

**UNIVERSIDADE FEDERAL DE SANTA MARIA
CENTRO DE TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE
PRODUÇÃO**

**AVALIAÇÃO DA QUALIDADE NO PROCESSO DE
LINGOTAMENTO CONTÍNUO NA PRESENÇA DE
CORRELAÇÃO CRUZADA**

DISSERTAÇÃO DE MESTRADO

Meire Mezzomo

Santa Maria, RS, Brasil

2013

**AVALIAÇÃO DA QUALIDADE NO PROCESSO DE
LINGOTAMENTO CONTÍNUO NA PRESENÇA DE
CORRELAÇÃO CRUZADA**

Meire Mezzomo

Dissertação apresentada ao Curso de Mestrado do Programa de Pós-Graduação em Engenharia de Produção, Área de Concentração em Gerência, da Universidade Federal de Santa Maria (UFSM, RS), como requisito parcial para obtenção do grau de **Mestre em Engenharia de Produção**

Orientador: Prof. Dr. Adriano Mendonça Souza

Santa Maria, RS, Brasil

2013

**Universidade Federal de Santa Maria
Centro de Tecnologia
Programa de Pós-Graduação em Engenharia de Produção**

A Comissão Examinadora, abaixo assinada,
aprova a Dissertação de Mestrado

**AVALIAÇÃO DA QUALIDADE NO PROCESSO DE LINGOTAMENTO
CONTÍNUO NA PRESENÇA DE CORRELAÇÃO CRUZADA**

elaborada por
Meire Mezzomo

Como requisito parcial para obtenção do grau de
Mestre em Engenharia de Produção

COMISSÃO EXAMINADORA:

Adriano Mendonça Souza, Dr.
(Presidente/Orientador)

Roselaine Ruviaro Zanini, Dra. (UFSM)

Vanusa Andrea Casarin, Dra. (URI – Santo Ângelo)

Santa Maria, 25 de Julho de 2013.

AGRADECIMENTOS

Agradeço,

Primeiramente a Deus, por todas as oportunidades que foram oferecidas em minha vida, pela saúde, pela coragem, por nos fazer acreditar que somos capazes de superar qualquer obstáculo, mesmo aqueles que em muitas vezes nós os consideramos impossíveis.

À Universidade Federal de Santa Maria pela singular oportunidade em fazer parte do Programa de Pós-Graduação em Engenharia de Produção e a CAPES pela grande ajuda financeira a mim conferida.

Ao meu Orientador Adriano Mendonça Souza pelos conhecimentos transmitidos, pela paciência e pelo incentivo ao meu trabalho.

Em especial, agradeço ao professor Jose Gomes Requeijo, pelos imensos ensinamentos a mim transmitidos, pela paciência, pelo carinho e pela grande contribuição para o desenvolvimento do trabalho.

Agradeço, com muito carinho, à professora Roselaine Ruviaro Zanini pela disponibilidade, presteza, dedicação e incentivo.

Agradeço também à professora Anaelena Bragança de Moraes pelos ensinamentos compartilhados, pelo carinho, pela confiança, estando sempre presente nos momentos difíceis em minha vida.

A professora Vanusa Andrea Casarin, pela gentileza de colaborar em meu trabalho, pela sua amizade, presteza e carinho.

Aos professores e toda equipe do PPGEP, que demonstraram competência em transmitir seus conhecimentos e contribuir para meu crescimento profissional e intelectual.

Ao meu irmão Régis Mezzomo pela confiança, pela atenção e carinho nos momentos de aflição e por me ajudar sempre!

As minhas estimadas amigas, Tonia Moraes Brum, Ysabel Fernández Campos, Angela Arrua, e Leda Freitas, pela paciência, carinho, estímulo e auxílio nos momentos difíceis, vocês são muito especiais e fizeram grande diferença em minha vida, obrigada!

Finalmente, gostaria de deparar meus sinceros agradecimentos aos meus queridos colegas do mestrado, Lizandra Salau, Tatiana Fernanda Mousquer e Acássio Feliciani pelo carinho, apoio, companheirismo e por todo aprendizado que eles me proporcionaram.

RESUMO

Dissertação de Mestrado
Programa de Pós-Graduação em Engenharia de Produção
Universidade Federal de Santa Maria

AValiação DA QUALIDADE DO PROCESSO DE LINGOTAMENTO CONTÍNUO NA PRESENÇA DE CORRELAÇÃO CRUZADA

AUTORA: MEIRE MEZZOMO

ORIENTADOR: ADRIANO MENDONÇA SOUZA

Data e Local da Defesa: Santa Maria, 25 de Julho de 2013.

