



Universidade de Brasília  
IE – Departamento de Estatística

# Controle Estatístico de Qualidade para indicadores do Instituto Federal de Brasília

Iran Barros Nunes Filho  
10/0031226

Brasília  
2015

Iran Barros Nunes Filho  
10/0031226

## Controle Estatístico de Qualidade para indicadores do Instituto Federal de Brasília

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

Orientador:  
Professor Dr. Gladston Luiz da Silva

Brasília  
2015

## Resumo

Neste trabalho foi proposto o uso de gráficos de controle para dois indicadores da área de suporte do Instituto Federal de Brasília (IFB). Estes indicadores foram selecionados por COELHO (2015), funcionário da área de suporte do Instituto Federal de Brasília (IFB) e aluno do Programa de Pós-graduação da Ciência da Computação da Universidade de Brasília. Inicialmente foi realizada a análise descritiva dos dados.

O próximo passo consistiu em ajustar as duas séries dos dois gráficos e propor um modelo. As séries dos dois indicadores se mostraram estacionárias, as suas funções de autocorrelação aparentam ser um Ruído Branco. Esta característica fez com que os gráficos de controle não utilizassem o resíduo de um modelo ajustado e sim a série original.

Alguns testes foram necessários para determinar qual seria o melhor gráfico de controle. Os testes de Kolmogorov-Smirnov e Shapiro-Wilk afastaram a hipótese de que os dados seguissem uma Normal, além de análises visuais como o QQPlot e o histograma dos dados. Tal característica contribuiu para que o gráfico de controle do tipo Médias Móveis Exponencialmente Ponderadas (MMEP) fosse sugerido para os dois indicadores, dado que este tipo de gráfico é o indicado quando há ausência de normalidade das observações. Posteriormente foram construídos os gráficos de controle variando o  $\lambda$  e o limite de controle.

Palavras-chaves: Gráficos de controle, Controle estatístico de qualidade, indicadores, Séries temporais, estacionariedade, normalidade, Komolgorov-Smirnov, Shapiro-Wilk, Médias Móveis Exponencialmente Ponderadas, MMEP

## Sumário

Capítulo 1 .....	6
1.1 Introdução .....	6
1.2 Objetivos .....	7
1.2.1 Objetivo Geral .....	7
1.2.2 Objetivos Específicos .....	7
1.3 Metodologia .....	8
1.4 Estrutura .....	9
Capítulo 2 .....	10
2.1 Análise de Séries Temporais .....	10
2.1.1 Modelos ARIMA .....	11
2.1.2 Função de Autocorrelação .....	13
2.1.3 Escolha do Modelo .....	13
2.1.4 Diagnóstico do Modelo .....	14
2.2 Gráficos de Controle .....	15
2.2.1 Gráfico de Shewhart .....	16
2.2.2 Média e amplitude .....	16
2.2.3 Média e Desvio Padrão .....	17
2.2.4 Somas Acumuladas (CUSUM) .....	18
2.2.5 Médias Móveis Exponencialmente Ponderadas .....	19
2.2.6 Eficiência Relativa .....	20
Capítulo 3 .....	22
3.1 Indicadores selecionados .....	22
3.2 Análise Descritiva dos Dados .....	23
3.3 Análise de Séries Temporais .....	28
3.3.1 Estacionariedade dos Indicadores .....	28
3.3.2 Autocorrelações dos indicadores .....	29
3.3.3 Normalidade dos indicadores .....	32
3.4 Construção dos Gráficos de Controle .....	35
3.5 Escolha dos gráficos de controle .....	38
3.6 Conclusões .....	39
Referências .....	41
Apêndice .....	42
1. Teste de Shapiro-Wilk .....	42

2. Teste de Kolmogorov-Smirnov .....	43
3. Programação utilizada .....	45

## Lista de Ilustrações e tabelas

Figura 1: Comportamento do “Tempo para a Instalação de Impressora” entre out/13 e set/14 .....	23
Tabela 1: Dados do “Tempo para a Instalação de impressora” .....	24
Figura 2: Comportamento do “Tempo para a Criação de e-mail de trabalho” entre out/13 e set/14.....	25
Tabela 2: Dados do “Tempo para a Criação de e-mail de trabalho” .....	26
Figura 3: Box plot do “Tempo para a Instalação de impressora” .....	26
Figura 4: Box plot do “Tempo para a Criação de e-mail de trabalho” .....	27
Tabela 3: Resultado do teste de Dickey-Fuller para verificar se a série do “Tempo para a Instalação de impressora” é estacionária. ....	28
Tabela 4: Resultado do teste de Dickey-Fuller para verificar se a série do “Tempo para a Criação de e-mail de trabalho” é estacionária. ....	29
Figura 5: Função de Autocorrelação do “Tempo para a Instalação de Impressora” .....	29
Figura 6: Função de Autocorrelação do “Tempo para a Criação de e-mail de trabalho” .....	30
Tabela 5: Resultado do teste de Ljung-Box para os dois indicadores .....	31
Tabela 6: Resultado do teste de Kolmogorov-Smirnov para os dois indicadores .....	32
Tabela 7: Resultado do teste de Shapiro-Wilk para os dois indicadores .....	32
Figura 7: QQPlot da série do “Tempo para a Criação de e-mail de trabalho” .....	33
Figura 8: Histograma da série do “Tempo para a Criação de e-mail de trabalho” .....	33
Figura 9: QQPlot da série do “Tempo para a Instalação de impressora” .....	34
Figura 10: Histograma da série do “Instalação de impressora” .....	34
Figura 11: Gráfico de Controle MMEP para o “Tempo para a Criação de e-mail de trabalho” com $\lambda=0,1$ e $L=2,7$ .....	36
Figura 12: Gráfico de Controle MMEP para o “Tempo para a Criação de e-mail de trabalho” com $\lambda=0,2$ e $L=2,8$ .....	36
Figura 13: Gráfico de Controle MMEP para o “Tempo para a Instalação de impressora” com $\lambda=0,1$ e $L=2,7$ .....	37
Figura 14: Gráfico de Controle MMEP para o “Tempo para a Instalação de impressora” com $\lambda=0,2$ e $L=2,8$ .....	37

# Capítulo 1

## 1.1 Introdução

O Controle Estatístico de Qualidade surgiu em meados do século passado quando Shewhart propôs os primeiros gráficos de controle no intuito de verificar variações em processos produtivos de empresas. Por meio desses gráficos passou a ser possível determinar o momento em que uma correção deveria ser aplicada a processos de produção. No primeiro momento, o Controle de Qualidade surgiu para verificar se a produção estava no rumo certo quanto ao processo de fabricação de um produto.

Neste trabalho será proposto o uso de gráficos de controle para indicadores da área de suporte do Instituto Federal de Brasília (IFB), o estudo desenvolvido neste trabalho poderá ser estendido a outros indicadores do IFB.

Para propor o gráfico de controle, é necessário um estudo adequado em Séries Temporais. Após a análise das séries dos indicadores um modelo ajustado pode ser proposto dependendo da estacionariedade, tendência e sazonalidade.

Com os resultados obtidos após a análise de séries temporais, é proposto o tipo de gráfico de controle mais adequado para cada situação.

Ao utilizar os conceitos de Controle de Qualidade as empresas se capacitam a entender melhor os seus processos, prever tendências, antecipar eventuais problemas e melhorar o processo de desenvolvimento de um produto ou serviço. Ademais, os Gráficos de Controle estão intimamente ligados às séries temporais, onde também é possível inferir os dados do futuro em alguns casos e prever futuras tendências e ações a serem tomadas a fim de melhorar o comportamento da série.

## **1.2 Objetivos**

### **1.2.1 Objetivo Geral**

Este trabalho tem por objetivo propor gráficos de controle a indicadores estratégicos selecionados pela área de tecnologia do Instituto Federal de Brasília (IFB).

### **1.2.2 Objetivos Específicos**

Neste sentido, o trabalho tem os seguintes objetivos específicos:

- Selecionar os indicadores da área de suporte de tecnologia do Instituto Federal de Brasília.
- Ajustar uma série temporal para os indicadores no intuito de identificar a presença de tendência ou de sazonalidade.
- Propor gráficos de controle para os indicadores selecionados, caso não sejam identificadas a presença de tendência ou de sazonalidade. Caso contrário, serão propostos gráficos para os resíduos do modelo ajustado aos indicadores selecionados.
- Parametrizar os gráficos de controle a partir do banco de dados fornecido pelo IFB.
- Apresentar metodologia para implantar os gráficos de controle aos indicadores selecionados.



### **1.3 Metodologia**

O primeiro passo do trabalho é 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 exponencialmente ponderadas e algoritmos para o cálculo da eficiência entre esses métodos. A partir deste levantamento, será possível identificar gráficos de controle que poderão ser utilizados neste estudo.

A seleção dos indicadores que serão analisados neste trabalho será realizada por Daniel Souza Coelho, funcionário do IFB e orientando do Prof. Gladston no programa de Pós-graduação em Computação Aplicada do Departamento de Ciência da Computação da UnB, cuja Dissertação tem como um dos objetivos propor metodologia para a seleção de indicadores dos processos estratégicos do IFB.

Definido os indicadores, Daniel irá formatar e disponibilizar o banco de dados contendo as informações pertinentes aos mesmos, para que seja, então, realizada uma análise exploratória nos dados para subsidiar a etapa seguinte, que consiste na definição dos gráficos de controle. Será realizada uma análise das séries temporais dos indicadores selecionados para verificação da existência de tendência ou sazonalidade. Caso seja verificada a presença de tendência ou de sazonalidade na série, os gráficos de controle a serem propostos serão baseados nos resíduos do modelo ajustado.

Com a definição dos gráficos de controle, o próximo passo consistirá em definir seus parâmetros. Como se trata do primeiro estudo com este banco de dados, 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 dos gráficos de controle propostos, relativamente ao gráfico de Shewhart.

Cumpridos estes passos, será apresentada uma metodologia para a implantação dos gráficos de controle no IFB, no intuito de contribuir para a melhoria da gestão nesta Instituição.

## **1.4 Estrutura**

A estrutura deste trabalho consiste em três capítulos. O Capítulo 1 contém a Introdução onde são apresentados o objetivo geral e os específicos para o problema em questão, a metodologia e a estrutura da monografia. O Capítulo 2 consiste na revisão bibliográfica, que contém conceitos de Séries Temporais e de Controle Estatístico de Qualidade, que são de fundamental importância para a realização do trabalho. O capítulo 3 consiste no estudo de caso, onde foram apresentados os indicadores e feita toda a análise de Séries Temporais e Controle Estatístico de Qualidade para cada indicador.

