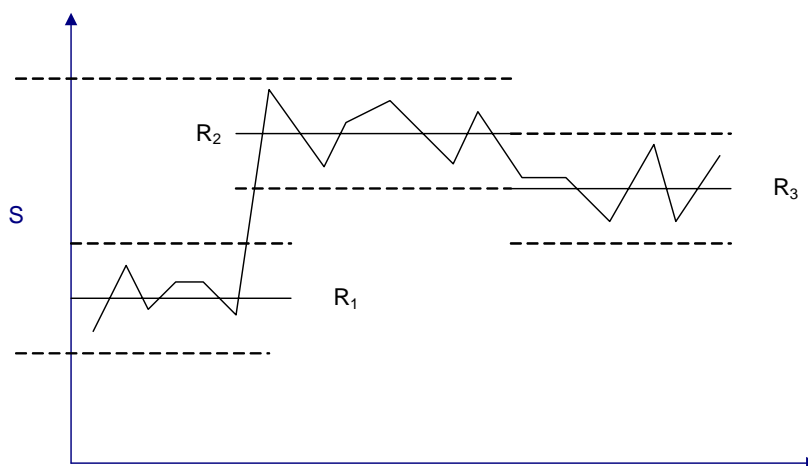


Figura 2 – Diferença entre a variabilidade de curto e longo prazo
Fonte: Shore (1997).

Portanto, o valor estimado por S é sempre maior que o estimado por $\hat{\sigma}$.

Esta diferença entre os dois estimadores descritos acima é um tema de grande discussão na literatura, pois quando se estima o desvio padrão por S , o índice de capacidade é chamado de índice de performance de processo (P_p e P_{pk}), definido no manual AIAG (1995), que serve de referência para muitos cursos de qualidade hoje em dia. É sugerido que se utilizem estes índices quando o processo não está em controle estatístico. Segundo o mesmo manual, se a estimação é feita por $\hat{\sigma}$, o índice é chamado de índice de capacidade de processos (C_p e C_{pk}) e utilizado em processos estáveis. Em

Kotz e Lovelace (1998, p. 252) é apontado que quando o processo está em controle estatístico $S = \hat{\sigma}$, portanto $C_p = P_p$ e $C_{pk} = P_{pk}$. Já Rodriguez em Kotz *et. al* (2002, p. 29) argumenta que mesmo um processo estando em controle estatístico, não é perfeitamente estável e existem variações de subgrupo para subgrupo que os gráficos de controle não detectam. Por isto recomenda que se utilize S como um estimador mais fiel de toda a variabilidade do processo. Pelas definições descritas acima o gráfico da Figura 2 representaria um processo instável, um exemplo bem diferente que o autor gostaria de demonstrar, quando abordou o assunto.

Todos os intervalos de confiança que serão apresentados na próxima seção são calculados com σ estimado por S . Caso se pretenda estimar σ por $\hat{\sigma}$, não se deve utilizar os intervalos de confiança que serão abordados a seguir.

2.3.4 Índices de Capacidade da Primeira Geração (C_p , C_{pk} , k , C_{pl} , C_{pu})

A primeira geração foi o desenvolvimento dos índices C_p , C_{pk} , k , C_{pl} , e C_{pu} no Japão e largamente empregados na indústria automotiva. Depois se tornaram os mais utilizados em outros tipos de indústrias também, e continuam até os dias de hoje sendo os mais populares. Isto se deve a sua simplicidade de cálculo, interpretação e aplicação no chão de fábrica, com bons resultados obtidos.

a) O Índice C_p .

O índice C_p foi o primeiro a ser desenvolvido, em 1974 por Juran, mas sua utilização só começou de fato no começo da década de 80. Sua forma pode ser vista na equação (2.35).

$$C_p = \frac{LSE - LIE}{6\sigma} \quad (2.35)$$

Onde LSE é o limite superior de especificação, LIE é o limite inferior de especificação e σ é o desvio padrão dos dados coletados. Este índice mede apenas a variabilidade do processo em comparação com as tolerâncias admitidas nas especificações de projeto (Figura 3) e por isto também é conhecido como índice de capacidade potencial, pois não leva em consideração o local da média em relação ao alvo do processo ou os limites de especificação.

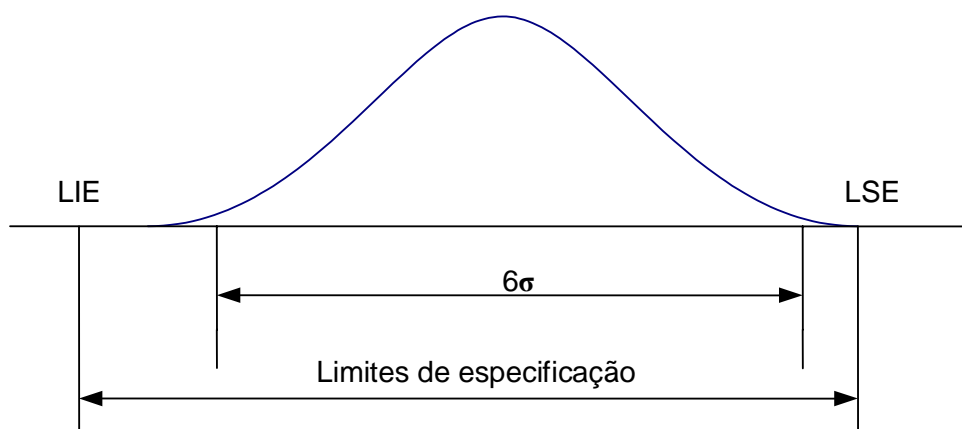


Figura 3 – Comparação dos limites de especificação com a variabilidade do processo

Se a média do processo estiver centrado no alvo e a distribuição dos dados for perfeitamente normal a relação do índice com a produção de não conformes pode ser vista na Tabela 3.

Tabela 3 - C_p e a produção de não conformes

C_p	Produção de não conformes por milhão
0,5	133.614,403
1	2.699,79606
1,33	66,0732953
1,5	6,79534627
1,67	0,54430051
2	0,00197318

A definição em que nível o processo é considerado capaz (valor de C_p) fica a cargo da escolha da empresa, pois cada processo vai admitir custos com a produção de defeituosos diferentes. Mas, em geral, **Montgomery (1996)** recomenda os seguintes valores mínimos para C_p e C_{pk} (que será apresentado no item b desta seção) apresentados na Tabela 4.

Tabela 4 – Recomendação de valores mínimos para os índices C_p e C_{pk}

	Especificações Bilaterais	Especificações Unilaterais
Processos existentes	1,33	1,25
Novos processos	1,50	1,45
Processos existentes que envolvam características de segurança, força, ou parâmetros críticos	1,50	1,45
Processos novos que envolvam características de segurança, força, ou parâmetros críticos	1,67	1,60

Fonte: **Montgomery (1996)**

Tanto o valor do índice quanto o percentual de não conformes produzidos são valores sujeitos as variações devido à amostragem. Neste caso o único valor estimado é σ , que se for por S, tem uma distribuição χ (qui). O intervalo de confiança para C_p foi desenvolvido por **Kane (1986)** e pode ser visto na equação (2.36).

$$\left(\frac{\chi_{n-1,\alpha/2}}{\sqrt{n-1}} \hat{C}_p, \frac{\chi_{n-1,1-\alpha/2}}{\sqrt{n-1}} \hat{C}_p \right) \quad (2.36)$$

Mas, na prática, o mais interessante não é o intervalo de confiança bilateral, e sim qual a probabilidade do índice ser menor do que aquele desejado como alvo ($C_p^{[c]}$). Então se utiliza apenas o limite inferior de confiança $(1 - \alpha)100\%$.

$$C_p^{[c]} = \hat{C}_p \sqrt{\frac{\chi_{n-1}^2(\alpha)}{n-1}} \quad (2.37)$$

Como utilizando esta expressão não é possível isolar n , que é o tamanho da amostra, foi desenvolvida uma aproximação. Franklin (1999) demonstra que:

$$\left(\frac{C_p^{|c|}}{\hat{C}_p}\right)^2 = \frac{1}{n-1} \left[(n-1) \left(Z(\alpha) \sqrt{\frac{2}{9(n-1)}} + 1 - \frac{2}{9(n-1)} \right)^3 \right] \quad (2.38)$$

Franklin (1999) isolou n e obteve:

$$n = 1 + \frac{2}{9} \frac{1}{\left[\frac{Z(\alpha)}{2} + \sqrt{1 + \left(\frac{Z(\alpha)}{2}\right)^2 - \left(\frac{C_p^{|c|}}{\hat{C}_p}\right)^{2/3}} \right]^2} \quad (2.39)$$

Agora é possível calcular o tamanho da amostra necessária para um intervalo de confiança, baseado no valor estimado de \hat{C}_p e o alvo $C_p^{|c|}$.

Exemplo 2.1: um processo com índice alvo $C_p^{|c|} = 1,33$ obteve uma estimativa de $\hat{C}_p = 1,5$ com uma amostra de tamanho $n = 47$. Com este tamanho de amostra existe uma probabilidade de 85% que o valor estimado seja maior ou igual a 1,33. Se o objetivo for ter uma confiança de 95%, aplicando a equação (2.39), a amostra necessária deve ser de $n = 109$. Se o objetivo for ter uma confiança de 95% com uma amostra de $n = 47$, o valor estimado deve ser $\hat{C}_p = 1.61$. Isto demonstra que se pode escolher entre aumentar o tamanho da amostra ou mudar o processo para que se tenha uma confiança no valor estimado.

b) O índice C_{pk} .

Como o índice C_p não leva em consideração a localização da média do processo foi desenvolvido o índice C_{pk} , que mede a distância entre a média do processo e os limites de especificação (superior e inferior) em relação a metade da variabilidade do processo (3σ). Dentre todos os índices existentes é o mais utilizado para medir a capacidade de processos. Quando o processo está centrado, isto é, o valor da média é igual ao alvo, $C_p = C_{pk}$. À medida que C_{pk} é menor que C_p , indica que o processo está se distanciando do valor alvo.

$$C_{pk} = \min\left(\frac{LSE - \mu}{3\sigma}, \frac{\mu - LIE}{3\sigma}\right) \quad (2.40)$$

A [Figura 4](#) representa o que o índice C_{pk} está medindo.

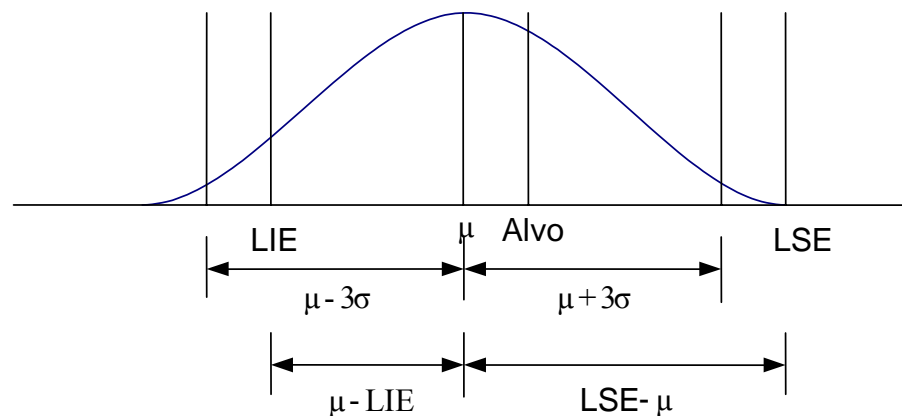


Figura 4 – Relação da variabilidade do processo em relação a média com os limites de especificação e a média.

Também foram desenvolvidos intervalos de confiança para este índice, mas a distribuição de C_{pk} envolve a distribuição conjunta de duas distribuições t não centradas, fazendo com que apenas aproximações fossem desenvolvidas até hoje. Uma

delas é a equação (2.41) Heavlin (1988) e o limite inferior de confiança $(1-\alpha)100\%$ é visto em (2.42) (FRANKLIN e WASSERMAN'S, 1992).

$$\left(\begin{array}{l} \hat{C}_{pk} - z_{1-\alpha/2} \left\{ \frac{n-1}{9n(n-3)} + \hat{C}_{pk}^2 \frac{1}{2(n-3)} \left(1 + \frac{6}{n-1} \right) \right\}^{1/2} \\ \hat{C}_{pk} + z_{1-\alpha/2} \left\{ \frac{n-1}{9n(n-3)} + \hat{C}_{pk}^2 \frac{1}{2(n-3)} \left(1 + \frac{6}{n-1} \right) \right\}^{1/2} \end{array} \right) \quad (2.41)$$

$$C_{pk}^{|c|} = \hat{C}_{pk} - Z(\alpha) \sqrt{\frac{1}{9n} + \frac{(\hat{C}_{pk})^2}{2n-2}} \quad (2.42)$$

A equação (2.42) é válida para $n \geq 30$, pois para valores menores a estimação é viesada. Franklin (1999) isolou n e obteve:

$$n = (Z(\alpha))^2 \frac{\left[\frac{1}{9(\hat{C}_{pk})^2} + 1/2 \right]}{\left[1 - \frac{C_{pk}^{|c|}}{\hat{C}_{pk}} \right]^2} \quad (2.43)$$

Exemplo 2.2: um processo com índice alvo $C_{pk} = 1,33$ obteve uma estimativa de $\hat{C}_{pk} = 1,5$ com uma amostra de tamanho $n = 47$. Com este tamanho de amostra existe uma probabilidade de 85% que o valor estimado seja maior ou igual a 1,33. Se o objetivo for ter uma confiança de 95%, aplicando a equação (2.43), a amostra necessária deve ser de $n = 116$. Se o objetivo for ter uma confiança de 95% com uma amostra de $n = 47$, o valor estimado deve ser $\hat{C}_{pk} = 1.615$.

c) Os índices C_{pu} e C_{pk} .

Para os casos onde só existe um limite de especificação, superior ou inferior, existem os índices C_{pu} e C_{pl} respectivamente.

$$C_{pu} = \frac{LSE - \mu}{3\sigma} \quad (2.44)$$

$$C_{pl} = \frac{\mu - LIE}{3\sigma} \quad (2.45)$$

Pode-se representar C_{pk} através destes dois índices.

$$C_{pk} = \min(C_{pu}, C_{pl}) \quad (2.46)$$

Intervalos de confiança e testes de hipótese podem ser encontrados em [Pearn e Chen \(2002\)](#) e [Pearn e Shu \(2003\)](#).

d) O índice k .

Por fim, último índice que faz parte dos cinco originais é o k . Nem C_p e C_{pk} medem a distância entre a média do processo e o alvo, por isto este foi desenvolvido. Quando $k = 0$ o processo está centrado no alvo. Quando $k = 1$ a média do processo está localizado em um dos limites de especificação. Quando $0 < k < 1$ a média está localizada entre o alvo e um limite de especificação.

$$k = \frac{2 \left| \mu - \frac{1}{2}(LSE + LIE) \right|}{LSE - LIE} \quad (2.47)$$

Existe uma relação direta entre C_p , C_{pk} e k , que pode ser vista na equação (2.48).

$$C_{pk} = (1-k)C_p \quad (2.48)$$

Na próxima seção será vista a segunda geração de índices, que é o índice C_{pm} .

2.3.5 Segunda Geração (C_{pm})

Os índices da primeira geração não consideravam o alvo do processo para a sua estimação. Para resolver este problema em [Hsiang e Taguchi, \(1985\)](#) foi desenvolvido a função de perda quadrática, que penaliza por estar fora do alvo T de especificação.

$$\tau^2 = E[(X - T)^2] = \sigma^2 + (\mu - T)^2 \quad (2.49)$$

É considerado que não basta apenas produzir peças dentro dos limites de especificação, mas sim também estar o mais próximo possível do alvo, desta maneira o processo se torna mais robusto, pois um desvio na média implicará em um número menor de defeituosos produzidos.

O índice desenvolvido a partir da função de perda foi o C_{pm} definido pela equação (2.50). Se o processo estiver centrado no alvo $C_{pm} = C_p = C_{pk}$.

$$C_{pm} = \frac{LSE - LIE}{6\tau} = \frac{LSE - LIE}{6\sqrt{\sigma^2 + (\mu - T)^2}} \quad (2.50)$$

A relação entre C_{pm} , C_{pk} e C_p pode ser vista na equação (2.51).

$$C_{pm} = \frac{C_p}{\sqrt{1+9(C_p - C_{pk})^2}} \quad (2.51)$$

Uma aproximação para o intervalo de confiança inferior $(1-\alpha)100\%$ para C_{pm} é visto na equação (2.52) .

$$C_{pm}^{cl} = \hat{C}_{pm} \sqrt{\frac{\chi_v^2(\alpha)}{\nu}} \quad (2.52)$$

Onde ν é estimado por:

$$\nu = \frac{2}{9} \frac{1}{\left[\frac{Z(\alpha)}{2} + \sqrt{1 + \left(\frac{Z(\alpha)}{2} \right)^2 - \left(\frac{C_{pm}^{cl}}{\hat{C}_{pm}} \right)^{2/3}} \right]^2} \quad (2.53)$$

E o tamanho da amostra para este limite inferior de confiança é (Franklin,1999):

$$n = \frac{(1+2\zeta^2)}{(1+\zeta^2)^2} \frac{2}{9} \frac{1}{\left[\frac{Z(\alpha)}{2} + \sqrt{1 + \left(\frac{Z(\alpha)}{2} \right)^2 - \left(\frac{C_{pm}^{cl}}{\hat{C}_{pm}} \right)^{2/3}} \right]^2} \quad (2.54)$$

E o valor de ζ é encontrado pela equação (2.55).

$$\zeta = \frac{(\mu - T)}{\sigma} \quad (2.55)$$

Restrições quanto ao uso do índice C_{pm} :

- O custo para manter o processo no alvo pode não compensar. Em processos em que a média se desloca ao longo do tempo (por desgaste de ferramentas por exemplo), pode ser interessante ela não estar centrada no alvo. O custo para se trocar a ferramenta pode ser maior que manter o processo no alvo.

- O índice não dá uma medida direta de o quanto o processo está fora do alvo.

- A interpretação do índice não é simples como a de C_p e C_{pk} .

2.3.6 Terceira Geração de índices

A terceira geração é classificada como a de índices desenvolvidos após o índice C_{pm} e começou no início da década de 90. O primeiro foi o C_{pmk} introduzido por *Pearn et al. (1992)* que é mais sensível a desvios do alvo do que C_{pm} , seguido de uma quantidade enorme de índices desenvolvidos principalmente para casos especiais de dados.

Em *Kotz e Lovelace (1998)* e *Kotz et al (2002)* existe uma grande revisão sobre estes índices, que muitas vezes são modificações dos índices da primeira e segunda geração para os casos de não normalidade, autocorrelação dos dados e processos de montagem. Mas segundo Hubele existe um grande problema com o desenvolvimento destes índices:

Um dos maiores problemas, até mesmo para o mais educado e bem intencionado usuário, é que existem tantos tipos de índices de capacidade de processo sendo utilizados por várias organizações que existe um estado de confusão e, conseqüentemente, um abandono. Pretendendo-se originalmente fazer a vida mais simples, agora nós temos uma estimativa conservativa do Drs. Kotz e Johnson que existem aproximadamente 20 variações de índices

univariados e 7 multivariados. Não existe mistério em porque a atenção está sendo gasta em escolher o índice “correto” e pouca atenção sendo gasta em apropriadamente usá-los em conjunto com uma estimativa de variabilidade. Nós devemos parar de desenvolver outros índices ou focar convencimento dos práticos para utilizar os já existentes mais responsabilmente? *Kotz et al* (2002, p. 20, tradução nossa).