No atual mercado competitivo, grande parte das empresas tem como principal objetivo a busca da melhoria contínua dos seus produtos e serviços. Assim, a aplicação de métodos estatísticos apresenta grande relevância na avaliação da qualidade, auxiliando na compreensão e monitoramento de processos. Nesse contexto, o presente estudo aborda a utilização de gráficos de controle multivariados na avaliação do processo produtivo na presença de correlação cruzada, cujo objetivo é verificar a estabilidade do processo de lingotamento contínuo na fabricação de tarugos de aço por meio do gráfico de controle multivariado T^2 de *Hotelling* aplicado nos resíduos estimados de modelos matemáticos lineares. Inicialmente, foi verificada a existência de autocorrelação nos dados, sendo necessária a utilização da modelagem ARIMA, pois quando isso ocorre, deve-se proceder à determinação dos resíduos e aplicar os gráficos de controle multivariados aos resíduos e não nas variáveis originais. A existência de correlação cruzada mostrou-se significativa entre as variáveis, sendo um dos pressupostos para a aplicação da estatística T^2 . Verificada a instabilidade no gráfico T^2 , buscaram-se identificar a variável ou conjunto de variáveis das temperaturas do aço no distribuidor e peso do distribuidor, responsáveis pela instabilidade. Posteriormente, os resíduos estimados foram decompostos em componentes principais, e com o auxílio da correlação entre as variáveis originais e as componentes principais, identificou-se as variáveis que mais contribuíram para a formação de cada componente. Assim, foi possível detectar as variáveis causadoras da instabilidade do sistema, sendo que para as temperaturas do aço no distribuidor foram as temperaturas T_4 e T_5 , seguidas de T_6 , T_3 , T_7 e T_2 e para o peso do distribuidor, PD_4 , PD_5 , PD_3 , PD_6 e PD_2 , respectivamente. Deste modo, os resíduos estimados oriundos dos modelos matemáticos, a aplicação dos gráficos de controle multivariados T^2 de *Hotelling* e a decomposição em componentes principais foram capazes de representar o processo produtivo. Esta metodologia possibilitou a compreensão do comportamento das variáveis e auxiliou no monitoramento do processo, bem como, na determinação das possíveis variáveis causadoras da instabilidade no processo de lingotamento contínuo.

Palavras Chave: Lingotamento Contínuo. Correlação Cruzada. Modelos ARIMA. Gráfico de Controle Multivariado T^2 de *Hotelling*. Componentes Principais.

ABSTRACT

Mastership Dissertation
Post-graduation Program in Production Engineering
Federal University of Santa Maria

QUALITY EVALUATION OF CONTINUOUS CASTING PROCESS IN PRESENCE OF CROSS-CORRELATION

AUTHOR: MEIRE MEZZOMO

ADVISOR: ADRIANO MENDONÇA SOUZA

Date and place of defense: Santa Maria, July 25th, 2013.

In the current competitive market, a great part of companies has as the main goal the search for continuous improvement of their products and services. Therefore, the application of statistical methods has great relevance in the quality evaluation, helping in the understanding and monitoring of the processes. In such context, the present study concerns to the use of multivariate control charts in the evaluation of the productive processes in the presence of cross-correlation, which the objective is to verify the continuous casting process stability in the production of still billets by means of Hotelling's T^2 multivariate control charts applied in the estimated residual mathematical linear models. Initially, the existence of data autocorrelation was verified, it is necessary the ARIMA modeling, because when it happens, it is necessary to determine the residues and apply multivariate control charts to the residues and not on the original variables. The existence of correlation showed to be meaningful among the variables, being one of the assumptions for the statistical application T^2 . When the T^2 chart instability is verified, it was necessary to identify the variable or the set of variables of steel temperatures in the distributor and in the distributor weight, which are responsible for the instability. Later, the estimated residues were decomposed into principal components, and with the help of the correlation of the original variables and the principal components, the variables which most contributed to the formation of each component were identified. Therefore, it was possible to detect the variables which caused the system instability, once for the steel temperature in the distributor were the T_4 and T_5 , followed by T_6 , T_3 , T_7 and T_2 and for the weight of the distributor, PD_4 , PD_5 , PD_3 , PD_6 and PD_2 , respectively. This way, the estimated residues from the mathematical models, the use of multivariate chart control Hotelling's T^2 and the decomposition into principal components which were able to represent the productive process. This methodology allowed the understanding of the behavior of the variables and helped the monitoring of this process, as well as, in the determination of the possible variables which caused the instability in the continuous casting process.

Key-Words: Continuous Casting. Cross-Correlation. ARIMA Models. Multivariate Control chart Hotelling's T^2 . Principal Components.