## Capítulo 2

Este capítulo apresenta os conceitos de Séries Temporais e Controle Estatístico de qualidade. Inicialmente são demonstrados os testes que se aplicam às séries e posteriormente os critérios para o ajuste de um modelo.

Os gráficos de controle estão ligados aos conceitos de séries temporais, pois os mesmos serão criados dependendo dos resultados obtidos. Provar estacionariedade, tendência e sazonalidade de uma série mostra se há a necessidade de uma série ajustada e se os resíduos são objetos de estudo ou não.

Após a análise de Séries Temporais, os gráficos de controle serão escolhidos usando os resultados obtidos. Após a escolha do gráfico, é realizada a parametrização das observações e definição de seus limites.

### 2.1 Análise de Séries Temporais

Segundo Morettin (2006) uma série temporal é qualquer conjunto de observações ordenadas no tempo, podendo ser discretas ou contínuas. Uma série temporal é contínua quando as observações são feitas continuamente ao longo de um tempo  $T$ , onde  $T = t_1 < t_2 < t_3 < \dots < t_n$ . Já uma série temporal é discreta quando os espaços entre as observações possuem pausas no tempo.

Dentre alguns objetivos da análise de séries temporais se encontra descobrir o mecanismo gerador da série, prever valores futuros, descrever o comportamento e procurar periodicidades relevantes nos dados.

Existem dois tipos de séries temporais: estacionárias e não-estacionárias. Séries estacionárias desenvolvem-se aleatoriamente ao redor de uma média constante, refletindo um equilíbrio estável, porém as séries geralmente são não-estacionárias, pois apresentam sazonalidade e tendências.

Pode-se definir o processo  $X(t)$  como:

$$X_t = T_t + S_t + a_t$$

Onde  $T_t$  é o componente de tendência,  $S_t$  o componente de sazonalidade e  $a_t$  um Ruído Branco.

A transformação de dados é utilizada para transformar a série em estacionária, segundo Morettin (2006) há duas razões para se transformar os dados originais: estabilizar a variância e tornar o efeito sazonal aditivo.

Uma transformação comum são diferenças sucessivas da série original, o início se dá com a diferença da última observação com a penúltima:

$$\Delta X_t = X_t - X_{t-1}$$

O próximo passo resulta em:

$$\Delta^2 X_t = \Delta[\Delta X_t] = \Delta X_t = \Delta[X_t - X_{t-1}]$$

Após isso, tem-se:

$$\Delta^2 X_t = X_t - 2X_{t-1} + X_{t-2}$$

Dessa maneira, a fórmula geral para n diferenças é dada por:

$$\Delta^n = \Delta[\Delta^{n-1} X_t]$$

Quando se utiliza séries econômicas e financeiras as vezes é necessário aplicar na série original alguma transformação não linear, podendo ser a mesma uma transformação logarítmica ou de Box-Cox(1964), representada da seguinte forma:

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

onde  $\lambda$  e  $c$  são os parâmetros a serem estimados. A transformação logarítmica é utilizada caso a medida de dispersão da série, como por exemplo o desvio padrão, for proporcional a sua média.

### 2.1.1 Modelos ARIMA

Como dito anteriormente, uma das vantagens de se fazer a análise de séries temporais é a capacidade de fazer projeções futuras com as observações já realizadas, um dos modelos capazes de se realizar essas projeções é o ARIMA, ou de Box e Jenkins (1970), sua metodologia tem como função ajustar modelos auto-regressivos integrados de médias móveis, ARIMA(p,d,q) a um conjunto de dados. Sendo p a ordem do modelo auto-regressivo, d o número de diferenças e q a ordem do modelo de médias móveis.

Para identificar um modelo ARIMA adequado, Morettin (2006) diz que devem se seguir algumas condições, sendo elas:

- É analisada uma classe geral de modelos
- Com base na autocorrelação, autocorrelação parcial e outros critérios, é sugerido um modelo
- Posteriormente se estima os parâmetros do modelo identificado
- E por fim é realizado o diagnóstico do modelo ajustado, sendo feito através de uma análise de resíduos

Caso o modelo não seja o adequado, repete-se o processo

## **Modelos Autoregressivos**

Considerando um modelo autoregressivo de ordem  $p$ , denota-se AR( $p$ ) e seu modelo é dado por:

$$Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \dots + \phi_p Z_{t-p} + a_t$$

## **Modelos de médias Móveis**

Considerando um modelo de médias móveis de ordem  $q$ , denota-se MA( $q$ ) e seu modelo é dado por:

$$Z_T = \mu + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q}$$

$\theta_1, \theta_2$  e  $\theta_q$  são constantes, assim como  $\mu$ .

## **Modelos autoregressivos e de médias móveis**

Os modelos autoregressivos são bastante utilizados em algumas áreas, principalmente em finanças, onde previsões são situações que demandam bastante interesse, os modelos autoregressivos e de médias móveis são muito utilizados quando se deseja um modelo com um número não muito grande de parâmetros. O modelo é denotado por ARMA( $p, q$ ) e sua forma é dada por:

$$Z_t = \phi_1 Z_{t-1} + \dots + \phi_p Z_{t-p} + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q}$$

De forma reduzida, o modelo é:

$$\phi(B)Z_t = \theta(B)a_t$$

## Modelos autoregressivos, integrados e de médias móveis

O Modelo autoregressivo, integrado e de médias móveis (ARIMA) é semelhante ao modelo regressivo e de médias móveis (ARMA(p,q)), porém é utilizado em séries não estacionárias, assim o  $d$  fica sendo o número de diferenças feitas até que a série passasse a ser estacionária.

### 2.1.2 Função de Autocorrelação

Segundo Morettin (2006), os processos AR(p), MA(q) e ARMA(p,q) apresentam função de autocorrelação (fac) com características especiais, sendo elas:

- Um processo AR(p) tem fac q decaí de acordo com exponenciais e/ou senóides amortecidas, infinita em extensão;
- Um processo MA(q) tem fac infinita, no sentido que ela apresenta um corte após o lag q;
- Um processo ARMA(p,q) tem fac infinita em extensão, a qual decaí de acordo com exponenciais e/ou senóides amortecidas após o lag q-p;

### 2.1.3 Escolha do Modelo

A escolha do modelo é uma representação teórica da série, sendo este modelo um dentre outros que poderiam ser utilizados, é uma modelagem que tenta de forma aproximada representar a série em estudo, não podendo assim ser considerado apenas um modelo com o certo.

O modelo que será utilizado segue alguns critérios de avaliação. São eles o Critério de Informação de Akaike (AIC), o Critério de Informação de Akaike corrigido (AICc) e o Critério de Informação Bayesiano (BIC).

Utiliza-se como  $L(\hat{\theta})$  a função de máxima verossimilhança do modelo.

## **Critério de Informação de Akaike**

É um método indicado para grandes amostras, Akaike define que se baseando nesse critério, atribui-se uma pontuação ao modelo utilizando como referência a adequação dos dados e a ordem do modelo. Sua estrutura é dada por

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

E  $p$  é o número de parâmetros que deve ser estimado no modelo.

## **Critério de Informação de Akaike Corrigido**

Este método é um melhoramento do Critério de Akaike, porém é utilizado para pequenas amostras, dado que reage de forma mais eficiente e precisa. Sua fórmula é dada por:

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

## **Critério de Informação Bayesiano**

Criado por Schwarz(1978), é uma forma alternativa em relação aos dois modelos apresentados anteriormente, utiliza uma ótica Bayesiana, utilizando a probabilidade a posteriori. É representado da seguinte forma:

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

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

### **2.1.4 Diagnóstico do Modelo**

Existem alguns testes para descobrir se o modelo é adequado, a maioria utilizando as autocorrelações estimadas dos resíduos. Um dos testes é o Box-Pierce-Ljung, que é uma variação do teste de Box-Pierce, apresentando melhores resultados do que o teste precursor. Inicialmente deve-se elaborar as hipóteses, sendo elas:

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

Primeiramente deve-se calcular as estimativas das autocorrelações, sendo estimado através da fórmula:

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

A estatística do teste é dada por:

$$Q(k) = n(n-2) \sum_{j=1}^K \frac{r_j^2}{(n-j)}$$

$Q(k)$  segue uma qui-quadrado com  $(k - p - q)$  graus de liberdade.

## 2.2 Gráficos de Controle

Os Gráficos de Controle são utilizados para verificar se processos estão sob controle estatístico ou não, isto é, se estão sujeitos a perturbações que provêm de causas especiais, além das perturbações inerentes aos próprios processos, isto é, das variações decorrentes de causas comuns, ou aleatórias.

Quando o processo encontra-se sob controle, segue uma distribuição Normal  $(\mu_0, \sigma)$  sendo esses parâmetros geralmente desconhecidos, devendo ser estimados.

Geralmente a representação do gráfico de controle é dado por um Limite Central (LC) que é a média do processo e os Limites Superiores (LSC) e Inferiores (LIC), que são dados em relação a um  $\sigma$ .

Sendo  $\mu$  a média e  $\sigma$  o desvio padrão, LC, LSC e LIC são dadas por:

$$LSC = \mu + L\sigma$$

$$LIC = \mu - L\sigma$$

$$LC = \mu$$

Assim, considera-se um processo sob controle se os pontos observados se encontram dentro desses limites, caso haja alguns fora, significa que não se encontra sob controle.



### 2.2.1 Gráfico de Shewhart

É o gráfico mais comumente utilizado devido a sua simplicidade e fácil análise de controle ou não do processo. Shewhart definiu alguns conceitos de decisão para verificação de que um processo encontra-se ou não sob controle, são eles:

- Um ou mais pontos fora dos limites de controle  $3\sigma$ .
- Dois ou três pontos consecutivos fora dos limites de alerta  $2\sigma$ .
- Quatro ou cinco pontos consecutivos além dos limites  $1\sigma$ .
- Uma sequência de 8 pontos consecutivos de um mesmo lado da linha central.
- Seis pontos em uma sequência sempre crescente ou decrescente.
- Quinze pontos em sequência na Zona C (tanto acima quanto abaixo da linha central)
- Quatorze pontos em sequência alternadamente para cima e para baixo.
- Oito pontos em sequência de ambos os lados da linha central com nenhum na Zona C.
- Um padrão não usual ou não aleatório nos dados
- Um ou mais pontos pertos de um limite de alerta ou de controle.