2.3.7 Relação entre os índices de capacidade e a produção de não conformes

A quantidade produzida de itens não conformes é o foco de estudo da capacidade de processos. Quer se produzir o máximo de itens de acordo com o especificado, tentando minimizar o número de produtos defeituosos. Para se calcular a produção de não conformes pode-se utilizar diretamente a distribuição de probabilidades dos dados coletados nas amostras.

$$p = \Phi\left(\frac{LIE - \mu}{\sigma}\right) + \left(1 - \Phi\left(\frac{LSE - \mu}{\sigma}\right)\right) \quad (2.56)$$

A primeira parte da equação (2.56) representa à cauda esquerda (percentual de defeitos abaixo do LIE) e a segunda a cauda direita (percentual de defeitos acima do LSE) da distribuição normal.

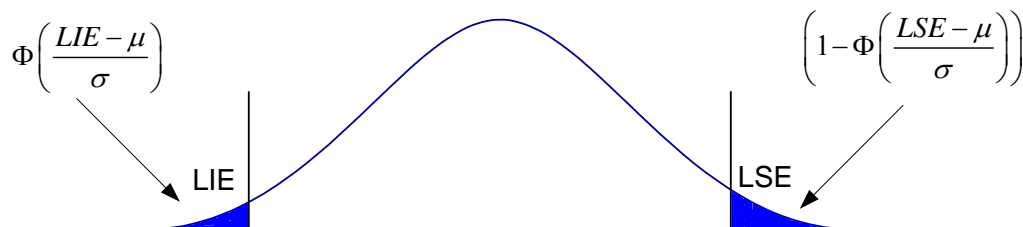


Figura 5 – Percentual de defeituosos

Se o número de defeituosos pode ser calculado diretamente assim, qual a vantagem em se utilizar os índices então? Os índices são adimensionais e representam também a quantidade de defeituosos produzidos. Por serem adimensionais pode-se comparar a capacidade de processos totalmente diferentes, seja na unidade ou grandeza de medida. Por outro lado utilizando a fórmula (2.56), a média e desvio padrão são características de cada processo, ficando difícil a comparação.

Os índices C_p e C_{pk} sozinhos não indicam a quantidade de defeituosos produzidos, mas em conjunto sim. O p da equação (2.57) representa o mesmo valor da equação (2.56). Uma demonstração para o desenvolvimento da equação pode ser encontrado em [Kotz e Lovelace \(1998, p.65\)](#)

$$p = \Phi \left[-3(2C_p - C_{pk}) \right] + \Phi(-3C_{pk}) \quad (2.57)$$

Na [Tabela 5](#) são apresentados alguns valores de C_p e C_{pk} e a quantidade de não conformes por milhão (ppm).

Tabela 5 – Relação de C_p e C_{pk} com a produção de não conformes por milhão

C_p/C_{pk}	1	1.15	1.2	1.25	1.3	1.35	1.4	1.45	1.5	1.55	1.6
1	2700										
1.15	1398	561									
1.2	1363	369	318								
1.25	1353	306	207	177							
1.3	1351	287	172	114	96						
1.35	1350	282	163	95	61	51					
1.4	1350	281	160	90	51	32	27				
1.45	1350	280	159	89	49	27	17	14			
1.5	1350	280	159	88	48	26	14	8	7		
1.55	1350	280	159	88	48	26	14	7	4	3	
1.6	1350	280	159	88	48	26	13	7	4	2	2

Outra forma ainda de se calcular a quantidade de defeituosos é através dos índices C_{pl} e C_{pu} . ([equação \(2.58\)](#)).

$$p = \Phi(-3C_{pl}) + [1 - \Phi(3C_{pu})] \quad (2.58)$$

Com o índice C_{pm} é necessário que o seu valor, o da média e o desvio padrão sejam calculados para se estimar o percentual de defeituosos produzidos. **Kotz & Johnson (1993)** encontraram esta relação de defeituosos com índice.

$$LSE = -LIE = d \quad (2.59)$$

$$p = \Phi\left(\frac{-d - \mu}{\sqrt{\lambda^2 - \mu^2}}\right) + \left[1 - \Phi\left(\frac{d - \mu}{\sqrt{\lambda^2 - \mu^2}}\right)\right] \quad (2.60)$$

$$\lambda = \frac{d}{3C_{pm}} \quad (2.61)$$

Neste capítulo foram vistos os índices de capacidade, como estimá-los, o cálculo de intervalos de confiança e a relação deles com a produção de não conformes. Mas, para isto, as condições apontadas na seção 2.3.2 devem ser satisfeitas. Quando isto não ocorre é preciso aplicar os conceitos que serão vistos no próximo capítulo.

3 ANÁLISE DA CAPACIDADE NA PRESENÇA DE DADOS NÃO NORMAIS E AUTOCORRELACIONADOS.

Neste capítulo será apresentado como realizar o estudo da capacidade de processos, utilizando índices de capacidade, para dois casos especiais: quando os dados não seguem a distribuição normal e quando são autocorrelacionados. Na [seção 3.1](#) será apresentado as causas para não normalidade, na [3.2](#) alguns testes de normalidade, tanto na forma gráfica como estatística, na [3.3](#) a transformação de Johnson (transformação dos dados para normalidade), na [3.4](#) e [3.5](#) como utilizar os índices para dados não normais (método de Clements) e autocorrelacionados respectivamente.

3.1 Causas para não normalidade

Como discutido na [seção 2.3.2](#) os dados devem seguir a distribuição normal para que se estimem os índices de capacidade da primeira geração corretamente. Existem várias causas que podem levar a não normalidade dos dados, mas na maioria das vezes é a limitação física (superior ou inferior) da característica que está sendo produzida. Esta limitação pode gerar assimetria nos dados, sendo este o problema mais encontrado nos processos. Em [Kotz e Lovelace \(1998, p. 133\)](#) são descritos alguns exemplos de características que podem ter sua distribuição não normal:

- Distância entre coordenadas, eixos e centros;
- Paralelismo, perpendicularidade e ângulo;
- Rugosidade, perfil, redondeza e nivelamento.

Por exemplo, a rugosidade tem como limite inferior de sua característica o valor zero, fazendo com que a distribuição tenha assimetria para a direita. Em [Deleryd \(1999b\)](#) são apontados alguns efeitos de distribuições não normais na estimação do

número de defeituosos. Considerando três distribuições diferentes, uma χ^2 com 4.5 graus de liberdade, uma t com 8 graus de liberdade e uma uniforme, a quantidade de itens não conformes é respectivamente (considerando $\pm 3\sigma$ de limite) 14.000, 4.000 e 0 unidades. Utilizando a distribuição normal a quantidade é de 2.700. Portanto, se a normalidade fosse considerada existente nos três primeiros processos, ocorreria um erro na estimação do número de não conformes.

Na próxima seção serão vistas técnicas para se testar a normalidade, depois a transformação de dados para distribuições não normais e a aplicação de índices de capacidade sob estas condições. Por fim, será tratado o caso de autocorrelação nos dados.

3.2 Teste de normalidade

A normalidade pode ser verificada tanto na forma gráfica (seção 3.2.1) como na forma de um teste de hipótese (seção 3.2.2). Neste capítulo serão vistas as duas formas e o indicado é que se utilizem ambas.

3.2.1 Método gráfico (histograma, gráfico normal de probabilidade e função de densidade kernel)

Nesta seção serão apresentadas três formas gráficas para se verificar a normalidade. São elas: o histograma, o gráfico normal de probabilidade e a função densidade kernel. A primeira é a forma mais básica de demonstração gráfica da distribuição de dados, a segunda já envolve alguns conhecimentos em estatística e a terceira é a forma mais complexa de se trabalhar.

a) Histograma

No histograma divide-se a série de dados em classes. O número de classes escolhido pode ser a raiz quadrada do número de observações, ficando entre 4 e 20 classes (MONTGOMERY, 1996, p.40). Por exemplo, se forem cem observações, dividir a série em 10 classes. No exemplo a seguir, que será utilizado ao longo do capítulo, existem 120 observações do ganho de amplificadores (dB) (Tabela 6), calculando a raiz quadrada obtiveram-se 10 classes.

Tabela 6 – Ganho de amplificadores (dB)

8.1	10.4	8.8	9.7	7.8	9.9	11.7	8	9.3	9	8.5	8.6
8.2	8.9	10.1	9.4	9.2	7.9	9.5	10.9	7.8	8.3	8.2	9.2
9.1	8.4	9.6	11.1	7.9	8.5	8.7	7.8	10.5	8.5	9	8.5
11.5	8	7.9	8.3	8.7	10	9.4	9	9.2	10.7	10.2	9.6
9.3	9.7	8.7	8.2	8.9	8.6	9.5	9.4	8.8	8.3	9.5	9
8.4	9.1	10.1	7.8	8.1	8.8	8	9.2	8.4	7.8	8.3	10.7
7.9	8.5	9.2	8.7	10.2	7.9	9.8	8.3	9	9.6	8.9	8.6
9.9	10.6	8.6	9.4	8.8	8.2	10.5	9.7	9.1	8	9.1	10
8.7	9.8	8.5	8.9	9.1	8.4	8.1	9.5	8.7	9.3	10.3	10.8
8.1	10.1	9.6	8.3	8	9.8	9	8.9	8.1	9.7	8.4	8.6

Fonte: (The Tools ...,1990)

Depois de determinada o número de classes (k), o próximo passo é determinar o tamanho de cada classe (h) e a frequência de ocorrência das observações dentro de cada uma.

$$h = \frac{(x_{\max} - x_{\min}) + 0,1}{k} \quad (3.1)$$

Substituindo os dados da Tabela 6 ($x_{\max} = 11,7$, $x_{\min} = 7,8$ e $k = 10$) na equação (3.1)

foi obtido $h = 0,4$.

Na Tabela 7 é apresentada a frequência de ocorrência dos valores dos amplificadores em cada classe.

Tabela 7 – Distribuição de frequência para o ganho de amplificadores

Classe	Ganho (dB)	Frequência	Frequência acumulada
1	$7,8 \leq x < 8,2$	24	24
2	$8,2 \leq x < 8,6$	22	46
3	$8,6 \leq x < 9,0$	21	67
4	$9,0 \leq x < 9,4$	17	84
5	$9,4 \leq x < 9,8$	12	96
6	$9,8 \leq x < 10,2$	12	108
7	$10,2 \leq x < 10,6$	5	113
8	$10,6 \leq x < 11,0$	4	117
9	$11,0 \leq x < 11,4$	1	118
10	$11,4 \leq x < 11,8$	2	120

Utilizando os dados calculados na Tabela 7, foi construído o histograma da Figura 6, onde no eixo x temos o intervalo de cada classe e no eixo y a frequência de ocorrência. Observando o histograma percebe-se que existe assimetria positiva nos dados (para a direita).

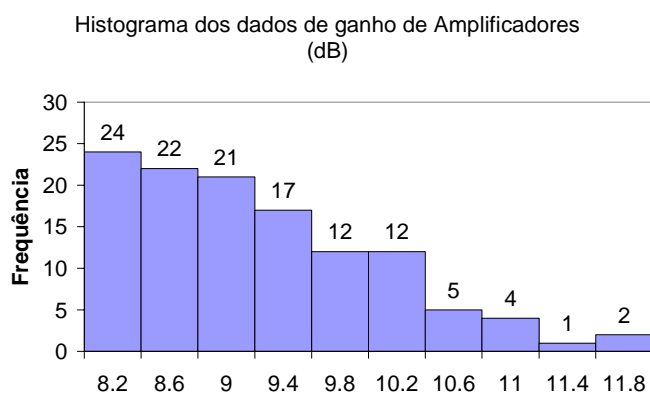


Figura 6 – Histograma do ganho de amplificadores

b) Gráfico normal de probabilidade.

A outra forma gráfica utilizada é o gráfico normal de probabilidade, no qual se compara o valor esperado normal e o valor ordenado das observações, esperando-se uma relação linear entre as duas variáveis, quando existir normalidade. Blom (1958) propôs uma aproximação para calcular o valor esperado normal, que pode ser obtido por meio da equação (3.2), onde i é a ordem dos dados observados, e vai de 1 até n , Φ^{-1} é

a distribuição cumulativa normal inversa. A [Figura 7](#) é um exemplo de gráfico normal de probabilidade feito com os dados da [Tabela 6](#) (p.52).

$$m_i = \Phi^{-1}\left(\frac{i - 3/8}{n + 1/4}\right) \quad (3.2)$$

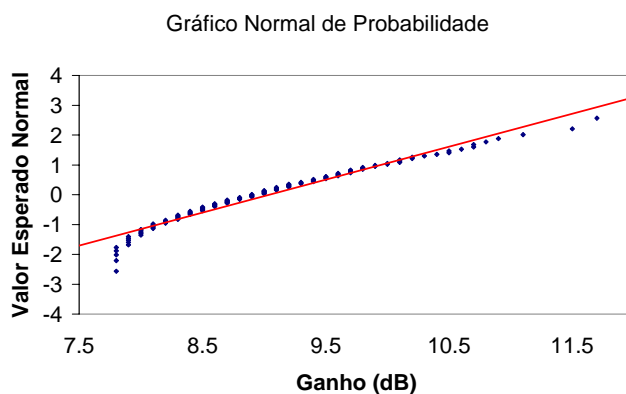


Figura 7 – Gráfico normal de probabilidade para o ganho de amplificadores

Nos dois gráficos mostrados acima é possível verificar que os dados não seguem a distribuição normal. Existe uma grande assimetria positiva (para a direita). Mas em outros casos não se chega a uma conclusão sobre a normalidade dos dados utilizando os gráficos, então se utiliza um teste estatístico, descrito na próxima seção.

c) Função de densidade kernel (FDK)

A função de densidade kernel é definida pela [equação \(3.3\)](#), onde n é o tamanho da amostra, h o parâmetro de suavização, $K(x)$ a função kernel e X_i a série de dados [\(SILVERMAN, 1986\)](#).

$$f(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \quad (3.3)$$

Por definição a função de densidade kernel deve satisfazer a seguinte condição:

$$\int_{-\infty}^{\infty} K(x) dx = 1 \quad (3.4)$$

Existem várias funções que podem ser escolhidas que satisfazem à equação (3.4) e a escolhida neste trabalho foi a normal.

$$K(x) = \frac{1}{\sqrt{2\pi}} e^{\left(-\frac{1}{2}x^2\right)} \quad (3.5)$$

Substituindo a equação (3.5) na equação (3.3) obtém-se:

$$f(x) = \frac{1}{nh\sqrt{2\pi}} \sum_{i=1}^n e^{\left(-\frac{1}{2}\left(\frac{x-X_i}{h}\right)^2\right)} \quad (3.6)$$

Depois de escolhida a função é necessário definir o número de pontos (M) e a faixa de valores para x que se quer utilizar para construção do gráfico. Os valores de x são calculados da seguinte maneira:

$$x_i = X_L + i \left(\frac{X_U - X_L}{M} \right), \quad \text{para } i = 0, 1, \dots, M-1 \quad (3.7)$$

O maior valor X_U foi calculado usando $\mu + 4S$ da série de dados X_i e o menor valor X_L usando $\mu - 4S$. O parâmetro h que controla a suavização da função de densidade estimada é calculado segundo a equação (3.8), onde R é amplitude entre quartis (3º quartil - 1ºquartil). Quanto maior o valor de h menor é a suavização da função (SILVERMAN, 1986).

$$h = 0,9n^{-1/5} \min\left(S, \frac{R}{1,34}\right) \quad (3.8)$$

Aplicando os procedimentos descritos acima aos dados da [Tabela 6 \(p. 52\)](#), obteve-se: $h = 0,9 \cdot 120^{-1/5} \min\left(0,876; \frac{1,125}{1,34}\right) = 0,3026$. Este valor encontrado é aplicado à [equação \(3.6\)](#). As duas curvas da Figura 8 mostram como é a função de densidade para os dados de ganho de amplificadores. A curva de linha cheia representa uma normal perfeita, com a média e desvio padrão estimados pelos dados observados. A curva de linha pontilhada representa a função de densidade kernel dos dados observados na [Tabela 6](#). Quando não se testa a normalidade, considera-se que os dados seguem a distribuição da curva 1, quando na realidade eles seguem a distribuição da curva 2. Assim como observado nos outros gráficos acima, percebe-se que existe uma assimetria dos dados para a direita. A vantagem deste gráfico é que a visualização de não normalidade é mais evidente que no histograma.

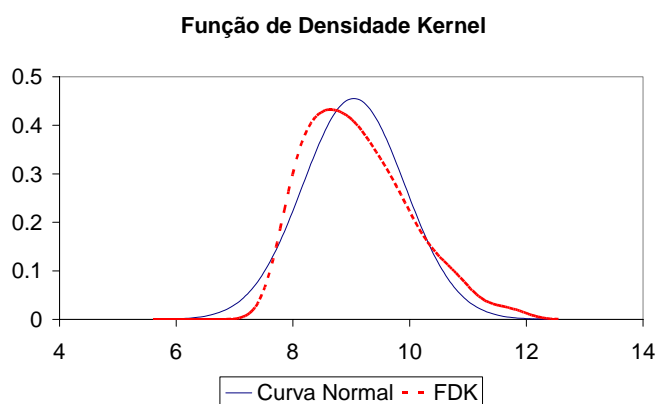


Figura 8 – Gráfico da Função de Densidade kernel para os dados de amplificadores

3.2.2 Método estatístico: o teste W de Shapiro-Wilk

Existem muitos testes estatísticos para se verificar a normalidade dos dados e basicamente podem ser divididos em três classes. Testes baseados na assimetria e curtose dos dados, na função de distribuição empírica (EDF) e regressão. Os testes baseados em assimetria e curtose comparam estatisticamente o valor destes parâmetros

da distribuição a ser testada, com o valor da distribuição normal, que deve ser de 0 para assimetria e 3 para curtose. Os testes baseados na EDF a comparam com a função cumulativa normal (CDF). Os testes baseados em regressão comparam estatisticamente a relação entre a distribuição a ser testada com os valores normais esperados de uma distribuição normal. Esta relação deve ser linear, como foi comentado na seção sobre o gráfico normal de probabilidade.

Em Seier (2002), foi realizado um estudo comparando os diversos métodos existentes, e o que se mostrou melhor em detectar não normalidade corretamente, na maioria das situações, foi o *Teste W*, criado por (SHAPIRO e WILK, 1965). O estudo realizou uma simulação de Monte Carlo com 50.000 amostras, de tamanho pequeno (20), médio (50) e grande (100), para diversos casos de não normalidade. Como este teste apresentou o melhor resultado e é o mais utilizado na área de qualidade foi escolhido para ser utilizado neste trabalho. Porém, o teste original só admitia amostras de tamanho no máximo igual a 50, por isto, Royston (1982) fez uma modificação no teste e ampliou o tamanho para 2000. É esta versão modificada que será apresentada a seguir.