LISTA DE FIGURAS

| | |
|--|----|
| Figura 1- Representação esquemática do processo de Lingotamento Contínuo . | 19 |
| Figura 2- Filtro linear com entrada at, saída Z_t | 23 |
| Figura 3 - Comportamento das funções de autocorrelação (FAC) e autocorrelação parcial (FACP) de um modelo AR(1). | 28 |
| Figura 4 - Comportamento da função de autocorrelação (FAC) e função de autocorrelação parcial (FACP) de um modelo MA (1). | 30 |
| Figura 5 – Representação genérica de gráficos de controle. | 45 |
| Figura 6 – Fase I e Fase II do controle estatístico do processo multivariado. | 51 |
| Figura 7 – Gráfico de controle multivariado T^2 dos resíduos originais estimados dos modelos ARIMA das variáveis temperaturas do aço no distribuidor, na Fase I do CEPM. | 76 |
| Figura 8 – Gráfico de controle multivariado T^2 dos resíduos substituídos pelas médias dos resíduos das variáveis temperaturas do aço no distribuidor, na Fase I do CEPM. | 77 |
| Figura 9 – Gráfico de controle multivariado T^2 dos resíduos excluídos fora dos limites de controle, das variáveis temperaturas do aço no distribuidor, na Fase I do CEPM. | 78 |
| Figura 10 – Gráfico de controle multivariado T^2 dos resíduos excluídos fora dos limites de controle, das variáveis temperaturas do aço no distribuidor, na Fase II do CEPM. | 79 |
| Figura 11 – Gráfico de controle multivariado T^2 dos resíduos originais estimados dos modelos ARIMA das variáveis pesos do distribuidor, na Fase I do CEPM. | 79 |
| Figura 12 – Gráfico de controle multivariado T^2 dos resíduos substituídos pelas médias dos resíduos das variáveis pesos do distribuidor, na Fase I do CEPM. | 80 |
| Figura 13 – Gráfico de controle multivariado T^2 dos resíduos excluídos fora dos limites de controle das variáveis pesos do distribuidor, na Fase I do CEPM. | 81 |
| Figura 14 – Gráfico de controle multivariado T^2 dos resíduos excluídos fora dos limites de controle das variáveis pesos do distribuidor, na Fase II do CEPM. | 82 |

LISTA DE TABELAS

| | |
|--|----|
| Tabela 1 – Valor mínimo de m | 57 |
| Tabela 2 – Análise descritiva das variáveis originais das temperaturas do aço no distribuidor em graus centígrados (°C), (n = 228)..... | 68 |
| Tabela 3 – Análise descritiva das variáveis originais dos pesos do distribuidor em toneladas (t), (n = 228). | 69 |
| Tabela 4 – Análise da correlação existente entre as variáveis das temperaturas do aço no distribuidor (n = 228)..... | 70 |
| Tabela 5 – Análise da correlação existente entre as variáveis dos pesos do distribuidor (n = 228)..... | 71 |
| Tabela 6 - Resultados dos testes de raiz unitária (ADF e KPSS) para a série em nível e primeira diferença das temperaturas do aço no distribuidor (n = 228)..... | 72 |
| Tabela 7 – Resultado dos testes de raiz unitária (ADF e KPSS) para a série em nível dos pesos do distribuidor (n = 228)..... | 73 |
| Tabela 8 – Modelos ARIMA selecionados para as variáveis temperaturas do aço no distribuidor em graus centígrados (°C), (n = 228). | 74 |
| Tabela 9 – Modelos ARIMA selecionados para as variáveis pesos do distribuidor em toneladas (t), (n = 228). | 75 |
| Tabela 10 – Autovalores e variância explicada de cada componente dos resíduos da variável temperatura do aço no distribuidor | 83 |
| Tabela 11 – Autovalores e variância explicada de cada componente dos resíduos da variável peso do distribuidor. | 83 |
| Tabela 12 – Carga fatorial de cada componente dos resíduos das variáveis temperaturas do aço no distribuidor e peso do distribuidor | 84 |

LISTA DE APÊNDICES

| | |
|---|-----------|
| APÊNDICE A – Gráficos das variáveis originais das temperaturas do aço no distribuidor..... | 93 |
| APÊNDICE B – Gráficos das variáveis originais dos pesos do distribuidor..... | 94 |
| APÊNDICE C – FAC e FACP das variáveis originais das temperaturas do aço no distribuidor..... | 96 |
| APÊNDICE D – FAC e FACP das variáveis originais dos pesos do distribuidor..... | 97 |

LISTA DE ABREVIATURAS E SIGLAS

ACP – Análise de Componentes Principais
ADF – Teste de Raiz Unitária *Dickey-Fuller* Aumentado (*Augmented Dickey-Fuller test statistic*)
AIC – Critério de Informação *Akaike* (*Akaike Information Criterion*)
ARL – Comprimento Médio de Sequências (*Average Runs Length*)
AR – Modelo Autorregressivo
ARI – Modelo Autorregressivo Integrado
ARMA – Modelo Autorregressivo e de Médias Móveis
ARIMA – Modelo Autorregressivo Integrado de Médias Móveis (*Auto Regressive Integrated Moving Averages*)
BIC – Critério de *Schwarz* (*Bayesian Information Criterion*)
CEP – Controle Estatístico do Processo
CEPM – Controle Estatístico do Processo Múltiplo ou Multivariado
CP – Componentes Principais
 d – Número de diferenças
DF – Teste de Raiz Unitária *Dickey-Fuller*
F – Estatística “F” de *Snedecor*
FAC – Função de Autocorrelação
FACP – Função de Autocorrelação Parcial
GC – Gráficos de Controle
IC – Intervalo de Confiança
i.i.d. – Independentes e Identicamente Distribuídas
IMA – Modelo de Médias Móveis Integradas
KPSS – Teste de Raiz Unitária proposto por *Kwiatkowski-Phillips-Schmidt-Schin*
LIC – Limite Inferior de Controle
LSC – Limite Superior de Controle
MA – Modelo de Médias Móveis
mm – Milímetros
 p – Número de Parâmetros da parte Autorregressiva
 q – Número de Parâmetros da parte de Médias Móveis
 $Q(k)$ – Estatística *Q* de *Ljung-Box*
RB – Ruído Branco
S – Matriz de Variância-Covariância
 S^{-1} – Matriz Inversa de Variância-Covariância
 T^2 – Estatística ou Gráfico Multivariado T^2 de *Hotelling*
 t – Toneladas
 χ^2 – Distribuição Qui-quadrado
 \bar{X} – Vetor Média