O gráfico de Shewhart se subdivide em dois tipos, nos quais são levados em consideração média e amplitude e no outro média e desvio padrão

### 2.2.2 Média e amplitude

É utilizado em um processo onde a variável segue uma observação contínua e busca o controle do valor médio.

Enquanto o gráfico de  $\bar{X}$  tem a função de controlar a média do processo, o gráfico de R tem o objetivo de controlar a variabilidade.

Para  $\bar{X}$  é calculada a média da amostra. É utilizada também a média de várias sub amostras, levando em consideração o somatório das mesmas. É observado em cada sub amostra um número n de observações. Neste tipo de gráfico o LC é a média e LIC e LSC são determinados de acordo com o desvio padrão.

Utilizando uma amplitude relativa e a denotando como W, definida como  $W = \frac{R}{\bar{\theta}}$ , e uma média constante denotada como  $d_2$ , é criado um estimador não viesado para  $\theta$ , sendo este igual a  $\hat{\theta} = \frac{\bar{R}}{d_2}$ ,  $\bar{R}$  sendo a média da amplitude.

Assim os limites de controle para a média são iguais a:

$$LIC = \bar{\bar{X}} - A_2\bar{R}$$

$$LC = \bar{\bar{X}}$$

$$LSC = \bar{\bar{X}} + A_2\bar{R}$$

Definido  $A_2 = \frac{3}{d_2\sqrt{n}}$   $d_2$  um fator de correção e seu valor tabelado de acordo com o tamanho da amostra.

Já os limites de controle para a amplitude são definidos como:

$$LIC = D_3\bar{R}$$

$$LC = \bar{R}$$

$$LSC = D_4\bar{R}$$

Sendo  $D_3$  e  $D_4$  constantes com valores tabelados com valor dependendo do tamanho da amostra.

### 2.2.3 Média e Desvio Padrão

O gráfico de Média e Desvio Padrão é semelhante ao gráfico apresentado anteriormente, porém  $\sigma$  deve ser estimado através de um estimador não viesado, estimador esse definido como:

$$\hat{\sigma} = \frac{S}{c_4}$$

$c_4$  é definido por:

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

Antes de definir os LSC, LC e LIC deste tipo de gráfico, segue que:

$$A_3 = \frac{3}{c_4\sqrt{n}}, B_3 = 1 - \frac{3}{c_4}\sqrt{1-c_4^2} \text{ e } B_4 = 1 + \frac{3}{c_4}\sqrt{1-c_4^2}$$

Assim, os limites desse gráfico são definidos para  $\bar{\bar{X}}$ :

$$LIC = \bar{\bar{X}} - A_3\bar{S}$$

$$LC = \bar{\bar{X}}$$

$$LSC = \bar{\bar{X}} + A_3\bar{S}$$

Para  $\bar{\bar{S}}$ :

$$LIC = B_3\bar{\bar{S}}$$

$$LC = \bar{\bar{S}}$$

$$LSC = B_4\bar{\bar{S}}$$

## 2.2.4 Somas Acumuladas (CUSUM)

Os gráficos de Controle Somas Acumuladas são uma alternativa para os Gráficos de Shewhart, a vantagem que estes apresentam informações sobre toda a amostra com cada amostra coletada tendo o mesmo peso, isto facilita na percepção se há ou não uma perturbação pequena na média do processo.

Existem dois tipos de gráfico de CUSUM (Cumulative Sum), sendo Tabular e Máscara V. Montgomery (2009) ressalta que o mais indicado é o Tabular, sendo este mais utilizado. A soma acumulada ( $C_i$ ) depois de  $n$  amostras é dada por:

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

Onde  $\mu$  é a média do processo e  $\bar{X}_j$  a média da  $j$ -ésima amostra.

É necessário saber se o processos está sob controle, para isto a ferramenta utilizada é um algoritmo da forma:

$$c_i^+ = \text{máx}[0, \bar{X}_t - (\mu + K) + C_{i-1}^+]$$

$$c_i^- = \text{máx}[0, (\mu - K) - \bar{X}_t + C_{i-1}^-]$$

$K$  é chamado de valor de referência e é a metade do valor que se detecta na média, intuitivamente sua fórmula é igual a:

$$K = \frac{|\mu_1 - \mu_0|}{2}$$

Após o cálculo do algoritmo, é criado um parâmetro para atuar como regra de decisão de saber se o processo está ou não sob controle, o H representa o intervalo para esta decisão:

$$LSC:H = h\sigma$$

$$LIC:H = -h\sigma$$

Após essa verificação, considera-se o processo fora de controle caso os algoritmos excedam o valor do intervalo de H.

Também é possível realizar a padronização do gráfico CUSUM onde a nova variável tem a forma:

$$Z_i = \frac{X_i - \mu}{\sigma}$$

$Z_i$  seguindo uma Normal.

Os limites de controle para o CUSUM padronizado:

$$c_i^+ = \max[0, z_i - k + c_{i-1}^+]$$

$$c_i^- = \max[0, -k - z_i + c_{i-1}^-]$$

### 2.2.5 Médias Móveis Exponencialmente Ponderadas

Médias Móveis Exponencialmente Ponderadas (MMEP) é uma outra alternativa para os gráficos de controle, apresenta uma eficiência aproximada ao CUSUM em detectar pequenas perturbações no processo. A principal diferença deste tipo de gráfico é que apresenta melhores resultados para dados que rejeitam a hipótese de normalidade, ou seja, quando não seguem uma normal, o MMEP é o método mais indicado.

Inicialmente é necessário uma transformação dos dados:

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

Onde  $\lambda$  tem um valor entre 0 e 1.

Utilizando o desvio padrão das observações, é possível calcular a variância de  $Z_i$ :

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

Os limites desse gráfico de controle são definidos por:

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

$$LC = \mu_0$$

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

Sendo  $0,05 \leq \lambda \leq 0,25$  e  $L=3$  mais comumente utilizadas, pois são capazes de detectar pequenas perturbações no processo.

### 2.2.6 Eficiência Relativa

É necessário saber qual gráfico se deve usar, um dos critérios de escolha é a comparação do Comprimento Médio da Sequência (CMS), este comprimento é o número médio de amostras que se utilizam para se detectar uma mudança. Esta mudança ocorre de acordo com o problema, podendo ser pequena, assim utilizando uma amostra maior, ou grande, podendo utilizar uma amostra menor. É possível determinar também a dimensão da mudança que se deseja detectar.

O valor do CMS é definido através da fórmula:

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

$$\Delta = \delta - k \text{ e } \delta = \frac{\mu_1 - \mu_0}{\sigma}$$

$$b = h + 1,166$$

$k$  e  $h$  têm seus valores como constantes iguais a 4,77 e 0,5 respectivamente, esses valores são os utilizados mais frequentemente.

Para se calcular o CSM de  $c_1^+$  e  $c_1^-$ :

$$\frac{1}{CMS} = \frac{1}{CMS^+} + \frac{1}{CMS^-}$$

Após encontrados os valores de CMS, o que apresentar menor valor é o mais indicado, pois será o gráfico que demora menos tempo em detectar perturbações no processo, em perceber que está fora de controle.

## Capítulo 3

Este Capítulo apresenta um estudo de caso motivado pelo trabalho desenvolvido por COELHO (2015), funcionário da área de suporte do Instituto Federal de Brasília (IFB) e aluno do Programa de Pós-graduação da Ciência da Computação da Universidade de Brasília, cujo objetivo geral é propor uma metodologia para a seleção de indicadores e uma ferramenta de TI a serem utilizadas na gestão da qualidade de processos estratégicos do IFB, baseada em Gráficos de Controle Estatístico.

Para o desenvolvimento de seu trabalho, COELHO (2015) precisa de gráficos de controle a serem adotados para os indicadores selecionados e parametrizá-los a partir de registros do banco de dados do IFB.

Neste sentido, este Capítulo apresenta as análises estatísticas realizadas para dois dos indicadores selecionados por COELHO (2015), que resultaram nos gráficos de controle parametrizados para cada indicador.

### 3.1 Indicadores selecionados

O primeiro passo para a elaboração do estudo foi a seleção dos indicadores relacionados ao desempenho de processos de responsabilidade da área de Tecnologia da Informação do IFB, que representassem aspectos relacionados à eficácia e à eficiência dos serviços oferecidos, e que estivessem ligados à satisfação dos clientes internos do IFB.

Diante do exposto, ficou clara a necessidade de se utilizar indicadores que retratassem a qualidade de serviços prestados aos clientes e que, de certa forma, estimassem sua satisfação. Como o IFB não realizou uma pesquisa de satisfação junto aos clientes internos dos serviços prestados por sua área de TI, foram selecionados indicadores que refletissem o nível de qualidade dos serviços prestados.

Analisando os registros disponibilizados por COELHO (2015), verificou-se que potenciais indicadores estavam relacionados a tempos para os atendimentos a serviços solicitados. Foram selecionadas para as análises estatísticas as variáveis *Tempo para a Instalação de Impressora* e *Tempo para a Criação de e-mail de trabalho*, por possuírem registros em todo o período de mensuração considerado, e por representarem dois tipos de serviços distintos: O primeiro relativo à instalação de *hardware*, e o outro relativo à procedimento administrativo no sistema.

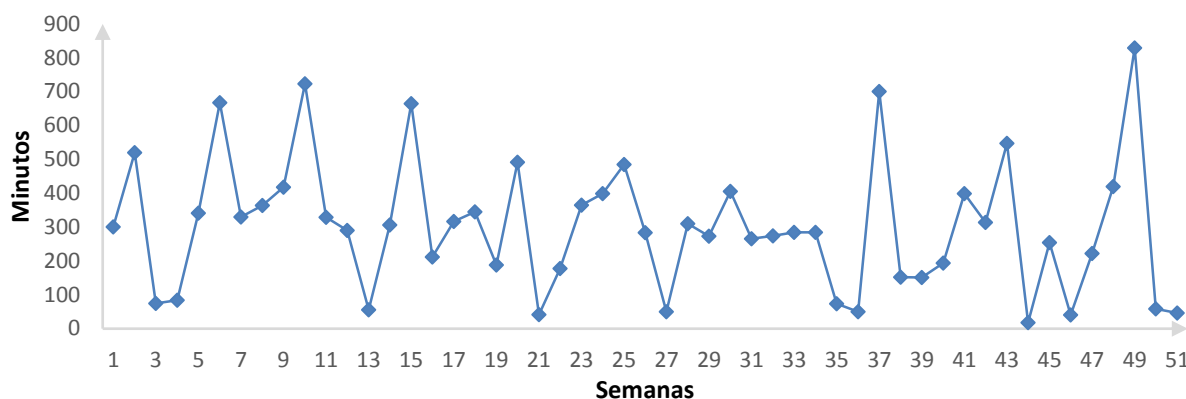
### 3.2 Análise Descritiva dos Dados

O banco de dados (BD) utilizado para as análises estatísticas realizadas neste trabalho foi fornecido por COELHO (2015). Possui registros de Ordens de Serviços (OS) que a área de Tecnologia da Informação (TI) recebeu de servidores do IFB, no período compreendido entre outubro de 2013 e setembro de 2014.