O teste W é baseado na correlação entre a variável de estudo e o valor esperado normal, que depende do tamanho da amostra. A variável $y \sim N(\mu, \sigma^2)$ pode ser expressa como $y = \mu + \sigma x$ onde $x \sim N(0,1)$. Esta é a forma que se utiliza no gráfico normal de probabilidade, e quer se testar a linearidade entre y e x . Quanto maior esta correlação, maior é a probabilidade de não se rejeitar a normalidade. As duas hipóteses neste caso são:

H_0 = os dados seguem a distribuição normal

H_1 = os dados não seguem a distribuição normal

O teste é expresso pela equação (3.9), onde $y_{(i)}$ é uma amostra aleatória ordenada em que se quer fazer o teste, $y_{(1)} < y_{(2)} < \dots < y_{(n)}$.

$$W = \frac{\left[\sum_{i=1}^n a_i y_{(i)} \right]^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (3.9)$$

Os coeficientes a_i são os “melhores coeficientes lineares, não viesados”, normalizados tabulados, que dependem do tamanho da amostra. Os procedimentos para o cálculo dos coeficientes são descritos abaixo. Primeiro calcular os valores de a_i^* .

$$a_i^* = \begin{cases} 2m_i, & i = 2, 3, \dots, n-1 \\ \left(\frac{\hat{a}_1^2}{1 - 2\hat{a}_n^2} \sum_{i=2}^{n-1} a_i^{*2} \right)^{1/2} & i = 1, i = n, \end{cases} \quad (3.10)$$

onde m_i é o mesmo da equação (3.2) e

$$\hat{a}_1^2 = \hat{a}_n^2 \begin{cases} g(n-1), & n \leq 20 \\ g(n), & n > 20 \end{cases} \quad (3.11)$$

e

$$g(n) = \frac{\Gamma\left(\frac{1}{2}[n+1]\right)}{\sqrt{2}\Gamma\left(\frac{1}{2}n+1\right)} \quad (3.12)$$

O valor do coeficiente a_i é calculado aplicando a equação (3.13).

$$a_i = \frac{a_i^*}{\left(\sum_{i=1}^n a_i^{*2} \right)^{1/2}} \quad (3.13)$$

Os cálculos dos coeficientes para $n = 120$ podem ser vistos no [Apêndice A](#).

Após o cálculo dos coeficientes é possível encontrar o valor de W . Para calcular o nível de significância do teste é necessário utilizar uma transformação para normalização de W (ROYSTON, 1982).

$$y = (1 - W)^\lambda \quad (3.14)$$

$$z = \frac{(y - \mu_y)}{\sigma_y} \quad (3.15)$$

Onde z é a variável padrão da distribuição normal, μ_y e σ_y são respectivamente a média e o desvio padrão de y . Para se estimar os valores de λ , μ_y e σ_y é necessário utilizar os coeficientes do polinômio da [Tabela 8](#) que têm a forma $\lambda = 0,48085 + 0,318828x - 0,0241665x^3 + 0,00879701x^4 + 0,002989646x^5$ para o caso de λ e o mesmo formato para $\ln(\mu_y)$ e $\ln(\sigma_y)$, onde o valor de x é calculado por $x = \ln(n) - d$. Se $n \leq 20$, $d = 3$ e se $21 \leq n \leq 2000$, $d = 5$.

Tabela 8 – Coeficientes para λ , μ_y e σ_y para o teste W .

Parâmetro	Valores de n	Coeficientes						
		0	1	2	3	4	5	6
λ	7-20	0,118898	0,133414	0,327907				
	21-2000	0,480385	0,318828		-0,0241665	0,00879701	0,002989646	
$\ln(\mu_y)$	7-20	-0,37542	-0,492145	-1,124332	-0,199422			
	21-2000	-1,91487	-1,37888	-0,04183209	0,1066339	-0,03513666	-0,01504614	
$\ln(\sigma_y)$	7-20	-3,15805	0,729399	3,01855	1,558776			
	21-2000	-3,73538	-1,015807	-0,331885	0,1773538	-0,01638782	-0,03215018	0,003852646

Fonte: Royston (1982)

Como exemplo, foram utilizados os dados da [Tabela 6](#). O valor de W encontrado foi de 0,9379 (demonstrado no Apêndice A). O tamanho da amostra é $n = 120$ e como $n > 20$ utilizar $x = \ln(n) - d = \ln(120) - 5 = -0.2125$. Com os coeficientes da [Tabela 8](#) os valores de encontrados para λ , μ_y e σ_y foram:

$$\lambda = 0,480385 + (0,318828)(-0,2125) + \dots + (0,002989646)(-0,2125)^5 = 0,4129$$

$$\ln(\mu_y) = -1,91487 + (-1,37888)(-0,2125) + \dots + (-0,01504614)(-0,2125)^5 = -1,6248$$

$$\mu_y = 0,1969$$

$$\ln(\sigma_y) = -3,73538 + (-1,015807)(-0,2125) + \dots + (0,003852646)(-0,2125)^6 = -3,5362$$

$$\sigma_y = 0,0291$$

Finalmente, é possível calcular o valor de p do teste.

$$y = (1 - W)^\lambda = (1 - 0,9379)^{0,4129} = 0,3175$$

$$z = \frac{(y - \mu_y)}{\sigma_y} = \frac{(0,3175 - 0,1969)}{0,0291} = 4,1414$$

$$p = 1 - \Phi(4,1414) = 1 - 0,999982742 = 1,7257E-05$$

Com este valor de p rejeita-se a hipótese nula de normalidade. Mas se os dados não seguirem a distribuição normal qual procedimento deve ser adotado então? Este é o assunto que será visto na próxima seção.

3.3 Transformação de Johnson.

Quando os dados não seguem a distribuição normal podem-se adotar dois procedimentos, o primeiro é descobrir a qual distribuição de probabilidades os dados pertencem, mediante o teste de aderência por exemplo, e o segundo, que será utilizado neste trabalho, é a transformação dos dados para a normalidade. No primeiro caso a desvantagem é que se devem testar várias possibilidades, até que se descubra a distribuição adequada (se existir), e a vantagem é que quando se sabe qual é a distribuição, é possível fazer todas as inferências estatísticas.

O outro procedimento é a transformação dos dados para a normalidade, que pode ser feita utilizando-se funções simples como a raiz quadrada ou o logaritmo, por exemplo. Existe um grande número de transformações possíveis, mas na prática não se sabe qual utilizar, tendo que testar várias até que se transformem os dados para a normalidade. Além disto, muitas vezes estas transformações não conseguem transformar os dados para a distribuição normal. Por isto será apresentado um método mais geral, que abrange um número maior de distribuições, chamado de transformação de Johnson (Johnson, 1949). O método consiste em transformar uma variável, X , para a *distribuição normal padronizada*. A transformação inclui três famílias de distribuição: S_b , S_l e S_u . A primeira pertence à família limitada (o b significa “*bounded*”, que em português é limitado), a segunda a lognormal e a terceira a ilimitada (o u significa “*unbounded*”, que em português é ilimitado). Na Tabela 9 são apresentadas as três equações de transformação para a normal padronizada.

Tabela 9 – Transformações associadas com as famílias de curvas de Johnson

Família de Johnson	Transformação	Condições de parâmetro	Condição de X
S _B	$Z = \gamma + \eta \ln\left(\frac{X - \varepsilon}{\lambda + \varepsilon - X}\right)$	$\eta, \lambda > 0, -\infty < \gamma < \infty,$ $-\infty < \varepsilon < \infty$	$\varepsilon < X < \varepsilon + \lambda$
S _L	$Z = \gamma + \eta \ln(X - \varepsilon)$	$\eta > 0, -\infty < \gamma < \infty,$ $-\infty < \varepsilon < \infty$	$X > \varepsilon$
S _U	$Z = \gamma + \eta \sinh^{-1}\left(\frac{X - \varepsilon}{\lambda}\right)$	Mesmo de S _B	$-\infty < X < \infty$

Fonte: Chou *et al.* (1998)

A dificuldade desta transformação é escolher qual família utilizar e estimar os parâmetros γ , η , ε e λ . Existem dois métodos para se fazer isto, o primeiro é o dos momentos e o segundo é o dos percentis. Neste trabalho só será abordado o segundo método. Slifker e Shapiro (1980) desenvolveram um método para a seleção da família e estimação dos parâmetros. Os passos para estimação estão descritos abaixo:

1º - Escolher quatro percentis simétricos da distribuição normal, $q_1 = \Phi(-sz)$, $q_2 = \Phi(-z)$, $q_3 = \Phi(z)$ e $q_4 = \Phi(sz)$, onde o valor s utilizado aqui é 3. Chou *et al.* (1998) realizou um estudo e verificou que faixa do valor de z varia de 0,25(0,01)1,25 para um valor de $p \geq 0,4$ no teste W . Portanto são 101 valores de z para serem testados e a escolha pode ser feita de modo que maximize o valor de p do teste W da variável transformada. Pode-se variar valores menores que 0,01, como por exemplo 0,001, mas serão testados agora 1001 valores. Na falta de um algoritmo para testar todos estes valores, Slifker e Shapiro (1980) sugerem utilizar $z = 0,524$. No exemplo abaixo (para os dados da Tabela 6 (p. 52)), o valor utilizado foi $z = 0,6839$ (este é o valor que maximiza o valor p do teste W). Obteve-se então:

$$q_1 = \Phi(-3 \cdot 0,6839) = 0,0201$$

$$q_2 = \Phi(-0,6839) = 0,2470$$

$$q_3 = \Phi(0,6839) = 0,7530$$

$$q_4 = \Phi(3 \cdot 0,6839) = 0,9799$$

2º - De acordo com os valores encontrados acima, encontrar os valores de j (posição do dado na amostra que representa aquele percentil). Isto pode ser feito utilizando a [equação \(3.16\)](#).

$$j_i = nq_i + 0,5 \quad (3.16)$$

Para os dados da [Tabela 6](#) foi obtido:

$$j_1 = 120 \cdot 0,0201 + 0,5 = 2,9124$$

$$j_2 = 120 \cdot 0,2470 = 30,1434$$

$$j_3 = 120 \cdot 0,7530 = 90,8566$$

$$j_4 = 120 \cdot 0,9799 = 118,0876$$

3º - Achar o valor correspondente de x_i da distribuição X original ([Tabela 6](#)).

Como nenhum valor encontrado de j_i foi inteiro deve-se interpolar para encontrar os valores de x_i .

$$\begin{aligned}x_1 &= 7,8 \\x_2 &= 8,31 \\x_3 &= 9,6 \\x_4 &= 11,13\end{aligned}$$

4° - Quando $s = 3$ e utilizando a razão de quartil (QR) definida na equação (3.17) pode-se identificar à qual família de Johnson os dados pertencem seguindo os critérios abaixo (SLIFKER e SHAPIRO, 1980).

Se X for uma distribuição S_b , então $QR < 1$;

Se X for uma distribuição S_l , então $QR = 1$;

Se X for uma distribuição S_u , então $QR > 1$.

$$QR = \frac{(x_4 - x_3)(x_2 - x_1)}{(x_3 - x_2)^2} \quad (3.17)$$

Para os dados calculados no 3° passo foi obtido $QR = 0,4777$. Com este resultado, verificou-se que a distribuição X é da família S_b .

5° - Depois de definido a que família os dados pertencem utilizar as equações abaixo para estimar o valor dos parâmetros desconhecidos. Estas equações foram encontradas por (Slifker e Shapiro, 1980). Seja $\hat{x}_L = \hat{x}_2 - \hat{x}_1$, $\hat{x}_M = \hat{x}_3 - \hat{x}_2$ e $\hat{x}_U = \hat{x}_4 - \hat{x}_3$ e para um dado valor de z escolhido é possível estimar $\hat{\gamma}$, $\hat{\eta}$, $\hat{\varepsilon}$ e $\hat{\lambda}$.

Para a distribuição S_b ,

$$\hat{\eta} = \frac{z}{\cosh^{-1} \left(\frac{1}{2} \left[(1 + \hat{x}_M / \hat{x}_U) (1 + \hat{x}_M / \hat{x}_L) \right]^{1/2} \right)} \quad (3.18)$$

$$\hat{\gamma} = \hat{\eta} \sinh \left[\frac{\left(\frac{\hat{x}_M}{\hat{x}_L} - \frac{\hat{x}_M}{\hat{x}_U} \right) \left\{ \left(1 + \frac{\hat{x}_M}{\hat{x}_U} \right) \left(1 + \frac{\hat{x}_M}{\hat{x}_L} \right) - 4 \right\}^{1/2}}{2 \left(\frac{\hat{x}_M^2}{\hat{x}_L \hat{x}_U - 1} \right)} \right] \quad (3.19)$$

$$\hat{\lambda} = \frac{\hat{x}_M \left[\left\{ \left(1 + \frac{\hat{x}_M}{\hat{x}_U} \right) \left(1 + \frac{\hat{x}_M}{\hat{x}_L} \right) - 2 \right\}^2 - 4 \right]^{1/2}}{\left(\frac{\hat{x}_M^2}{\hat{x}_L \hat{x}_U - 1} \right)} \quad (3.20)$$

$$\hat{\varepsilon} = \frac{1}{2} \left(\hat{x}_2 + \hat{x}_3 - \hat{\lambda} + \frac{\hat{x}_M (\hat{x}_M / \hat{x}_L - \hat{x}_M / \hat{x}_U)}{(\hat{x}_M^2 / (\hat{x}_L \hat{x}_U) - 1)} \right) \quad (3.21)$$

Para distribuição S_l ,

$$\hat{\eta} = \frac{2z}{\ln(\hat{x}_U / \hat{x}_M)} \quad (3.22)$$

$$\hat{\gamma} = \hat{\eta} \ln \left[\frac{\hat{x}_U / \hat{x}_M - 1}{(\hat{x}_U \hat{x}_M)^{1/2}} \right] \quad (3.23)$$

$$\hat{\varepsilon} = \frac{1}{2} \left[\hat{x}_2 + \hat{x}_3 - \hat{x}_M \left(\frac{\hat{x}_U / \hat{x}_M + 1}{\hat{x}_U / \hat{x}_M - 1} \right) \right] \quad (3.24)$$

Para a distribuição S_u ,

$$\hat{\eta} = \frac{2z}{\cosh^{-1} \left[\frac{1}{2} \left(\hat{x}_U / \hat{x}_M + \hat{x}_L / \hat{x}_M \right) \right]} \quad (3.25)$$

$$\hat{\gamma} = \hat{\eta} \sinh^{-1} \left[\frac{\left(\hat{x}_L / \hat{x}_M - \hat{x}_U / \hat{x}_M \right)}{2 \left(\hat{x}_U \hat{x}_L / \hat{x}_M^2 - 1 \right)^{1/2}} \right] \quad (3.26)$$

$$\hat{\lambda} = \frac{2 \hat{x}_M \left(\hat{x}_U \hat{x}_L / \hat{x}_M^2 - 1 \right)^{1/2}}{\left(\frac{\hat{x}_U}{\hat{x}_M} + \frac{\hat{x}_L}{\hat{x}_M} - 2 \right) \left(\frac{\hat{x}_U}{\hat{x}_M} + \frac{\hat{x}_L}{\hat{x}_M} + 2 \right)^{1/2}} \quad (3.27)$$

$$\hat{\varepsilon} = \frac{1}{2} \left(\hat{x}_2 + \hat{x}_3 + \frac{\hat{x}_M \left(\hat{x}_L / \hat{x}_M - \hat{x}_U / \hat{x}_M \right)}{\left(\hat{x}_U / \hat{x}_M + \hat{x}_L / \hat{x}_M - 2 \right)} \right) \quad (3.28)$$

Para os dados calculados do exemplo acima foi obtido:

$\hat{\eta} = 0,9543$, $\hat{\gamma} = 0,9602$, $\hat{\lambda} = 4,6482$ e $\hat{\varepsilon} = 7,6101$ e a função de transformação

$$Z = 0,9602 + 0,9543 \ln \left(\frac{X - 7,6101}{4,6482 + 7,6101 - X} \right).$$

Aplicando esta transformação aos dados da [Tabela 6](#), obtiveram-se os resultados (mostrados na tabela do [Apêndice B](#)) para a construção dos gráficos da [Figura 9](#) (p. 67). Nesta figura observa-se que agora os dados seguem a distribuição normal, tanto o histograma, gráfico normal de normalidade e a função de densidade kernel apontam esta situação.

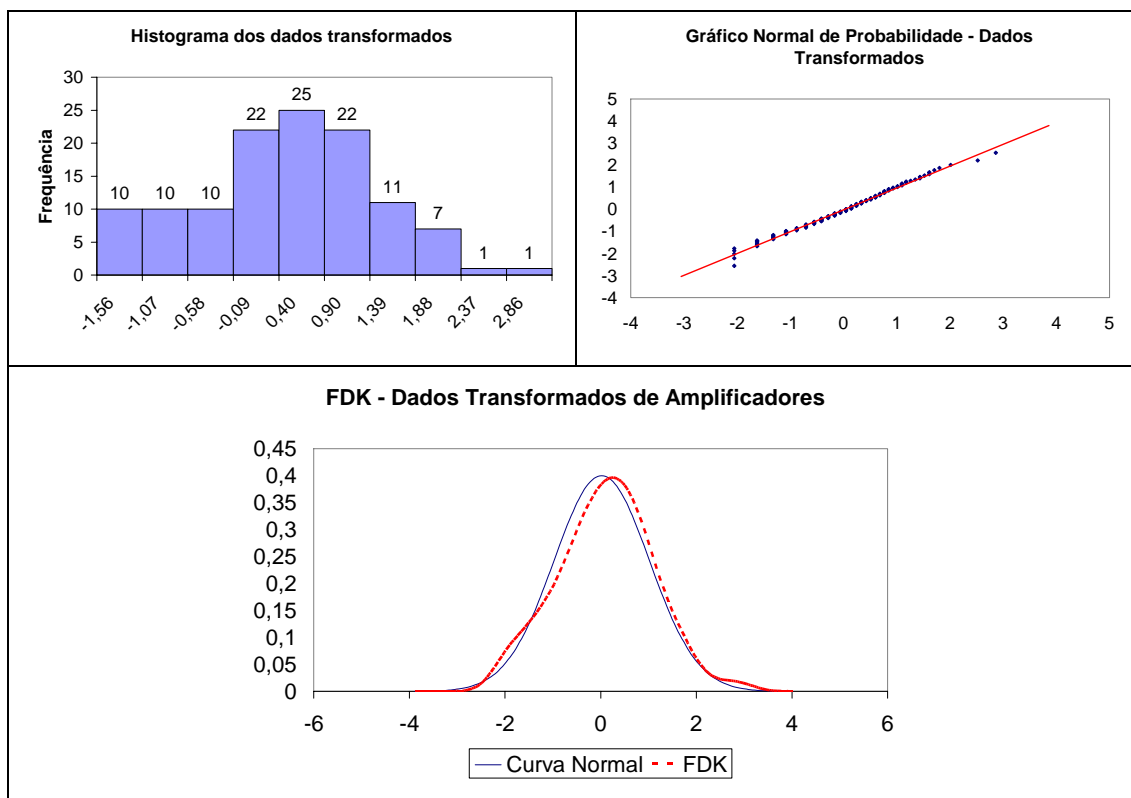


Figura 9 – Histograma, gráfico normal de probabilidade e função de densidade kernel para os dados transformados do ganho de amplificadores.

O valor do teste W foi $W = 0,9789$ e $p = 0,4120$, mostrando que os dados transformados seguem a distribuição normal. Após a transformação dos dados é possível utilizar os índices mostrados nas seções anteriores e isto é o que será visto na próxima seção.

