



Universidade de Brasília

IE – Departamento de Estatística

Estágio Supervisionado em Estatística 2

Lívia Rachel Sant' Anna Monteiro Rocha Lopes

**GRÁFICO DE CONTROLE ESTATÍSTICO DE QUALIDADE PARA
INDICADOR ESTRATÉGICO DA SECRETARIA DA FAZENDA DO
GOVERNO DO PIAUÍ**

Brasília - DF

2013

Lívia Rachel Sant' Anna Monteiro Rocha Lopes

**GRÁFICOS DE CONTROLE ESTATÍSTICO DE QUALIDADE PARA
INDICADOR ESTRATÉGICO DA SECRETARIA DA FAZENDA DO
GOVERNO DO PIAUÍ**

Relatório apresentado à disciplina Estágio Supervisionado 2 do curso de graduação em Estatística, Instituto de Ciências Exatas, Universidade de Brasília, como parte dos requisitos necessários para o grau de Bacharel em Estatística.

Orientador: Prof. Dr. Gladston Luiz da Silva

Brasília - DF

2013

Lívia Rachel Sant' Anna Monteiro Rocha Lopes

**GRÁFICOS DE CONTROLE ESTATÍSTICO DE QUALIDADE PARA
INDICADOR ESTRATÉGICO DA SECRETARIA DA FAZENDA DO
GOVERNO DO PIAUÍ**

A Comissão Examinadora, abaixo identificada, aprova o Trabalho de
Conclusão de Curso de Estatística da Universidade de Brasília da aluna

Lívia Rachel Sant' Anna Monteiro Rocha Lopes

Dr. Gladston Luiz da Silva
Professor-Orientador

Msc. Luís Gustavo do Amaral Vinha
Professor-Examinador

PhD, Antônio Eduardo Gomes
Professor-Examinador

Brasília, 17 de dezembro de 2013.

"Effective SPC is 10% statistics and 90% management action."

John Hradesky

Agradecimentos

Ao Senhor Jesus Cristo, autor da minha fé, por ter guiado os meus passos e por nunca ter me abandonado nesta caminhada;

Aos meus pais Ida e William por todo o suporte, incentivos e preocupações que me foram dados durante toda a minha vida. Aos meus irmãos William Daniel e Luísa Cristina por sempre ficarem ao meu lado;

Ao meu marido Arthur, por todo o apoio e incentivo que não me deixaram desistir, e que mesmo distante, sempre esteve ao meu lado;

Às minhas amigas de curso Bruna e Tatiana, que durante (muitos) momentos de desespero eram capazes de me animar, pelas horas de estudos e momentos que passamos juntas;

Ao meu orientador Prof., Dr. Gladston por todas as respostas dadas aos meus questionamentos, disposição a me ajudar nesta caminhada e por toda orientação para a execução deste trabalho;

À Maria José por fornecer toda a base de dados, por ter paciência comigo durante todo o período de conclusão deste trabalho respondendo a todos os meus questionamentos;

E a todos que de forma direta ou indiretamente contribuíram com este estudo.

Obrigada.

Resumo

A Secretaria da Fazenda do Estado do Piauí, objeto deste estudo, é uma entidade pública que prima por “adotar uma gestão orientada para resultados, com foco na melhoria dos serviços aos cidadãos”. Uma das ferramentas bastante utilizadas no meio gerencial são as Cartas de Controle, mais especificamente os Gráficos de Controle Estatístico introduzidos por Shewhart. Assim, este trabalho teve como objetivo a proposição de gráficos de controle estatístico de qualidade para um dos indicadores estratégicos da Secretaria da Fazenda do Estado do Piauí. Devido à natureza da série apresentar tendência e sazonalidade, foi necessária a aplicação de modelagem de séries temporais ao indicador escolhido, O modelo ARIMA foi escolhido para este fim, com o efeito de obterem-se os resíduos da série e assim aplicar a metodologia de gráficos de controle Médias Móveis Exponencialmente Ponderadas aos resíduos.

Palavras-chaves: Cartas de Controle, CEP, Controle de Qualidade, CEQ, Autoregressivo Integrado e de Médias Móveis, ARIMA, Médias Móveis Exponencialmente Ponderadas, MMEP.

Abstract

The Treasury Office of the State of Piauí, the study object, is a public unit that cares about “adopt a management guided to the results, focus on the improvement of the citizens services”. One tool very used in the management field is the Charts Control, more specifically the Statistics Control Graphics introduced by Shewhart. Based on that, this work had like its goal the proposition of the standard statistic control graphics for one of the strategic index of Treasury Office of the State of Piauí. Due to the serie nature shows trends and seasonal, it was necessary to apply the time series model to the chosen index. The ARIMA model was chosen to this end, to get the serie residuals and this way to apply the Exponentially Weighted Moving Average, a control graphics methodology, to the residuals.

Key words: Statistician Quality Control, Chart Control, Auto-Regressive Integrated Moving Averages, Exponentially Weighted Moving Average.

Sumário

Lista Símbolos	ix
Lista de Abreviaturas e Siglas	xi
Lista de Ilustrações e Tabelas	xii
Capítulo 1	1
1.1 Introdução	1
1.2 Justificativa	2
1.3 Objetivos	4
1.3.1 Objetivo Geral	4
1.3.2 Objetivos Específicos	4
1.4 Metodologia	5
Capítulo 2	6
2.1 Análise de Séries Temporais	6
2.1.1 Funções de Autocorrelação	7
2.1.2 Modelos ARIMA	8
2.1.3 Escolha do Modelo	11
2.1.4 Diagnóstico de modelo	12
2.2 Gráficos de Controle	13
2.2.1 Gráficos de Shewhart	14
2.2.2 Somas Acumuladas	18
2.2.3 Combinado entre Shewhart e Somas Acumuladas	21
2.2.3.1 Eficiência Relativa	21
2.2.4 Médias Móveis Exponencialmente Ponderadas	23
Capítulo 3	25
3.1 Resultados e Discussões	25
3.1.1 Reconhecimento Base de Dados	25
3.1.2 Construção dos Indicadores	26
3.1.3 Problemas encontrados durante a execução do estudo	27
3.1.4 Análise descritiva dos dados e Inferências	28
3.1.5 Análise de Séries Temporais	31
3.1.6 Construção dos gráficos de controle	37
3.1.7 Implantação dos gráficos de controle	38
3.2 Conclusões	39

Referências Bibliográficas.....	41
Apêndice.....	44
1.Os Quatorze Pontos do Método Deming.....	44
2. Teste de Shapiro-Wilk.....	45
3. Teste de Kolmogorov-Smirnov.....	45
4. Programação Utilizada.....	47
Anexo	48
1. Tabela da distribuição X^2	48

Lista Símbolos

- A_2 - Coeficiente para a construção de gráficos de Shewhart a_t - Ruído Branco
- B - Operador de translação para o passado
- B_3 - Coeficiente para a construção de gráficos de Shewhart
- B_4 - Coeficiente para a construção de gráficos de Shewhart
- C_i - Soma Acumulada incluindo a i-ésima amostra
- C_i^+ - Soma acumulada dos valores acima do pretendido
- C_i^- - Soma acumulada dos valores abaixo do pretendido
- C_0 - Soma acumulada (valor inicial)
- c_4 - Valor dependente de n para a construção de gráficos de Shewhart
- d - Número de diferenciações realizadas nos modelos ARIMA
- d_2 - Fator de correção para a estatística \bar{R}
- d_3 - Fator de correção para a estatística \bar{R}
- D_3 - Coeficiente para a construção de gráficos de Shewhart
- D_4 - Coeficiente para a construção de gráficos de Shewhart
- F - Operador de translação para o futuro
- H - Intervalo de decisão
- K - Valor de referência
- k - Número de defasagens
- L - Distância em função de desvios padrões dos limites de controle para a linha central
- n - Tamanho da amostra
- \bar{n} - Tamanho médio da amostra
- p - Ordem do modelo AR
- Q - Estatística do teste de Ljung-Box
- q - Ordem do modelo MA

\hat{r} - Autocorrelação estimada

R - Amplitude de um processo

\bar{R} - Amplitude média

S - Desvio padrão amostral

\bar{S} - Média dos desvios padrões amostral

S_t - Componente da sazonalidade da série temporal

t - Tempo de observação de uma série

T_t - Componente da tendência de uma série temporal

X_i - i-ésima observação

\bar{X} - Média das observações

$\bar{\bar{X}}$ - Média das médias as observações

Z_t - Nível de uma série temporal

$\gamma_{t,s}$ - Covariância entre as variáveis Z_t e Z_s

Δ - Desvio do valor nominal

δ - Tamanho da mudança da média em unidades de desvio padrão

θ - Parâmetro do modelo de um processo MA (q)

λ - Fator de suavização da média móvel

μ_0 - valor médio desejado do processo, média sob controle

μ_1 - valor médio do processo para o estado fora controle

ρ - Correlação entre variáveis

σ - Desvio padrão da população

$\hat{\sigma}$ - Desvio padrão estimado

Φ - Parâmetro do modelo de um processo AR (p)

φ_{kk} - Coeficiente das equações de Yule-Walker

Lista de Abreviaturas e Siglas

AIC - Critério de Informação de Akaike

AICc - Critério de Informação de Akaike corrigido

AR - Autoregressivo

ARMA - Autoregressivo e de Médias Móveis

ARIMA - Autoregressivo Integrado e de Médias Móveis ou *Auto-Regressive Integrated Moving Averages*

BIC - Critério de Informação Bayesiano

CEP - Controle Estatístico de Processos

CMS - Comprimento Médio da Sequência

CMS₀ - CMS para o desvio (δ) nulo

CUSUM - Somas Acumuladas ou *Cumulative Sum*

FAC - Função de Autocorrelação

FACP - Função de Autocorrelação Parcial

LC - Linha Central

LIC - Limite Inferior de Controle

LSC - Limite Superior de Controle

MA - Médias Móveis

MMEP - Médias Móveis Exponencialmente Ponderadas ou *Exponentially Weighted Moving Average*

SEFAZ/PI - Secretaria da Fazenda do Governo do Piauí

Lista de Ilustrações e Tabelas

Figura 1: Fluxograma das possíveis relações entre as situações fiscais e cadastrais	27
Gráfico 1: Comportamento do Indicador $P_{irregularidades}$ entre 2008 e 2013	29
Gráfico 2: Comportamento do indicador $P_{irregularidades}$ em cada mês entre 2008 e 2013	30
Gráfico 3: Função de Autocorrelação da série do indicador $P_{irregularidades}$	31
Gráfico 4: Função de Autocorrelação Parcial da série do indicador $P_{irregularidades}$	32
Gráfico 5: Comportamento da série ajustada pelo modelo ARIMA(1,1,0) - 2008 e 2013	33
Gráfico 6: Comportamento dos resíduos provenientes do ajuste do modelo - 2008 e 2013	34
Gráfico 7: Comparação entre a série original e o modelo ajustado à série do indicador $P_{irregularidades}$ entre os anos de 2008 e 2013	34
Gráfico 8: Histograma dos resíduos provenientes do ajuste do modelo de série temporal ao indicador $P_{irregularidades}$	36
Gráfico 9: QQPlot dos resíduos provenientes do ajuste do modelo de série temporal ao indicador $P_{irregularidades}$	36
Gráfico 10: Gráfico de Controle MMEP para os resíduos ($\lambda=0,1$ e $L=2,703$)	37
Gráfico 11: Gráfico de Controle MMEP para os resíduos($\lambda=0,2$ e $L=2,86$)	37
Tabela 1: Situações cadastrais possíveis utilizadas no estudo.....	25
Tabela 3: Situação Cadastral versus Situação Fiscal	27
Tabela 4: Estimativa de cada ano do Indicador $P_{irregularidades}$	30
Tabela 5: Valores dos critérios de informação para os modelos gerados para a série do indicador $P_{irregularidades}$	32
Tabela 6: Parâmetros estimados do modelo ARIMA(1,1,0) para os resíduos	33
Tabela 7: Estatística descritiva para os resíduos provenientes da aplicação da metodologia ARIMA(1,1,0) ao indicador $P_{irregularidades}$	35

Tabela 8: Resultado do teste de Ljung-Box para o diagnóstico de modelo	35
Tabela 9: Resultado do teste de Shapiro-Wilk e Kolmogorov-Smirnov para a normalidade dos resíduos provenientes da aplicação da metodologia ARIMA(1,1,0) ao indicador <i>P_{irregularidades}</i>	35

Capítulo 1

1.1 Introdução

O controle estatístico da qualidade teve seu início na década de 30, quando Dr. W. A. Shewhart desenvolveu o que é chamado hoje de gráfico de controle. Tal metodologia foi desenvolvida para verificar variações em processos produtivos da Bell Telephones. Por meio desse gráfico foi possível determinar o momento em que uma correção deveria ser aplicada ao processo de produção.