Para a variável *Tempo para a Instalação de Impressora* dispõe-se de 361 observações relativas a registros diários, que foram agrupadas nas semanas. Considerando os registros ocorridos as segundas-feiras e sexta-feiras, verificou-se que o agrupamento dos dados resultou em registros relativos a 51 semanas.

A série representada na Figura 1 apresenta o comportamento do indicador *Tempo para a Instalação de Impressora* entre outubro de 2013 a setembro de 2014. Cada observação representa o tempo médio (em minutos) dos serviços realizados na semana.

Figura 1: Comportamento do “Tempo para a Instalação de Impressora” entre out/13 e set/14



Verificou-se que, no período analisado, o tempo médio decorrido entre a solicitação do cliente e a instalação da impressora em seu destino foi de 299 minutos, isto é, aproximadamente 5 horas.

Vale dizer que os funcionários responsáveis por este serviço também realizam outras tarefas e que, portanto, o atendimento a este tipo de demanda não é necessariamente automático. Pelo Gráfico 1, percebe-se que em algumas semanas as solicitações ocorreram quase que imediatamente após o registro da OS, enquanto que em outras semanas o tempo médio para o atendimento às solicitações foi de algumas horas.

Observa-se ainda que, aparentemente, os registros relativos ao indicador apresentam um comportamento estacionário no período analisado, e que os valores oscilam em torno de um



tempo médio segundo um comportamento aleatório, o que será verificado na Análise de Séries Temporais subsequente.

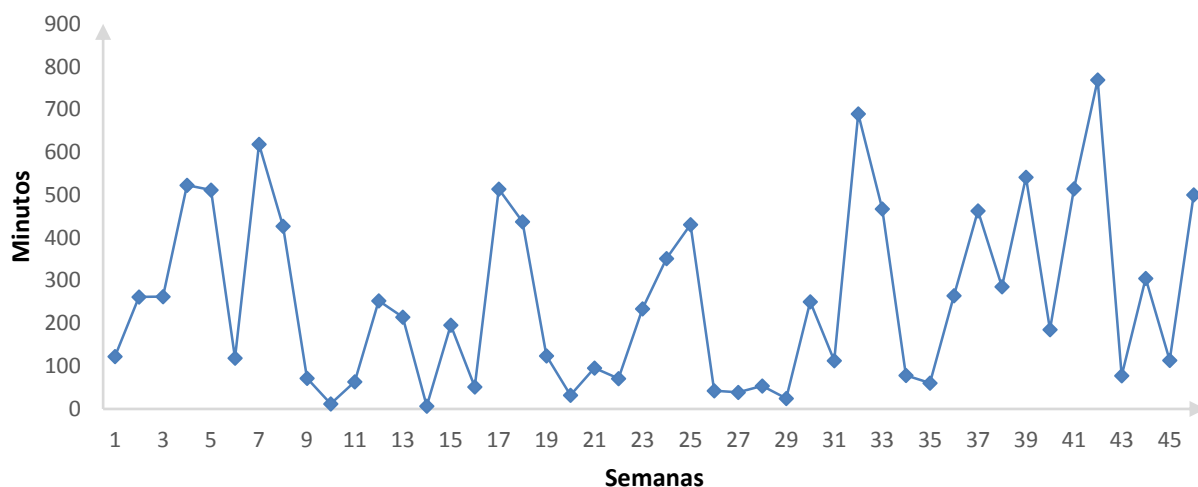
Pela Tabela 1, verifica-se que os tempos mínimos de atendimento variaram entre cinco e quinze minutos, enquanto os tempos máximos registrados alcançaram valores superiores a trinta horas. Isto representa que, de acordo com a demanda por serviços na área que instala impressoras, o nível de serviço pode ser avaliado de péssimo a ótimo. Vários fatores devem explicar isso, cabendo ao(s) responsável(is) pela área identificar os motivos dessa variação no intuito de minimizá-la.

*Tabela 1: Dados do “Tempo para a Instalação de impressora”*

Mês	Menor valor	Maior valor	Desvio Padrão	Média
out/13	00:11	27:12	06:13	04:52
nov/13	00:08	33:55	10:22	07:41
dez/13	00:12	20:12	07:46	08:39
jan/14	00:12	32:20	07:14	05:11
fev/14	00:05	34:00	07:01	05:17
mar/14	00:06	19:28	05:50	04:54
abr/14	00:10	33:26	08:24	04:43
mai/14	00:12	24:09	06:06	04:37
jun/14	00:07	22:06	06:24	04:36
jul/14	00:05	28:46	06:23	03:51
ago/14	00:09	33:43	08:42	07:22
set/14	00:15	29:34	08:38	05:55
de out/13 a set/14	00:05	34:00	07:34	04:59

A mesma análise foi feita para o indicador *Tempo para a Criação de e-mail de trabalho*, que contém 314 observações na base de dados. O Gráfico 2 apresenta sua evolução ao longo do período que compreende 46 semanas.

Figura 2: Comportamento do “Tempo para a Criação de e-mail de trabalho” entre out/13 e set/14



Verificou-se que, no período analisado, o tempo médio decorrido entre a solicitação do cliente e a criação de e-mail de trabalho foi de 270 minutos, isto é, 4 horas e 30 minutos.

Assim como o indicador passado, a demanda de solicitações varia e a resolução da OS depende de uma fila de obrigações dos servidores. Algumas vezes são resolvidas em poucos minutos e outras vezes demoram algumas horas.

Aparentemente a série também é estacionária no período analisado, porém apenas com a Análise de Séries Temporais é possível determinar isso. Esta suposição parte do princípio de que os valores oscilam perto de um tempo médio.

Na Tabela 2 é possível notar que o tempo mínimo de resolução da OS tem sempre observações inferiores a 10 minutos, com exceção do mês de setembro de 2014. Este fato ocorre provavelmente devido a poucas observações no mês em questão. Já os tempos máximos registrados apresentam em alguns valores superiores a trinta horas.

A seguir, são apresentadas análises descritivas para os dados. A Tabela 2 apresenta informações da série no formato mensal.

Tabela 2: Dados do “Tempo para a Criação de e-mail de trabalho”

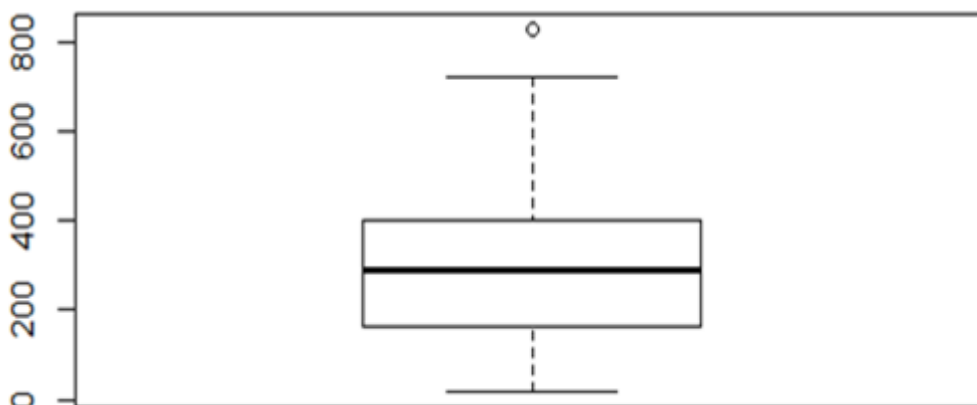
Mês	Menor valor	Maior valor	Desvio Padrão	Média
out/13	00:07	26:32	06:09	04:39
nov/13	00:03	29:01	08:12	06:57
dez/13	00:02	02:58	01:00	00:53
jan/14	00:03	18:51	04:59	03:00
fev/14	00:03	37:54	09:39	04:48
mar/14	00:03	27:00	05:29	02:51
abr/14	00:02	36:19	07:41	03:35
mai/14	00:02	24:22	07:36	04:46
jun/14	00:06	25:31	07:18	04:30
jul/14	00:04	35:04	09:24	06:37
ago/14	00:04	34:19	10:52	08:50
set/14	01:09	25:23	09:13	10:24
de out/13 a set/14	00:02	37:54	08:04	04:30

Percebe-se um pequeno aumento na média nos últimos meses observados, mas no geral, a série varia entre valores próximos.

Analisando o gráfico de Box Plot dos dois indicadores é possível observar valores extremos, 1º e 3º quartil e sua mediana.

O Box Plot para o indicador *Tempo para a Instalação de Impressora* é ilustrado no Gráfico 3.

Figura 3: Box Plot do “Tempo para a Instalação de impressora”

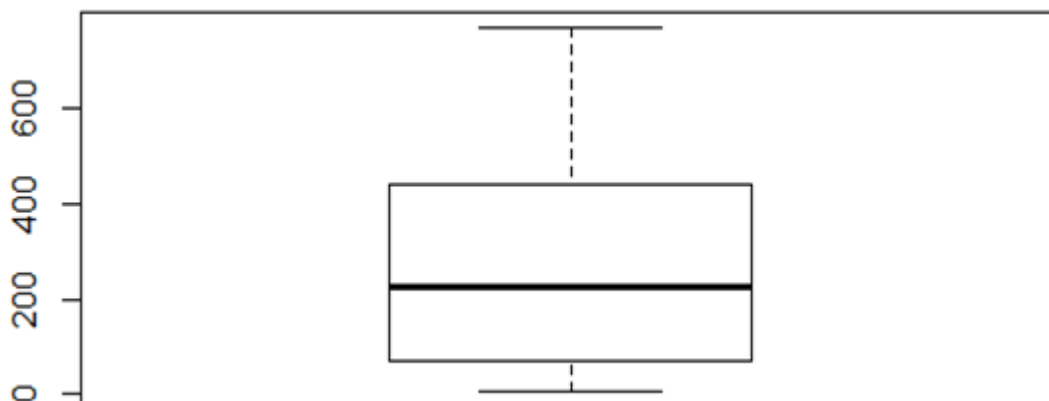


Nota-se que o primeiro quartil é aproximadamente 2 horas e 40 minutos, a mediana igual a 4 horas e 50 minutos e o terceiro quartil aproximadamente 6 horas e 40 minutos. Representado por um círculo no gráfico, percebe-se um *outlier* entre as observações, ou seja, um valor discrepante.