3.4 Utilização dos índices de capacidade para dados não normais

O método que será descrito aqui foi desenvolvido por [Clements \(1989\)](#) e sua aplicação pode ser encontrada em diversos artigos [Spedding e Rawlings \(1994\)](#), [Pearn e Chen \(1997\)](#), [Pan e WU \(1997\)](#), [Tong e Chen \(1998\)](#), [Pearn et. al \(1999\)](#), [Krishnamoorthi e Khatwani \(2000\)](#), [Kotz et. al \(2002\)](#) e no capítulo 5 do livro de [Kotz e Lovelace \(1998\)](#). Considerando os dados de uma distribuição normal com média μ e desvio padrão σ , podem-se fazer as seguintes substituições:

$$x_{0,00135} = \mu - 3\sigma$$

$$x_{0,50} = \mu$$

$$x_{0,99865} = \mu + 3\sigma$$

$$x_{0,99865} - x_{0,00135} = 6\sigma$$

Quando a distribuição não segue a distribuição normal, devem-se encontrar os quatro valores dos percentis de x demonstrados acima e substituir nos índices de capacidade. Desta maneira então, os índices C_p e C_{pk} são definidos da seguinte maneira:

$$C_{p(q)} = \frac{LSE - LIE}{x_{0,99865} - x_{0,00135}} \quad (3.29)$$

$$C_{pk(q)} = \min \left(\frac{LSE - x_{0,50}}{x_{0,99865} - x_{0,50}}, \frac{x_{0,50} - LIE}{x_{0,50} - x_{0,00135}} \right) \quad (3.30)$$

A forma como se encontra os quatro valores de x pode variar, ou por transformação de dados para a normalidade ou encontrando um modelo de distribuição de probabilidades apropriado para os dados em estudo. Por exemplo, [Clements \(1989\)](#) utilizou as curvas de Pearson para encontrar os quatro percentis, método baseado nos valores de assimetria e curtose. Neste trabalho será utilizada a transformação de Johnson mostrada na seção anterior. Como exemplo, para os dados da [Tabela 6 \(p. 52\)](#) o $LSE = 12,2 \text{ dB}$ e $LIE = 7,75 \text{ dB}$. Para encontrar os quatro percentis deve-se isolar X em função de Z nas equações da [Tabela 9 \(p. 62\)](#) e substituir Z por 0 ($x_{0,50}$), 3 ($x_{0,99865}$) e -3 ($x_{0,00135}$). As equações são:

$$X = \frac{e^{\frac{Z-\gamma}{\eta}} (\lambda + \varepsilon) + \varepsilon}{\left(1 + e^{\frac{Z-\gamma}{\eta}}\right)} \quad \text{família } S_b \quad (3.31)$$

$$X = \sinh\left(\frac{Z-\gamma}{\eta}\right) \lambda + \varepsilon \quad \text{família } S_u \quad (3.32)$$

$$X = e^{\frac{Z-\gamma}{\eta}} + \varepsilon \quad \text{família } S_l \quad (3.33)$$

Os valores encontrados para o exemplo foram:

$$x_{0,50} = 8,85, \quad x_{0,99865} = 11,77 \quad \text{e} \quad x_{0,00135} = 7,68.$$

Então, substituindo estes valores nas equações (3.29) e (3.30), obtive-se:

$$C_{p(q)} = \frac{12,20 - 7,75}{11,77 - 7,98} = 1,089$$

$$C_{pk(q)} = \min\left(\frac{12,20 - 8,85}{11,77 - 8,85}, \frac{8,85 - 7,75}{8,85 - 7,68}\right) = 0,9422$$

Se a normalidade não tivesse sido testada e, portanto, não se fazendo a transformação dos dados, os índices calculados seriam:

$$C_p = \frac{12,2 - 7,75}{6 \cdot 0,8760} = 0,8466$$

$$C_{pk} = \min\left(\frac{12,2 - 9,0441}{3 \cdot 0,8760}, \frac{9,0441 - 7,75}{3 \cdot 0,8760}\right) = 0,4924$$

Quando se transformou os dados e foram aplicados os índices, o número de defeituosos estimado foi de 2456 *ppm*. Já quando não se considerou a possibilidade de não normalidade o número de defeituosos estimado foi de 69969 *ppm*. Portanto, corre-se o risco, neste caso, de considerar o processo pior do que ele é na realidade, quando não considerada a hipótese de não normalidade.

Alguns problemas podem ser encontrados quando os dados não são normais e se utiliza a transformação de Johnson. São apontados por [Kotz e Lovelace \(1998, p. 145\)](#):

- o método dos momentos e método dos percentis não permite estimar a variabilidade e viés nos índices estimados, não sendo possível calcular intervalos de confiança.
- Nem todas as distribuições de dados podem ser descritas pelas famílias de curvas de Johnson.

Ainda não existem soluções encontradas até o momento para o primeiro problema apontado. Em [Tong e Chen \(1998\)](#) é proposto um método para o cálculo do limite inferior de confiança, mas exige uma amostra muito grande (>740) não sendo prática sua utilização. Para o segundo problema, podem-se utilizar os diversos índices que foram criados para lidar com a não normalidade, um resumo pode ser encontrado em [Kotz e Lovelace \(1998\)](#), capítulo 4. Mais recentemente foi desenvolvido um índice que também trabalha com distribuições não normais e também calcula o número de não conformes produzidos ([Chen e Ding, 2001](#)).

3.5 Estimação dos índices de capacidade na presença de dados autocorrelacionados

Assim como a não-normalidade, outro problema comum que ocorre nos dados é a autocorrelação. As causas para esta ocorrência geralmente estão associadas a

processos de amostragem (em que o intervalo entre as amostras é pequeno), desgaste de ferramentas e variáveis cíclicas como temperatura, por exemplo. Os principais problemas encontrados na estimação dos índices são apontados em [Jagadeesh e Subash \(1994\)](#) e [Shore \(1997\)](#) e, basicamente, estão relacionados à estimação do desvio padrão. Quando os dados são autocorrelacionados positivamente existe um viés na estimação do desvio padrão, ele é subestimado. O valor esperado da variância na presença de autocorrelação é:

$$E(S^2) = \left(\frac{1}{(n-1)n^2} \right) \left(\sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n a_{ij} a_{ik} \text{Cov}(x_j, x_k) \right) \quad (3.34)$$

onde $\text{Cov}(x_i, x_j)$ é a covariância das observações entre x_j e x_k e

$$a_{ij} = \begin{cases} n-1, & j=1 \\ -1, & j \neq 1 \end{cases} \quad (3.35).$$

Em gráficos de controle isto causa um número maior de alarmes falsos (os limites do gráfico irão ficar mais “apertados”). Nos índices de capacidade ocorrerá uma superestimação, o processo parecerá ser melhor do que realmente é. Na [seção 3.5.1](#) será apresentado como medir e um teste estatístico para a autocorrelação e na [seção 3.5.2](#) uma solução para este problema.

3.5.1 Teste estatístico para a autocorrelação

A função de autocorrelação pode ser medida segundo a [equação \(3.36\)](#), que é correlação de x_t com x_{t-k} (sendo k o número de períodos passados).

$$r_k = \frac{\sum_{t=k+1}^n (x_t - \bar{x})(x_{t-k} - \bar{x})}{\sum_{t=1}^n (x_t - \bar{x})} \quad (3.36)$$

Um teste estatístico para verificar se o coeficiente de autocorrelação estimado é significativamente diferente de zero é o de [Ljung e Box \(1978\)](#).

$$Q^* = n(n+2) \sum_{k=1}^h (n-k)^{-1} r_k^2 \quad (3.37)$$

Para este teste a hipótese nula e alternativa é:

H_0 : os dados não são autocorrelacionados

H_1 : os dados são autocorrelacionados

O valor de Q^* é distribuído segundo a distribuição χ^2 com $(h-m)$ graus de liberdade, onde h é o valor máximo de atrasos escolhido e m é o número de parâmetros estimados de um modelo de série temporal. Nos gráficos de controle de Shewhart é considerado que o processo tenha o seguinte modelo de série temporal:

$$x_t = \mu + e_t \quad t = 1, 2, \dots \quad (3.38)$$

O valor da média é estimado com amostras retiradas do processo e o resíduo e_t deve ser $e_t \sim N(0, \sigma_e^2)$. Portanto, para o modelo de Shewhart, o valor de m é 1.

Depois de realizar o teste descrito acima e se verificar a existência de autocorrelação é necessário à utilização de um modelo autoregressivo de série temporal para eliminar a autocorrelação nos resíduos, este tópico será visto na próxima seção. Antes será apresentado um exemplo numérico para a aplicação dos conceitos vistos

acima. Este exemplo será utilizado ao longo desta seção e foi realizado com os dados do diâmetro de um furo em uma peça. Os dados foram coletados em 10 subgrupos de 5 unidades cada e são apresentados na [Tabela 10](#).

Tabela 10 – Diâmetro de eixos de uma peça

Subgrupo	Diâmetro do furo em milímetros				
1	50	51	50.5	49	50
2	43	42	45	47	49
3	46	50	52	52.5	51
4	52	50	49	54	51
5	52	46	42	43	45
6	46	42	44	43	46
7	42	43	42	45	49
8	50	51	52	54	51
9	49	50	49.5	51	50
10	52	50	48	49.5	49

Fonte: Shore (1997)

Os valores individuais das observações são apresentados na Figura 10.

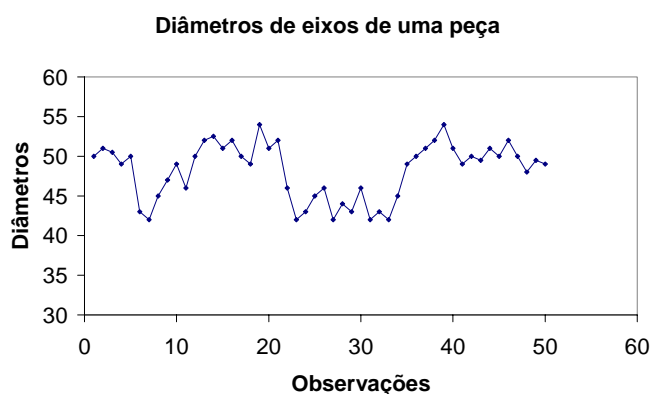


Figura 10 – Gráfico das observações individuais do diâmetro de eixos de uma peça

Calculando a função de autocorrelação ([equação \(3.36\)](#)) para $k = 10$ e aplicando o teste de Q^* ([equação \(3.37\)](#)) foram obtidos os resultados apresentados na [Tabela 11](#).

Tabela 11 – Função autocorrelação e teste de Ljung-Box para os dados de diâmetros do furo de uma peça

k (atraso)	r_k	Q^*	p
1	0.741	29.113	0,000
2	0.540	44.898	0,000
3	0.337	51.199	0,000
4	0.236	54.337	0,000
5	0.129	55.299	0,000
6	-0.021	55.325	0,000
7	-0.086	55.771	0,000
8	-0.206	58.408	0,000
9	-0.207	61.124	0,000
10	-0.308	67.271	0,000

Todos os valores na tabela apresentaram autocorrelação, segundo o teste de Ljung-Box.

3.5.2 Solução para o problema de autocorrelação

A solução para o problema de autocorrelação é utilizar modelos autoregressivos, demonstrados na equação (3.39), onde ϕ pode ser estimado pela minimização da soma dos erros quadráticos. O modelo apresentado é de ordem k , pois a variável independente está k períodos atrasada em relação a variável dependente.

$$x_t = \xi + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots + \phi_k x_{t-k} + e_t \quad (3.39)$$

Para os dados da Tabela 10, aplicando o modelo da equação (3.39) obteve-se a equação (3.40). Este modelo é autoregressivo de ordem 2, com os resultados da regressão apresentados na Tabela 12 e a função de autocorrelação na Tabela 13.

$$x_t = 12,437 + 0,741x_{t-1} \quad (3.40)$$

Tabela 12 – Resultados do modelo autoregressivo para os dados de diâmetros do furo de uma peça.

Variável	Coefficiente	Erro Padrão	Estatística t	p
ξ	12.43701	4.705847	2.642885	0.0111
x_{t-1}	0.741460	0.097403	7.612292	0.0000
R^2	0.552155			
$R^2_{ajustado}$	0.542626			
S.E regressão	2.410299			
Soma dos quadrados dos resíduos	273.0485			

Tabela 13 - Função de autocorrelação e teste de Ljung-Box para os resíduos do modelo autoregressivo dos dados de diâmetros do furo de uma peça

k (atraso)	r_k	Q^*	p
1	0.004	0.0007	0.979
2	0.085	0.3825	0.826
3	-0.114	1.0905	0.779
4	0.033	1.1506	0.886
5	0.140	2.2650	0.811
6	-0.129	3.2331	0.779
7	0.074	3.5606	0.829
8	-0.238	7.0233	0.534
9	0.127	8.0252	0.532
10	-0.155	9.5733	0.479

Observa-se que agora os resíduos do modelo não apresentam autocorrelação (Tabela 13).

Depois de aplicado o modelo, e se os resíduos apresentarem a forma $e_t \sim N(0, \sigma_e^2)$, utilizá-los nos gráficos de controle para monitorar a estabilidade do processo.

No caso da estimação da capacidade de processo, deve-se estimar a média e o desvio padrão de x_t do modelo da equação (3.39) (SHORE, 1997). Para esta situação, a estimação do desvio padrão é:

$$\sigma = \frac{\sigma_e}{(1 - r_1\phi_1 - r_2\phi_2 - \dots - r_k\phi_k)^{1/2}} \quad (3.41)$$

Calculando σ para os dados da [Tabela 10](#) utilizando a [equação \(3.40\)](#) encontrada a partir do modelo da [equação \(3.39\)](#) e dos dados de autocorrelação da [Tabela 11](#) obteve-se:

$$\hat{\sigma} = \frac{2.385060}{(1-0,74 \cdot 0,74)^{1/2}} = 12,7$$

Agora, com este valor estimado, podem-se utilizar os índices de capacidade C_p e C_{pk} de forma correta. Se o desvio padrão dos dados da [Tabela 10](#) fosse estimado por S o valor seria $S = 3,54$, um valor bem abaixo do encontrado acima. Os valores dos índices estimados desta maneira são menores que os estimados sem que se considere a autocorrelação, mas é um retrato mais fiel do percentual de defeituosos produzidos.

3.6 Considerações finais

Neste capítulo foi apresentado como se calcular os índices de capacidade de processo para dados não normais e autocorrelacionados. No caso de não normalidade ainda existem muitas dificuldades teóricas, principalmente quando se transformam os dados para a normalidade. Em gráficos de controle pode-se trabalhar com os dados transformados sem nenhuma restrição enquanto os dados pertencerem àquela população que gerou a função de transformação. Se por algum motivo os dados coletados pertencerem à outra população a função de transformação utilizada não é mais válida.

Quando aplicada a transformação no cálculo de índices de capacidade de processo não é possível trabalhar com os dados transformados, é necessário que eles estejam na sua escala original. Esta necessidade se deve ao fato de não se poder transformar os limites de especificação na função encontrada, pois não existe nenhuma relação entre os limites e a população de dados que gerou a transformação. Se fosse

utilizado os dados transformados poder-se-ia calcular os intervalos de confiança para a média e desvio padrão, assim como é feito para os gráficos de controle, e aplicar nos índices estimados.

No caso de dados autocorrelacionados a dificuldade é a de se estimar o desvio padrão corretamente, pois precisa-se estimar primeiro os parâmetros de um modelo autoregressivo, calcular o desvio padrão dos resíduos e depois aplicar a equação (3.41). Mas os resultados obtidos são mais próximos da realidade do que se não tomar estas medidas.

4 MODELOS DE ANÁLISE DA CAPACIDADE DE PROCESSOS

Neste capítulo serão apresentados dois modelos de análise da capacidade de processos, um deles ([modelo 1](#)) para quando os dados seguem a distribuição normal e são independentes e outro ([modelo 2](#)) para dados não normais e/ou autocorrelacionados. Estes modelos se baseiam nas revisões teóricas apresentadas nos [capítulos 2 e 3](#). A diferença entre um e outro é que no primeiro podem-se aplicar intervalos de confiança para a estimação dos índices e conseqüentemente determinar o tamanho de amostras necessárias e fazer teste de hipótese. No segundo é necessário aplicar os conceitos do [capítulo 3](#), transformação de dados no caso de não normalidade e modelos autoregressivos no caso de autocorrelação, porém não é possível calcular intervalos de confiança para os índices.

4.1 Modelo da análise da capacidade de processos para dados normais e independentes (Modelo 1)

Quando as condições de normalidade e independência são satisfeitas pode-se aplicar o modelo 1 da [Figura 11](#). Neste modelo são quatro as variáveis que se podem modificar na análise da capacidade de processos. São elas: o tamanho das amostras, o valor do índice estimado, o intervalo de confiança desejado e o valor escolhido como capacidade alvo. Serão exploradas várias situações que a empresa pode se encontrar e como atuar sobre cada variável. As situações que a empresa pode se encontrar são:

Situação1: O índice estimado é menor que o estabelecido como alvo.

- Solução 1: realizar modificações no processo. Ajuste da média do processo ao alvo e redução da variabilidade. Geralmente os custos com a redução da variabilidade são maiores que os do ajuste da média do processo ao alvo. Portanto, primeiro se faz o

ajuste da média do processo em relação ao alvo e, depois, se tenta reduzir a variabilidade.

- Solução 2: reduzir o valor especificado como alvo. Quando não se podem realizar modificações no processo, rever se o especificado como alvo não está muito além das reais capacidades da empresa. Outra opção é rever os limites de especificação de projeto e adequá-los à realidade do processo.

Situação2: O índice estimado é maior que o estabelecido como alvo. Dentro desta situação pode ocorrer:

- Situação 2.1 – O intervalo de confiança escolhido é maior que o desejado. Propicia a possibilidade da redução do tamanho das amostras, com conseqüentes reduções de custo.
- Situação 2.2 – O intervalo de confiança escolhido é menor que o desejado.
 - Solução 1: aumentar o tamanho das amostras. Com esta medida a empresa está diminuindo o risco de concluir que o processo é capaz, quando na realidade ele é incapaz. Deve-se avaliar qual é o custo de amostragem em relação ao risco que se quer correr.
 - Solução 2: na impossibilidade de aumentar o tamanho das amostras, realizar modificações no processo para aumentar o valor estimado do índice. Se as modificações realizadas resultarem em um índice estimado maior, a empresa está diminuindo o risco de concluir que o processo é capaz, quando na realidade ele é incapaz. Deve-se avaliar qual é o custo das modificações em relação ao risco que se quer correr.
 - Solução 3: na impossibilidade de aumentar o tamanho das amostras e de se modificar o processo, reduzir o valor do intervalo de confiança escolhido.

Com esta medida a empresa está correndo um risco maior de concluir que o processo é capaz, quando na realidade ele é incapaz.

- Solução 4: na impossibilidade de aumentar o tamanho das amostras, de se modificar o processo e reduzir o valor do intervalo de confiança escolhido, rever se o especificado como capacidade alvo não está muito além das reais capacidades da empresa. Ou rever os limites de especificação de projeto.

Definidas as situações em que empresa pode se encontrar são apresentadas as etapas do modelo da Figura 11:

1º. Definir o processo a ser avaliado.