Conhecendo os estudos de Shewhart, Edwards W. Deming, um dos maiores nomes no que diz respeito à qualidade, defendia que para uma organização manter a ênfase na qualidade era necessário que os gestores mantivessem de forma contínua uma gestão de alto nível. Deming enumerou 14 passos (vide apêndice) para que os gestores de uma instituição alcançassem a qualidade.

Seguindo fundamentalmente a mesma linha de pensamento, surge o nome de Kaoru Ishikawa. Ele absorveu a linha de pensamento de Deming e expandiu o conceito para estratégia japonesa de qualidade. Ishikawa contribuiu nessa temática criando um diagrama de fácil uso que pode ser usada para analisar e resolver problemas. Essa ferramenta, mais tarde ficou conhecida como o Diagrama de Ishikawa. Para Kaoru Ishikawa

“Praticar um bom controle de qualidade é desenvolver, projetar, produzir e comercializar um produto de qualidade que é mais econômico, mais útil e sempre satisfatório para o consumidor.”

(Ishikawa, 1915)

1.2 Justificativa

Baseado neste breve histórico, Montgomery e Runger (2009) observaram que atualmente uma das preocupações das empresas tem sido em relação à melhoria e o monitoramento da qualidade. Assim, Montgomery (2009) destacou algumas das razões pelas quais os gráficos de controle tornaram-se tão comuns ao meio organizacional:

- Os gráficos de controle são métodos comprovados de melhoria da produtividade;
- São eficientes na prevenção de defeitos;
- Evitam ajustes desnecessários nos processos;
- Fornecem informações confiáveis para diagnóstico sobre o desempenho e para a capacidade dos processos.

Com a popularização do controle estatístico da qualidade as cartas de controle foram incluídas ao planejamento estratégico. Desta forma, segundo Florac(1999) “(...) empresas teriam uma ferramenta capaz de entender melhor os problemas e capacidades de suas organizações, prever tendências, antecipar eventuais problemas e finalmente melhorar o processo de desenvolvimento de um produto ou serviço”.

O planejamento estratégico “(...) tem como preocupação essencial sistematizar o processo de escolha de estratégias, que são expressas nas diretrizes estratégicas. Sob esse aspecto, os planos constituem a base para a formulação, implementação, acompanhamento e controle das ações estratégicas” (Planejamento Estratégico-SEFAZ/PI, 2012).

Com efeito, o Controle Estatístico da Qualidade é uma ferramenta capaz de estabilizar, controlar e melhorar um processo ao longo do tempo. Aliada a esta ferramenta pode-se utilizar técnicas de séries temporais a um processo, uma vez que um CEP é uma série temporal.

Por meio dos procedimentos de séries temporais é possível fazer previsões sobre o comportamento futuro da série podendo-se assim verificar se haverá a necessidade de aplicar melhorias ao processo. Indo de acordo a ideia de Florac(1999) sobre a importância de um controle estatístico eficaz a utilização do

artifício de séries temporais também se torna muito útil num planejamento estratégico.

Este trabalho consistirá na proposição de ferramentas de Controle Estatístico de Processo para indicadores relacionados no planejamento estratégico da Secretaria da Fazenda do Estado do Piauí – SEFAZ/PI, cujo produto final é a monografia referente ao Estágio Supervisionado em Estatística 2 do Departamento de Estatística da UnB.

1.3 Objetivos

1.3.1 Objetivo Geral

Este estudo tem como objetivo geral a proposição de ferramentas de Controle Estatístico da Qualidade para o indicador relativo à porcentagem de contribuintes em situação irregular, considerado estratégico para a Secretaria de Fazenda do Piauí - SEFAZ/PI.

1.3.2 Objetivos Específicos

Neste sentido, este estudo tem os seguintes objetivos específicos a serem alcançados:

- i. Selecionar, dentre os indicadores relacionados no Planejamento Estratégico da SEFAZ/PI, o que será objeto deste trabalho.
- ii. Ajustar um modelo de série temporal ao indicador selecionado.
- iii. Construir o gráfico de controle para o monitoramento dos resíduos do modelo ajustado ao indicador selecionado.
- iv. Parametrizar o gráfico de controle a partir do banco de dados fornecido pela SEFAZ/PI.
- v. Propor as etapas a serem cumpridas pela SEFAZ/PI para a implantação dos gráficos de controle propostos neste trabalho.

1.4 Metodologia

O início desse estudo dar-se-á com a revisão bibliográfica, cujo objetivo é fazer um levantamento de textos relacionados à temática do Controle Estatístico da Qualidade, mais especificamente sobre os gráficos de Shewhart, gráficos de somas acumuladas, gráfico de médias móveis e algoritmos para o cálculo da eficiência entre esses métodos. Também serão levantados textos a respeito de Planejamento Estratégico com o objetivo de uma melhor contextualização do meio em que os gráficos de controle serão aplicados. Por meio deste levantamento será possível identificar metodologias que poderão ser utilizadas neste estudo.

Em seguida, será selecionado um indicador da Secretaria da Fazenda do estado do Piauí. Após esta definição, ocorrerá a proposição do gráfico de controle. Para a elaboração dos gráficos do primeiro tipo, será utilizada a metodologia padrão 3σ . Para o segundo caso serão adotadas as sugestões de Montgomery (2009).

Com a definição dos gráficos de controle, o próximo passo consistirá em definir os parâmetros dos gráficos de controle do tipo Shewhart, CUSUM ou MMEP. Como se trata de um estudo piloto, os valores da média e da variância são desconhecidos e portanto serão estimados, conforme proposto por Montgomery (2009).

Para fins de verificação da eficiência na detecção de perturbações na média será criado um estudo comparativo entre os gráficos de controle de Shewhart e CUSUM, caso seja possível. E para obter um bom desempenho será calculado o Comprimento Médio da Sequência (CMS), onde o gráfico mais eficiente será aquele que possuir o menor CMS.

Cumpridos estes passos, o estudo será concluído por meio da proposição de procedimentos para implantação do controle da qualidade na SEFAZ/PI, com o objetivo oferecer uma ferramenta apoio para a implantação do planejamento.

No capítulo seguinte é composto de um referencial teórico capaz de auxiliar a construção deste estudo. Em cada seção serão descritas quais metodologias poderão ser empregadas ao estudo com o objetivo final de construir gráficos de controle eficazes para a base de dados da SEFAZ/PI. Contudo, não necessariamente todos os métodos descritos no próximo capítulo serão aplicados a este estudo.

Capítulo 2

2.1 Análise de Séries Temporais

Segundo Ehlers (2009), uma série temporal é um conjunto de observações coletadas ao longo de um determinado tempo. Estas observações podem ter caráter discreto ou contínuo, possuindo como característica principal o fato de que observações vizinhas são dependentes.

Desta forma, seja $\{X(t), t \in T\}$ uma série temporal, onde T é o tempo de observação da série $T = t_1, t_2, \dots, t_n$. Uma série temporal é dita contínua quando as observações são feitas continuamente ao longo de um tempo T , onde $T = t_1 < t_2 < \dots < t_n$. E uma série temporal é dita discreta quando os espaços entre as observações são os mesmos e/ou são específicos.

Morettin e Tolo (2006) definem que a análise de uma série temporal tem como objetivos: "(...) descobrir qual foi o mecanismo que gerou esta série, fazer previsões de valores futuros, descobrir o comportamento da série e procurar periodicidade relevante nos dados. Além disto, a análise também tem como objetivo fazer o controle de processos, onde os valores da série temporal medem a qualidade de um processo de manufatura e o objetivo é o controle do processo".

Existem dois tipos de séries temporais: estacionárias e não-estacionárias. Uma série temporal $\{X(t), t \in T\}$ dita estacionária de segunda ordem (que é a mais utilizada) ou fracamente estacionária possui as seguintes características:

- $E[X(t)] = \mu$
- $Cov[X(t), X(t + \tau)]$ depende unicamente da diferença entre os tempos, $(t_2 - t_1)$ e onde $\tau \in \mathbb{Z}$.

Assim, séries estacionárias desenvolvem-se aleatoriamente ao redor de uma média constante, apresentando certo equilíbrio. Entretanto é mais comum que as séries que não sejam estacionárias possuindo assim, tendências e sazonalidades. Desta forma, um processo $X(t)$ pode ser escrito como:

$$X_t = T_t + S_t + a_t, \quad (1)$$

onde a_t é um Ruído Branco, uma componente aleatória que possui distribuição $RB \sim (0, \sigma_a^2)$, T_t é a componente da tendência e S_t é a componente da sazonalidade.

Entretanto, as componentes de T_t e S_t são bastante correlacionadas e, segundo Pierce (1979), os métodos de estimação de S_t podem ser afetados caso seja ignorada a tendência e a especificação de S_t depende da especificação de T_t . Portanto, não se pode isolar uma componente da outra.

Uma solução para transformar a série em estacionária é a transformação dos dados. Este procedimento acarreta numa estabilização da variância e o efeito sazonal torna-se aditivo. A transformação mais comum é de diferenças sucessivas da série original da seguinte maneira:

$$\Delta X_t = X_t - X_{t-1}, \quad (2)$$

é a primeira diferença. O cálculo seguinte é dado por:

$$\Delta^2 X_t = \Delta[\Delta X_t] = \Delta X_t = \Delta[X_t - X_{t-1}] \quad (3)$$

resultando na seguinte expressão:

$$\Delta^2 X_t = X_t - 2X_{t-1} + X_{t-2} \quad (4)$$

e realizando n diferenças sucessivas, obtém-se uma fórmula geral:

$$\Delta^n = \Delta[\Delta^{n-1} X_t] \quad (5)$$

Às vezes, nos casos de séries econômicas e financeiras, antes de realizar esse tipo transformação é necessário aplicar uma transformação não-linear à série original. Geralmente utiliza-se a transformação de Box-Cox (1964) dada por:

$$X_t^\lambda = \begin{cases} \frac{X_t^\lambda - c}{\lambda}, & \text{se } \lambda \neq 0 \\ \log X_t, & \text{se } \lambda = 0 \end{cases}, \quad (6)$$

onde c e λ são parâmetros a serem estimados. Tal transformação é adequada quando o desvio padrão da série for proporcional à média.

2.1.1 Funções de Autocorrelação

2.1.1.1 Função de Autocorrelação

A função de autocorrelação verifica quão correlacionadas as variáveis estão, em um dado instante, consigo mesma, em um instante de tempo posterior. Isto é, a FAC fornece uma medida útil do grau de dependência entre os valores de uma série temporal em diferentes períodos.

Desta forma, a FAC pode ser definida como a razão entre a autocovariância e a variância de uma sequência de dados. Assim, a FAC é dada por:

$$\rho_{t,s} = \text{Corr}(Z_t, Z_s) = \frac{\gamma_{t,s}}{\sqrt{\gamma_{t,t}\gamma_{s,s}}} \quad (7)$$

onde $\gamma_{t,s} = \text{Cov}(Z_t, Z_s)$, $\gamma_{t,t} = \text{Var}(Z_t)$ e $\gamma_{s,s} = \text{Var}(Z_s)$. A principal utilidade da FAC é identificar se o modelo em estudo, pois ao gerar-se o correlograma do modelo o número de *Lag* será igual à ordem do tipo Média Móvel.

2.1.1.2 Função de Autocorrelação Parcial

A FACP é definida como a sequência de correlações entre $(Z_t$ e $Z_{t-1})$, $(Z_t$ e $Z_{t-2})$, ... , desde que os efeitos de defasagens k anteriores sobre Z_t permaneçam constantes. Assim, ela torna-se bastante útil no momento de decidir-se a ordem p de um processo autoregressivo de um modelo ARIMA.