Isso significa que metade das observações foram resolvidas antes de 5 horas, um tempo considerado bom levando em conta que um dia útil de trabalho tem 8 horas. Apenas 25% das observações levaram mais de 6 horas e 40 minutos para serem solucionadas.

O mesmo gráfico pode ser feito para o indicador *Tempo para a Criação de e-mail de trabalho*. Este está representado na Figura 4.

Figura 4: Box Plot do “Tempo para a Criação de e-mail de trabalho”



É possível perceber que os dados para este indicador variam menos comparando com o indicador *Tempo para a Instalação de Impressora*, os dados estão dispostos de forma mais compacta, com valores mais próximos. Além disso, as observações não possuem *outliers* como no caso anterior.

Nota-se que primeiro quartil é igual a 1 hora e 13 minutos, a mediana igual a 3 horas e 45 minutos e o terceiro quartil igual a 7 horas e 16 minutos.

Os valores são inferiores ao indicador anterior, 50% das solicitações demoraram menos de 4 horas para serem resolvidas, mais de 75% das OS abertas foram resolvidas em menos de um dia de trabalho.

### 3.3 Análise de Séries Temporais

#### 3.3.1 Estacionariedade dos Indicadores

Para dar início a análise de séries temporais dos indicadores é preciso testar a estacionariedade dos mesmos. Visualizando os Gráficos 1 e 2 apresentados anteriormente, nota-se que aparentemente os dois são estacionários. No entanto, como a análise visual não é suficiente para comprovar tal característica, foi realizado o teste de Dickey-Fuller para se concluir quanto a este aspecto.

Pelo teste, as hipóteses são as seguintes:

$H_0$ : A série necessita de diferenças para se tornar estacionária

$H_1$ : A série é estacionária

Verifica-se que a hipótese a ser confirmada é a da não estacionariedade da série. O critério de decisão usado para esse teste é o *p-valor*, que representa o nível descritivo do teste. Conforme a escala de significância proposta por Fisher, BUSSAB *et al.* (2013), rejeita-se a hipótese nula para valores iguais ou inferiores a 0,05.

Conforme apresentado na Tabela 3, que apresenta os resultados do teste para o indicador *Tempo para a Instalação de Impressora*, o *p-valor* é inferior a 0,01, representando que há uma evidência muito forte para a rejeição da hipótese nula, ou seja, a série desse indicador tem um comportamento estacionário.

*Tabela 3: Resultado do teste de Dickey-Fuller para verificar se a série do “Tempo para a Instalação de impressora” é estacionária.*

Teste	Estatística do teste	P-valor
Dickey-Fuller	-5,2217	>0,01

A Tabela 4 apresenta o resultado do teste de Dickey-Fuller para o indicador *Tempo para a Criação de e-mail de trabalho*.

Tabela 4: Resultado do teste de Dickey-Fuller para verificar se a série do “Tempo para a Criação de e-mail de trabalho” é estacionária.

Teste	Estatística do teste	P-valor
Dickey-Fuller	-3,5228	0,0497

Neste caso, o *p-valor* encontrado é de 0,0497, representando que a hipótese de que a série relativa ao *Tempo para a Criação de e-mail de trabalho* deva ser estacionária. Sendo assim, os dois indicadores possuem comportamentos estacionários.

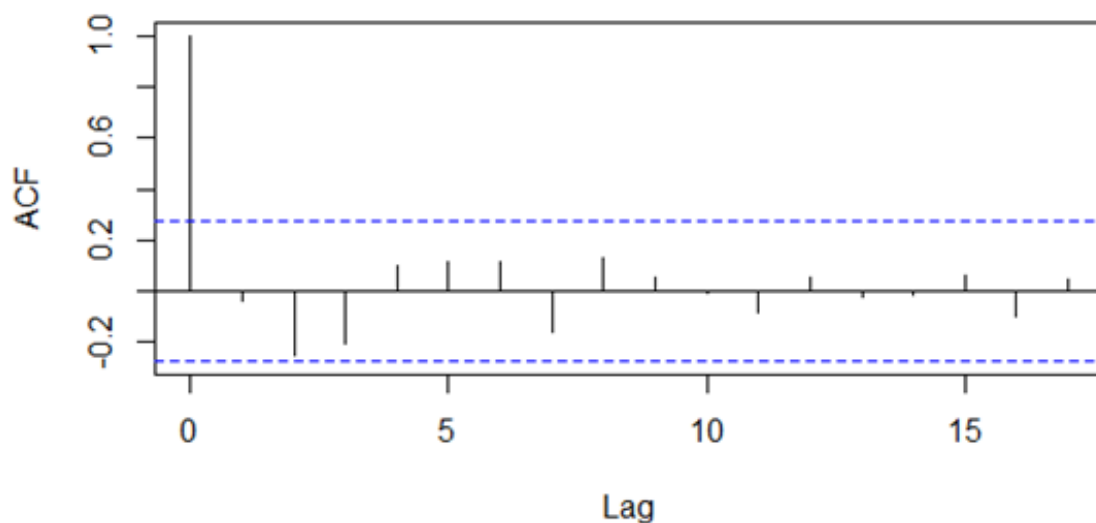
### 3.3.2 Autocorrelações dos indicadores

Além de testar a estacionariedade da série, seria preciso analisar as autocorrelações e autocorrelações parciais dos indicadores para propor um modelo de ajuste.

Segundo Morettin (2006), “... a identificação do modelo é feita com base na análise de autocorrelações, autocorrelações parciais e outros critérios, posteriormente seus parâmetros são estimados e é feita a análise dos resíduos para saber se o modelo sugerido é adequado, caso o modelo não seja aprovado o ciclo se repete e outro modelo é selecionado.”.

A Figura 5 ilustra a autocorrelação para o indicador *Tempo para a Instalação de Impressora*.

Figura 5: Função de Autocorrelação do “Tempo para a Instalação de Impressora”



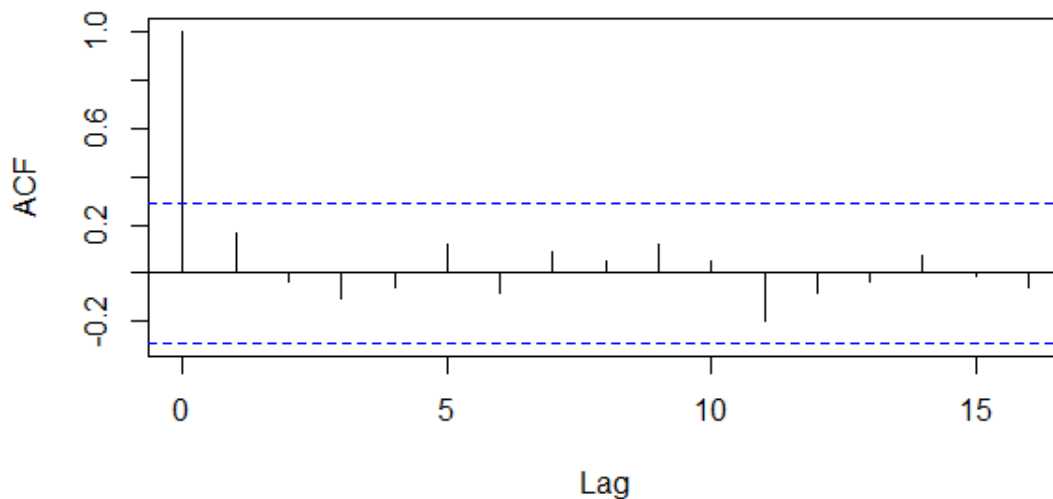
Neste caso, é possível perceber que apenas o *Lag* 0 corta a linha do limite, nenhum outro *lag* corta o limite superior. Usando o conceito de seleção dos modelos representados em Morettin (2006) onde, segundo ele:

- Um processo AR(p) tem fac que decai de acordo com exponenciais e/ou senóides amortecidas sendo infinita em extensão;
- Um processo MA(q) tem fac finita apresentando um corte após o *lag* q;
- Um processo ARMA(p,q) tem fac infinita em extensão, a qual decai de acordo com exponenciais e/ou senóides amortecidas após o *lag* “q-p”.

A FAC apresentada no Gráfico 5 não apresenta nenhuma das características citadas, apresenta na verdade o formato da FAC de um Ruído Branco.

A série do indicador *Tempo para a Criação de e-mail de trabalho* também deve passar por esse processo de identificação do modelo e sua função de autocorrelação demonstrada.

Figura 6: Função de Autocorrelação do “Tempo para a Criação de e-mail de trabalho”



Semelhante ao indicador anterior, o indicador *Tempo para a Criação de e-mail de trabalho* apresenta apenas o *Lag* 0 superior ao limite. É possível notar a semelhança das funções de autocorrelação para os dois indicadores, sugerindo que a série do indicador *Tempo para a Criação de e-mail de trabalho* também é um Ruído Branco.

Caso essas funções de autocorrelação se identificassem com algum tipo de modelo, a etapa seguinte seria validar os mesmos levando em consideração os valores da verossimilhança, do Critério de Akaike (AIC), Critério de Akaike Corrigido (AICc) e o Critério de Informação Bayesiano (BIC). O modelo que apresentasse o menor valor para os critérios e o maior para a verossimilhança seria indicado.

Com a definição de que as duas séries são um Ruído Branco algumas etapas serão descartadas e o ajuste da série para o gráfico de controle que antes seria realizado nos seus resíduos será feita na sua série original.

Não há a necessidade da análise de resíduos pois não houve a criação de um modelo ajustado, também não serão estimados os parâmetros do modelo.

Para se definir qual gráfico de controle usar deve-se realizar alguns testes, entre eles o de independência e de normalidade dos dados. O teste de Ljung-Box testa as seguintes hipóteses neste caso:

$H_0$ : As observações da série não são correlacionadas

$H_1$ : As observações da série são correlacionadas

A tabela 5 traz os resultados do teste para os dois indicadores.