SUMÁRIO

| | |
|---|----|
| INTRODUÇÃO | 12 |
| Tema da pesquisa | 13 |
| Justificativa e importância da pesquisa | 14 |
| Objetivos | 14 |
| Objetivo geral..... | 15 |
| Objetivos específicos | 15 |
| Delimitação da pesquisa | 15 |
| Estrutura do trabalho | 16 |
| REVISÃO DE LITERATURA | 17 |
| Características da empresa | 17 |
| Processo de lingotamento contínuo | 18 |
| Análise de séries temporais | 20 |
| Modelos de <i>Box & Jenkins</i> | 22 |
| Modelo autorregressivo (AR)..... | 26 |
| Modelo de média móvel (MA)..... | 28 |
| Modelo autorregressivo e de médias móveis (ARMA)..... | 30 |
| Processo estocástico não estacionário - modelo autorregressivo integrado e de médias móveis (ARIMA)..... | 32 |
| Procedimentos para a análise de séries temporais | 33 |
| Análise gráfica..... | 34 |
| Teste de <i>Ljung-Box</i> , função de autocorrelação (FAC) e função de autocorrelação parcial (FACP)..... | 34 |
| Teste Dickey-Fuller..... | 37 |
| Teste Dickey-Fuller aumentado..... | 39 |
| Teste de Kwiatkowski-Phillips-Schmidt-Shin (KPSS).... | 40 |
| Crterios de informação | 42 |
| Diagnóstico de resíduo | 43 |
| Gráficos estatísticos de controle | 44 |
| Gráficos de controle com dados autocorrelacionados | 48 |
| Gráfico de controle multivariado T^2 de Hotelling | 48 |
| Gráficos de controle multivariados baseados em observações individuais..... | 52 |
| Interpretação dos gráficos de controle T^2 de Hotelling | 58 |
| Comentários gerais do capítulo | 60 |
| METODOLOGIA | 62 |
| Caracterização do estudo | 62 |
| Coleta dos dados | 62 |
| Característica e ajuste dos dados | 63 |
| Etapas da metodologia | 64 |
| ANÁLISE E DISCUSSÃO DOS RESULTADOS | 68 |
| Análise exploratória dos dados | 68 |
| Análise da correlação e autocorrelação nas variáveis | 70 |
| Modelagem de <i>Box-Jenkins</i> das variáveis | 71 |
| Gráfico de controle multivariado T^2 de Hotelling | 76 |
| Decomposição dos resíduos em componentes principais | 82 |
| Comentários gerais do capítulo | 85 |

| | |
|--|-----------|
| CONSIDERAÇÕES FINAIS E RECOMENDAÇÕES..... | 87 |
| Considerações Finais..... | 87 |
| Recomendações para estudos futuros..... | 88 |
| REFERÊNCIAS BIBLIOGRÁFICAS..... | 90 |

1 INTRODUÇÃO

A crescente competitividade mundial induziu as organizações a investirem, cada vez mais, em métodos e tecnologias modernas e eficazes visando aprimorar os seus processos e serviços. Em decorrência desses novos investimentos, ocorreram significativas mudanças no cenário industrial, sendo que os processos e serviços passaram por avaliações mais específicas e criteriosas. Com essa nova exigência a utilização do controle de qualidade passou a ser um elemento chave para as organizações, aumentando assim, a complexidade dos processos produtivos e conseqüentemente, o controle da qualidade.

Nesse contexto, o controle da qualidade passou a ter destaque e tornou-se um fator de decisão básico para o crescimento e sucesso de uma empresa ou negócio.

De acordo com Pereira e Requeijo (2008), de maneira geral, podem-se mencionar três fatores relevantes na competitividade de uma empresa: qualidade, produtividade e os custos de operação, que influenciam diretamente no custo do produto final e conseqüentemente no seu preço de venda. Nesse sentido, as técnicas de controle estatístico do processo (CEP), têm sido amplamente utilizadas pelas empresas de sucesso, com o intuito de aprimorar, desenvolver e produzir produtos e serviços com melhor qualidade a custos compatíveis para que os produtos se tornem mais competitivos.

Uma das ferramentas mais utilizadas no controle estatístico do processo são os gráficos de controle de *Shewhart*, os quais tornam presumível distinguir os possíveis tipos de variação em um processo e conseqüentemente a identificação de diagnósticos e admissíveis correções para a solução de problemas que, eventualmente, podem advir nos sistemas produtivos.