Desta forma, a FACP é denotada por φ_{kk} , onde φ_{kk} é valor do coeficiente das equações de Yule-Walker.

$$\varphi_{kk} = \frac{\begin{vmatrix} 1 & \rho_1 & \dots & \rho_{k-2} & \rho_1 \\ \rho_1 & 1 & \dots & \rho_{k-3} & \rho_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \dots & \rho_1 & \rho_k \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \dots & \rho_{k-2} & \rho_{k-1} \\ \rho_1 & 1 & \dots & \rho_{k-3} & \rho_{k-2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \dots & \rho_1 & \rho_{t,s} \end{vmatrix}} \quad (8)$$

2.1.2 Modelos ARIMA

Uma das vantagens de ter-se uma série temporal é poder fazer previsões sobre o comportamento da série h passos à frente. Para que esta ferramenta seja utilizada, segundo Makridakis et al. (1998), três condições devem ser atendidas:

1. Possuir informações passadas sobre a série;
2. Capacidade de interpretação das previsões apresentadas;
3. Supor que o padrão da série seja o mesmo.

Os principais modelos capazes de fazer previsões sobre uma série temporal são os modelos de Box-Jenkins (ARIMA), os modelos de decomposição clássica e os modelos de suavizações exponenciais (simples, Holt e Holt-Winters).

Os modelos ARIMA em especial consiste em ajustar modelos autoregressivos integrados de médias móveis a uma amostra do tipo

ARIMA(p,d,q), onde p é a ordem modelo auto-regressivo, d é o número de diferenciações e q é a ordem do modelo de média móvel.

Para a seleção do modelo ARIMA mais adequado segundo Morettin e Tolo (2006) se utiliza os seguintes passos:

1. Selecionar um modelo geral para análise;
2. Identificar o modelo mais específico observando-se as autocorrelações, autocorrelações parciais e outras informações;
3. Estimar os parâmetros do modelo;
4. Diagnosticar o modelo ajustado.

Caso o modelo não esteja ajustado adequadamente, repete-se o ciclo. Com isso, nos tópicos seguintes serão levantados alguns modelos possíveis. Para isto, abaixo seguem algumas informações que serão utilizadas nos modelos em questão:

- Operador de translação para o passado (B)

$$BX_t = X_{t-1}, B^m X_t = X_{t-m} \quad (9)$$

- Operador de translação para o futuro (F)

$$FX_t = X_{t+1}, F^m X_t = X_{t+m} \quad (10)$$

2.1.2.1 Modelos Autoregressivos

Seja $\{X(t), t \in T\}$ uma série temporal, um processo autoregressivo de ordem p . O processo é descrito como $X_t \sim AR(p)$ e segue a seguinte lei de formação:

$$X_t = \phi_0 + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + a_t \quad (11)$$

onde $\phi_0, \phi_1, \phi_2, \dots, \phi_p$ são os parâmetros do modelo e a_t é o ruído branco associado. Desta forma, pode-se definir o operador autoregressivo $\phi(B)$ como:

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 + \dots + \phi_p B^p \quad (12)$$

ou ainda

$$\phi(B)\check{X}_t = a_t \quad (13)$$

Uma característica de um processo $AR(p)$ é que a sua função de autocorrelação é infinita em extensão que decai sob o formato de senóides amortecidas e/ou exponencialmente. E a função de autocorrelação parcial apresenta um corte no *Lag* p .

2.1.2.2 Modelos Médias Móveis

Seja $\{X(t), t \in T\}$ um processo de médias móveis de ordem q é denotado por $X_t \sim MA(q)$ possui a seguinte lei de formação:

$$X_t = \mu + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q} \quad (14)$$

onde μ e $\theta_1, \theta_2, \dots, \theta_p$ são constantes. Escrevendo em função do operador de médias móveis de ordem q obtém-se:

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q \quad (15)$$

ou também

$$\check{X}_t = \theta(B)a_t \quad (16)$$

A FAC de um processo $MA(q)$ é finita, apresentando um corte no *Lag* q . E a FACP possui comportamento semelhante à FAC de um processo $AR(p)$, decaimento exponencial e/ou senóide amortecida.

2.1.2.3 Modelos Autoregressivo e de Médias Móveis

Este tipo de modelo geralmente é utilizado quando se tem o objetivo de uma modelagem com poucos parâmetros. Como o próprio nome sugere, ele é uma junção dos modelos anteriores na seguinte forma:

$$X_t = \phi_0 + \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q} \quad (17)$$

Assim, esse modelo é denotado por $ARMA(p, q)$ e possui o seguinte operador:

$$\phi(B)\check{X}_t = \theta(B)a_t \quad (18)$$

A FAC deste modelo é infinita, decaindo exponencialmente e/ou em forma de senóide apresentando um corte no *Lag* $q-p$. A FACP apresenta um comportamento como o de um processo $MA(q)$.

2.1.2.4 Modelos Autoregressivos, Integrados e de Média Móveis

Semelhante ao modelo $ARMA(p, q)$ este tipo de modelo une o processo $AR(p)$ ao $MA(q)$. Entretanto este modelo é indicado para séries não estacionárias possuindo assim a ordem d que refere-se à quantidade de diferenciações feitas até que a série fique estacionária. Desta forma, a lei de formação do operador é dada por:

$$\phi(B)\Delta^d\tilde{X}_t = \theta(B)a_t \quad (19)$$

E denotado por ARIMA(p,d,q), onde p é a ordem de $\phi(B)$ e q de $\theta(B)$.

2.1.3 Escolha do Modelo

Quando se faz um modelo de previsão deve-se ter a ciência de que o modelo gerado é teórico. Assim, qualquer modelo gerado com base numa série temporal trata-se de um modelo aproximado. Desta forma, existem infinitos modelos possíveis para uma série qualquer.

Para eleger qual dos modelos é aquele que mais se assemelha da realidade geralmente utiliza-se o Critério de Informação de Akaike (AIC), o Critério de Informação de Akaike corrigido (AICc) e o Critério de Informação Bayesiano (BIC). O objetivo aqui é escolher o critério de informação que possua menores valores.

Para a construção de todos os critérios utiliza-se a função de máxima verossimilhança do modelo como $L(\hat{\theta})$.

2.1.3.1 AIC

Este método foi proposto com Akaike em 1974 com o conceito de que será atribuído ao modelo uma pontuação baseando-se na adequação aos dados e na ordem do modelo. Entretanto em outros estudos prova-se que estes critério é assintoticamente eficiente, mas não é assintoticamente consistente. Sua estrutura é dada por

$$AIC = -2 \log L(\hat{\theta}) + 2(p) \quad (20)$$

onde p é o número de parâmetros que serão estimados no modelo. Entretanto, este método não é indicado para pequenas amostras. Assim, foi criada uma correção para esta limitação, o AICc.

2.1.3.2 AICc

Com base nos estudos de Akaike, o AIC sofreu uma derivação que gerou o AICc de forma que ele é possui um desempenho melhor que o AIC para pequenas amostras. Para este método utiliza-se a seguinte fórmula:

$$AICc = -2 \log L(\hat{\theta}) + 2(p) + 2 \frac{p(p+1)}{n-p-1} \quad (21)$$

2.1.3 BIC

O BIC foi proposto também por Akaike em 1978 como uma alternativa ao AIC e AICc, sob uma perspectiva Bayesiana, onde utiliza-se a probabilidade a posteriori. Diferentemente do AIC, este método parte do princípio que o modelo real tem dimensão infinita. Seu ajuste é dado por:

$$BIC = -2 \log f(x_n | \hat{\theta}) + p \log n \quad (22)$$

onde $f(x_n | \hat{\theta})$ é o modelo escolhido.

2.1.4 Diagnóstico de modelo

Após definir o modelo deve-se verificar se o mesmo é adequado ou não aos dados apresentados. Para isto utiliza-se comumente o teste de Ljung-Box para verificar a independência dos resíduos. Utiliza-se este teste pelo motivo de que o Ljung-Box apresenta melhores resultados do que o teste de Box-Pierce. Assim testa-se as seguintes hipóteses:

$$\begin{cases} H_0 = \text{Os resíduos não são correlacionados} \\ H_1 = \text{Os resíduos são correlacionados} \end{cases}$$

Desta forma, calcula-se as estimativas das autocorrelações por meio das fórmulas:

$$\hat{r}_k = \frac{\sum_{t=k+1}^n \hat{a}_t \hat{a}_{t-k}}{\sum_{t=1}^n \hat{a}_t^2} \quad (23)$$

E utiliza-se a seguinte estatística do teste:

$$Q(k) = n(n-2) \sum_{j=1}^K \frac{\hat{r}_j^2}{(n-j)} \quad (23)$$

onde k é número de defasagens tomadas e $Q(k) \sim \chi^2$ com $(K - p - q)$ graus de liberdade. Assim, a hipótese nula será rejeitada se $Q(k) > \chi_{1-\alpha, k-p-q}^2$.

2.2 Gráficos de Controle

O conceito de gráficos de controle proposto por Shewhart é simples e de fácil aplicação, voltado à atividade de inspeção, direcionado ao “chão de fábrica” (Montgomery, 2009), realizado por pessoas que diariamente monitoram e conhecem bem o processo.

Gráficos de Controle são ferramentas capazes de identificar se um processo está sob controle estatístico ou não, isento de perturbações alheias ao processo. Além disto, havendo sinais de presença perturbações no processo, medidas corretivas devem ser aplicadas com o objetivo de minimizar tais variações. Eles podem ser aplicados a quaisquer características do produto ou operações que a gestão acredita ser importante.

É importante destacar que quando um processo está sob controle as variáveis envolvidas no processo seguem uma distribuição Normal (μ_0, σ). Geralmente μ_0 e σ não são conhecidos e, portanto, devem ser estimados baseados numa amostra piloto tomadas em subgrupos racionais¹.

Em geral, gráficos de controle possuem a seguinte estrutura: a Linha Central (LC) que corresponde à média do processo, o Limite Superior de Controle (LSC) geralmente dado em função de σ e o Limite Inferior Controle (LIC) que também é dado em função de σ .

¹Definição introduzida por Shewhart teve como objetivo designar uma maneira de construir subgrupos com as observações coletadas. Cada subgrupo é uma amostra onde as variações possam ser atribuídas apenas a causas aleatórias; mas entre as amostras as variações sejam devidas a causas identificáveis, cuja presença seja suspeitada ou considerada possível. Para os subgrupos racionais é possível também calcular as estatísticas média, amplitude ou desvio padrão, de acordo com a necessidade do gráfico de controle.

Um modelo geral de gráficos de controle é: seja w uma estatística a qual se deseja observar. Desta forma, μ_w é a média de w e σ_w o desvio padrão. A LC, LSC e LIC são dadas por:

$$LSC = \mu_w + L\sigma_w \quad (25)$$

$$LSC = \mu_w \quad (26)$$

$$LIC = \mu_w - L\sigma_w \quad (27)$$

onde L é a distância (em função de σ) que LSC e LIC se encontra em relação à LC.

Assim, classifica-se, de uma maneira geral, que um processo está sob controle quando os pontos observados possuem um padrão aleatório e localizam-se dentro dos limites. Caso haja a presença de um dos pontos fora dos limites de controle há evidência de que o processo está fora de controle. Mais a frente serão definidos como os processos encontram-se fora de controle.

2.2.1 Gráficos de Shewhart

Os gráficos do tipo Shewhart são mais indicados quando se quer detectar grandes perturbações na média do processo, ou de aumentos significativos na variância ou da fração defeituosa. Devido à sua simplicidade, este tipo de gráfico alcançou grande sucesso dada a facilidade da regra de decisão. Shewhart definiu os passos que indicam se um processo está fora de controle:

1. Um ou mais pontos fora dos limites de controle 3σ .
2. Dois ou três pontos consecutivos fora dos limites de alerta 2σ .
3. Quatro ou cinco pontos consecutivos além dos limites 1σ .
4. Uma sequência de 8 pontos consecutivos de um mesmo lado da linha central.
5. Seis pontos em uma sequência sempre crescente ou decrescente.
6. Quinze pontos em sequência na zona C (tanto acima quanto abaixo da linha central)
7. Quatorze pontos em sequência alternadamente para cima e para baixo.

8. Oito pontos em sequência de ambos os lados da linha central com nenhum na zona C.
9. Um padrão não usual ou não aleatório nos dados
10. Um ou mais pontos pertos de um limite de alerta ou de controle.

Este tipo de gráfico se subdivide em dois tipos, que sequeem a mesma estrutura, alterando-se somente um dos parâmetros base.

2.2.1.1 Média e Amplitude

Os gráficos de controle para Média e para a Amplitude (\bar{x} e R) são usados para monitorar um processo onde a variável em interesse é expressa em uma escala contínua de medida e tem como objetivo controlar o valor médio da característica da qualidade e sua variabilidade.