O processo a ser avaliado deve ser aquele que produza produtos com características consideradas fundamentais para a empresa.

2º. Definir se o estudo é de máquina ou processo

A diferença do estudo de máquina e processo, é que no primeiro tenta-se isolar todas as fontes de variação do processo, deixando apenas a variabilidade da máquina. Troca de ferramentas, troca de operadores, e outras variáveis que estão presentes no processo produtivo não são feitas neste tipo de estudo. Além disto, são retiradas peças consecutivas da máquina. Quando o estudo é de processo quer se medir toda a variabilidade presente, então as amostras são espaçadas ao longo de um dia de produção por exemplo, para que a troca de turno, temperatura e outras variáveis influenciam nos resultados obtidos. Como se realizar um estudo da capacidade de máquinas pode ser encontrado em (Roth II, 2003).

3º. Garantir o sistema de medição

A empresa deve ter todos os seus instrumentos de medição calibrados e a sua escala deve ser adequada à grandeza que se quer medir. Os efeitos do erro de medição na avaliação da capacidade podem ser encontrados [Kotz e Lovelace \(1998, p. 238\)](#).

4º. Escolher a capacidade alvo do processo

Esta escolha pode ser feita por contrato entre cliente-fornecedor, de referência do setor onde a empresa trabalha ou de estudos anteriores sobre capacidade. Quando a empresa está iniciando o estudo da capacidade de processos pode escolher o valor por *benchmarking* de outras do mesmo setor ou da literatura existente. Com o passar do tempo a empresa reavalia se o valor escolhido é viável ou não de ser atingido.

5º. Escolher tamanho da amostra

Para se conduzir o estudo é recomendado que as amostras sejam de pelo menos 100 unidades. Na prática isto pode ser inviável, mas deve-se em um primeiro momento escolher o maior número de amostras possíveis. Dependendo do valor estimado e do intervalo de confiança escolhido este número pode ser reduzido em estudos futuros.

6º. Escolher intervalo de confiança

O valor alvo de capacidade é largamente indicado na literatura, como sendo por exemplo 1,33. Já na escolha de intervalos de confiança não existem ainda recomendações quanto a um valor. Recomenda-se utilizar pelo menos 95%, com este valor corre-se apenas 5% de chance de concluir que o processo é capaz incorretamente.

7º. Coletar amostras

Deve ocorrer na presença de toda a variabilidade possível, mudanças de turno, ferramentas, matéria prima entre outras. Desta forma a avaliação do processo será a mais fiel possível da realidade.

8º. Verificar se os dados são normais e independentes

Nesta etapa utilizar os métodos gráficos e o teste de normalidade apresentados [seção 3.2 \(p. 51\)](#).

9º. Verificar a estabilidade do processo

Nesta etapa utilizar os gráficos de controle apresentados na [seção 2.2 \(p. 27\)](#).

10º. Estimar índice

A estimação dos índices deve ser feita de acordo com o que foi revisado na [seção 2.3](#).

11º. Identificar em qual situação a empresa se encontra.

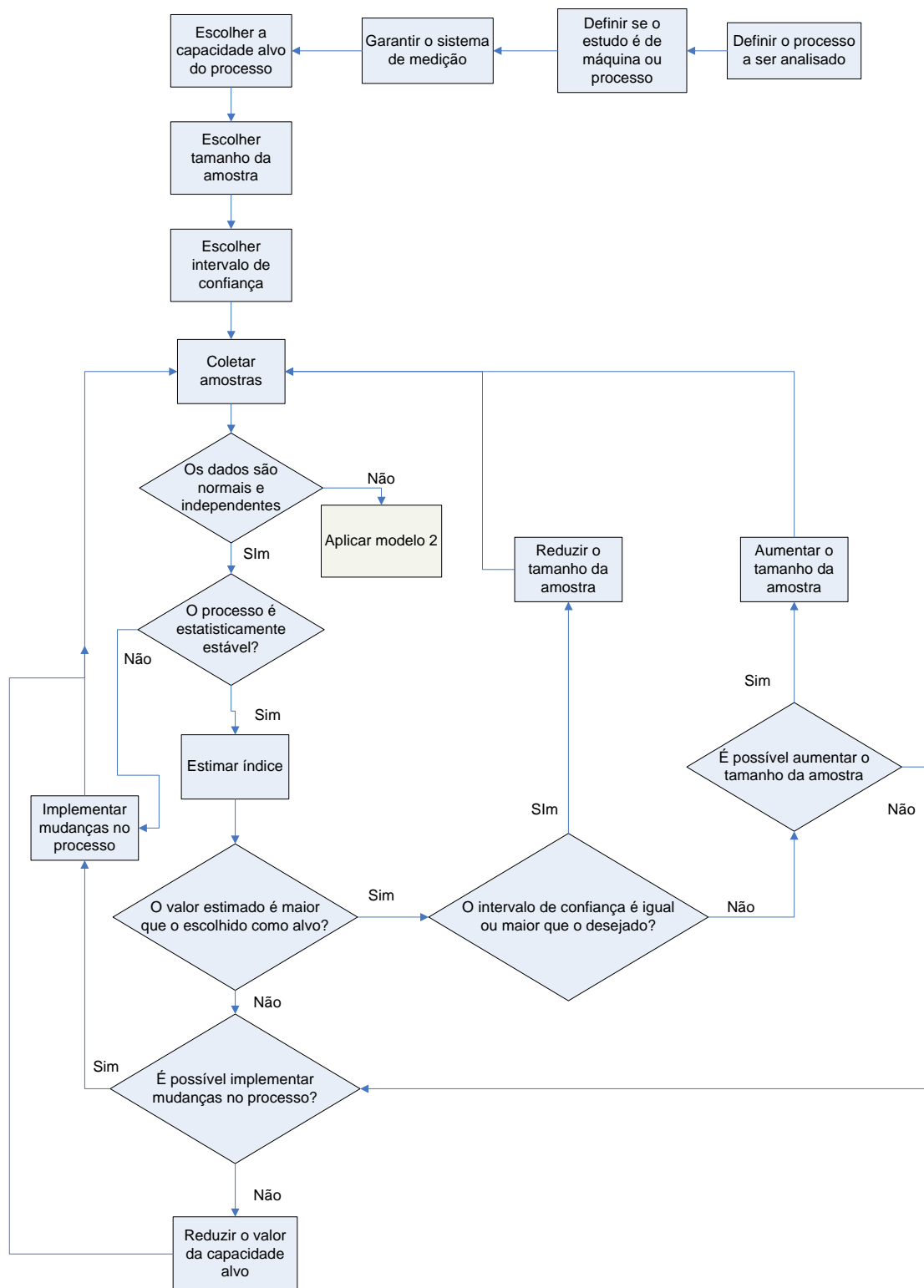


Figura 11 – Modelo de análise da capacidade de processos para dados normais e independentes (modelo 1)

Observando o modelo da [Figura 11](#), percebe-se que ele tem um início mas não tem um fim, a avaliação da capacidade é contínua. É muito mais confiável a estimação

de vários índices em vários momentos diferentes, do que apenas em um. Os dados coletados para verificar a estabilidade do processo estão disponíveis e podem ser utilizados na estimação dos índices. Enquanto o processo for estável o valor estimado entre um e outro deve ser muito próximo, dentro do intervalo de confiança, mas quando o processo não está sob controle estatístico o resultado irá ser diferente. Esta idéia pretende diminuir as chances de manipulação de dados (tentar estimar os índices até que se obtenha o valor desejado) e serve para guardar um histórico da capacidade do processo ao longo do tempo.

4.2 Modelo da análise da capacidade de processos para dados não normais e autocorrelacionados (Modelo 2)

Quando as condições de normalidade e independência não são satisfeitas deve-se aplicar o modelo 2 da [Figura 12](#). Neste modelo é necessário realizar transformações na variável de estudo, no caso de não normalidade a transformação de Johnson e para a autocorrelação aplicar o modelo autorregressivo. A restrição deste modelo é a não aplicação de intervalos de confiança para os índices estimados. Os passos para a aplicação do modelo são parecidos com o modelo anterior. As etapas de 1 a 5 são as mesmas, o que difere é da 6 em diante.

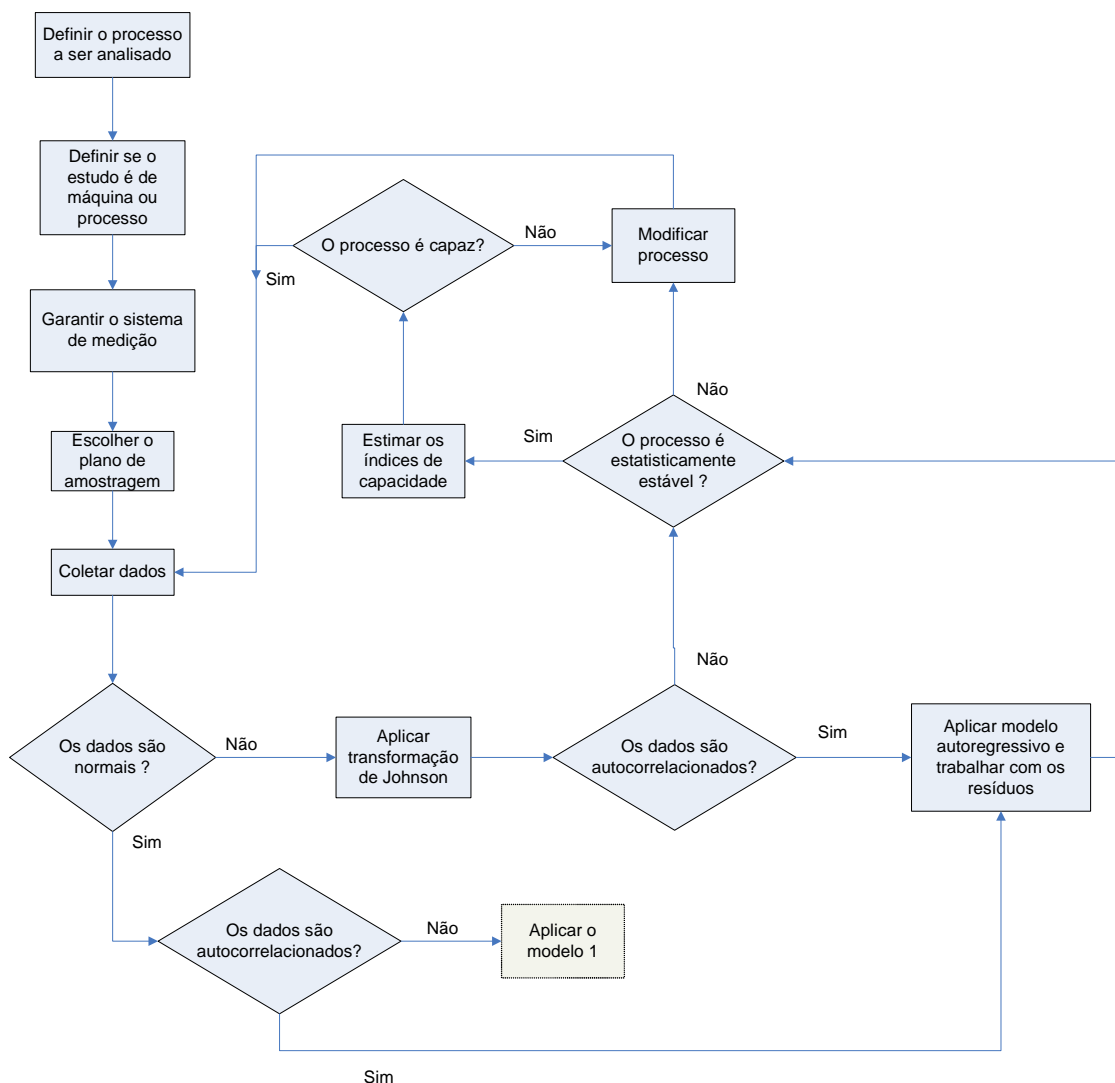


Figura 12 – Modelo da análise da capacidade de processos para dados não normais e autocorrelacionados (Modelo 2)

Estes dois modelos apresentados neste capítulo serão aplicados em uma empresa do setor automotivo de Santa Catarina, os processos analisados em questão foram o de usinagem de peças onde se mediu o diâmetro interno (processo 1) e a distância entre faces (processo 2). Foi necessária a utilização destes dois modelos para estimação da capacidade de processos.

5 APLICAÇÃO DO MODELO EM UMA FÁBRICA DO SETOR AUTOMOTIVO.

Neste capítulo serão aplicados os modelos apresentados no Capítulo 4 em uma empresa do setor automotivo de Santa Catarina. A empresa em questão é uma fornecedora de peças para montadoras de automóveis como Audi, Ford e Renault. Neste setor o fornecedor tem que garantir a capacidade do processo para os clientes com valores fixados em contrato. Por isto é de fundamental importância que capacidade avaliada represente a realidade o mais fielmente possível. Este estudo serve tanto para fornecedores que desejam melhorar o desempenho de seus processos quanto para os clientes que querem avaliar seus fornecedores.

O estudo analisou o processo de usinagem de duas linhas de produtos diferentes e por pedido da empresa a descrição destes não será apresentada aqui. As definições das etapas 1 a 7 do modelo 1 ficaram a cargo da empresa, pois não fazem parte do escopo deste trabalho definir qual o processo a ser analisado, definir se o estudo é de máquina ou processo, garantir o sistema de medição, escolher a capacidade alvo do processo, escolher o tamanho da amostra, escolher o intervalo de confiança e a coleta de amostras. Além disto, os processos precisavam ser estáveis para que a avaliação fosse realizada, portanto na época do estudo, novembro de 2004, eles apresentaram estabilidade. Isto por que não foi objetivo do trabalho estabilizar processos produtivos, mas apenas analisar capacidade de processos e propor modificações quando necessário. A seguir é apresentado o estudo em cada processo.

5.1 Processo 1 – Característica avaliada: diâmetro interno

As etapas de análise são descritas abaixo e seguem o modelo 1.

1º. Definir o processo a ser avaliado.

O processo em questão é o de usinagem de peças e a característica medida é o diâmetro interno (mm). Esta medida de diâmetro é definida pelo cliente.

2º. Definir se o estudo é de máquina ou processo.

O estudo é de processo.

3º. Garantir o sistema de medição.

Foi considerado que o sistema de medição da empresa é perfeito. Neste setor geralmente as empresas do setor possuem um bom sistema.

4º. Escolher a capacidade alvo do processo.

O valor da capacidade alvo é definido em contrato entre o fornecedor e o cliente. Os valores são $C_p = 1,33$, $C_{pk} = 1,33$, $P_p = 1,67$ e $P_{pk} = 1,67$. Como discutido na seção 2.3.3 (pág. 35) em processos estáveis esta diferença de valores não fazem o menor sentido, já que $C_p = P_p$ e $C_{pk} = P_{pk}$. Portanto, o valor adotado como alvo neste estudo foi 1,33, equivalente a 66 *ppm*.

5º. Escolher tamanho da amostra.

Quando a empresa quer avaliar a sua capacidade, retira uma amostra de cada dez produzidas, em um total de trinta amostras. O valor destas amostras é apresentado na [tabela 1C do apêndice C](#).

6°. Escolher intervalo de confiança.

A empresa atualmente não trabalha com a metodologia de encontrar intervalos de confiança para os seus índices estimados. Para este estudo se escolheu o valor inicial de confiança de 95% tanto para C_p quanto para C_{pk} .

7°. Coletar amostras

A empresa já pratica o controle de qualidade a bastante tempo e domina a técnica de coleta de amostras.

8°. Verificar se os dados são normais e independentes.

Utilizar os passos descritos no capítulo 3 das seções (3.2 e 3.5.1). Os gráficos gerados para visualizar se os dados são normais encontrados foram:

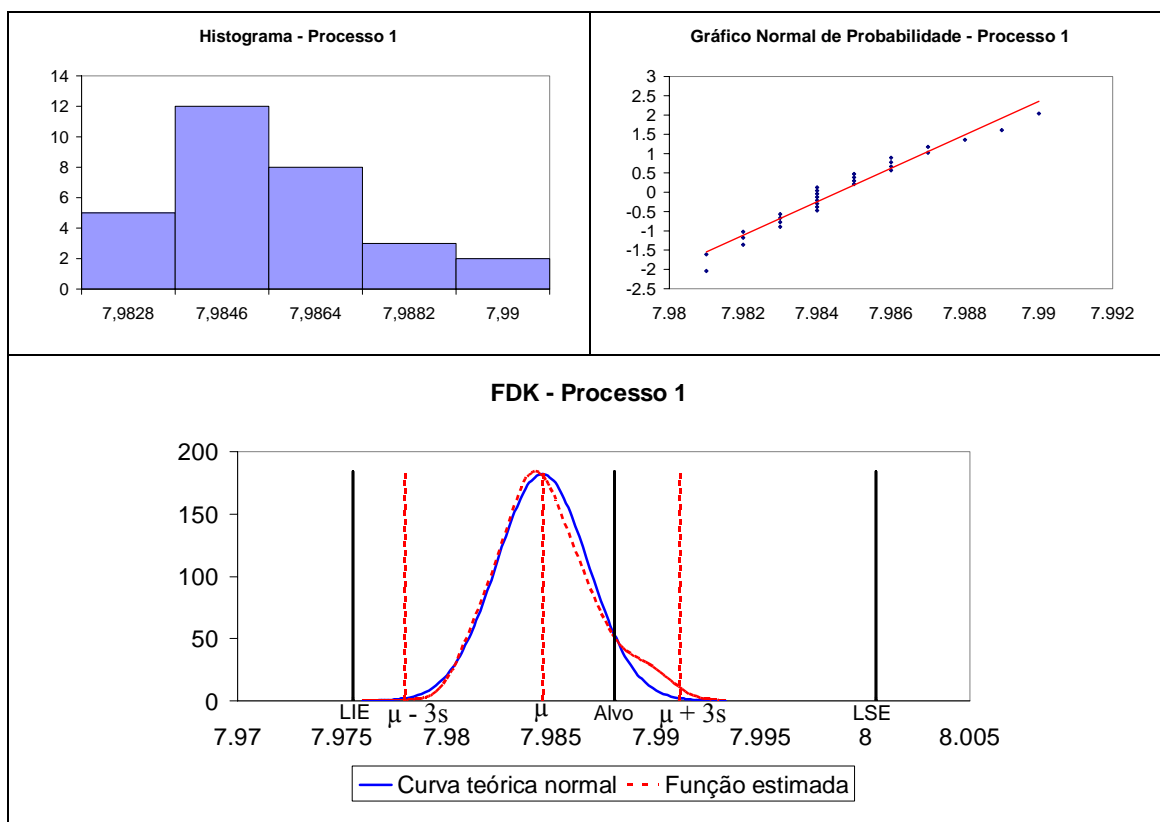


Figura 13 – Histograma, Gráfico normal de probabilidade e FDK dos dados do processo 1.

O teste W apresentou o resultado $W = 0,9526$ e $p = 0,2306$ levando a não rejeição da hipótese nula de normalidade. A [Figura 13](#) também demonstra este resultado.

O teste de autocorrelação é apresentado na [Tabela 14](#) e apresenta autocorrelação significativa nos quatro primeiros períodos.