*Tabela 5: Resultado do teste de Ljung-Box para os dois indicadores*

Indicador	Estatística do teste	P-valor
<i>Tempo para Criação de e-mail de trabalho</i>	1,4002	0,2367
<i>Tempo para Instalação de impressora</i>	0,095792	0,7569

Com estes resultados verifica-se que, segundo o teste de Ljung-Box, os níveis descritivos para os dois indicadores são superiores a 5%, representando que não há evidências para a rejeição das hipóteses nulas, indicado que as séries são não correlacionadas. Vale dizer que níveis descritivos correspondentes a 10% já seriam considerados marginais.



### 3.3.3 Normalidade dos indicadores

Outros dois testes serão aplicados, o de Kolmogorov-Smirnov e o de Shapiro-Wilk, estes dois testes tem o objetivo de provar que a série segue ou não uma distribuição Normal, suas hipóteses são:

$H_0$ : A amostra segue uma distribuição normal.

$H_1$ : A amostra não segue uma distribuição normal.

Os resultados encontrados são vistos na tabela 6.

*Tabela 6: Resultado do teste de Kolmogorov-Smirnov para os dois indicadores*

Teste	Indicador	Estatística do teste	P-valor
Kolmogorov-Smirnov	<i>Tempo para Criação de e-mail de trabalho</i>	3,245	2.2e-16
Kolmogorov-Smirnov	<i>Tempo para Instalação de impressora</i>	0,095792	2.2e-16

O *p-valor* encontrado é muito inferior 0,001, indicando evidência fortíssima contra a hipótese de normalidade para a série dos dois indicadores. O teste de Shapiro-Wilk serve para reforçar essa decisão, tendo seus resultados apresentados na tabela 7.

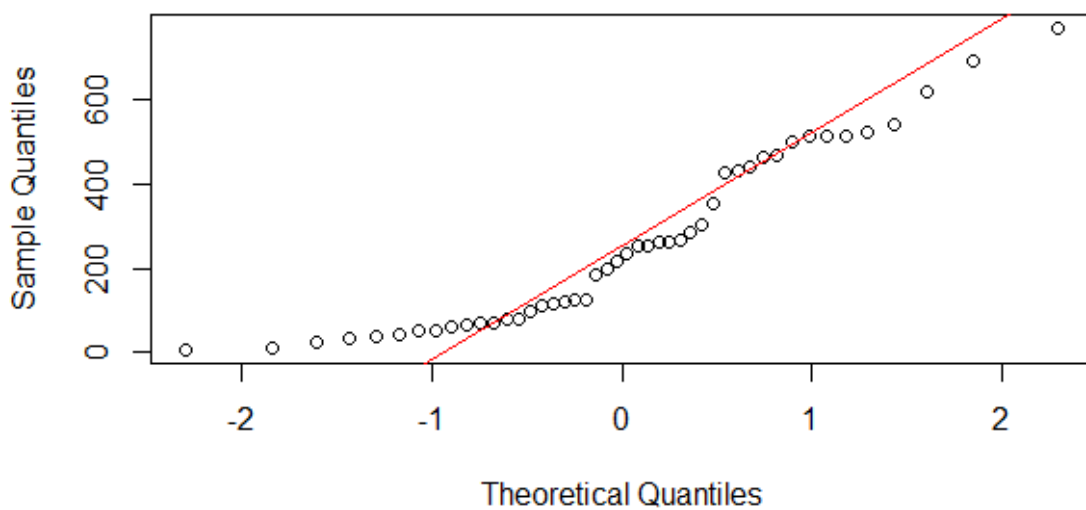
*Tabela 7: Resultado do teste de Shapiro-Wilk para os dois indicadores*

Teste	Indicador	Estatística do teste	P-valor
Shapiro-Wilk	<i>Tempo para Criação de e-mail de trabalho</i>	0,90652	0,001334
Shapiro-Wilk	<i>Tempo para Instalação de impressora</i>	0,94005	0,01234

Novamente os valores do *p-valor* apresentam resultados muito baixo e menor que o  $\alpha$  igual a 0,05, reforçando a decisão de rejeitar  $H_0$  e considerar que as séries não apresentam normalidade.

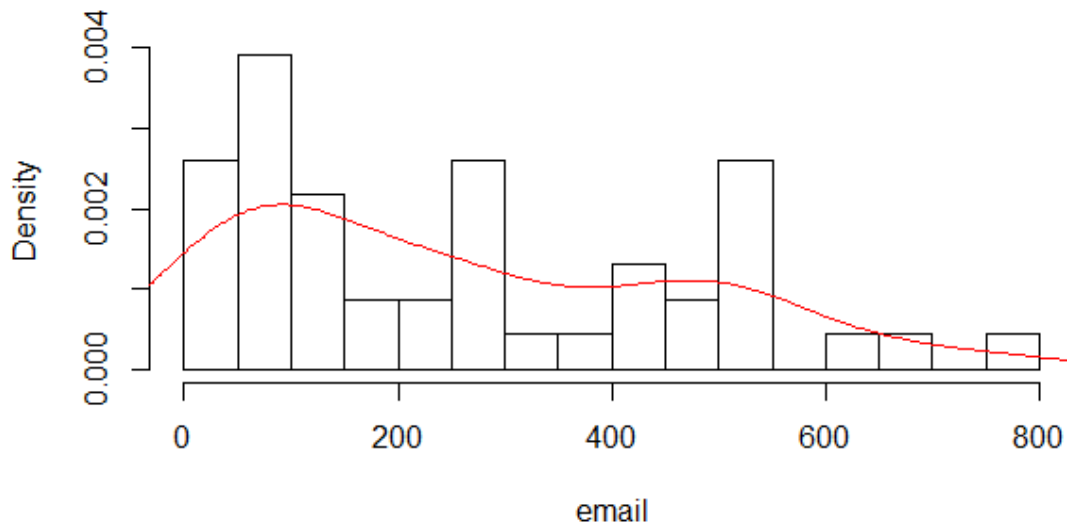
A maneira visual de reforçar essa hipótese é apresentando o QQPlot de cada série e também seu histograma, a figura 7 mostra como é o QQPlot do indicador *Tempo para a Criação de e-mail de trabalho*.

Figura 7: QQPlot da série do “Tempo para a Criação de e-mail de trabalho”



Para que os dados seguissem uma Normal seria necessário que as observações estivessem sobre a linha traçada na diagonal, a dispersão destas observações mostra que a série não apresenta normalidade.

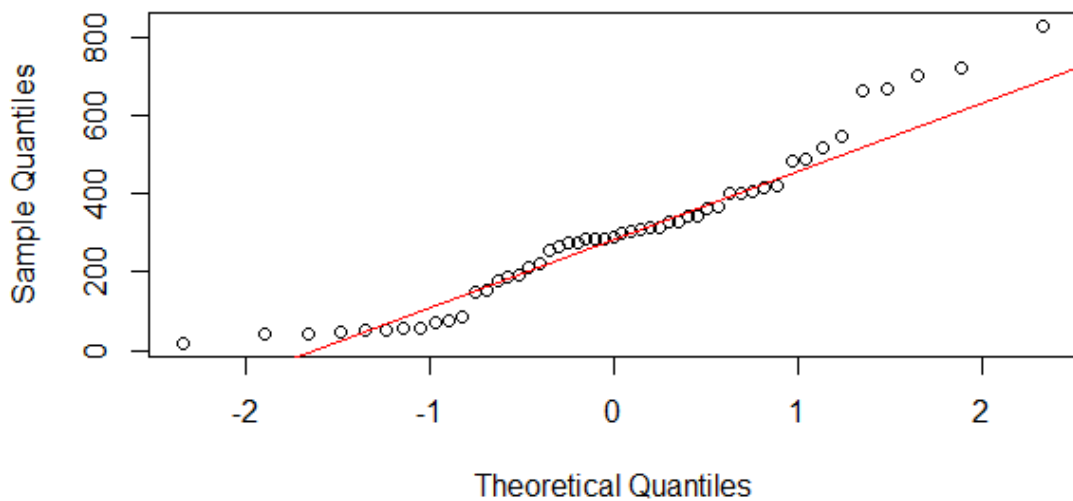
Figura 8: Histograma da série do “Tempo para a Criação de e-mail de trabalho”



O histograma não apresenta o formato da distribuição Normal. Seria possível através de outras análises verificar a qual distribuição tende essa série, no entanto o objetivo para a escolha do gráfico de controle é apenas a rejeição da normalidade, não sendo interessante a suposição de uma distribuição.

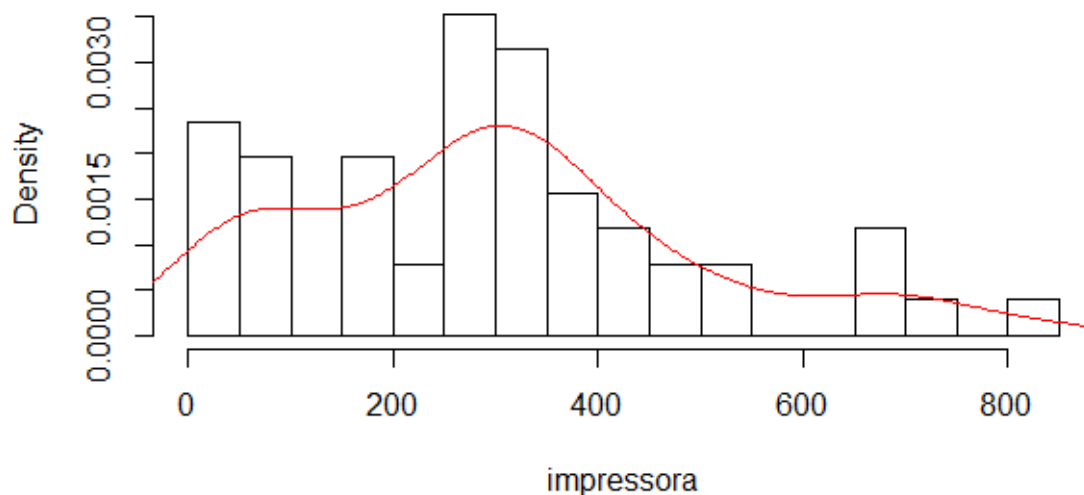
Para o indicador *Tempo para a Instalação de impressora* o processo é repetido e seu QQPlot demonstrado na Figura 9.

Figura 9: QQPlot da série do “Tempo para a Instalação de impressora”



O gráfico QQPlot para este indicador apresenta as observações mais próximas da linha do que o indicador anterior, porém ainda longe do que se define como uma distribuição Normal. Há poucas observações sobre a linha diagonal, sendo a maioria concentrada acima e abaixo.