O gráfico de \bar{X} representa o valor médio de um subgrupo e é utilizado visando controlar a média do processo, enquanto o gráfico de R controla a variação. O cálculo dos limites para estes pares de gráficos são explicados a seguir.

Seja X_1, X_2, \dots, X_n uma amostra aleatória de tamanho n . Desta forma, a média desta amostra é dada por:

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n} \quad (28)$$

Entretanto, para a metodologia de gráficos de controle não é utilizado apenas uma amostra, e sim um conjunto de m sub-amostras e dentro de cada uma há um conjunto com n observações. Então a média do processo é dada pela fórmula:

$$\bar{\bar{X}} = \frac{\bar{X}_1 + \bar{X}_2 + \dots + \bar{X}_m}{m} = \sum_{j=1}^m \frac{\bar{X}_j}{m} \quad (29)$$

No caso do gráfico de controle de \bar{X} a LC equivale à média do processo \bar{X} . Para o cálculo dos outros limites, deve-se ser levado em consideração o desvio padrão do processo.

Desta forma, seja R a amplitude do processo e W uma variável aleatória denominada amplitude relativa, onde $W = \frac{R}{\sigma}$ e a principal propriedade de W é possuir uma média constante igual a d_2 que depende do tamanho da amostra. Assim, um estimador não-viesado para σ é:

$$\hat{\sigma} = \frac{\bar{R}}{d_2} \quad (30)$$

onde \bar{R} é média da amplitude do processo.

$$\bar{R} = \frac{R_1 + R_2 + \dots + R_m}{m} \quad (31)$$

Assim, segundo Montgomery (2009), tem-se que os limites de controle para média são dados por:

$$LSC = \bar{X} + A_2 \bar{R} \quad (32)$$

$$LC = \bar{X} \quad (33)$$

$$LIC = \bar{X} - A_2 \bar{R}, \quad (34)$$

onde $A_2 = \frac{3}{d_2 \sqrt{n}}$ e d_2 é uma constante dita fator de correção cujo valor é tabelado em função do tamanho n de cada amostra.

E os limites de controle para a amplitude são:

$$LSC = D_4 \bar{R} \quad (35)$$

$$LC = \bar{R} \quad (36)$$

$$LIC = D_3 \bar{R} \quad (37)$$

onde $D_3 = 1 - 3 \frac{d_3}{d_2}$ e $D_4 = 1 + 3 \frac{d_3}{d_2}$ e d_3 é uma constante, assim como d_2 é um valor tabelado que depende do tamanho da amostra.

É importante frisar que os gráficos de média e amplitude devem ser utilizados simultaneamente, pois as funções se complementam, de forma que se um ponto ultrapassar as linhas de controle ou houver indício de não aleatoriedade, o processo como um todo estará fora de controle.

2.2.1.2 Média e Desvio Padrão

Este tipo de gráfico (\bar{X} e S) é bem semelhante ao anterior, onde a única diferença refere-se ao cálculo da estimativa de σ . Estima-se σ de forma direta, ou seja, através do cálculo do desvio padrão amostral, onde um estimador não viesado para σ é dado por:

$$\hat{\sigma} = \frac{S}{c_4} \quad (38)$$

E o valor de c_4 é encontrado baseado no seguinte cálculo:

$$c_4 = \left(\frac{2}{n-1} \right)^{\frac{1}{2}} \frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{(n-1)}{2}\right)} \quad (39)$$

Ainda seguindo a linha de estudo da última referência, tem-se os seguintes limites de controle para \bar{X} :

$$LSC = \bar{\bar{X}} + A_3 \bar{S} \quad (40)$$

$$LC = \bar{\bar{X}} \quad (41)$$

$$LIC = \bar{\bar{X}} - A_3 \bar{S} \quad (42)$$

E para \bar{S} :

$$LSC = B_4 \bar{S} \quad (43)$$

$$LC = \bar{S} \quad (44)$$

$$LIC = B_3 \bar{S} \quad (45)$$

onde $A_3 = \frac{3}{c_4 \sqrt{n}}$, $B_3 = 1 - \frac{3}{c_4} \sqrt{1 - c_4^2}$ e $B_4 = 1 + \frac{3}{c_4} \sqrt{1 - c_4^2}$, dependendo unicamente do valor do tamanho da amostra.

Os gráficos \bar{X} e S são mais indicados quando o tamanho da amostra n é grande ($n > 10$ ou 12) ou quando o tamanho da amostra é variável.

2.2.2 Somas Acumuladas

Gráficos de controle do tipo Somas Acumuladas (CUSUM) foram introduzidos por E.S. Page em 1954 como uma alternativa aos gráficos de controle propostos por Shewhart. Os gráficos do tipo CUSUM (*Cumulative Sum*) podem ser construídos tanto para observações individuais, como para as médias de subgrupos racionais.

Segundo Cruz et al.(2009) e Alves (2003) a principal vantagem das cartas de controle do tipo CUSUM em relação aos gráficos de Shewhart é que esses acumulam informações de toda a amostra ponderando-as igualmente, de forma que todas as amostras coletadas possuam o mesmo peso. Por este motivo essa ferramenta é mais eficiente em captar pequenas e persistentes perturbações na média de um processo.

Existem dois tipos de gráficos de controle do tipo Somas Acumuladas: Tabular e Máscara V. Montgomery (2009) cita alguns motivos para não se utilizar o este procedimento. Eles são:

- “A máscara V é bilateral. Para o monitoramento de processos unilaterais essa ferramenta não pode ser utilizada;
- A ferramenta Hedstart² não pode ser implementada;

²Também chamado de Resposta Inicial Rápida foi proposto por Lucas e Croiser (1982) com o objetivo de aumentar a sensibilidade do CUSUM no início do processo. Aumentando esta característica, o ajuste no processo será detectado mais rapidamente.

- Algumas vezes, é difícil determinar até onde se deve ir, para trás, com os braços da máscara V, de forma que a interpretação do procedimento torna-se difícil;
- É provável que o maior problema com a máscara V seja a ambiguidade associada aos parâmetros α e β presentes na fórmula.”

Com base nestas informações, para este trabalho será utilizada a ferramenta CUSUM Tabular. Então, seja C_i a soma acumulada após a i -ésima amostra, cujo valor será dado por:

$$C_i = \sum_{j=1}^i (\bar{X}_j - \mu_0) \quad (46)$$

onde μ_0 é o valor desejado para a média do processo e \bar{x}_j é a média da j -ésima amostra. É importante ressaltar que C_i acumula a informação dos valores anteriores.

Caso o processo esteja sob controle para o valor desejado μ_0 as C_i formam um processo aleatório com média zero. Para saber se o processo está sob controle ou não, utiliza-se o seguinte algoritmo:

$$C_i^+ = \max [0, \bar{X}_i - (\mu_0 + K) + C_{i-1}^+] \quad (47)$$

$$C_i^- = \max [0, (\mu_0 - K) - \bar{X}_i + C_{i-1}^-] \quad (48)$$

onde C_i^+ e C_i^- são estatísticas denominadas CUSUM unilaterais superior e inferior respectivamente. Para iniciar o algoritmo, utiliza-se $C_0^- = C_0^+ = 0$.

O parâmetro K é denominado valor de referência, ou valor de tolerância ou de folga. Corresponde a aproximadamente metade do valor que se tem interesse em detectar a mudança da média, determinado valores entre o valor pretendido μ_0 e o valor da média fora de controle μ_1 . Assim, o valor de K é dado por:

$$K = \frac{|\mu_1 - \mu_0|}{2} = \frac{\delta}{2} \sigma \text{ ou } K = k\sigma \quad (49)$$

Com a definição do algoritmo, deve-se agora determinar qual parâmetro utilizar para saber se o processo está fora de controle. Adota-se um valor H como intervalo de decisão, dado por:

$$LSC: H = h\sigma \quad (50)$$

$$LIC: H = -h\sigma \quad (51)$$

Assim, H é denominado o erro do tipo II, interpretar que o processo está sob controle sendo que o mesmo está fora controle, ou a probabilidade de não se detectar a variação média do processo.

Em geral, é recomendado que esses dos parâmetros k e h parâmetros sejam escolhidos com o objetivo de fornecer um bom desempenho do gráfico CUSUM. Para tanto será calculado o Comprimento Médio da Sequência (CMS) ou *Average Run Length* (ARL), que será utilizado e explicado mais a frente na metodologia.

Definido os parâmetros, o processo é considerado fora de controle quando o valor de C_i^+ ou C_i^- excede o valor de decisão do intervalo H .

É possível, também, gerar gráficos de controle CUSUM utilizando uma padronização da variável x_i , onde a nova variável em observação será:

$$z_i = \frac{x_i - \mu_0}{\sigma} \quad (52)$$

onde μ_0 é a média do processo e σ é o desvio padrão e a variável z_i segue uma distribuição Normal (0,1).

Segundo Montgomery (2009), ao realizar este cálculo observam-se duas vantagens: “Primeiro, muitos gráficos de CUSUM podem agora ter os mesmos valores de k e h , e as escolhas desses parâmetros não dependem de escala, isto é, não dependem de σ . Em segundo lugar, o CUSUM padronizado conduz, naturalmente, a um CUSUM para controle da variabilidade”.

Desta forma, os limites de controle para o CUSUM Bilateral Padronizado serão:

$$C_i^+ = \text{máx} [0, z_i - k + C_{i-1}^+] \quad (53)$$

$$C_i^- = \text{máx} [0, -k - z_i + C_{i-1}^-] \quad (54)$$

2.2.3 Combinado entre Shewhart e Somas Acumuladas

Alguns autores como Lucas (1976), Duncan (1986) e Hawkins e Olwell (1998), defendem que os gráficos CUSUM são mais eficientes que os gráficos de Shewhart, no que se refere a pequenas perturbações no processo. Entretanto, quando se trata de grandes variações a ferramenta CUSUM é menos eficiente.

Desta forma, foi desenvolvida outra metodologia capaz de aumentar a sensibilidade a variações, sejam elas grandes ou pequenas. Este procedimento, nomeado combinado CUSUM-Shewhart, foi citado à primeira vez no trabalho de Westgard et al. (1977). Tal ferramenta é uma modificação no CUSUM, onde são adicionados a este tipo de gráfico de controle os Limites de Controle dos gráficos de Shewhart.

Um gráfico combinado Shewhart-CUSUM incorpora num mesmo eixo os valores observados, os limites de controle Shewhart aproximadamente igual a 3.5σ , a estatística CUSUM e o limite de controle do CUSUM.

2.2.3.1 Eficiência Relativa

Para saber qual dos gráficos de controle deve ser implantado faz-se necessário um estudo comparativo entre os gráficos. Para tanto, será calculado o Comprimento Médio da Sequência (CMS) de todos os gráficos e, ao final, esses valores serão comparados par a par.

O CMS é o número médio de amostras necessárias para que seja detectada uma mudança. O Comprimento Médio da Sequência depende do tamanho de amostra n e da dimensão da mudança que se deseja detectar, isto é, número de desvios padrões que a média pode variar.

Caso o objetivo seja detectar mudanças de menor magnitude será necessário utilizar tamanhos de amostras maiores ($n \geq 15$). Caso contrário, basta

tomar o tamanho da amostra como 4 ou 5. Este parâmetro depende do tamanho da amostra n , do limite de decisão h e da dimensão da mudança.

O valor de CMS é dado aproximadamente pela seguinte fórmula:

$$CMS = \begin{cases} \frac{e^{-2\Delta b} - 2\Delta b - 1}{2\Delta^2} & , \quad \Delta \neq 0 \\ \frac{1}{b^2} & , \quad \Delta = 0 \end{cases} \quad (55)$$

e para $\Delta \neq 0$ tem-se $\Delta = \delta - k$ e $\delta = \frac{\mu_1 - \mu_0}{\sigma}$ é a mudança da média em unidades de desvio padrão para o qual o CMS é calculado. O valor da variável b é igual a $b = h + 1,166$. Os valores de k e h foram definidos como $h = 4,77$ e $k = 0,5$, pois são os valores mais utilizados.

Quando $\delta = 0$ tem-se o CMS_0 calculado pela equação acima. Da mesma forma acontece quando $\delta \neq 0$, tem-se o CMS correspondente ao tamanho da variação.

Como o CUSUM é uma estatística unilateral, deve ser calculado o CMS para C_i^+ e C_i^- , por meio da seguinte equação:

$$\frac{1}{CMS} = \frac{1}{CMS^+} + \frac{1}{CMS^-} \quad (56)$$

Porém, por simetria os valores de $CMS^+ = CMS^-$.