Tabela 14 – Teste de autocorrelação para os dados do processo 1

k (atraso)	r_k	Q^*	p
1	0.489	7.9224	0.005
2	0.241	9.9102	0.007
3	-0.032	9.9472	0.019
4	-0.167	10.975	0.027
5	0.038	11.03	0.051
6	0.046	11.115	0.085
7	0.216	13.06	0.071

9°. Aplicar modelo autoregressivo

Como existe autocorrelação nos dados, aplicar os procedimentos descritos na [seção 3.5.2 \(p. 74\)](#). O modelo encontrado foi $x_t = 4,059 + 0,4916x_{t-1}$. Os resíduos resultantes do modelo foram normalmente distribuídos e não apresentaram autocorrelação ([Apêndice D](#)).

10°. Verificar a estabilidade do processo

As amostras do processo são de tamanho igual a um, portanto os gráficos utilizados foram o de amplitude móvel e medidas individuais. A aplicação dos gráficos pode ser verificada na [Figura 14](#), que contém tanto os dados originais quanto os resíduos do modelo autoregressivo. Quando não considerada a autocorrelação, aparentemente o processo se mostrou fora do controle estatístico, pois no gráfico da média existe um ponto fora dos limites de controle e as observações 16 e 17 estão além de dois desvios padrão. Mas quando aplicado o gráfico de controle nos resíduos o processo se mostrou estável, tanto na variabilidade quanto na tendência central.

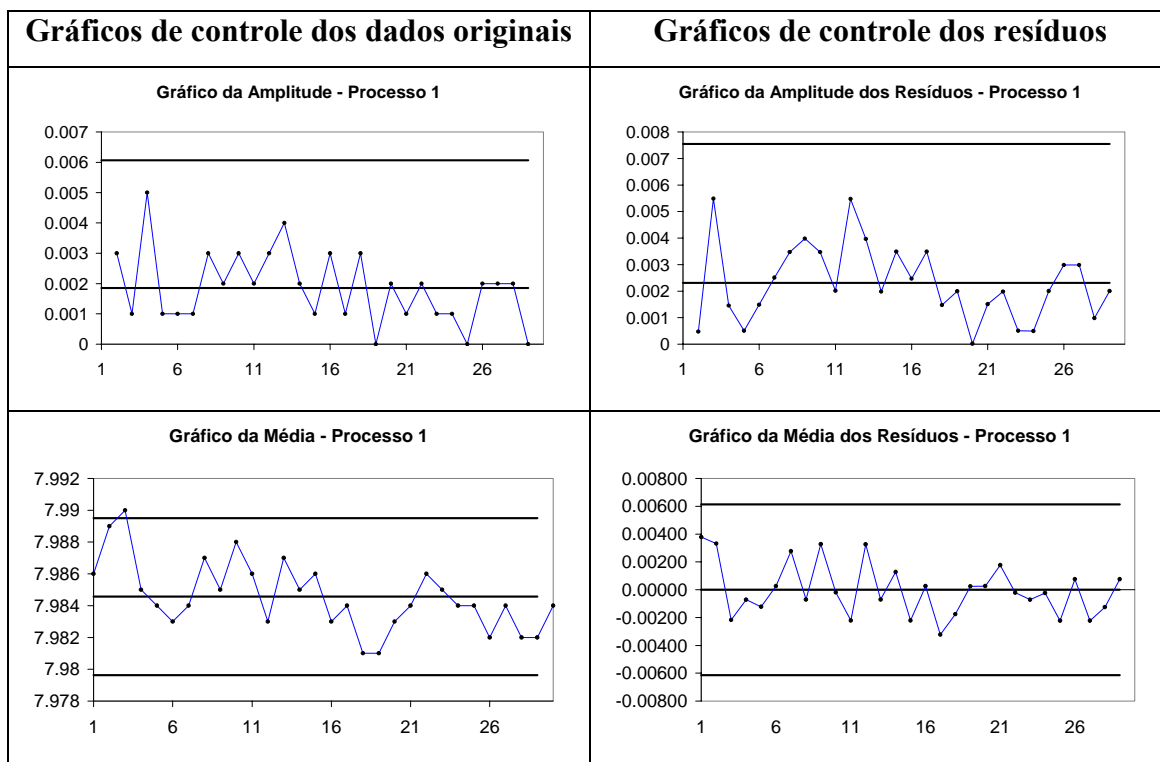


Figura 14 – Gráficos de controle para o processo 1

11°. Estimar índice

O alvo, limite superior e inferior de especificação, média e desvio padrão para esta peça são:

$Alvo = 7,988$, $LSE = 8,001$, $LIE = 7,976$, $\bar{x} = 7,985$ e $S = 0,002$. Todas as medidas estão em mm.

Como o processo é autocorrelacionado, o valor do desvio padrão deve ser corrigido de acordo com o explicado na seção 3.5.2. O desvio padrão encontrado foi:

$$\sigma = \frac{\sigma_e}{(1 - r_1 \phi_1)^{1/2}} = \frac{0,002}{(1 - 0,489 \cdot 0,492)^{1/2}} = 0,0022$$

O valor dos índices estimados e o percentual de defeituosos produzidos para as duas situações, considerando e não considerando autocorrelação, é apresentado na [Tabela 15](#).

Tabela 15 – Estimação dos índices de capacidade do processo 1

Sem considerar autocorrelação	Considerando autocorrelação
$\hat{C}_p = 1,90$ e $\hat{C}_{pk} = 1,38$	$\hat{C}_p = 1,88$ e $\hat{C}_{pk} = 1,35$
<i>ppm</i> acima do <i>LSE</i> = 1,82E-07	<i>ppm</i> acima do <i>LSE</i> = 1,94E-07
<i>ppm</i> abaixo do <i>LIE</i> = 17,68	<i>ppm</i> abaixo do <i>LIE</i> = 25,76
<i>ppm</i> total = 17,68	<i>ppm</i> total = 25,76

A autocorrelação pouco afetou a estimação dos índices, o seu principal impacto aconteceu na verificação da estabilidade do processo. Como comentado no capítulo 3 aumenta o número de alarmes falso. Para aplicar a utilização de intervalos de confiança foi desconsiderada a autocorrelação para estimação da média e desvio padrão.

12º. Identificar em qual situação a empresa se encontra.

A empresa se encontra na situação 2 (p. 78), índice estimado é maior que o estabelecido como alvo, mas o intervalo de confiança escolhido é menor que o desejado.

Aplicando as equações (2.38) e (2.42) os intervalos inferiores de confiança encontrados foram de 99,03% para \hat{C}_p e 60,20% para \hat{C}_{pk} . Valores abaixo do que foi especificado.

Para atingir o resultado de 95% podem-se aplicar as soluções apresentadas na seção 4.1.

A primeira seria aumentar o tamanho das amostras, mas para o valor estimado de $\hat{C}_{pk} = 1,38$ seriam necessárias 1275, um valor muito elevado. A segunda é realizar

modificação no processo. Neste caso a média do processo está deslocada em relação ao alvo (Figura 13,p. 88), seria necessário um deslocamento da média do processo em +

0,003 (mm) para centrar o processo. Com este procedimento $\hat{C}_p = \hat{C}_{pk} = 1,90$ e o

intervalo de confiança para C_{pk} será de 98,80%. A terceira solução é não fazer

nenhuma modificação e ficar com um intervalo de confiança de 60,20%. Porém nesta

situação corre-se o risco de considerar o processo capaz em 39,80%, quando na realidade ele não é. A quarta solução é reavaliar o valor de capacidade de processo escolhido como alvo. Para 95% de intervalo de confiança e um valor estimado de $\hat{C}_{pk} = 1,38$ o valor alvo deverá ser de 1,07, equivalente a 1339 *ppm*.

Para a escolha de alguma das soluções apontadas devem-se analisar os custos envolvida com cada uma. A recomendada é a solução 2, modificar o processo, pois centrar o processo no alvo não é muito oneroso, ao mesmo tempo em que se está melhorando o processo.

5.2 Processo 2 – Característica avaliada: distância entre faces

As etapas de análise são descritas abaixo e seguem o modelo 2.

1º. Definir o processo a ser avaliado.

O processo em questão é o de usinagem de peças e a característica medida é a distância entre faces (mm). Esta medida de distância é definida pelo cliente. As etapas 2, 3 e 4 são as mesmas do processo anterior.

5º. Escolher tamanho da amostra.

Foram retiradas 130 amostras consecutivas do processo para avaliar a capacidade. O valor destas amostras é apresentado na tabela 2C do apêndice C. As etapas 6 e 7 são as mesmas do processo anterior.

8º. Verificar se os dados são normais e independentes.

Utilizar os passos descritos no capítulo 3 das seções (3.2 e 3.5.1). O teste W apresentou o resultado $W = 0,9260$ e $p = 7,04E-08$ levando a rejeição da hipótese nula de normalidade.

9º. Aplicar transformação de Johnson.

Aplicando a transformação de Johnson (seção 3.3, pág. 61) foi obtida a seguinte equação de transformação: para um $z = 0,7$ a família foi S_u , $QR = 1,601$, $\hat{\gamma} = 1,042$, $\hat{\eta} = 1,688$, $\hat{\lambda} = 0,114$ e $\hat{\varepsilon} = 122,421$. A tabela 3D com os resultados da transformação é apresentada no apêndice D.

$$Z = 1,042 + 1,688 \sinh^{-1} \left(\frac{X - 122,421}{0,144} \right)$$

Para esta função de transformação o teste W apresentou o resultado $W = 0,9899$ e $p = 0,9445$ levando a não rejeição da hipótese nula de normalidade. Na Figura 15 são apresentados os gráficos para visualizar se os dados são normais, tanto dos dados transformados quanto dos originais.

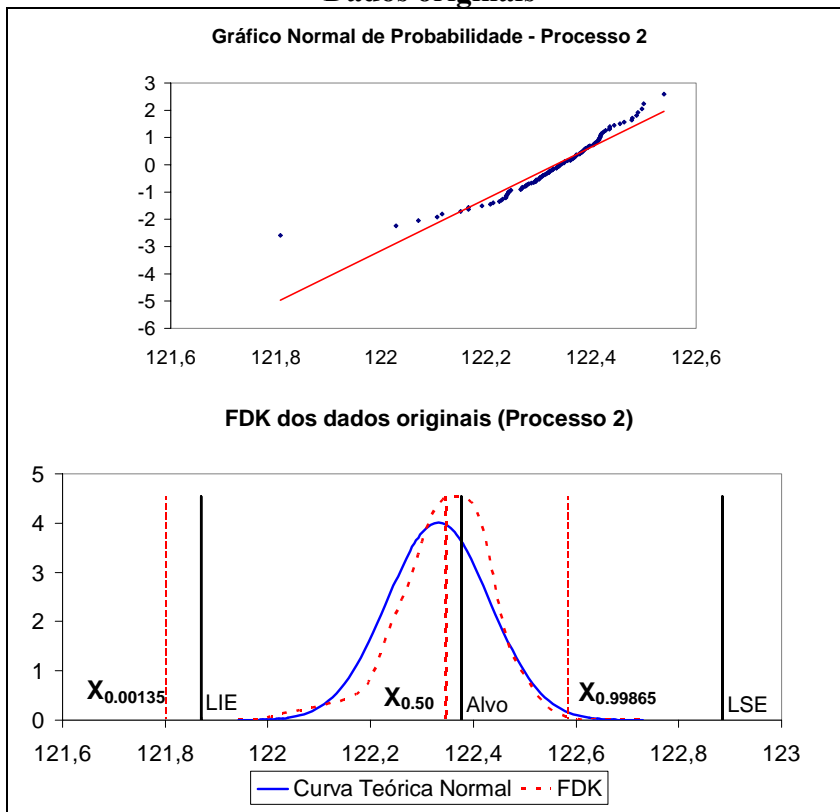
10º. Testar autocorrelação.

O teste de autocorreção é apresentado na [Tabela 16](#) e apresenta uma correlação significativa no primeiro período.

Tabela 16 – Teste de autocorrelação para os dados transformados do processo 2

k (atraso)	r_k	Q^*	P
1	0,419	23,33	0,000
2	0,365	41,16	0,000
3	0,232	48,44	0,000
4	0,172	52,46	0,000
5	0,064	53,03	0,000
6	0,056	53,46	0,000
7	0,076	54,27	0,000

Dados originais



Dados transformados

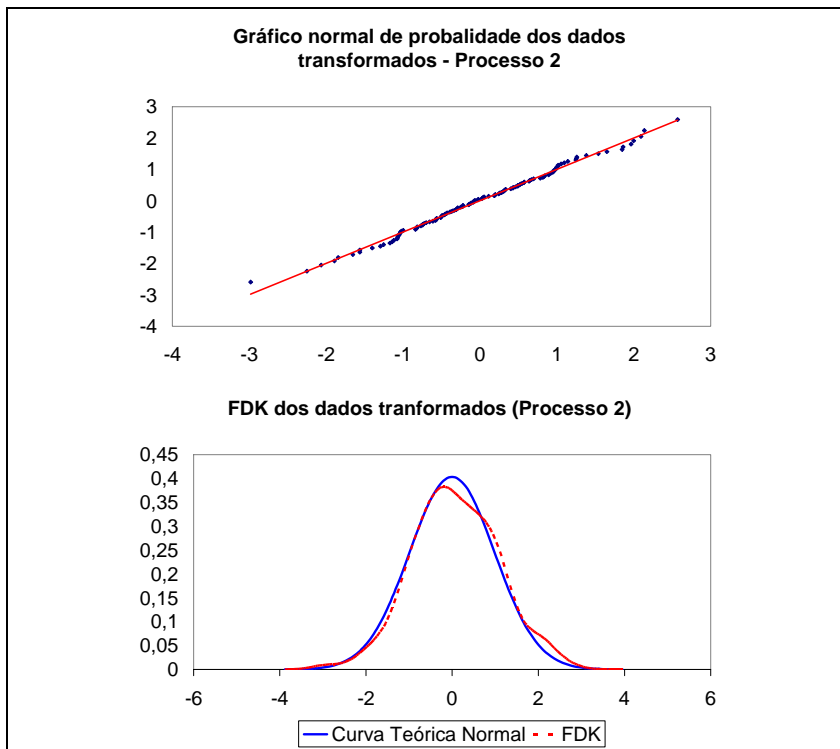


Figura 15 – Gráfico normal de probabilidade e FDK dos dados originais e transformados do processo 2

11°. Aplicar modelo autoregressivo

Como existe autocorrelação nos dados transformados, aplicar os procedimentos descritos na seção 3.5.2. O modelo encontrado foi $x_t = 0,5015x_{t-1}$. Os resíduos resultantes do modelo foram normalmente distribuídos e não apresentaram autocorrelação.

12°. Verificar a estabilidade do processo

As amostras do processo são de tamanho igual a um, portanto os gráficos utilizados foram o de amplitude móvel e medidas individuais. A aplicação dos gráficos pode ser verificada na [Figura 16](#), que contém tanto os dados originais quanto os resíduos do modelo autoregressivo. Quando não considerada a autocorrelação, aparentemente o processo se mostrou fora do controle estatístico, com vários pontos fora dos limites de controle, tanto no gráfico da média quanto no da amplitude. Com a transformação dos dados e a aplicação do modelo autoregressivo, o processo na verdade se mostrou estável, com apenas um ponto fora dos limites de controle no gráfico da amplitude, porém com a transformação de variável fica difícil determinar uma causa especial.

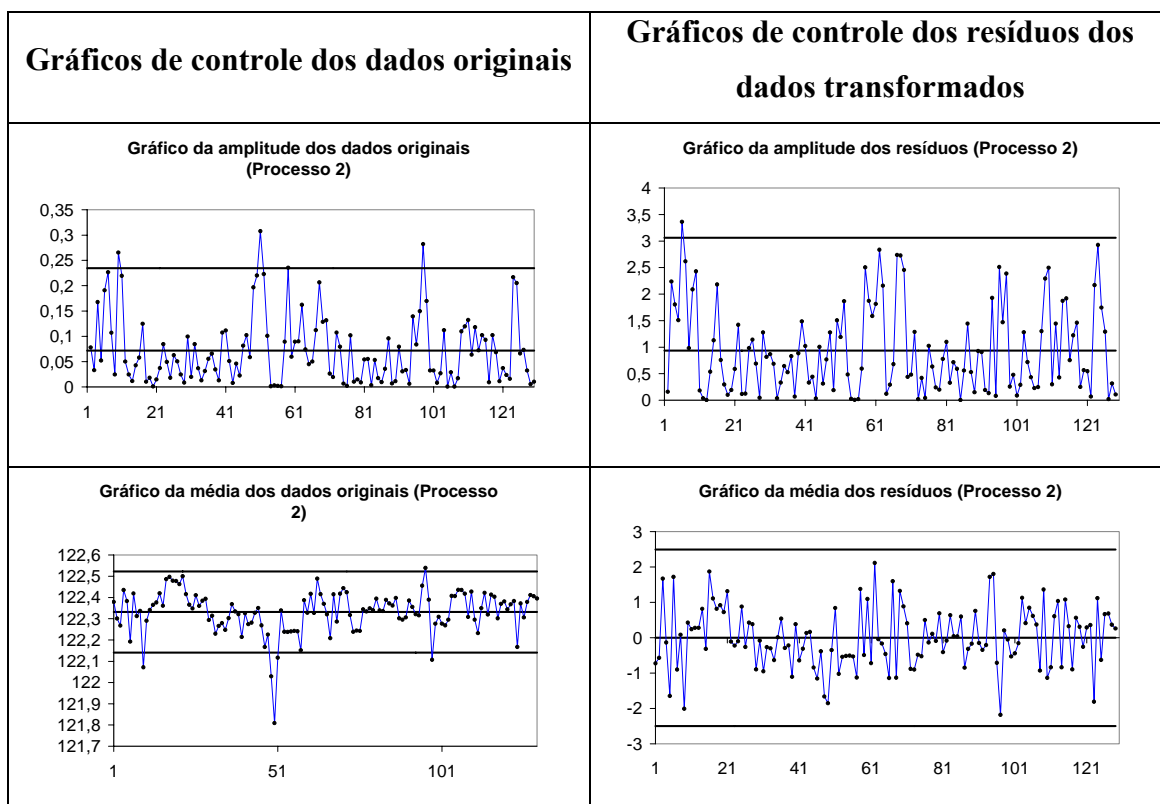


Figura 16 – Gráficos de controle para o processo 2

13º. Estimar índice

O alvo, limite superior e inferior de especificação para esta peça são:

$$Alvo = 122,3772, LSE = 122,8852, LIE = 121,8692$$

Para se estimar os índices nestas condições é preciso aplicar as técnicas vistas na seção

3.4. Os valores encontrados de x foram:

$$x_{0,00135} = 121,8020$$

$$x_{0,50} = 122,3462555$$

$$x_{0,99865} = 122,5848087$$

Com estes valores encontrados dos percentis o valor dos índices estimados são os apresentados na [Tabela 17](#).

Tabela 17 – Estimação dos índices de capacidade do processo 2

Sem considerar a não normalidade	Considerando a não normalidade
$\hat{C}_p = 1,70$ e $\hat{C}_{pk} = 1,55$	$\hat{C}_p = 1,30$ e $\hat{C}_{pk} = 0,88$
ppm acima do $LSE = 1,587$	ppm acima do $LSE = 0,125$
ppm abaixo do $LIE = 0,013$	ppm abaixo do $LIE = 4272$
ppm total = 1,600	ppm total = 4272

Observando a [Tabela 17](#) percebe-se que a estimação dos índices de capacidade sem se considerar a não normalidade dos dados causou um grande viés. Neste caso o processo era considerado muito melhor do a realidade. A diferença de defeituosos produzidos é muito diferente do que encontrado na realidade, de 1,6 ppm para 4272 ppm. É isto que se observa a [Figura 15](#), o limite inferior de especificação está acima está do percentil $x_{0,00135}$.