Conforme Montgomery (1997), dentre as diversas variedades de gráficos de controle, os gráficos univariados são os mais conhecidos e, habitualmente, utilizados no setor industrial, devido a sua simplicidade e facilidade de operacionalização. Entretanto, grande parte dos panoramas de controle de processos envolvem uma ou mais variáveis de interesse. Nesse contexto, torna-se necessário a utilização de métodos estatísticos multivariados, que consideram conjuntamente todas as variáveis envolvidas em um processo. Essa ferramenta do controle estatístico do processo utiliza como pressuposto básico, a correlação existente entre as características da qualidade, que são monitoradas de forma conjunta, permitindo assim, a

otimização do controle do processo, devido à detecção mais eficiente e rápida de possíveis causas de instabilidade num sistema produtivo.

Um dos pressupostos fundamentais para a aplicação do CEP é a independência das observações tanto para o estudo univariado quanto para o multivariado. Dessa forma, quando se verifica a existência de correlação cruzada significativa nos dados, deve-se proceder, a uma análise criteriosa do processo e utilizar formas alternativas para avaliá-lo.

A estatística multivariada possibilita analisar a relação existente em grandes quantidades de dados industriais sistematicamente. Nas situações onde existe autocorrelação significativa nos dados do processo, Montgomery (1997), sugere modelar a autocorrelação de modo que seja possível construir gráficos mais confiáveis e eficazes, sendo que, a abordagem habitual utilizada consiste em modelar o processo por meio de modelos matemáticos, denominados de *Box-Jenkins*.

No processo de lingotamento contínuo, para que se obtenham produtos com a qualidade desejada, faz-se necessário, o monitoramento constante, bem como a determinação dos parâmetros do processo, de maneira que, todas as condições críticas sejam satisfeitas.

Assim, inicialmente, busca-se modelar um conjunto de variáveis como a temperatura do aço no distribuidor e o peso do distribuidor no processo de lingotamento contínuo por meio de modelos de previsão e realizar o monitoramento utilizando-se gráficos de controle multivariado T^2 de Hotelling, levando-se em consideração a média do processo e a análise conjunta das variáveis em estudo. Posteriormente, caso se verifique pontos fora de controle no gráfico T^2 , um diagnóstico auxiliar será aplicado aos resíduos estimados, por meio da decomposição em componentes principais (CP). Desse modo, será possível analisar o processo produtivo de fundição de aço semi-acabado e verificar a influência de cada variável em relação às demais, bem como, avaliar a estabilidade do processo produtivo de lingotamento contínuo.

1.1 Tema da pesquisa

O tema da presente pesquisa é a utilização de técnicas de gráficos de controle em processos produtivos que contemplem características multivariadas que apresentem características de autocorrelação nas observações e correlação significativa entre as variáveis e busca-se avaliar as características relativas à qualidade simultaneamente. Dessa forma, o

estudo se concentra na aplicação de gráficos de controle nos resíduos estimados por meio da modelagem ARIMA das variáveis envolvidas no processo.

1.2 Justificativa e importância da pesquisa

O controle estatístico do processo é uma metodologia habitualmente utilizada nas organizações e atua preventivamente nos processos produtivos, utilizando a estatística como instrumento básico. Neste contexto, torna-se necessário desenvolver métodos eficientes para avaliá-los, e conseqüentemente reduzir significativamente os desperdícios de mão-de-obra, matéria-prima, e custos adicionais, quando os produtos ou processos não estiverem de acordo com as especificações exigidas.

O propósito primordial da pesquisa é avaliar a estabilidade do processo produtivo, tratando tanto a autocorrelação revelada pela dependência entre os instantes de tempo e a correlação atrelada pela relação existente entre as variáveis.

Com isso, considera-se este estudo relevante, tanto para a área acadêmica, quanto para os profissionais que atuam nas indústrias. Na área acadêmica, por utilizar uma metodologia eficiente na aplicação das técnicas de controle de qualidade multivariada e nas indústrias como uma nova ferramenta de avaliação do processo produtivo, por meio da monitoração contínua, possibilitando assim, a identificação das variáveis que se encontram fora de controle em um sistema produtivo.

A metodologia sugerida propicia um controle eficaz no monitoramento da qualidade utilizando simultaneamente um conjunto de variáveis, possibilitando a redução dos custos de produção, bem como, o número de itens produzidos não conformes, além de detectar as possíveis variáveis responsáveis pela instabilidade no sistema. Nesse sentido, a metodologia em estudo poderá ser utilizada como uma forma alternativa de controle e monitoramento de processos produtivos, quando existir a presença de correlação nos dados, refletindo em produtos com maior qualidade e em indústrias mais eficientes e competitivas.

1.3 Objetivos

1.3.1 Objetivo geral

Avaliar a estabilidade do processo produtivo de lingotamento contínuo na presença de correlação cruzada na etapa de fabricação de tarugos de aço por meio de gráficos de controle multivariados.

1.3.2 Objetivos específicos

- Analisar a estabilidade do processo de lingotamento contínuo, por meio de gráficos de controle multivariados utilizando as variáveis originais e os resíduos estimados dos modelos de *Box- Jenkins*;
- Detectar as possíveis variáveis causadoras da instabilidade no processo produtivo, por meio dos resíduos estimados aplicados nos gráficos de controle multivariados T^2 de *Hotelling* e na decomposição em componentes principais;
- Desenvolver uma metodologia estatística multivariada para auxiliar no monitoramento de processos produtivos de lingotamento contínuo.