Comparando os CMS para Shewhart e CUSUM, o mais indicado será aquele que possuir um menor CMS, pois fornecerá de forma mais rápida sobre o momento em que é emitido um sinal de alerta de que o processo estaria saindo de controle.

Definindo o referencial teórico até aqui, no capítulo seguinte serão apresentados os resultados obtidos através das aplicações das metodologias descritas. E também serão levantadas possíveis discussões para os resultados apresentados juntamente com as suas soluções plausíveis para implantação dos gráficos de controle.

2.2.4 Médias Móveis Exponencialmente Ponderadas

O gráfico de Médias Móveis Exponencialmente Ponderadas foi introduzido por Roberts em 1959 como uma alternativa aos gráficos de controle existentes.

Assim como os gráficos de controle do tipo CUSUM os gráficos de MMEP são eficientes relativamente aos gráficos de Shewhart para detecção de pequenas perturbações na média. Apresenta desempenho equivalente ao CUSUM.

O grande diferencial deste gráfico é que o MMEP é robusto a dados que não apresentam distribuição Normal. Isto faz com que seja indicado a observações individuais e que rejeitam a hipótese de normalidade.

Para a construção deste gráfico inicialmente deve-se fazer uma “transformação” dos dados por meio da seguinte fórmula:

$$z_i = \lambda x_i + (1 - \lambda)z_{i-1} \quad (57)$$

onde $0 < \lambda \leq 1$ é uma constante e o valor inicial do processo é o alvo, onde $z_0 = \mu_0$. Outra maneira de calcular z_i seria:

$$z_i = \lambda \sum_{j=0}^{i-1} (1 - \lambda)^j x_{i-j} + (1 - \lambda)^i z_0 \quad (58)$$

É importante frisar que a variável que aparecerá na carta de controle será z_i . Para o cálculo da variância de z_i utiliza-se a fórmula abaixo, baseando-se no desvio padrão σ das observações e supondo-se que as observações são variáveis aleatórias independentes:

$$\sigma_{z_i}^2 = \sigma^2 \left(\frac{\lambda}{2-\lambda} \right) [1 - (1 - \lambda)^{2i}] \quad (59)$$

As fórmulas apresentadas são utilizadas para amostras de tamanho $n=1$. Entretanto, caso o estudo trate de subgrupos racionais de tamanho $n>1$ devem ser feitas as seguintes substituições: x_i por \bar{x}_i e σ por $\sigma_{\bar{x}_i} = \frac{\sigma}{\sqrt{n}}$

Quando o valor alvo do processo μ_0 é desconhecido, adota-se como alvo o valor de \bar{x}_i , onde está é a média amostral da variável estudada.

Assim, as linhas de controle e a linha central são calculadas da seguinte maneira:

$$LSC = \mu_0 + L\sigma\sqrt{\frac{\lambda}{2-\lambda}[1 - (1 - \lambda)^{2i}]} \quad (60)$$

$$LC = \mu_0 \quad (61)$$

$$LIC = \mu_0 - L\sigma\sqrt{\frac{\lambda}{2-\lambda}[1 - (1 - \lambda)^{2i}]} \quad (62)$$

onde L é distância em função do desvio padrão que separa a linha central dos limites de controle.

Para a seleção de λ e L deve-se definir qual CMS deseja-se obter. Entretanto, é importante selecionar estes parâmetros de forma que o CMS possa ser comparado com o desempenho dos gráficos de controle de Shewhart e CUSUM, com o objetivo de se verificar pequenas perturbações na média do processo. Assim, é comum adotar-se valores de λ dentro do intervalo $0,05 \leq \lambda \leq 0,25$, onde valores menores de λ acarretam em detecções de menores mudanças. É comum utilizar-se $L=3$.

Capítulo 3

3.1 Resultados e Discussões

3.1.1 Reconhecimento Base de Dados

Os dados utilizados neste estudo foram obtidos da SEFAZ/PI. Trata-se de dados atualizados diariamente em um computador da Secretaria. Ao final de cada dia é realizada uma contabilização da quantidade de contribuintes pertencentes ao estado e da quantidade de contribuintes irregulares.

O *status* dos contribuintes são apresentados abaixo na Tabela 1, bem como uma breve explicação sobre cada uma das situações.

Tabela 1: Situações cadastrais possíveis utilizadas no estudo

Situação Cadastral	Descrição
Ativo	Situação de cadastro normalizado
Suspenso	Suspende as atividades a pedido do contribuinte
Baixado	O contribuinte solicita baixar sua inscrição estadual, finalizando suas atividades
Em Processo de Baixa	Situação anterior à Suspenso até a análise final do pedido do contribuinte
Em Processo de Suspensão	Situação anterior à de Baixa até a análise final do pedido do contribuinte
Cancelado	A SEFAZ atribui esta situação devido à irregularidade cometida pelo contribuinte
Baixado de Ofício	A SEFAZ faz a baixa da inscrição estadual devido à irregularidade cometida pelo contribuinte

A situação de irregularidade é do tipo fiscal, de forma que o contribuinte entra para esta classe após 90 dias sem quitar uma dívida. É importante ressaltar que este estudo não levou em consideração o tipo de irregularidade que o contribuinte se encontra. Também é importante frisar que um contribuinte em situação irregular pode possuir mais de uma irregularidade.

Para a construção da base de dados utilizada neste estudo foram consideradas as seguintes informações:

Tabela 2: Variáveis extraídas do servidor da SEFAZ/PI para a construção de um indicador

Campo	Descrição
Data	Dia da observação
Total_Contribuintes	Número total de contribuintes no dia
Total_irregulares_dia	Número total de contribuintes irregulares do dia
Quant_irregularidades_entra	Número de irregularidades que entraram no dia
Quant_Irregularidades_Sai	Número de irregularidades que saíram no dia
Quant_Irregularidades_Acumulado	Número de Irregularidades acumuladas no dia

3.1.2 Construção dos Indicadores

Para a elaboração deste estudo foi selecionado o indicador relacionado a irregularidades. Trata-se de um indicador diário, obtido da divisão do total de contribuintes irregulares no dia pelo total de contribuintes, que resulta na proporção diária de contribuintes irregulares do estado do Piauí, uma estimativa da fração de contribuintes irregularidades.

$$\hat{P}_{irregularidades} = \frac{\text{Total de Contribuinte Irregulares}}{\text{Total de Contribuintes}}$$

É importante destacar que este indicador é estratégico para a SEFAZ/PI. Com sua construção, a fração de contribuintes irregulares será monitorada de forma direta e indicará o momento em que uma ação fiscal deverá ser realizada.

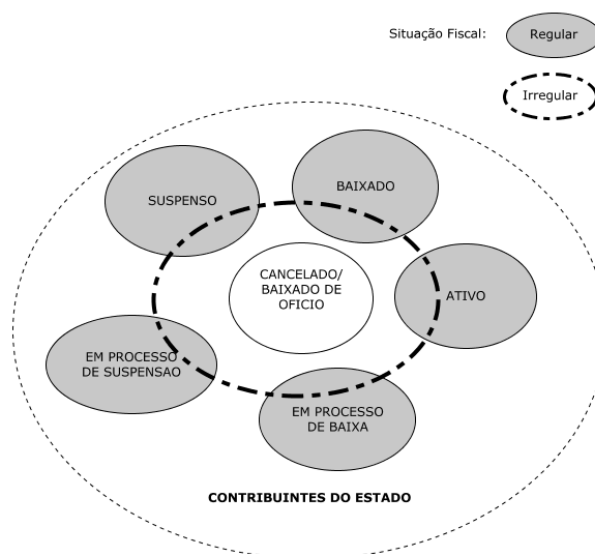
Para fins de classificação do contribuinte, é necessário considerar as seguintes possibilidades:

Tabela 3: Situação Cadastral versus Situação Fiscal

Situação Cadastral	Situação Fiscal
Ativo	Regular ou Irregular
Suspenso	Regular ou Irregular
Baixado	Regular ou Irregular
Em Processo De Baixa	Regular ou Irregular
Em Processo De Suspensão	Regular ou Irregular
Cancelado	Apenas Irregular
Baixado De Oficio	Apenas Irregular

Para uma melhor compreensão das possíveis relações entre as situações fiscais e cadastrais, segue abaixo um fluxograma.

Figura 1: Fluxograma das possíveis relações entre as situações fiscais e cadastrais



3.1.3 Problemas encontrados durante a execução do estudo

A base de dados utilizada é alimentada diariamente, manualmente e mecanicamente, representando que não existe um padrão de entrada dos dados na base. Há dias em que apenas uma irregularidade entra no sistema, em outros mais de 5 mil irregularidades são registrados no sistema. A não padronização da

digitação de novos registros no sistema pode gerar inconsistências na base de dados, tais como a criação de uma sazonalidade inexistente na inclusão de novos contribuintes irregulares.

Outro problema encontrado foi o fato de que não havia um estudo prévio sobre esta base de dados. Definido o indicador a ser utilizado neste trabalho, foi necessária a realização de um estudo para identificação das variáveis da base que deveriam compor o indicador. Para tanto, a base de dados deste trabalho foi adequada diversas vezes.

Em virtude de não haver um estudo prévio da base de dados pela SEFAZ/PI, o comportamento do indicador era totalmente desconhecido. Por este motivo a metodologia deste estudo foi revista algumas vezes, no sentido de adequar-se à realidade da base de dados.

Assim como a base de dados, as variáveis que compõem o indicador foram alteradas. De princípio, se levou em consideração apenas as irregularidades lançadas no dia. Entretanto, devido à forma como são computados os dados no sistema, um contribuinte que esteja em situação irregular pode não necessariamente ser lançado no sistema no momento em que torna-se irregular. Este problema faz com que existam picos frequentes durante o mês, visto que grande quantidade de irregularidades são lançadas em determinado dia, quando de fato foram geradas em outros dias. Outro problema é a inexistência de algumas datas por motivos desconhecidos.

Uma forma encontrada para corrigir estes problemas foi trabalhar com o total consolidado de irregularidades ao final de cada mês. Assim o indicador que seria diário, torna-se mensal, possuindo a estrutura descrita abaixo:

$$\hat{P}_{irregularidades} = \frac{\text{Total de Contribuinte Irregulares no final do mês}}{\text{Total de Contribuintes no final do mês}}$$

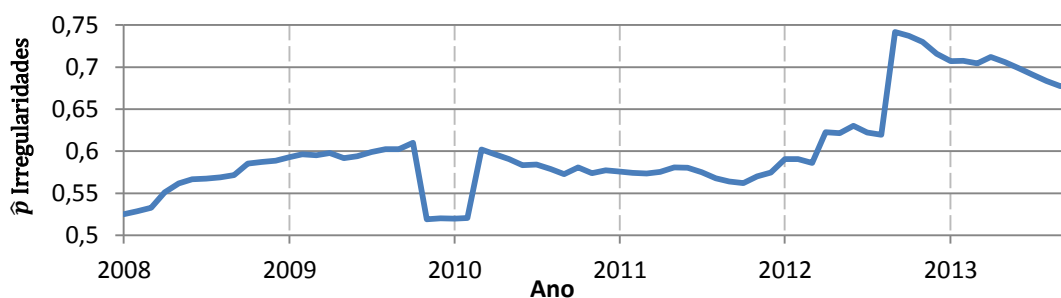
3.1.4 Análise descritiva dos dados e Inferências

A base de dados é constituída por registros referentes ao período entre janeiro de 2008 e outubro de 2013. Contém 2.094 registros que foram lançados no sistema diariamente, independentemente do dia observado ser útil ou não. É

importante dizer que qualquer alteração no total de contribuintes só ocorre em dias úteis, enquanto o número de irregularidade oscila independentemente do dia.

O passo seguinte consistiu na criação do indicador. Após seu cálculo, foi gerado um gráfico que informa qual é o comportamento do indicador ao longo do período.

Gráfico 1: Comportamento do Indicador $P_{irregularidades}$ entre 2008 e 2013



Este sistema foi implantado em Janeiro de 2008, quando foram lançadas no sistema informações a respeito da situação fiscal dos contribuintes. Na nova base dados deste estudo o número de observações foi alterado para 70 devido à mudança na estrutura do indicador.

Pelo Gráfico 1, observa-se que em dois períodos o indicador apresenta comportamentos diferenciados decorrentes de ações fiscais realizadas. É importante destacar que o último pico deu-se em decorrência do acréscimo de um novo tipo de irregularidade no sistema. Isto fez com que o nível da série alterasse bastante, aumentando o número de irregularidades no sistema.