Figura 10: Histograma da série do “Instalação de impressora”



O histograma também não apresenta a forma da Normal, junto com o QQPlot e com os testes de Kolmogorov-Smirnov e Shapiro-Wilk conclui-se que as duas séries não apresentam normalidade.

Esta definição é de suma importância para a escolha do gráfico de controle, rejeitando normalidade o gráfico mais indicado é o de Médias Móveis Exponencialmente Ponderadas (MMEP). Os gráficos MMEP e também os do tipo CUSUM são tão eficientes quanto os gráficos de Shewhart para detecção de pequenas perturbações na média, tendo como vantagem a não necessidade dos dados seguirem uma Normal.

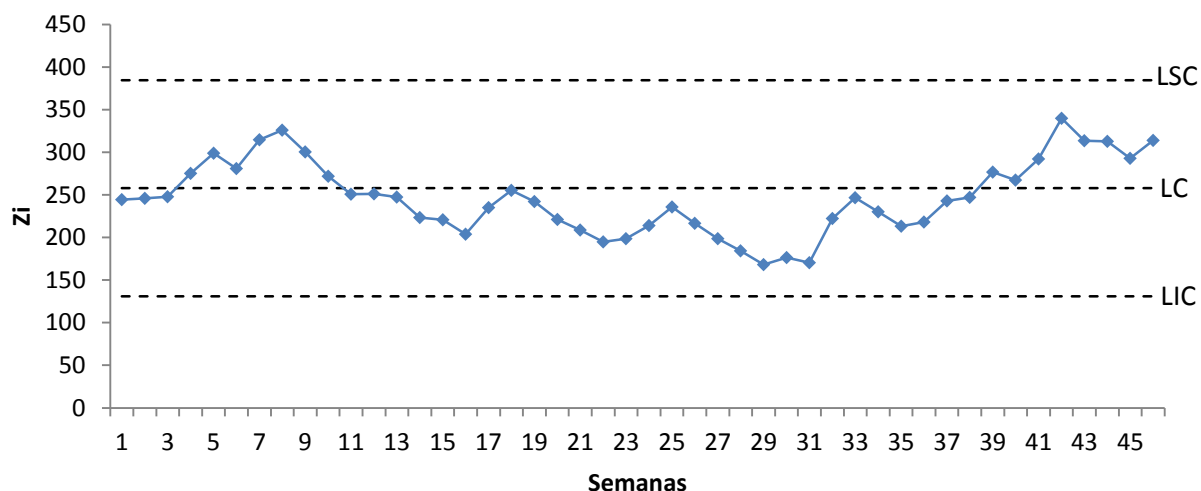
### **3.4 Construção dos Gráficos de Controle**

O gráfico de controle Médias Móveis Exponencialmente Ponderadas é o mais indicado para os dois indicadores, visto que são livres de distribuição. O primeiro passo neste sentido refere-se à parametrização dos gráficos. Montgomery(2009) sugere que o  $\lambda$  esteja entre 0 e 1. Comumente utiliza-se  $\lambda=0,1$  ou  $\lambda=0,2$ , enquanto que para os parâmetros relativos aos limites de controle os parâmetros sugeridos são  $L=2,7$  e  $L=2,8$  respectivamente.

Estes valores são determinantes para o estabelecimento dos limites superior e inferior. Para o indicador *Tempo para a Criação de e-mail de trabalho* foram avaliados dois gráficos, o primeiro utilizando  $\lambda=0,1$  e  $L=2,7$  e o segundo  $\lambda=0,2$  e  $L=2,8$ .

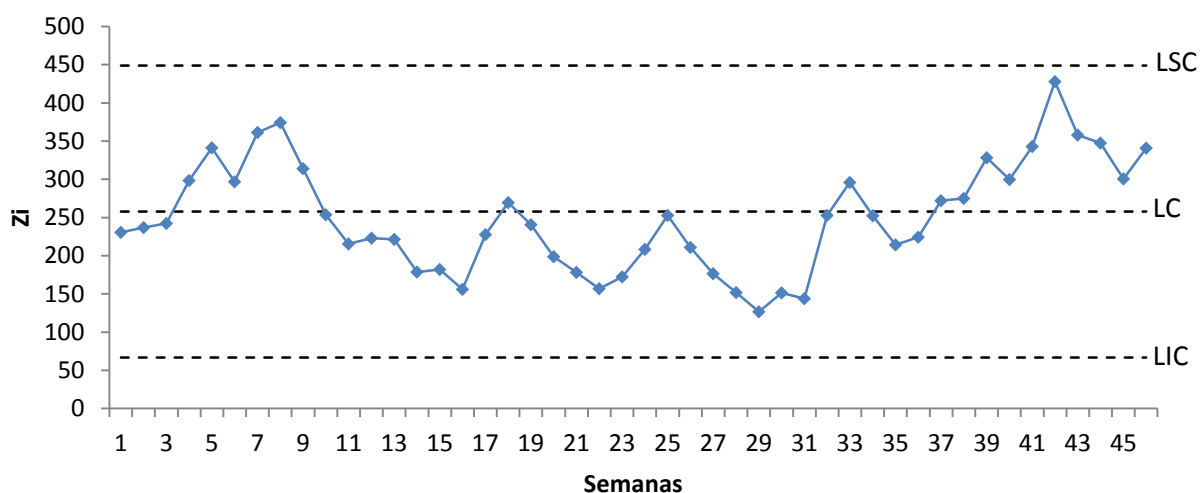
Pela Figura 11, verifica-se que os valores relativos aos Limites Superior e Inferior de Controle são, respectivamente, 384,51 e 131,04, sendo a Linha Central igual a 257,8. Conforme pode-se observar, todas as observações encontram-se dentro dos limites, representando que o processo apresenta-se, aparentemente, sob controle estatístico.

Figura 11: Gráfico de Controle MMEP para o “Tempo para a Criação de e-mail de trabalho” com  $\lambda=0,1$  e  $L=2,7$



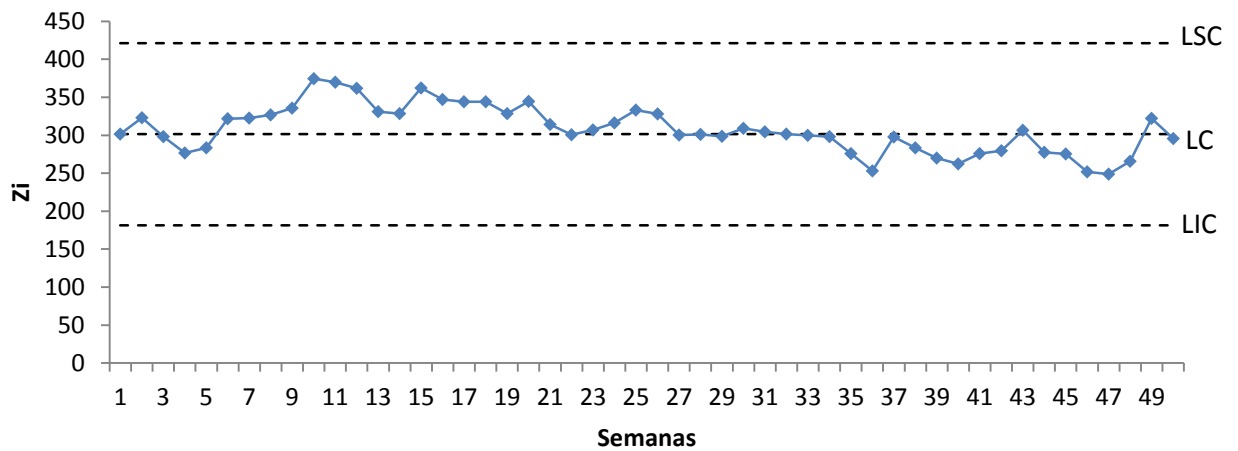
A Figura 12 foi construída considerando a seguinte parametrização:  $\lambda=0,2$  e  $L=2,8$ . Neste caso também os valores observados encontram-se dentro dos limites estabelecidos, a saber: Limite Superior é igual a 448,74, Limite Inferior igual a 66,82 e Linha Central igual a 257,8.

Figura 12: Gráfico de Controle MMEP para o “Tempo para a Criação de e-mail de trabalho” com  $\lambda=0,2$  e  $L=2,8$



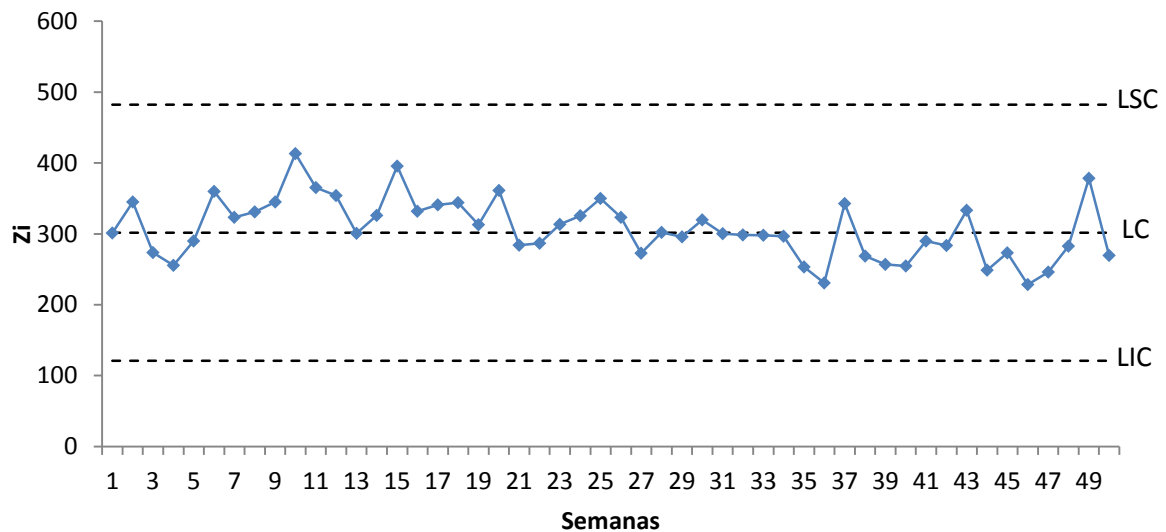
O procedimento anterior foi repetido para o indicador *Tempo para a Instalação de impressora*, mediante a variação de  $\lambda$  e de  $L$ .