1.4 Delimitação da pesquisa

A presente pesquisa restringe-se ao estudo de gráficos de controle multivariados aplicados no processo de lingotamento contínuo na fabricação de tarugos de aço, sendo que, por questões éticas, o nome e a identificação da empresa não serão revelados. Entretanto, a análise dos custos inferidos aos diferentes níveis das características analisadas, não se apresentam inclusos neste trabalho.

Assim, espera-se que, por meio da análise da estabilidade e do monitoramento do processo, atinja-se uma melhor qualidade dos produtos produzidos e, conseqüentemente, uma redução nos custos de produção. Ressalta-se que, o sistema produtivo, no todo, é composto por diversas variáveis, não citadas e avaliadas no presente estudo, sendo utilizados apenas

dois conjuntos de variáveis, as temperaturas do aço no distribuidor e os pesos do distribuidor, para validar a técnica proposta.

1.5 Estrutura do trabalho

Para atingir os objetivos propostos, a pesquisa foi estruturada em cinco capítulos, descritos a seguir.

O primeiro capítulo apresenta a introdução do estudo, a qual contempla o tema, a justificativa e a importância, os objetivos geral e específicos, a delimitação e a estrutura da pesquisa.

O segundo capítulo, apresenta o referencial teórico que descreve as características da empresa, o processo de lingotamento contínuo, as técnicas de controle de qualidade e da modelagem matemática bem como, o gráfico de controle multivariado T^2 de Hotelling e a Análise de Componentes Principais (ACP).

O terceiro capítulo aborda a metodologia do trabalho e as técnicas utilizadas para o seu desenvolvimento.

O quarto capítulo aborda a análise e discussão dos resultados obtidos.

O quinto e último capítulo apresenta as conclusões desse trabalho, além de sugestões para futuras pesquisas.

Na sequência, apresentam-se as referências bibliográficas.

2 REVISÃO DE LITERATURA

Neste capítulo descrevem-se os referenciais teóricos relativos às metodologias utilizadas para a elaboração da pesquisa, com o intuito de alcançar os objetivos propostos. São apresentados conceitos relevantes como, as características da empresa em estudo, o processo de lingotamento contínuo, as técnicas de análise de séries temporais, os modelos de *Box-Jenkins*, os procedimentos para a análise de séries temporais, os critérios de informação, o diagnóstico de resíduos, os gráficos estatísticos de controle, os gráficos de controle com dados autocorrelacionados, os gráficos de controle multivariados T^2 de *Hotelling* e a decomposição de T^2 de *Hotelling* em componentes principais (CP).

2.1 Características da empresa

A empresa em estudo utiliza o processo de lingotamento contínuo e é líder no segmento de aços longos nas Américas, sendo uma das principais fornecedoras mundiais de aços longos especiais. Conforme descrito no *site* oficial da empresa, com mais de 45 mil colaboradores, possui operações industriais em 14 países nas Américas, na Europa e na Ásia, as quais somam uma capacidade instalada superior a 25 milhões de toneladas por ano. É a maior recicladora da América Latina e, no mundo, transforma, anualmente, milhões de toneladas de sucata em aço, reforçando seu compromisso com o desenvolvimento sustentável das regiões onde atua.

Com uma ampla linha de produtos, comercializados para os cinco continentes, a empresa atende os setores da construção civil, indústria, e agropecuário. Além disso, é líder mundial no fornecimento de aços longos especiais para a indústria automotiva. No Brasil, o aço da empresa faz parte da construção e modernização de oito estádios de futebol para Copa do Mundo de 2014, bem como de importantes obras de infraestrutura para o País, como ferrovias, usinas eólicas, portos e estradas.

Na posição de maior recicladora da América Latina, a empresa transforma anualmente mais de 15 milhões de toneladas de sucata ferrosa em aço, cuja principal matéria-prima é a sucata. A utilização de sucata para o processo produtivo do aço contribui para a preservação

do meio ambiente, à medida que reduz o consumo de energia e de outros insumos, como o minério de ferro e o coque, e, por consequência, reduz as emissões de CO₂ no meio ambiente. Grande parte da sucata reciclada provém de materiais obsoletos de consumo doméstico e de processos produtivos industriais.

A empresa tem construído ao longo de sua história, uma prática permanente da responsabilidade social em todos os seus processos. O tema está presente na atitude dos seus colaboradores, que seguem valores éticos, e como cidadãos, contribuem de forma efetiva com a busca de soluções para os desafios da sustentabilidade. Com isso, desenvolve programas sociais e parcerias com entidades representativas da sociedade. Sua atuação está focada na educação formal, na educação pela qualidade, produtividade e competitividade, na educação pela cultura e esporte, educação ambiental e mobilização solidária.

2.2 Processo de lingotamento contínuo

O lingotamento contínuo é o processo pelo qual o metal fundido é solidificado em um produto semi-acabado no formato de tarugo, bloco ou placa. O produto final do processo de lingotamento contínuo é chamado de lingote (GARCIA et al., 2006).