Pela Tabela 4 percebe-se que pouco menos da metade dos contribuintes do estado do Piauí está em situação irregular. Ainda, pode-se comparar a média do ano de 2008 com a média de 2012 (pois o ano de 2013 encontra-se incompleto) e assim pode-se observar um aumento considerável no valor médio de irregularidades por ano.

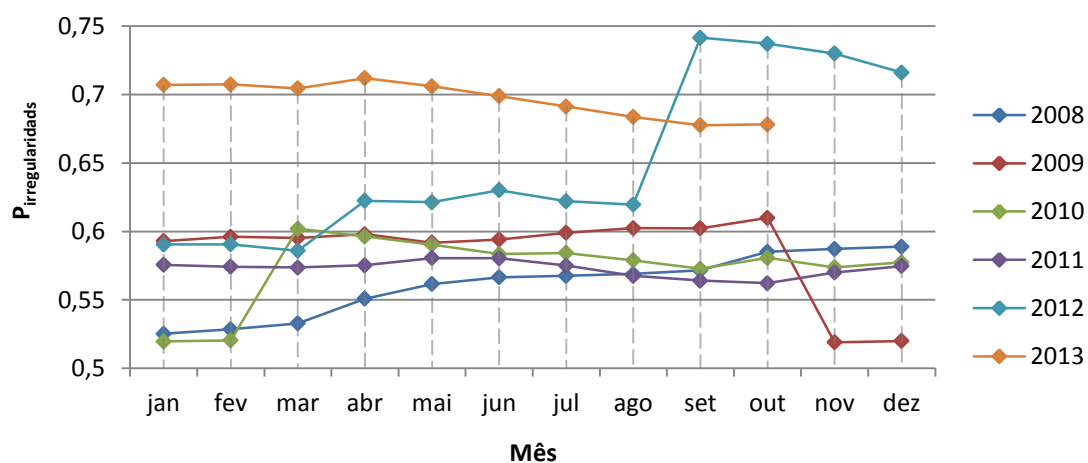
Tabela 4: Estimativa de cada ano do Indicador $P_{irregularidades}$

Ano	Estatística			
	Mínimo	Máximo	Desvio Padrão	Média
2008	0,5328	0,5961	0,0187	0,5725
2009	0,5190	0,6099	0,0310	0,5851
2010	0,5196	0,6019	0,0264	0,5733
2011	0,5622	0,5805	0,0058	0,5728
2012	0,5860	0,7416	0,0615	0,6506
2013	0,6776	0,7120	0,0130	0,6967
Todos	0,5190	0,7416	0,0576	0,6040

Com o aumento do número de irregularidades acumuladas acredita-se que a arrecadação da SEFAZ/PI está menor do que deveria ser, pois mais pessoas estão devendo à receita a cada ano que passa. Este é um dos motivos que leva à entidade realizar ações de fiscais, o aumento do número de regularidades de contribuintes.

Pelo Gráfico 2 é possível ver claramente a variação do indicador ao longo de cada ano. É fácil ver o efeito da inclusão de um novo tipo de irregularidade no comportamento do indicador ocorrido entre agosto e setembro de 2012.

Gráfico 2: Comportamento do indicador $P_{irregularidades}$ em cada mês entre 2008 e 2013

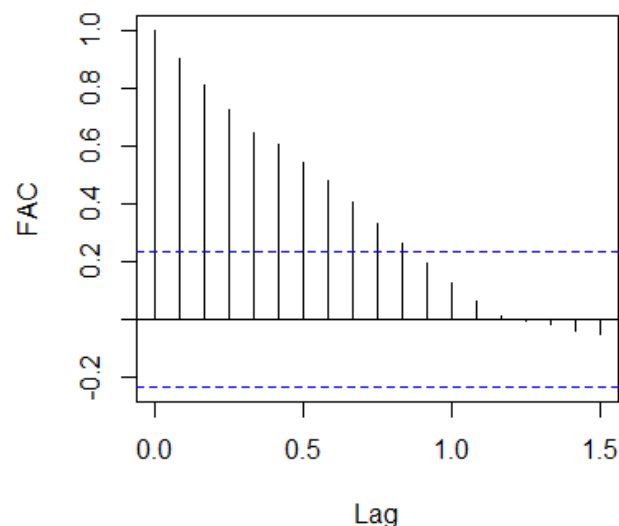


3.1.5 Análise de Séries Temporais

Durante a execução dos passos contidos nesta seção, foi utilizado o software R Development Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org/>.

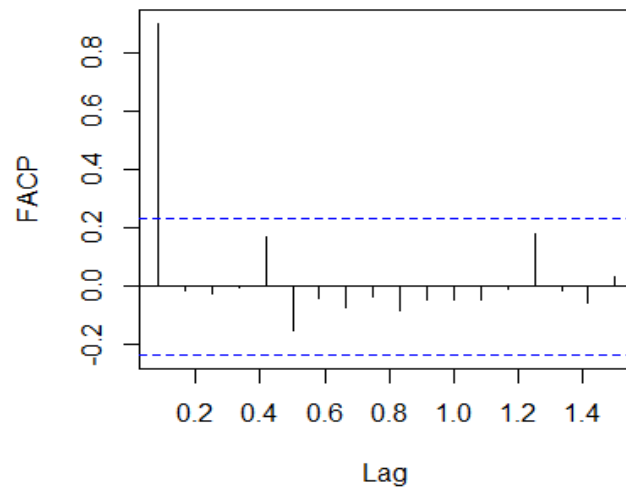
Para dar início ao estudo da série temporal, foram gerados dois gráficos: o primeiro refere-se à Função de Autocorrelação (FAC) e o segundo à Função de Autocorrelação Parcial (FACP), com o objetivo de indicar qual o modelo melhor para o indicador $P_{irregularidades}$.

Gráfico 3: Função de Autocorrelação da série do indicador $P_{irregularidades}$



Pelo gráfico da FAC é possível observar o decaimento dos $Lags$. Este comportamento indica que um provável modelo para o indicador $P_{irregularidades}$ seria um AR(1). Esta hipótese fica mais robusta ao analisar-se o Gráfico 4 da FACP, onde ocorre um corte no primeiro Lag .

Gráfico 4: Função de Autocorrelação Parcial da série do indicador $P_{irregularidades}$



A partir destes resultados foram gerados alguns modelos possíveis para o indicador $P_{irregularidades}$. Na Tabela 5 pode-se observar os valores do AIC, AICc e BIC para cada um dos modelos.

Tabela 5: Valores dos critérios de informação para os modelos gerados para a série do indicador $P_{irregularidades}$

Modelo	AIC	AICc	BIC
AR(1)	-329,34	-328,97	-322,59
ARMA(1,1)	-327,34	-326,72	-318,34
ARIMA(1,1,0)	-326,66	-326,48	-322,19
MA(1)	-262,67	-262,3	-255,92
ARIMA(0,1,1)	-326,67	-326,49	-322,2

Deseja-se um modelo que apresente os menores valores para o critério de informação. Entretanto os valores do AIC, AICc e BIC são muito semelhantes. Assim, devido a não estacionariedade da série e como deseja-se o modelo mais simples possível, a modelagem mais indicada é o ARIMA(1,1,0), que apresenta os seguintes parâmetros:

Tabela 6: Parâmetros estimados do modelo ARIMA(1,1,0) para os resíduos

	ARIMA(1,1,0)
Parâmetro	-0,0390
Erro padrão	0,1195

E, segundo a equação (19) possui lei de formação:

$$X_t = -0,0390X_{t-1} + a_t$$

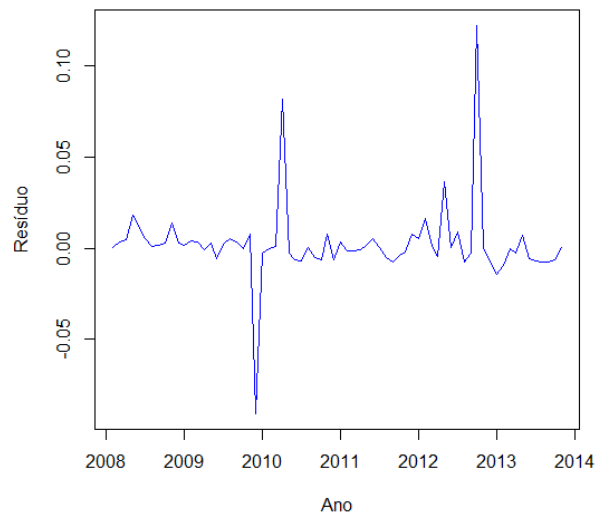
No Gráfico 5 pode-se analisar como a série ajustada se comportou no intervalo de tempo em estudo, bem como observar que o picos apresentados assemelham-se aos reais.

Gráfico 5: Comportamento da série ajustada pelo modelo ARIMA(1,1,0) - 2008 e 2013



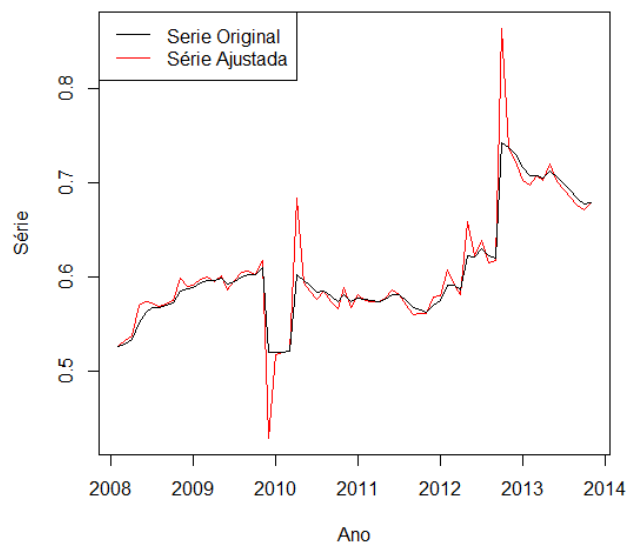
Por trata-se de um modelo teórico, naturalmente existe um resíduo entre o valor observado do indicador e o valor esperado. Os resíduos são apresentado no Gráfico 6:

Gráfico 6: Comportamento dos resíduos provenientes do ajuste do modelo - 2008 e 2013



Assim pode-se analisar que o modelo aparenta estar bem ajustada à série, uma vez que os resíduos, em sua maioria estão bem próximos de zero. Com o efeito de verificar o ajuste do modelo, gerou-se um gráfico onde se encontra a série original e o modelo.

Gráfico 7: Comparação entre a série original e o modelo ajustado à série do indicador *Pirregularidades* entre os anos de 2008 e 2013



Por meio do Gráfico 8 é possível analisar que a exceção dos dias de ação fiscal, a série ajustou-se bem aos dados apresentados.

Dando continuidade ao estudo dos resíduos, torna-se importante realizar uma análise exploratória dos dados para entender melhor o comportamento dos resíduos acima.

Tabela 7: Estatística descritiva para os resíduos provenientes da aplicação da metodologia ARIMA(1,1,0) ao indicador *Pirregularidades*

Mínimo	Máximo	Média	Desvio Padrão
-0,09057	0,12190	0,02277	0,021917

Para os resíduos calculados foram feitos testes de independência e de normalidade com objetivo final de definir qual gráfico de controle seria aplicado. Os testes aplicados foram o teste de Shapiro-Wilk e Kolmogorov-Smirnov, cujos resultados são apresentados a seguir.

A Tabela 8 apresenta os resultados do teste de Ljung-Box, que indica não haver evidência estatística para a rejeição da hipótese de que os resíduos são não correlacionados, para $\alpha=0,05$.

Tabela 8: Resultado do teste de Ljung-Box para o diagnóstico de modelo

Teste	Estatística do Teste	P-valor
Ljung-Box	0,0126	0,9105

Contudo, os resultados apresentados na Tabela 9 indicam que os resíduos não possuem uma distribuição Normal, uma vez que os valores do P-valor para os testes de Shapiro-Wilk e Kolmogorov-Smirnov foram muito inferiores ao nível de significância de $\alpha=0,05$.

Tabela 9: Resultado do teste de Shapiro-Wilk e Kolmogorov-Smirnov para a normalidade dos resíduos provenientes da aplicação da metodologia ARIMA(1,1,0) ao indicador *Pirregularidades*

Teste	Estatística do teste	P-valor
Shapiro-Wilk (W)	0,5429	1,98E-13
Kolmogorov-Smirnov (D)	0,4857	7,49E-08

A hipótese de não normalidade residual é confirmada ao analisar-se o histograma dos resíduos e o qqplot, apresentados nos Gráfico 8 e 9.