Figura 13: Gráfico de Controle MMEP para o “Tempo para a Instalação de impressora” com  $\lambda=0,1$  e  $L=2,7$



Assim como o indicador anterior, as observações se mantêm dentro dos limites superior e inferior de controle. Este procedimento foi repetido utilizando o  $\lambda=0,2$  e  $L=2,8$  e pode ser observado na Figura 14.

Figura 14: Gráfico de Controle MMEP para o “Tempo para a Instalação de impressora” com  $\lambda=0,2$  e  $L=2,8$



A série continua dentro dos limites e um pouco menos suavizada do que a anterior, mesmo com essa pequena mudança na parametrização ainda continua dentro dos limites estabelecidos.

### **3.5 Escolha dos gráficos de controle**

A variação do  $\lambda$  faz com que a parametrização e os limites sofram alterações quanto aos valores e limites. Analisando os gráficos dos dois indicadores percebe-se que em nenhum caso os limites estabelecidos foram rompidos. Assim, tanto para  $\lambda=0,1$  e  $\lambda=0,2$  os indicadores apresentam-se sob controle estatístico. A variação dos parâmetros  $\lambda$  e  $L$  faz com que os limites de controle se tornem mais ou menos restritivos e sua escolha depende do interesse em se dar um peso maior ou menor para as observações mais recentes.

A escolha de  $\lambda$  é feita conforme o interesse do responsável pela parametrização.

### 3.6 Conclusões

Este trabalho foi elaborado utilizando dados do Instituto Federal de Brasília e usando conceitos de Séries Temporais e Controle Estatístico de Qualidade, os dois indicadores foram escolhidos considerando a qualidade das observações e a importância na área de informática do IFB e consistem no tempo médio da resolução de uma OS.

O banco de dados inicial considerava dias e horas corridas, foi realizada uma filtragem onde se levou em consideração apenas o tempo útil, o que é efetivamente trabalhado pelo servidor do Instituto. Os dados foram convertidos em minutos e iniciou-se a análise descritiva dos indicadores *Tempo para a Instalação de Impressora* e *Tempo para a Criação de e-mail de trabalho*.

Ao analisar inicialmente as duas séries notou-se que visualmente pareciam estáveis, entretanto apenas a classificação visual não é suficiente. Foi aplicado então o teste de Dickey-Fuller onde observando o *p-valor* para cada série e as hipóteses propostas concluiu-se que eram estacionárias, utilizando um nível de significância de 5%.

Após a hipótese de estacionaridade ser comprovada, o passo seguinte foi construir os gráficos de autocorrelação e autocorrelação parcial para cada série e propor um modelo ajustado. Este gráfico apresentou resultados semelhantes para os dois indicadores e a conclusão que as séries se tratavam de um Ruído Branco.

Com essa definição, alguns passos que seriam necessários como propor um modelo ajustado e trabalhar nos seus resíduos foram descartados, o gráfico de controle então seria construído utilizando a série inicial. Foi realizado ainda o teste de Ljung-Box com o objetivo de verificar a hipótese de que as séries eram Ruído Branco, hipótese essa que ficou comprovada.

Após a comprovação de que a série a ser utilizada na construção do gráfico de controle para os dois indicadores seria a inicial, foi realizado os testes de Shapiro-Wilk e Kolmogorov-Smirnov, estes testes tem o objetivo de comprovar ou não a normalidade dos dados. É necessária tal avaliação, pois os gráficos de controle Shewhart e CUSUM necessitam dessa condição. Após a rejeição de normalidade optou-se pelo gráfico de Médias Móveis Exponencialmente Ponderadas (MMEP).

Neste tipo de gráfico inicialmente deve-se parametrizar as observações utilizando um  $\lambda$  entre 0 e 1, foram testados dois gráficos, um com  $\lambda=0,1$  e  $L=2,7$  e outro com  $\lambda=0,2$  e  $L=2,8$ , a



observação seguinte depende da observação anterior. As fórmulas utilizadas para essa parametrização das observações e a construção dos limites foram dadas no Capítulo 2.

O gráfico com  $\lambda=0,1$  apresenta os limites mais próximos, ou seja, é um gráfico mais rígido, porém para as duas séries foram elaborados os dois gráficos variando o  $\lambda$  e mesmo assim a série ficou sob controle.

A escolha de  $\lambda$  e conseqüentemente a escolha do melhor gráfico de controle é feita conforme o interesse do responsável pela parametrização.

Esta metodologia aplicada nesse estudo é extensiva a outros indicadores, mas cada um deve ser tratado de maneira individual, para este estudo calhou dos dois indicadores em análise apresentarem resultados e aplicações semelhantes, mas toda a análise de séries temporais e controle estatístico de qualidade foi desenvolvida individualmente.

## Referências

- ALVES, C. *A estatística Média Móvel Exponencialmente Ponderada para o controle preditivo, monitoramento e ajuste de processos*, Congresso Latino-Ibericoamericano de Investigación Operativa. Rio de Janeiro. 2012
- BOX, G.E.P. & Jenkins, *Time Series Analysis, Forecasting and Control*, San Francisco, Holden-Day .1970, (revised edition, 1976).
- COELHO, D. *Utilizando Controle Estatístico de Qualidade em Um Processo Estratégico de um Instituto Federal de Educação*. 2015. 113 p. Dissertação (Mestrado em Computação Aplicada) – Programa de Pós-graduação da Ciência da Computação, Universidade de Brasília. 2015.
- DUNCAN, A. J. *Quality Control and Industrial Statistics*, 5th edition, Chicago: Richard D.Irwin, Inc.1986.
- 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.
- ISHIKAWA, K. 1915. *Controle da qualidade total: à maneira japonesa*, tradução Iliana Torres, 2ª Ed., Rio de Janeiro: Campus. 2000.
- LOPES, L. *Gráfico de Controle Estatístico de Qualidade para indicador estratégico da secretaria da Fazenda do governo do Piauí*, 2013. 48 p. Monografia (Bacharelado em Estatística) – Universidade de Brasília, 2013.
- 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.
- MORETTIN, P.A.; BUSSAB W. O. *Estatística Basica*. 8ª ed. São Paulo. 2013
- MORETTIN, P. A.; TOLOI C. M. de C. *Análise de Séries Temporais*. São Paulo: EdgardBlücher. 2006.

## Apêndice

### 1. Teste de Shapiro-Wilk

O teste Shapiro-Wilk foi proposto em 1965 e é baseado na estatística  $W$  dada por:

$$W = \frac{b^2}{\sum_{i=1}^n (x_{(i)} - \bar{x})^2}$$

Onde  $x_{(i)}$  são os valores da amostra ordenados em ordem crescente.

A constante  $b$ , numerador da equação acima, é determinada da seguinte forma:

$$b = \begin{cases} \sum_{i=1}^{n/2} a_{n-i+1} (x_{(n-i+1)} - x_{(i)}) & \text{se } n \text{ é par} \\ \sum_{i=1}^{(n+1)/2} a_{n-i+1} (x_{(n-i+1)} - x_{(i)}) & \text{se } n \text{ é ímpar} \end{cases}$$

Em que  $a_{n-i+1}$  são constantes geradas pelas médias, variâncias e covariâncias das estatísticas de ordem de uma amostra de tamanho  $n$  de uma distribuição Normal.

As hipóteses do teste são dadas por:

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

E a regra de decisão é rejeitar  $H_0$  ao nível de significância  $\alpha$  se  $W_{\text{calculado}} < W_{\text{tabelado}}$  ou se o  $p$ -valor encontrado é menor que o nível de significância.

## 2. Teste de Kolmogorov-Smirnov

O teste de Kolmogorov-Smirnov, assim como o teste de Shapiro-Wilk, tem a função de averiguar se uma amostra pode ser considerada como proveniente de uma população com uma determinada distribuição. O teste é particularmente indicado para distribuições contínuas. Este teste observa a máxima diferença absoluta entre a função de distribuição acumulada assumida para os dados, no caso a Normal, e a função de distribuição empírica dos dados.

As hipóteses do teste são:

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

Para realizar o cálculo da estatística do teste define-se  $F_n(x)$  como sendo uma função empírica que cresce  $\frac{1}{n}$  a cada observação.

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

Logo, a estatística do teste é dada por:

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

$F(x)$  é a função de distribuição acumulada. Sendo a função empírica descontínua, deve-se desconsiderar as estatísticas D:

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

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

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

A estatística do teste é dada por:

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

E a regra de decisão é rejeitar  $H_0$  ao nível de significância  $\alpha$  se  $D_n$  for superior ao valor crítico ou se o *p-valor* encontrado é menor que o nível de significância.

### 3. Programação utilizada

```
require(forecast)
```

```
require(fpp)
```

```
require(astsa)
```

```
require(MuMIn)
```

```
require(qcc)
```

```
impressora<-
```

```
c(300,520,74,84,341,667,330,364,418,723,329,290,55,306,664,211,316,345,188,491,41,  
177,365,399,485,283,50,310,273,405,265,274,284,284,73,50,700,152,151,193,399,314,  
547,18,254,40,222,419,829,58,46)
```

```
email<-
```

```
c(123,262,263,523,512,119,619,427,72,12,64,253,215,7,196,52,514,438,124,32,96,71,234,35  
2,431,43,39,54,25,251,113,690,468,79,61,265,463,286,542,186,515,769,78,305,114, 501)
```

```
plot.ts(impressora)
```

```
plot.ts(email)
```

```
boxplot(impressora)
```

```
boxplot(email)
```

```
acf(impressora)
```

```
pacf(impressora)
```

```
acf(email)
```

```
pacf(email)
```

```
##The Augmented Dickey-Fuller (ADF) t-statistic test: small p-values suggest the data is  
stationary and doesn't need to be differenced stationarity.##
```

```
adf.test(impressora, alternative = "stationary")
```

```
adf.test(email, alternative = "stationary")
```

##The Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test; here accepting the null hypothesis means that the series is stationarity, and small p-values suggest that the series is not stationary and a differencing is required.##

```
kpss.test(impressora)
```

```
kpss.test(email)
```

```
Box.test(impressora, type="Ljung")
```

```
Box.test(email, type="Ljung")
```

```
shapiro.test(impressora)
```

```
ks.test(impressora)
```

```
shapiro.test(email)
```

```
ks.test(email)
```

```
qqnorm(impressora)
```

```
qqline(impressora, col="red")
```

```
hist(impressora, prob=TRUE, 12)
```

```
lines(density(impressora),col="Red")
```

```
qqnorm(email)
```

```
qqline(email, col="red")
```

```
hist(email, prob=TRUE, 12)
```

```
lines(density(email),col="Red")
```