Inicialmente a sucata é transportada para o forno panela onde o aço é fundido. Após a fundição, o aço escorre para o distribuidor sendo que, nesta etapa, podem-se efetuar as medições, tanto de velocidade, como de temperatura do aço fundido.

O processo de lingotamento contínuo é constituído por um sistema de fluxo que transporta o aço líquido aquecido com a vazão desde a panela, passando pelo distribuidor até o molde, que além de atribuir a forma ao lingote, possui a função de absorver o calor do metal. Assim, a massa de aço líquida é resfriada ao passar por um molde de cobre refrigerado a água. Em seguida, a massa é extraída em um sistema de rolos até deixar o molde, onde segue para a segunda etapa de resfriamento, que é feita por aspensão de água em *sprays* e pela perda de calor por radiação, até que toda a massa se solidifique (GARCIA et al., 2006).

Na Figura 1 é demonstrado o processo descrito anteriormente:

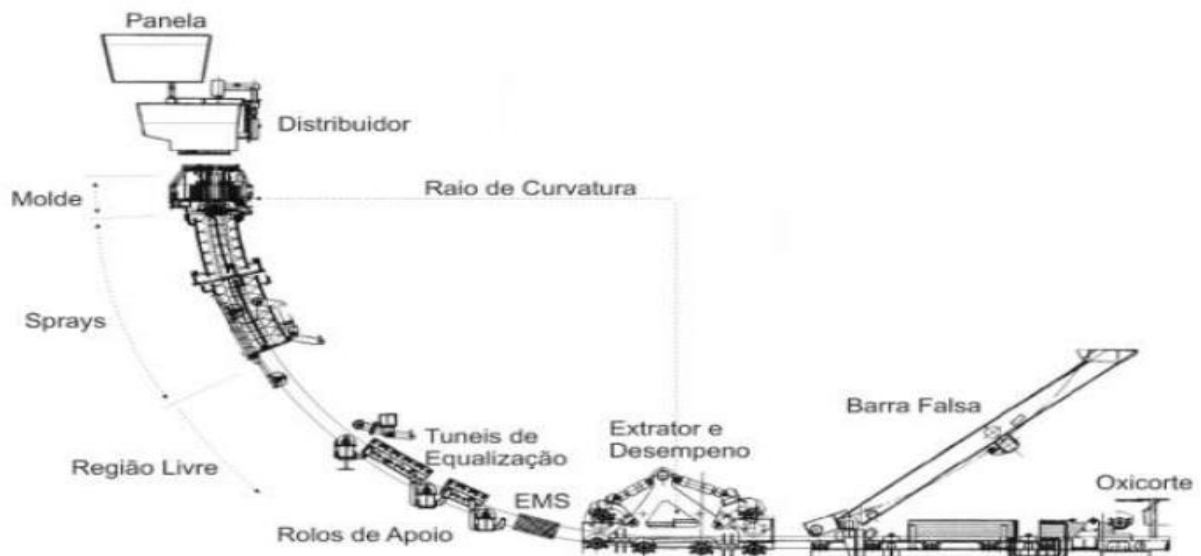


Figura 1- Representação esquemática do processo de Lingotamento Contínuo .

Fonte: CASARIN, 2012a, p. 26.

O tarugo ideal é um material homogêneo físico e quimicamente, sem inclusões, cavidades, porosidade, entre outros. Entretanto, para obter um tarugo ideal no processo de solidificação, deve-se avaliar e conhecer as características do mesmo, a fim de reduzir ao máximo a ocorrência de defeitos, como defeitos superficiais, como trincas e/ou dobras.

Com o intuito de verificar se o tarugo lingotado é produzido com a qualidade desejada, torna-se necessário conhecer e controlar as etapas do processo, desde a entrada das sucatas (caracterizadas por sobras da própria empresa e/ou de outras empresas), até a saída do tarugo no oxicorte e, inclusive toda a mecânica de funcionamento dos equipamentos.

Como a principal finalidade do estudo é avaliar a estabilidade do processo produtivo de lingotamento contínuo na etapa de fabricação de tarugos, é importante ressaltar que os defeitos podem ser ocasionados por problemas mecânicos e/ou metalúrgicos durante o lingotamento. Assim, as falhas no processo de fabricação do aço nas áreas de refino, podem ser causadas em função de um conjunto de variáveis, como a temperatura do aço no distribuidor, o peso do distribuidor, a velocidade dos veios, a composição química, dentre outras.

Os principais defeitos podem ser ocasionados pelo efeito de fatores de composição do aço como, a concentração de carbono ou elementos de liga; a falta de uniformidade no resfriamento do lingote na região do molde; a oscilação do molde envolvendo o movimento do mesmo, o pó fluxante, a transferência de calor e nível do menisco; o alinhamento dos rolos onde ocorre o final da solidificação entre inúmeros outros fatores (CASARIN et al., 2012b).

Neste contexto, os defeitos resultantes do processo de lingotamento contínuo podem refletir em grandes dispêndios à empresa. Assim, faz-se necessário avaliar e analisar as possíveis variáveis causadoras dessas não conformidades, pois as mesmas refletem diretamente na instabilidade do processo produtivo.