14°. Modificar o processo.

Neste caso o processo é incapaz, e modificações no processo são necessárias. Se a mediana do processo for deslocada em + 0,229 mm o processo reduzirá a produção de defeituosos para 99 ppm. Este é apenas o ajuste na tendência central do processo. Se a empresa realmente quer ter o processo capaz com 1,33, deverá reduzir também a variabilidade.

Para finalizar a apresentação deste trabalho no próximo capítulo serão apresentadas as conclusões, se os objetivos apresentados no capítulo 1 foram alcançados e quais as foram as dificuldades encontradas.

6 CONCLUSÕES

O objetivo principal deste trabalho, que foi a criação de modelos para a avaliação da capacidade de processos foi atingindo, a medida que estes foram aplicados com sucesso em um estudo de caso real em uma empresa do setor automotivo. Nos dois processos analisados os dois modelos conseguiram lidar com situações que não são consideradas ideais para a estimação de índices de capacidade de processos, autocorrelação e não normalidade. O modelo 1, para dados normais e independentes, se adequou ao processo 1, enquanto que o modelo 2, para dados não normais e autocorrelacionado, ao processo 2. Nos dois processos analisados acredita-se que se tenha chagado a uma avaliação da capacidade mais próxima da realidade quando foram considerados aspectos teóricos que geralmente as empresas ignoram. Estes aspectos ignorados estimulam a descrença na utilização dos índices e da estatística.

Por isto, foi realizada uma pesquisa bibliográfica a fim de se conhecer e organizar este conteúdo teórico em dois capítulos (2 e 3) desta dissertação. Nesta revisão da literatura foram abordados assuntos como: estimação de índices de capacidade, intervalos de confiança para estes índices, uso dos índices para dados não normais, uso dos índices para dados autocorrelacionados. Toda revisão teórica feita com exemplos numéricos serviram de base para a construção dos modelos e a aplicação em um estudo de caso.

No processo 1 os dados eram normais, porém apresentaram autocorrelação. Para resolver este problema utilizaram-se modelos autoregressivos, e se trabalhou com os resíduos deste modelo. Esta autocorrelação afetou muito mais o monitoramento da estabilidade do processo do que a estimação da capacidade do processo. Sem considerá-la o processo parecia ser instável, mas quando utilizado os resíduos do modelo

verificou-se a estabilidade. Os valores estimados dos índices pouco se alteraram com a autocorrelação, fato este inesperado, pois a princípio se achava que eles seriam superestimados.

O valor estimado dos índices para este processo estavam acima do firmado em contrato, mas a confiança sob o valor calculado foi abaixo da estabelecida, que era de 95%, enquanto que o valor encontrado foi de 60%. Isto quer dizer que existe 40 % de chance de se concluir incorretamente que o processo é capaz. Para resolver este problema foram sugeridas várias soluções, que incluíam a possibilidade de mudanças no processo, aumento do tamanho das amostras e até mesmo a reavaliação do valor escolhido como alvo.

No processo 2 os dados apresentaram não normalidade e foi necessária a transformação dos dados para utilização dos gráficos de controle e estimação da capacidade. Esta estimação apresentou resultados muito diferentes, quando considerada e não considerada a normalidade. O processo parecia ser melhor do que na realidade era. A situação do processo 2 foi mais grave que a do processo 1, pois o processo se mostrou incapaz, de acordo com o que foi definido $C_p = 1,33$ e $C_{pk} = 1,33$, por isto, foram sugeridas mudança, que em conjunto com a experiência dos funcionários, podem resultar em melhorias para o processo.

As dificuldades encontradas na aplicação dos modelos se deram na medida em que muitas técnicas utilizadas exigem a utilização de recursos computacionais, dificultando a implantação das técnicas no chão de fábrica.. Além disto, alguns problemas teóricos encontrados ainda não foram resolvidos ou não fizeram parte do escopo deste trabalho e ficam como sugestões para trabalhos futuros.

Recomendações para trabalhos futuros.

- Aplicar o estudo da capacidade de processos para dados discretos.
- Aplicar o estudo da capacidade de processo para dados multivariados
- Encontrar intervalos de confiança para dados não normais.
- Incluir os erros de medição na estimação dos índices de capacidade
- Criar um modelo de custos para auxiliar na tomada de decisões, quando se tem que fazer modificações no processo.

REFERÊNCIA BIBLIOGRÁFICA

- AIAG. **Statistical Process Control Reference Manual**. Automotive Industry Action Group, Southfield, ML, 1995.
- BLOM, G. **Statistical Estimates and Transformed Beta-variables**. New York: Wiley, 1958. 69-71 p.
- BOYLES, R. A. The Taguchi Capability Index. **Journal of Quality Technology**, vol. 21, n° 1, p. 17-26, 1991.
- CHEN, J. P.; DING C. G. A new process capability index for non-normal distributions. **The International Journal of Quality & Reliability Management**, v. 18, n° 6/7, p. 162-170, 2001.
- CHOU, Y. M.; POLANSKY, A. M.; MASON, R. L. Transforming Non-Normal Data to Normality in Statistical Process Control. **Journal of Quality Technology**, v. 30, n° 2, p. 133-141, 1998.
- CLEMENTS, J. A. Process Capability Calculations for Non-Normal Distributions. **Quality Progress**, 22, p. 95-100, 1989.
- DELERYD, M. On The Gap between Theory and Practice of Process Capability Studies. **International Journal of Quality & Reliability Management**, v. 15, n° 2, p. 178-191, 1998.
- DELERYD, M. A Pragmatic View on Process Capability Studies. **International Journal of Production Economics**, n° 58, p. 319-330, 1999a.
- DELERYD, M. The Effect of Skewness on Estimates of Some Process Capability Indices. **International Journal of Applied Quality Management**, v. 2, n° 2, p. 153-186, 1999b.
- FRANKLIN, L. A.; WASSERMAN, G. S. A Note on Conservative Nature of The Tables Of Lower Confidence Limits For C_{pk} With A Suggested Correction. **Communications in Statistics: Simulation and Computation**, vol. 21, n° 4, p. 1165-1169, 1992.
- FRANKLIN, L. A. Sample size determination for lower confidence limits for estimating process capability indices. **Computers & Industrial Engineering**, 36, p.603-614, 1999.

- HEAVLIN, W. D. **Statistical properties of capability indices**. Technical report n° 320, C.A: Thech. Library, Advanced Micro Devices, Inc., Sunnyvale, 1988.
- HSIANG, T. C.; TAGUCHI, G. A Tutorial on Quality Control and Assurance – The Taguchi Methods. **ASA Annual Meeting**. Las Vegas, Nevada, 1985.
- JAGADEESH, R.; SUBASH, B. A. Process Capability Assessment with Tool Wear: An Investigative Study. **The International Journal of Quality & Reliability Management**, v. 11, n° 2, p. 51-62, 1994.
- JOHNSON, N. L. Systems of Frequency Curves Generated by Methods of Translation. **Biometrika**, 36, p. 149-176, 1949.
- JURAN, J. M.; GRZYNA, F. M.; BINGHAM, R. S. Jr. **Quality Control Handbook**. New York: McGraw-Hill, 1974.
- KANE, V. E. Process Capability Indices. **Journal of Quality Technology**, vol. 18, n° 1, p. 41-52, 1986.
- KOTZ, S.; JOHNSON, N. L. **Process Capability Indices**. London: Chapman & Hall, 1993.
- KOTZ, S.; LOVELACE, C. R. **Process Capability Indices in Theory and Practice**. London: Arnold, 1998.
- KOTZ, S.; JOHNSON, N. L.; HUBELE, N. F.; SPIRING, F, et al. Process capability indices - a review, 1992-2000 / Discussion / Response. **Journal of Quality Technology**, v. 34, n° 1, p. 2-52, 2002.
- KRISHNAMOORTHY, K. S.; KHATWANI, S. A Capability Index for All Occasions. **Quality Congress. ASQ's ... Annual Quality Congress Proceedings**, p. 77-81, 2000.
- LJUNG, G. M.; BOX, G. E. P. On A Measure of Lack of Fit in Time Series Models. **Biometrika**, 65, p. 297-303, 1978.
- MONTGOMERY, D. C.; RUNGER, G. C. **Applied Statistics and Probability for Engineers**. 3. ed. John Wiley & Sons, 2003.
- MONTGOMERY, D. C. **Introduction to Statistical Quality Control**. 3. ed. John Wiley & Sons, 1996.
- PAN, J. N.; WU, S. L. Process Capability Analysis for Non-Normal Relay Test Data. **Microelectronics and Reliability**, v. 37, n° 3, p. 421-428, 1997.
- PEARN W. L.; SHU, M.H. An algorithm for calculating the lower confidence bounds of CPU and CPL with application to low-drop-out linear regulators. **Microelectronics Reliability**, n° 43, 495-502, 2003.

- PEARN W. L.; KOTZ, S.; JOHNSON N. L. Distributional and Inferential Properties of Process Capability Indices. **Journal of Quality Technology**, vol. 4, nº 4, p. 216-231, 1992.
- PEARN, W. L.; CHEN, K. S. Capability Indices for Non-Normal Distributions With an Application in Electrolytic Capacitor Manufacturing. **Microelectronics and Reliability**, v. 37, nº 12, p. 1853-1858, 1997.
- PEARN, W. L.; CHEN K. S.; LIN, G. H. A Generalization of Clements' Method for Non-Normal Pearsonian Processes with Asymmetric Tolerances. **The International Journal of Quality & Reliability Management**, v. 16, nº 5, p. 507-521, 1999.
- PEARN, W. L.; CHEN, K. S. One-sided capability indices C(PU) and C(PL): Decision making with sample information. **The International Journal of Quality & Reliability Management**; vol. 19, nº 2-3, 221-245.
- PYZDEK, T. Process Capability Analysis Using Personal Computers. **Quality Engineering**, v. 4, p. 419-440, 1992.
- ROTH G. II. How to Perform a Machine Capability Study. **ASQ Six Sigma Forum Magazine**, p. 22-31, Novembro de 2003.
- ROYSTON, J. P. An extension of Shapiro and Wilk's W test for normality to large samples. **Applied Statistics**, 31, p. 115-124, 1982.
- SEIER, E. Comparison of Tests for Univariate Normality. **InterStat**. Janeiro de 2002. Disponível em: <<http://interstat.statjournals.net/>>. Acesso em: Dezembro de 2003.
- SHAPIRO, S. S.; WILK, M. B. An analysis of variance test for normality (complete samples). **Biometrika**, 52, p. 591-611, 1965.
- SHORE, H. Process Capability Analysis When Data Are Autocorrelated. **Quality Engineering**, v. 9, nº 4, p. 615-626, 1997.
- SILVERMAN, B. W. **Density Estimation for Statistics and Data Analysis**. Chapman & Hall, 1986.
- SLIFKER, J. F.; SHAPIRO, S. S. The Johnson System: Selection and Parameter Estimation. **Technometrics**, v. 22, nº 2, p. 239-246, 1980.
- SPEDDING, T. A.; RAWLINGS, P. L. Non-Normality in Statistical Process Control Measurements. **International Journal of Quality & Reliability Management**, v. 11, nº 6, p. 27-37, 1994.
- THE TOOLS of Quality, Part IV: Histograms. **Quality Progress**, September, p. 75-78, 1990.

TONG, L. I.; CHEN, J. P. Lower Confidence Limits Of Process Capability Indices For Non-Normal Process Distributions. **The International Journal of Quality & Reliability Management**, v. 15, n° 8/9, p. 907-919, 1998.

ANEXO A – Coeficientes para os gráficos de controle de Shewhart

Tabela 1A - Valores dos coeficientes para construção do gráfico da média.

Observações na amostra, n	Coeficientes para os limites de controle		
	A	A ₂	A ₃
2	2,121	1,880	2,659
3	1,732	1,023	1,954
4	1,500	0,729	1,628
5	1,342	0,577	1,427
6	1,225	0,483	1,287
7	1,134	0,419	1,182
8	1,061	0,373	1,099
9	1,000	0,337	1,032
10	0,949	0,308	0,975
11	0,905	0,285	0,927
12	0,866	0,266	0,886
13	0,832	0,249	0,850
14	0,802	0,235	0,817
15	0,775	0,223	0,789
16	0,750	0,212	0,763
17	0,728	0,203	0,739
18	0,707	0,194	0,718
19	0,688	0,187	0,698
20	0,671	0,180	0,680
21	0,655	0,173	0,663
22	0,640	0,167	0,647
23	0,626	0,162	0,633
24	0,612	0,157	0,619
25	0,600	0,153	0,606

Fonte: Montgomery(1996)

Tabela 2A - Valores dos coeficientes para construção do gráfico do desvio padrão.

Observações na amostra, n	Coeficientes para linha central		Coeficientes para os limites de controle			
	c ₄	1/c ₄	B ₃	B ₄	B ₅	B ₆
2	0,7979	1,2533	0,000	3,267	0,000	2,606
3	0,8862	1,1284	0,000	2,568	0,000	2,276
4	0,9213	1,0854	0,000	2,266	0,000	2,088
5	0,9400	1,0638	0,000	2,089	0,000	1,964
6	0,9515	1,0510	0,030	1,970	0,029	1,874
7	0,9594	1,0423	0,118	1,882	0,113	1,806
8	0,9650	1,0363	0,185	1,815	0,179	1,751
9	0,9693	1,0317	0,239	1,761	0,232	1,707
10	0,9727	1,0281	0,284	1,716	0,276	1,669
11	0,9754	1,0252	0,321	1,679	0,313	1,637
12	0,9776	1,0229	0,354	1,646	0,346	1,610
13	0,9794	1,0210	0,382	1,618	0,374	1,585
14	0,9810	1,0194	0,406	1,594	0,399	1,563
15	0,9823	1,0180	0,428	1,572	0,421	1,544
16	0,9835	1,0168	0,448	1,552	0,440	1,526
17	0,9845	1,0157	0,466	1,534	0,458	1,511
18	0,9854	1,0148	0,482	1,518	0,475	1,496
19	0,9862	1,0140	0,497	1,503	0,490	1,483
20	0,9869	1,0133	0,510	1,490	0,504	1,470
21	0,9876	1,0126	0,523	1,477	0,516	1,459
22	0,9882	1,0119	0,534	1,466	0,528	1,448
23	0,9887	1,0114	0,545	1,455	0,539	1,438
24	0,9892	1,0109	0,555	1,445	0,549	1,429
25	0,9896	1,0105	0,565	1,435	0,559	1,420

Fonte: Montgomery(1996)

Tabela 3A - Valores dos coeficientes para construção do gráfico da amplitude.

Observações na amostra, n	Coeficientes para linha central		Coeficientes para os limites de controle				
	d_2	$1/d_2$	d_3	D_1	D_2	D_3	D_4
2	1,128	0,8865	0,853	0,000	3,686	0,000	3,267
3	1,693	0,5907	0,888	0,000	4,359	0,000	2,575
4	2,059	0,4857	0,880	0,000	4,698	0,000	2,282
5	2,326	0,4299	0,864	0,000	4,918	0,000	2,115
6	2,534	0,3946	0,848	0,000	5,078	0,000	2,004
7	2,704	0,3698	0,833	0,204	5,204	0,076	1,924
8	2,847	0,3512	0,820	0,388	5,306	0,136	1,864
9	2,970	0,3367	0,808	0,547	5,393	0,184	1,816
10	3,078	0,3249	0,797	0,687	5,469	0,223	1,777
11	3,173	0,3152	0,787	0,811	5,535	0,256	1,744
12	3,258	0,3069	0,778	0,922	5,594	0,283	1,717
13	3,336	0,2998	0,770	1,025	5,647	0,307	1,693
14	3,407	0,2935	0,763	1,118	5,697	0,328	1,672
15	3,472	0,2880	0,756	1,203	5,741	0,347	1,653
16	3,532	0,2831	0,750	1,282	5,782	0,363	1,637
17	3,588	0,2787	0,744	1,356	5,820	0,378	1,622
18	3,640	0,2747	0,739	1,424	5,856	0,391	1,608
19	3,689	0,2711	0,734	1,487	5,891	0,403	1,597
20	3,735	0,2677	0,729	1,549	5,921	0,415	1,585
21	3,778	0,2647	0,724	1,605	5,951	0,425	1,575
22	3,819	0,2618	0,720	1,659	5,979	0,434	1,566
23	3,858	0,2592	0,716	1,710	6,006	0,443	1,557
24	3,895	0,2567	0,712	1,759	6,031	0,451	1,548
25	3,931	0,2544	0,708	1,806	6,056	0,459	1,541

Fonte: Montgomery(1996)

Apêndice A – Cálculo dos coeficientes a_i e do W do Teste W de Shapiro-Wilk para o exemplo do capítulo 3.

A1. Cálculo dos coeficientes a_i

1º passo: calcular os valores de a_i^* utilizando a equação (3.10) para $i = 2, 3, \dots, n-1$.

$$\text{Para } i = 2 \text{ tem-se: } a_2^* = 2m_2 = 2 \cdot \Phi^{-1}\left(\frac{i-0,375}{n+0,25}\right) = 2 \cdot \Phi^{-1}\left(\frac{2-0,375}{120+0,25}\right) = -4,422$$

⋮
⋮

$$\text{Para } i = n-1 = 119 \text{ tem-se: } a_{119}^* = 2m_{119} = 2 \cdot \Phi^{-1}\left(\frac{119-0,375}{120+0,25}\right) = 4,422$$

2º passo: calcular os valores de \hat{a}_1^2 e \hat{a}_n^2 utilizando a equação (3.12).

$$\hat{a}_1^2 = \hat{a}_n^2 = \frac{\Gamma\left(\frac{1}{2}[120+1]\right)}{\sqrt{2}\Gamma\left(\frac{1}{2} \cdot 120+1\right)} = 0,0911$$

3º passo: calcular o somatório $\sum_{i=2}^{n-1} a_i^{*2}$ com os valores encontrados no primeiro passo.

$$\sum_{i=2}^{n-1} a_i^{*2} = 414,627$$

4º passo: calcular os valores de a_i^* utilizando a equação (3.10) para $i = 1, i = n$.

$$-a_1^* = a_{120}^* = \left(\frac{\hat{a}_1^2}{1-2\hat{a}_n^2} \sum_{i=2}^{120-1} a_i^{*2}\right)^{1/2} = \left(\frac{0,0911}{1-0,0911} \cdot 414,627\right)^{1/2} = 6,796$$

5º passo: calcular os coeficientes a_i utilizando a equação (3.13).

$$a_1 = \frac{a_1^*}{\left(\sum_{i=1}^{120} a_i^{*2}\right)^{1/2}} = \frac{-6,796}{507^{0,5}} = -0,3018, \quad a_2 = \frac{a_2^*}{\left(\sum_{i=1}^{120} a_i^{*2}\right)^{1/2}} = \frac{-4,422}{557^{0,5}} = -0,1964$$

Na tabela 1A estão os coeficientes para o exemplo do capítulo 3 (os dados de ganho de amplificadores da [Tabela 6 \(p. 52\)](#)). Para o cálculo dos coeficientes precisa-se apenas saber qual é o tamanho da amostra, que no caso é de 120 observações.