Gráfico 8: Histograma dos resíduos provenientes do ajuste do modelo de série temporal ao indicador $P_{irregularidades}$

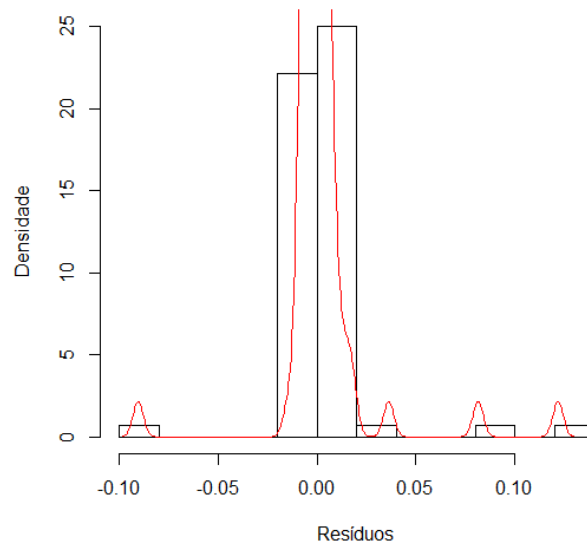
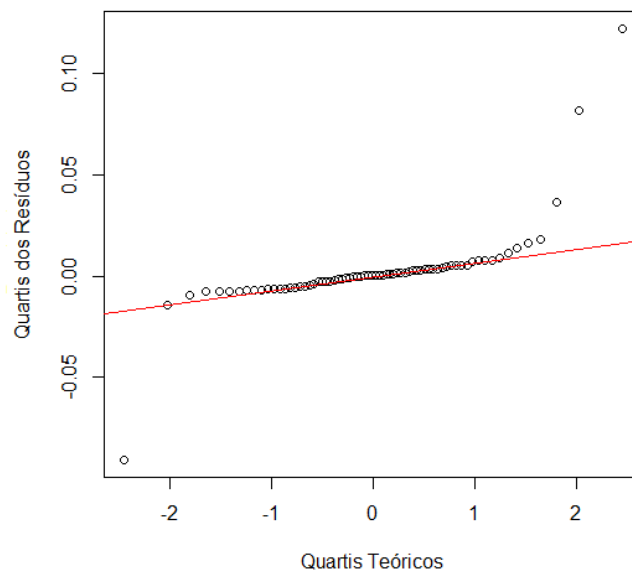


Gráfico 9: QQPlot dos resíduos provenientes do ajuste do modelo de série temporal ao indicador $P_{irregularidades}$



Pelo histograma é possível analisar-se que a curva não se assemelha à curva da Normal. E no Gráfico 9 observa-se que a linha da Normal encontra-se abaixo da “linha” dos quartis da amostra. Desta forma, o gráfico indicado para os resíduos do

modelo ajustado é o gráfico de controle MMEP, pois ser um gráfico livre de distribuição.

3.1.6 Construção dos gráficos de controle

Uma vez feita a escolha pelo gráfico de controle Médias Móveis Exponencialmente Ponderadas, o próximo passo consistiu em parametriza-lo, para receber os resíduos do modelo ajustado. Para isto foram consideradas as recomendações de Montgomery (2009), que sugere: $\lambda = 0,1$ e $L=2,703$, ou $\lambda = 0,2$ e $L=2,86$. Para a construção dos limites de controle foram utilizadas as fórmulas (60), (61) e (62).

Gráfico 10: Gráfico de Controle MMEP para os resíduos
($\lambda=0,1$ e $L=2,703$)

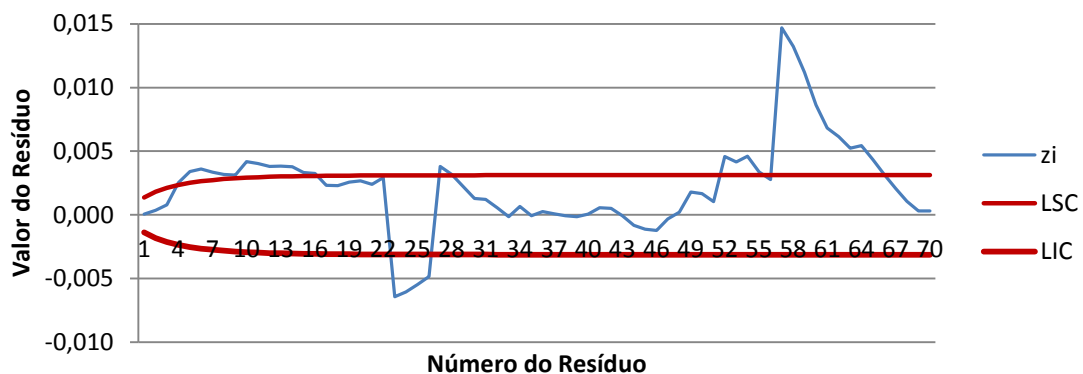
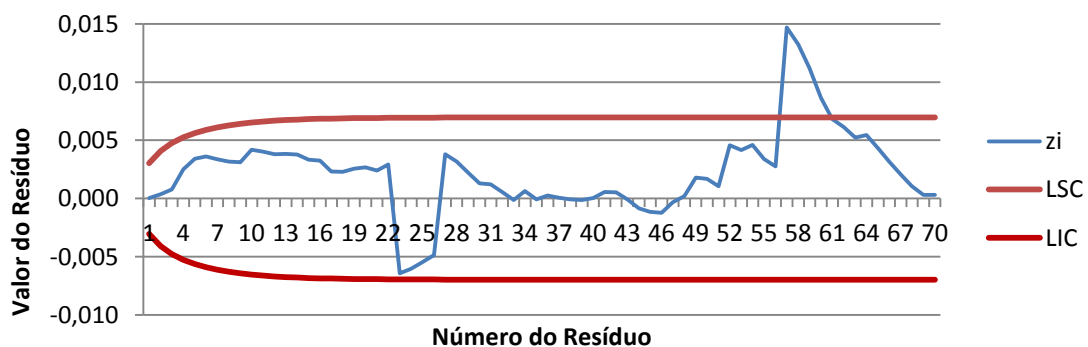


Gráfico 11: Gráfico de Controle MMEP para os resíduos
($\lambda=0,2$ e $L=2,86$)



Pelos gráficos acima, verifica-se que ambos os modelos a maioria dos pontos estão dentro de controle para os valores de λ estipulados, destacando-se apenas os dois picos em ambos os gráficos. Tais picos resultaram das ações fiscais realizadas pela SEFAZ/PI, conforme mencionado anteriormente. Pode ser observado nos dois gráficos à medida que o valor de λ aumentou, os limites de controle tornaram-se menos restritivos.

O Gráfico 10 apresentou mais pontos fora de controle do que o segundo tipo. Como não houve ações fiscais nestes dois dias seria interessante verificar o porquê de o gráfico ter saído de controle, uma vez que houve uma perturbação na média por motivos desconhecidos até então.

Outro ponto que deve ser considerado na implantação do gráfico refere-se aos limites do gráfico MMEP, que podem ser mais restritivos ou não em virtude da variação de λ .

3.1.7 Implantação dos gráficos de controle

Por não possuir ainda uma cultura de gráficos de controle na SEFAZ/PI o gráfico mais indicado é aquele que apresenta a priori menos perturbações na média. Desta forma, o modelo indicado para a implantação é o gráfico que possui os parâmetros são $\lambda = 0,2$ e $L=2,86$.

Entretanto, quando a SEFAZ/PI já estiver familiarizada com a cultura de controle e a importância dos gráficos, o modelo indicado é o primeiro, pois ele mostrou-se mais sensível a variações na média do processo.

Vale lembrar que segundo Montgomery (2009), tais parâmetros devem ser escolhidos com o objetivo de obter CMS_0 próximos aos CMS de Shewhart.

Foi observado neste estudo que toda vez que uma ação fiscal for capaz de mudar o patamar do indicador *Pirregularidades*, deve ser feita uma revisão dos parâmetros do gráfico de controle.

3.2 Conclusões

Este estudo foi realizado no contexto do Estágio Supervisionado do curso de Bacharelado em Estatística da Universidade de Brasília. Apresentou como objeto de trabalho o uso de ferramentas de Controle Estatístico da Qualidade para um indicador considerado estratégico para a Secretaria de Fazenda do Piauí - SEFAZ/PI.

No primeiro momento, foi necessário definir o indicador que seria objeto deste trabalho, o que foi possível após estudo do Planejamento Estratégico da SEFAZ/PI e de um estudo exaustivo da base de dados.

Tal estudo foi importante para a detecção de inconsistências na base de dados, que serviram para a identificação e escolha dos gráficos de controle propostos para o indicador selecionado.

É importante ressaltar que este estudo foi pioneiro para a SEFAZ/PI, o que explica, em parte, as inconsistências encontradas da base de dados.

Das análises descritivas realizadas inferenciais observou-se claramente a tendência de crescimento da série relativa ao indicador selecionado. Desta forma, a proposição da aplicação dos gráficos de controle diretamente ao indicador não seria possível, visto que para isso seria necessário que a série escolhida fosse estável. Com efeito, foi aplicada a metodologia de Box e Jenkins para série temporal como passo anterior à proposição dos gráficos de controle.

A aplicação da metodologia de séries temporais resultou em num vetor de resíduos, decorrente da diferença entre os valores resultantes do modelo e os valores observados da série. A adequação do modelo foi verificada por meio do teste de Ljung-Box, que revelou a independência dos resíduos.

Ainda sobre os resíduos, foram realizados os teste de Shapiro-Wilk e Kolmogorov-Smirnov, com o objetivo de verificar-se a existência de normalidade residual, visto que os gráficos de controle de Shewhart e CUSUM só podem ser aplicados caso a hipótese de normalidade atendida. Como isso foi verificado, optou-se pela adoção dos gráficos MMEP.

Assim, foram construídos dois gráficos de controle, um gerado pelos parâmetros $\lambda = 0,1$ e $L=2,703$ e o outro com $\lambda = 0,2$ e $L=2,86$. O interessante em gerar estes dois gráficos foi observar que os limites de controle entre os métodos possuem uma variação pequena, porém significativa ao gráfico devido à magnitude dos dados.

Entre os modelos testados, verificou-se que o gráfico mais indicado foi com $\lambda = 0,2$ e $L=2,86$ uma vez que a SEFAZ/PI não possui uma cultura de gráficos de controle nesta área. Com a criação desta cultura, o Gráfico 11 torna-se mais interessante uma vez que é mais sensível às perturbações.

A metodologia proposta neste trabalho pode ser aplicada a outros indicadores da SEFAZ/PI. Para tanto, basta que os parâmetros sejam selecionados adequadamente. Caso haja estabilidade no indicador, recomenda-se aplicar o gráfico de controle diretamente aos dados e não aos resíduos, por ser uma metodologia mais simples.

Referências Bibliográficas

ALVES, C. C.; et al. *A estatística Média Móvel Exponencialmente Ponderada para o controle preditivo, monitoramento e ajuste de processos*, Congresso Latino-Iberoamericano de Investigación Operativa. Rio de Janeiro. 2012

_____; et al. *A Utilização de Gráficos de Controle de Soma Acumulada (CUSUM) para Monitoramento de um Processo de Usinagem*, Cadernos IME – Série Estatística. Universidade do Rio de Janeiro – UERJ, v. 27, p 45 -48, 2009.

_____. *Gráficos de controle CUSUM: um enfoque dinâmico para a análise estatística de processos*. Florianópolis, 2003. 119 p. Dissertação (Mestrado em Engenharia de Produção) – Programa de Pós-Graduação em Engenharia de Produção – PPGEP, Universidade Federal de Santa Catarina, 2003.

BOX, G. E. P. & COX, D. R. *An analysis of transformations*, Journal of the Royal Statistical Society, Series B, 26, 211-252. 1964.

BOX, G.E.P. & Jenkins, *Time Series Analysis, Forecasting and Control*, San Francisco, Holden-Day .1970, (revised edition, 1976).

BRASIL. Governo do Estado do Piauí. Secretaria da Fazenda. *Planejamento Estratégico da Secretaria de Fazenda do Estado do Piauí – 2012-2015*. Teresina. Disponível

em:<<http://www.sefaz.pi.gov.br/arquivos/Relatorio%20Final%20Plano%20Estrategico%20SEFAZ-PI%202012-2015.pdf>> Acesso em 27/05/2013.