Tabela 4A – Coeficientes a_i para o teste W de Shapiro-Wilk

i	a_i	i	a_i	i	a_i	i	a_i	i	a_i	i	a_i
1	-0,3018	21	-0,0842	41	-0,0372	61	0,0009	81	0,0392	101	0,0872
2	-0,1964	22	-0,0814	42	-0,0351	62	0,0028	82	0,0412	102	0,0902
3	-0,1792	23	-0,0786	43	-0,0332	63	0,0046	83	0,0433	103	0,0934
4	-0,1669	24	-0,0759	44	-0,0312	64	0,0065	84	0,0454	104	0,0967
5	-0,1571	25	-0,0732	45	-0,0292	65	0,0083	85	0,0475	105	0,1001
6	-0,1490	26	-0,0707	46	-0,0273	66	0,0102	86	0,0497	106	0,1036
7	-0,1419	27	-0,0682	47	-0,0253	67	0,0121	87	0,0519	107	0,1074
8	-0,1356	28	-0,0657	48	-0,0234	68	0,0139	88	0,0541	108	0,1114
9	-0,1300	29	-0,0633	49	-0,0215	69	0,0158	89	0,0563	109	0,1155
10	-0,1248	30	-0,0609	50	-0,0196	70	0,0177	90	0,0586	110	0,1200
11	-0,1200	31	-0,0586	51	-0,0177	71	0,0196	91	0,0609	111	0,1248
12	-0,1155	32	-0,0563	52	-0,0158	72	0,0215	92	0,0633	112	0,1300
13	-0,1114	33	-0,0541	53	-0,0139	73	0,0234	93	0,0657	113	0,1356
14	-0,1074	34	-0,0519	54	-0,0121	74	0,0253	94	0,0682	114	0,1419
15	-0,1036	35	-0,0497	55	-0,0102	75	0,0273	95	0,0707	115	0,1490
16	-0,1001	36	-0,0475	56	-0,0083	76	0,0292	96	0,0732	116	0,1571
17	-0,0967	37	-0,0454	57	-0,0065	77	0,0312	97	0,0759	117	0,1669
18	-0,0934	38	-0,0433	58	-0,0046	78	0,0332	98	0,0786	118	0,1792
19	-0,0902	39	-0,0412	59	-0,0028	79	0,0351	99	0,0814	119	0,1964
20	-0,0872	40	-0,0392	60	-0,0009	80	0,0372	100	0,0842	120	0,3018

A2. Cálculo de W

Com os coeficientes calculados anteriormente e os valores ordenados da [Tabela 6 \(p.52\)](#) é possível agora encontrar o valor de W utilizando a [equação \(3.9\)](#).

$$W = \frac{\left[\sum_{i=1}^n a_i y_{(i)}\right]^2}{\sum_{i=1}^n (y_i - \bar{y})^2} = \frac{\left[(-0,3018 \cdot 7,8) + (-0,1964 \cdot 7,8) + \dots + (0,3018 \cdot 11,7)\right]^2}{(7,8 - 9,044)^2 + (7,8 - 9,044)^2 + \dots + (11,7 - 9,044)^2}$$

$$W = \frac{85,64}{91,31} = 0,9379$$

Apêndice B – Tabela com os valores da transformação de Johnson do exemplo do capítulo 3.

Os dados da tabela abaixo foram calculados com a expressão

$$Z = 0,9602 + 0,9543 \ln \left(\frac{X - 7,6101}{4,6482 + 7,6101 - X} \right),$$

que foi encontrada utilizando os dados da

Tabela 6 (p.52) e a transformação de Johnson apresentada no capítulo 3.

Tabela 1B – Valores obtidos aplicando a transformação de Johnson aos dados de ganho de amplificadores do capítulo 3

i	x	z	i	x	z	i	x	z	i	x	z
1	7,8	-2,052	31	8,4	-0,553	61	8,9	0,047	91	9,6	0,684
2	7,8	-2,052	32	8,4	-0,553	62	9	0,147	92	9,6	0,684
3	7,8	-2,052	33	8,4	-0,553	63	9	0,147	93	9,7	0,767
4	7,8	-2,052	34	8,4	-0,553	64	9	0,147	94	9,7	0,767
5	7,8	-2,052	35	8,4	-0,553	65	9	0,147	95	9,7	0,767
6	7,9	-1,626	36	8,5	-0,414	66	9	0,147	96	9,7	0,767
7	7,9	-1,626	37	8,5	-0,414	67	9	0,147	97	9,8	0,850
8	7,9	-1,626	38	8,5	-0,414	68	9,1	0,243	98	9,8	0,850
9	7,9	-1,626	39	8,5	-0,414	69	9,1	0,243	99	9,8	0,850
10	7,9	-1,626	40	8,5	-0,414	70	9,1	0,243	100	9,9	0,932
11	8	-1,321	41	8,5	-0,414	71	9,1	0,243	101	9,9	0,932
12	8	-1,321	42	8,6	-0,287	72	9,1	0,243	102	10	1,014
13	8	-1,321	43	8,6	-0,287	73	9,2	0,336	103	10	1,014
14	8	-1,321	44	8,6	-0,287	74	9,2	0,336	104	10,1	1,097
15	8	-1,321	45	8,6	-0,287	75	9,2	0,336	105	10,1	1,097
16	8,1	-1,081	46	8,6	-0,287	76	9,2	0,336	106	10,1	1,097
17	8,1	-1,081	47	8,7	-0,169	77	9,2	0,336	107	10,2	1,179
18	8,1	-1,081	48	8,7	-0,169	78	9,3	0,426	108	10,2	1,179
19	8,1	-1,081	49	8,7	-0,169	79	9,3	0,426	109	10,3	1,263
20	8,1	-1,081	50	8,7	-0,169	80	9,3	0,426	110	10,4	1,348
21	8,2	-0,880	51	8,7	-0,169	81	9,4	0,514	111	10,5	1,434
22	8,2	-0,880	52	8,7	-0,169	82	9,4	0,514	112	10,5	1,434
23	8,2	-0,880	53	8,8	-0,058	83	9,4	0,514	113	10,6	1,523
24	8,2	-0,880	54	8,8	-0,058	84	9,4	0,514	114	10,7	1,613
25	8,3	-0,707	55	8,8	-0,058	85	9,5	0,599	115	10,7	1,613
26	8,3	-0,707	56	8,8	-0,058	86	9,5	0,599	116	10,8	1,707
27	8,3	-0,707	57	8,9	0,047	87	9,5	0,599	117	10,9	1,804
28	8,3	-0,707	58	8,9	0,047	88	9,5	0,599	118	11,1	2,013
29	8,3	-0,707	59	8,9	0,047	89	9,6	0,684	119	11,5	2,520
30	8,3	-0,707	60	8,9	0,047	90	9,6	0,684	120	11,7	2,860

Apêndice C – Dados de medidas de peças dos dois processos de usinagem do capítulo 5.

Tabela 1C – Diâmetro interno de uma peça usinada no processo 1 (mm)

Amostra	Diâmetro	Amostra	Diâmetro	Amostra	Diâmetro
1	7,986	11	7,986	21	7,984
2	7,989	12	7,983	22	7,986
3	7,990	13	7,987	23	7,985
4	7,985	14	7,985	24	7,984
5	7,984	15	7,986	25	7,984
6	7,983	16	7,983	26	7,982
7	7,984	17	7,984	27	7,984
8	7,987	18	7,981	28	7,982
9	7,985	19	7,981	29	7,982
10	7,988	20	7,983	30	7,984

Tabela 2C – Distância entre faces de uma peça usinada no processo 2 (mm)

Amostra	Distância	Amostra	Distância	Amostra	Distância	Amostra	Distância	Amostra	Distância
1	122,3788	27	122,3606	53	122,2386	79	122,3489	105	122,4068
2	122,3008	28	122,3851	54	122,2374	80	122,3401	106	122,4357
3	122,2678	29	122,3936	55	122,2403	81	122,3943	107	122,4349
4	122,4354	30	122,2941	56	122,2425	82	122,3392	108	122,4178
5	122,3832	31	122,3139	57	122,2410	83	122,3359	109	122,3082
6	122,1925	32	122,2293	58	122,1517	84	122,3889	110	122,4278
7	122,4191	33	122,2662	59	122,3873	85	122,3714	111	122,2957
8	122,3121	34	122,2790	60	122,3278	86	122,3619	112	122,2319
9	122,3367	35	122,2475	61	122,4171	87	122,3976	113	122,3497
10	122,0712	36	122,3030	62	122,3270	88	122,3020	114	122,4220
11	122,2907	37	122,3689	63	122,4891	89	122,2951	115	122,3198
12	122,3408	38	122,3344	64	122,4153	90	122,3063	116	122,4130
13	122,3652	39	122,3213	65	122,3703	91	122,3858	117	122,4037
14	122,3768	40	122,2138	66	122,3203	92	122,3551	118	122,3015
15	122,4193	41	122,3251	67	122,2082	93	122,3217	119	122,3699
16	122,3615	42	122,2740	68	122,4147	94	122,3156	120	122,3812
17	122,4864	43	122,2817	69	122,2862	95	122,4551	121	122,3441
18	122,4966	44	122,3278	70	122,4179	96	122,5388	122	122,3676
19	122,4786	45	122,3504	71	122,4442	97	122,3893	123	122,3835
20	122,4776	46	122,2689	72	122,4248	98	122,1071	124	122,1667
21	122,4630	47	122,1667	73	122,3176	99	122,2768	125	122,3718
22	122,5003	48	122,2253	74	122,2382	100	122,3091	126	122,3057
23	122,4157	49	122,0287	75	122,2446	101	122,2767	127	122,3790
24	122,3662	50	121,8087	76	122,2425	102	122,2686	128	122,4114
25	122,3483	51	122,1166	77	122,3444	103	122,2956	129	122,4059
26	122,4109	52	122,3394	78	122,3341	104	122,4076	130	122,3957

Apêndice D – Resíduos dos modelos dos estudos de caso dos processos (1 e 2) apresentados no capítulo 5.

D1. Processo 1

Na tabela 1D são apresentados os resíduos do modelo autoregressivo $x_t = 4,059 + 0,4916x_{t-1}$ aplicado aos dados do processo 1 (tabela 1C do apêndice C) do capítulo 5.

Tabela 1D – Resíduos do modelo autoregressivo do processo 1 do capítulo 5.

t	Resíduo	t	Resíduo	t	Resíduo
1		11	-0,000195412	21	0,000262516
2	0,003787759	12	-0,002212241	22	0,00177093
3	0,003313002	13	0,003262516	23	-0,000212241
4	-0,002178583	14	-0,000703827	24	-0,000720655
5	-0,000720655	15	0,001279345	25	-0,00022907
6	-0,00122907	16	-0,002212241	26	-0,00222907
7	0,000262516	17	0,000262516	27	0,000754101
8	0,00277093	18	-0,00322907	28	-0,00222907
9	-0,000703827	19	-0,001754313	29	-0,001245899
10	0,003279345	20	0,000245687	30	0,000754101

Na tabela 2D é apresentada a função de autocorrelação para os resíduos do modelo acima.

Tabela 2D – Função de autocorrelação dos resíduos do modelo autoregressivo do processo 1 do capítulo 5.

k (atraso)	r_k	Q^*	p
1	-0,060	0,1167	0,733
2	0,054	0,2128	0,899
3	-0,068	0,3713	0,946
4	-0,275	3,0941	0,542
5	0,167	4,1382	0,530
6	-0,093	4,4768	0,612
7	0,208	6,2380	0,512
8	0,053	6,3603	0,607
9	-0,134	7,1678	0,620
10	0,088	7,5374	0,674

O teste W apresentou um $W = 0,9365$ e um $p = 0,094$.

D2. Processo 2

Na tabela 3D está o resultado da transformação de Johnson aplicada aos dados do processo 2 (tabela 2C do apêndice C). A transformação encontrada foi :

$$Z = 1,042 + 1,688 \sinh^{-1} \left(\frac{X - 122,421}{0,144} \right)$$

para um $z = 0,7$ a família foi S_u , $QR = 1,601$, $\hat{\gamma} = 1,042$, $\hat{\eta} = 1,688$, $\hat{\lambda} = 0,114$ e $\hat{\epsilon} = 122,421$.

Tabela 3D – Transformação de Johnson aplicada aos dados do processo 2 capítulo 5.

i	x	z	i	x	z	i	x	z	i	x	z
1	122.3788	0.4287	36	122.303	-0.4897	71	122.4442	1.3825	106	122.4357	1.2582
2	122.3008	-0.5123	37	122.3689	0.2931	72	122.4248	1.0972	107	122.4349	1.2464
3	122.2678	-0.8256	38	122.3344	-0.1434	73	122.3176	-0.3346	108	122.4178	0.9934
4	122.4354	1.2538	39	122.3213	-0.2936	74	122.2382	-1.0719	109	122.3082	-0.4356
5	122.3832	0.4902	40	122.2138	-1.2543	75	122.2446	-1.0211	110	122.4278	1.1416
6	122.1925	-1.4006	41	122.3251	-0.2509	76	122.2425	-1.0379	111	122.2957	-0.5637
7	122.4191	1.0126	42	122.274	-0.7701	77	122.3444	-0.0229	112	122.2319	-1.1207
8	122.3121	-0.3942	43	122.2817	-0.6990	78	122.3341	-0.1469	113	122.3497	0.0430
9	122.3367	-0.1161	44	122.3278	-0.2201	79	122.3489	0.0329	114	122.422	1.0556
10	122.0712	-2.0660	45	122.3504	0.0518	80	122.3401	-0.0753	115	122.3198	-0.3103
11	122.2907	-0.6129	46	122.2689	-0.8158	81	122.3943	0.6485	116	122.413	0.9223
12	122.3408	-0.0668	47	122.1667	-1.5638	82	122.3392	-0.0862	117	122.4037	0.7853
13	122.3652	0.2435	48	122.2253	-1.1705	83	122.3359	-0.1256	118	122.3015	-0.5051
14	122.3768	0.4010	49	122.0287	-2.2509	84	122.3889	0.5710	119	122.3699	0.3066
15	122.4193	1.0156	50	121.8087	-2.9821	85	122.3714	0.3269	120	122.3812	0.4621
16	122.3615	0.1946	51	122.1166	-1.8445	86	122.3619	0.1999	121	122.3441	-0.0266
17	122.4864	1.9640	52	122.3394	-0.0838	87	122.3976	0.6963	122	122.3676	0.2756
18	122.4966	2.0926	53	122.2386	-1.0688	88	122.302	-0.5000	123	122.3835	0.4944
19	122.4786	1.8622	54	122.2374	-1.0782	89	122.2951	-0.5696	124	122.1667	-1.5638
20	122.4776	1.8490	55	122.2403	-1.0554	90	122.3063	-0.4555	125	122.3718	0.3324
21	122.463	1.6503	56	122.2425	-1.0379	91	122.3858	0.5269	126	122.3057	-0.4618
22	122.5003	2.1380	57	122.241	-1.0498	92	122.3551	0.1115	127	122.379	0.4315
23	122.4157	0.9623	58	122.1517	-1.6525	93	122.3217	-0.2892	128	122.4114	0.8986
24	122.3662	0.2569	59	122.3873	0.5482	94	122.3156	-0.3564	129	122.4059	0.8176
25	122.3483	0.0254	60	122.3278	-0.2201	95	122.4551	1.5392	130	122.3957	0.6687
26	122.4109	0.8913	61	122.4171	0.9830	96	122.5388	2.5701			
27	122.3606	0.1828	62	122.327	-0.2293	97	122.3893	0.5767			
28	122.3851	0.5170	63	122.4891	1.9986	98	122.1071	-1.8932			
29	122.3936	0.6384	64	122.4153	0.9563	99	122.2768	-0.7445			
30	122.2941	-0.5795	65	122.3703	0.3120	100	122.3091	-0.4261			
31	122.3139	-0.3748	66	122.3203	-0.3048	101	122.2767	-0.7454			
32	122.2293	-1.1404	67	122.2082	-1.2939	102	122.2686	-0.8185			
33	122.2662	-0.8397	68	122.4147	0.9475	103	122.2956	-0.5647			
34	122.279	-0.7242	69	122.2862	-0.6564	104	122.4076	0.8426			
35	122.2475	-0.9977	70	122.4179	0.9949	105	122.4068	0.8308			

Na tabela 4D são apresentados os resíduos do modelo autoregressivo aplicado aos dados do processo 2 (tabela 1C do apêndice C) do capítulo 5.

$$x_t = 0,5015x_{t-1}$$

Tabela 4D - Resíduos do modelo autoregressivo do processo 2 do capítulo 5.

t	resíduos	t	resíduos	t	resíduos	t	resíduos
1		36	0.0106	71	0.8836	106	0.8416
2	-0.7272	37	0.5387	72	0.4039	107	0.6155
3	-0.5687	38	-0.2904	73	-0.8847	108	0.3684
4	1.6678	39	-0.2218	74	-0.9041	109	-0.9337
5	-0.1385	40	-1.1071	75	-0.4836	110	1.3600
6	-1.6464	41	0.3781	76	-0.5259	111	-1.1361
7	1.7150	42	-0.6442	77	0.4976	112	-0.8380
8	-0.9020	43	-0.3129	78	-0.1354	113	0.6049
9	0.0816	44	0.1304	79	0.1066	114	1.0341
10	-2.0078	45	0.1622	80	-0.0918	115	-0.8397
11	0.4230	46	-0.8418	81	0.6863	116	1.0779
12	0.2405	47	-1.1547	82	-0.4114	117	0.3228
13	0.2771	48	-0.3863	83	-0.0824	118	-0.8989
14	0.2788	49	-1.6640	84	0.6340	119	0.5599
15	0.8146	50	-1.8534	85	0.0406	120	0.3084
16	-0.3147	51	-0.3492	86	0.0359	121	-0.2583
17	1.8664	52	0.8412	87	0.5960	122	0.2890
18	1.1078	53	-1.0268	88	-0.8491	123	0.3562
19	0.8129	54	-0.5422	89	-0.3189	124	-1.8117
20	0.9152	55	-0.5147	90	-0.1699	125	1.1166
21	0.7232	56	-0.5087	91	0.7553	126	-0.6285
22	1.3104	57	-0.5294	92	-0.1527	127	0.6630
23	-0.1098	58	-1.1260	93	-0.3451	128	0.6823
24	-0.2256	59	1.3768	94	-0.2114	129	0.3670
25	-0.1034	60	-0.4950	95	1.7180	130	0.2587
26	0.8785	61	1.0934	96	1.7982		
27	-0.2641	62	-0.7222	97	-0.7121		
28	0.4253	63	2.1135	98	-2.1824		
29	0.3791	64	-0.0458	99	0.2049		
30	-0.8997	65	-0.1675	100	-0.0528		
31	-0.0842	66	-0.4612	101	-0.5317		
32	-0.9525	67	-1.1411	102	-0.4447		
33	-0.2678	68	1.5963	103	-0.1542		
34	-0.3031	69	-1.1315	104	1.1257		
35	-0.6345	70	1.3240	105	0.4083		

O teste *W* apresentou um $W = 0,9808$ e um $p = 0,5025$.