_____. Poder Judiciário. Justiça do Trabalho. Tribunal Regional do trabalho da 10ª Região. *1ª revisão técnica do Plano Estratégico do Tribunal Regional Trabalho - 10ª Região - 2013*. Brasília. Disponível em:<http://gestaoestrategica.trt10.jus.br/portal/images/stories/diest/livreto%20trt10%2017_05_2013.pdf> Acesso em 27/05/2013.

CRUZ , A. C. da.; et al.. *A aplicação de gráficos de controle de Soma Acumulada (CUSUM) para monitoramento de um processo de usinagem*. Bauru-SP: 16a

Simpósio de Engenharia de Produção, 2009.

EHLERS, R.S. (2009) *Análise de Séries Temporais*, <<http://www.icmc.usp.br/~ehlers/stemp/stemp.pdf>> Acessado em 18/06/2013.

DEMING, W. E. *Qualidade: A Revolução da Administração*. Rio de Janeiro: Marques Saraiva. 1990

DUNCAN, A. J. *Quality Control and Industrial Statistics*, 5th edition, Chicago: Richard D.Irwin, Inc.1986.

FACCIO, K.; WERNER, L. *Comparando gráfico CUSUM e Redes Neurais Artificiais no contexto do Controle Estatístico de Qualidade*. Revista INGEPRO. v. 02, n. 11. 2010.

FOLLADOR, F. A. C.; et al. *Controle de qualidade da água medido através de cartas de controle de Shewhart, CUSUM e MMEP*. Engenharia Ambiental (Online), v. 9, p. 183-197, 2012.

FLORAC, W.A. and CARLETON A.D. *Measuring the Software Process: Statistical Process Control for Software Process Improvement*, Addison-Wesley, 1999.

GOMES, Paulo J. *A evolução do conceito de qualidade: dos bens manufacturados aos serviços de informação*, Cadernos BAD. Lisboa: APBAD, v. 2, p. 6-18.2004.

GUJARATI, D.N. *Econometria Básica*. Rio de Janeiro. Editora Campus; 2000.

HAWKINS, D.M. and OLWELL, D.H. *Cumulative Sum Charts and Charting for Quality Improvement, Engineering and Physical Science*. Springer, New York. 1998.

ISHIKAWA, K. 1915. *Controle da qualidade total: à maneira japonesa*, tradução Iliana Torres, 2ª Ed., Rio de Janeiro: Campus. 2000.

LUCAS, J.M. *The Design and Use of V-Mask Control Schemes*. Journal of Quality Technology, 8(1) p.1-12.1976.

MAKRIDAKIS, S.; et al. *Forecasting Methods and Applications*. 3ª Ed. New York: John Wiley & Sons, 1998.

MONTGOMERY, Douglas C. *Introdução ao Controle Estatístico de Qualidade*, tradução Ana Maria Lima de Farias, Vera Regina Lima de Farias e Flores; revisão técnica Luiz da Costa Laurencel. 4ª Ed., Rio de Janeiro: LTC. 2009.

MONTGOMERY, D. C.; RUNGER, G. C. *Estatística Aplicada e Probabilidade Para Engenheiros*. 4ª Ed. Rio de Janeiro: LTC. 2009.

MORETTIN, P. A.; TOLOI C. M. de C. *Análise de Séries Temporais*. São Paulo: Edgard Blücher. 2006.

PIERCE, D. A. *Seasonal adjustment when both deterministic and stochastic seasonality are present*. In: Seasonal analysis of economic time series. Arnold Zellner, ed. Washington D.C., United States Department of Commerce, Bureau of the Census. 242-269p. 1979.

WESTGARD, J. et al. 1977. *Combined Shewhart-CUSUM Control Chart for Improved Quality Control in Clinical Chemistry*. Clinical Chemistry, v. 23, n. 10, p. 1881-1887.

Apêndice

1. Os Quatorze Pontos do Método Deming

1º: Estabeleça constância de finalidade para a melhoria do produto e do serviço.

2º: Adote a nova filosofia.

3º: Deixe de depender da inspeção para atingir a qualidade.

4º: Cesse a prática de avaliar transações apenas com base no preço.

5º: Melhore sempre e constantemente o sistema de produção e de prestação de serviços.

6º: Institua treinamento e retreinamento no local de trabalho.

7º: Institua liderança.

8º: Elimine o medo.

9º: Elimine as barreiras entre os departamentos.

10º: Elimine slogans, exortações e metas para a os empregados.

11º: Elimine padrões de trabalho (cotas numéricas) na linha de produção.

12º: Remova as barreiras que privam o operário horista de seu direito de orgulhar-se de seu desempenho.

13º: Institua um sólido programa de educação e retreinamento.

14º: Engaje todos da empresa no processo de realizar a transformação. A transformação é da competência de todo mundo.

2. Teste de Shapiro-Wilk

O teste foi proposto em 1965 e sua estatística do teste é:

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

onde b é uma constante dada por:

$$b = \begin{cases} \sum_{i=1}^{n/2} a_{n-i+1} (X_{(n-i+1)} - X_i) & \text{para n par} \\ \sum_{i=1}^{(n+1)/2} a_{n-i+1} (X_{(n-i+1)} - X_i) & \text{para n ímpar} \end{cases} \quad (2)$$

em que X_i são os valores obtidos na amostra ordenados do menor para o maior valor. E a_{n-i+1} são constantes calculadas da seguinte forma:

$$(a_1, a_2, \dots, a_n) = \frac{\mathbf{m}^T \mathbf{V}^{-1}}{(\mathbf{m}^T \mathbf{V}^{-1} \mathbf{V}^{-1} \mathbf{m})^{1/2}} \quad (3)$$

e m é uma matriz 1xn composta pelos valores esperados da estatística de ordem da amostra e V é matriz de covariância dessas estatísticas.

As hipóteses do teste são dadas por:

$$\begin{cases} H_0: a amostra segue uma distribuição Normal \\ H_1: a amostra não segue uma distribuição Normal \end{cases}$$

E a regra de decisão é: dado um nível de significância α , rejeitar H_0 se $W_{\text{calculado}} < W_{\text{tabelado}}$ ou caso o p-valor encontrado seja menor que o nível de significância.

3. Teste de Kolmogorov-Smirnov

O teste D de Kolmogorov-Smirnov é uma estatística do tipo Função de Distribuição Empírica (FDE) e baseia-se na comparação da curva da frequência cumulativa dos dados com a função de distribuição teórica em hipótese. O teste é mais indicado para variáveis contínuas. Assim o teste é dado pelas seguintes hipóteses:

$$\begin{cases} H_0: a amostra segue uma distribuição Normal \\ H_1: a amostra não segue uma distribuição Normal \end{cases}$$

Para o cálculo da estatística do teste, seja $F_n(x)$, onde $F_n(x)$ é uma função empírica cresce em $1/n$ a cada observação, é definida como:

$$F_n(x) = \begin{cases} 0, & x < X(1) \\ \frac{i}{n}, & X(i) \leq x \leq X(i+1) \\ 1, & X(n) \leq x \end{cases} \quad (4)$$

A estatística do teste é feita da seguinte forma:

$$D_n = \text{Sup}_x |F(x) - F_n(x)| \quad (5)$$

onde $F(x)$ é a função de distribuição acumulada. Como a função empírica é descontínua deve-se considerar também as estatísticas D :

$$D_+ = \text{Sup}_{x_i} |F(x_i) - F_n(x_i)| \quad (6)$$

$$D_- = \text{Sup}_{x_i} |F(x_i) - F_n(x_i)| \quad (7)$$

onde D_+ é maior distância vertical entre a distribuição empírica e a função de distribuição quando o FDE é maior que a função de distribuição, e D_- é a maior distância vertical quando o FDE for menor que a função de distribuição.

Desta forma, obtém-se a estatística do teste:

$$D_n = \max_x (D_+, D_-) \quad (8)$$

Como regra de decisão, deve-se rejeitar a hipótese nula caso D_n seja superior ao valor crítico.

4. Programação Utilizada

```
setwd("C:\\Users\\Arthur.Arthur-PC\\Documents\\Livia")

library(tseries)

dados.df<-read.csv("Base.csv")

names(dados.df)

#Cria a série temporal

serie<-ts(dados.df$Indicador,start=c(2008,2),frequency=12)

plot(serie)

acf(serie)

pacf(serie)

Ar <- arima(serie, order = c(1,1,0))

residuos<-residuals(Ar)

ajustada<-serie+residuos

plot(ajustada,col=2)

lines(serie)

legend('topleft', c("Série Original","Série Ajustada"), col=c(1,2), lty=1)

Box.test (residuos, type="Ljung")

hist(residuos, prob=TRUE, 12)

lines(density(residuos),col="Red")

qqnorm(residuos)

qqline(residuos,col="Red")

normal<-rnorm(70)

shapiro.test(residuos)

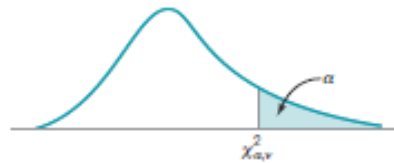
ks.test(residuos,normal)
```

Anexo

1. Tabela da distribuição χ^2

■ APPENDIX III

Percentage Points of the χ^2 Distribution^a



ν	α								
	0.995	0.990	0.975	0.950	0.500	0.050	0.025	0.010	0.005
1	0.00 +	0.00 +	0.00 +	0.00 +	0.45	3.84	5.02	6.63	7.88
2	0.01	0.02	0.05	0.10	1.39	5.99	7.38	9.21	10.60
3	0.07	0.11	0.22	0.35	2.37	7.81	9.35	11.34	12.84
4	0.21	0.30	0.48	0.71	3.36	9.49	11.14	13.28	14.86
5	0.41	0.55	0.83	1.15	4.35	11.07	12.38	15.09	16.75
6	0.68	0.87	1.24	1.64	5.35	12.59	14.45	16.81	18.55
7	0.99	1.24	1.69	2.17	6.35	14.07	16.01	18.48	20.28
8	1.34	1.65	2.18	2.73	7.34	15.51	17.53	20.09	21.96
9	1.73	2.09	2.70	3.33	8.34	16.92	19.02	21.67	23.59
10	2.16	2.56	3.25	3.94	9.34	18.31	20.48	23.21	25.19
11	2.60	3.05	3.82	4.57	10.34	19.68	21.92	24.72	26.76
12	3.07	3.57	4.40	5.23	11.34	21.03	23.34	26.22	28.30
13	3.57	4.11	5.01	5.89	12.34	22.36	24.74	27.69	29.82
14	4.07	4.66	5.63	6.57	13.34	23.68	26.12	29.14	31.32
15	4.60	5.23	6.27	7.26	14.34	25.00	27.49	30.58	32.80
16	5.14	5.81	6.91	7.96	15.34	26.30	28.85	32.00	34.27
17	5.70	6.41	7.56	8.67	16.34	27.59	30.19	33.41	35.72
18	6.26	7.01	8.23	9.39	17.34	28.87	31.53	34.81	37.16
19	6.884	7.63	8.91	10.12	18.34	30.14	32.85	36.19	38.58
20	7.43	8.26	9.59	10.85	19.34	31.41	34.17	37.57	40.00
25	10.52	11.52	13.12	14.61	24.34	37.65	40.65	44.31	46.93
30	13.79	14.95	16.79	18.49	29.34	43.77	46.98	50.89	53.67
40	20.71	22.16	24.43	26.51	39.34	55.76	59.34	63.69	66.77
50	27.99	29.71	32.36	34.76	49.33	67.50	71.42	76.15	79.49
60	35.53	37.48	40.48	43.19	59.33	79.08	83.30	88.38	91.95
70	43.28	45.44	48.76	51.74	69.33	90.53	95.02	100.42	104.22
80	51.17	53.54	57.15	60.39	79.33	101.88	106.63	112.33	116.32
90	59.20	61.75	65.65	69.13	89.33	113.14	118.14	124.12	128.30
100	67.33	70.06	74.22	77.93	99.33	124.34	129.56	135.81	140.17

ν = degrees of freedom.

^aAdapted with permission from *Biometrika Tables for Statisticians*, Vol. 1, 3rd ed., by E. S. Pearson and H. O. Hartley, Cambridge University Press, Cambridge, 1966.

Fonte: MONTGOMERY, Douglas C. *Introdução ao Controle Estatístico de Qualidade*, 2009.