O lingotamento contínuo é considerado o maior processo responsável pela produção de aço no mundo. Este processo siderúrgico apresenta relevante importância na atualidade, pelo fato de ser responsável por 90% da produção global de aço e ao seu alto grau de rendimento. Assim, conhecer os mecanismos do processo de lingotamento contínuo é fundamental, pois este processo é de extrema importância no mercado atual.

2.3 Análise de séries temporais

Conforme enfoque de Morettin e Toloí (2004), uma série temporal é qualquer conjunto de observações ordenadas ou equidistantes no tempo, que devido a essa ordenação cronológica, apresentam o efeito da autocorrelação entre as observações. As séries temporais podem ser classificadas como discretas (normalmente as séries são observadas em instantes de tempo igualmente espaçados) ou contínuas (utiliza-se geralmente a amostragem em intervalos de tempo iguais, transformando a série de intervalar para discreta).

A análise de séries temporais tem como objetivo principal a realização de previsões, sendo que essa análise permite que valores futuros de uma série sejam previstos, tomando por base apenas seus valores presentes e/ou passados, por meio da correlação temporal que existe geralmente entre os valores exibidos pela série. A análise permite também, a investigação do mecanismo gerador da série temporal, descrevendo o comportamento da série.

O estudo das características das séries temporais, baseados na análise de seus valores passados, explicam os valores futuros em função das suas autocorrelações, sendo este o diferencial em relação às demais modelagens, como por exemplo, análise de regressão.

Conforme Morettin e Toloí (2004), uma das pressuposições mais frequentes para o estudo de uma série temporal é que a série a ser analisada seja estacionária, ou seja, a série deve se distribuir aleatoriamente em torno de uma média constante, admitindo certa estabilidade ao longo do tempo. Dessa forma, garante-se que os parâmetros estimados para essa série sejam válidos para todo o período.

Em 1976 George Box e Gwilyn Jenkins desenvolveram um método de análise de séries temporais que se tornou mundialmente conhecido como metodologia de *Box & Jenkins*. A relação temporal considerada pelo enfoque de *Box-Jenkins* é representada formalmente por um conjunto de processos estocásticos, sendo que a previsão é um dos principais fatores da popularidade dos Modelos *Box-Jenkins*, genericamente denominados por Modelos Autorregressivos Integrados de Médias Móveis – ARIMA (SOUZA et al., 2011).

Os modelos ARIMA resultam da combinação de três componentes denominados de “filtros”: o componente Autorregressivo (AR), o componente de Integração (I) e o componente de Médias Móveis (MA). Uma série temporal pode conter os três filtros ou apenas um subconjunto deles, possibilitando várias alternativas de modelos passíveis de análise por meio dessa metodologia. Uma alternativa da combinação dos filtros é o modelo autorregressivo e de médias móveis (ARMA), e também, o modelo autorregressivo integrado de médias móveis (ARIMA). Dessa forma, encontra-se o processo gerador da série em estudo e conforme as pressuposições básicas para o modelo torna-se viável fazer previsões para a série de dados (VASCONCELLOS e ALVES, 2002).

A metodologia de *Box-Jenkins* (1976) consiste em ajustar os modelos ARIMA (p,d,q) a um conjunto de observações ao longo do tempo, onde p representa o número de parâmetros da parte autoregressiva, d representa o número de diferenças aplicados na série para torná-la estacionária e q representa o número de parâmetros da parte de médias móveis.

Segundo Morettin e Tolo (2004), a construção dos modelos de *Box-Jenkins* é baseada em etapas, sendo que estas etapas são altamente dependentes entre si, e a escolha do modelo é feita com base nos próprios dados. Os passos relativos ao desenvolvimento desta metodologia encontram-se descritos a seguir:

1º) Identificação: consiste em determinar os valores adequados de p (número de parâmetros autorregressivos), q (número de parâmetros de médias móveis) e d (número de diferenciações da série caso seja não estacionária) que compõem o processo gerador da série, bem como quais são as suas respectivas ordens, por meio do gráfico da função de autocorrelação (FAC) e do gráfico da função de autocorrelação parcial (FACP);

2º) Estimação: uma vez determinados os valores adequados de p , d e q , passa-se para a estimação dos p parâmetros ϕ (autorregressivos) e dos q parâmetros θ e (médias móveis) incluídos no modelo;

3º) Verificação: esta etapa consiste em verificar se o modelo estimado é adequado. Em caso positivo, pode-se adotá-lo para fazer previsões, em caso negativo, outras especificações devem ser adotadas para modelar a série, o que implica em refazer as etapas de identificação e

estimação novamente. Em suma, esta etapa verifica se o modelo escolhido se ajusta razoavelmente aos dados. As formas de verificação comumente utilizadas são a análise de resíduos e a avaliação da ordem do modelo.

2.4 Modelos de *Box & Jenkins*

Conforme aborda Gujarati (2000), um processo é estocástico ou aleatório quando se tem um conjunto de variáveis aleatórias ordenadas no tempo. Denomina-se um processo estocástico estacionário aquele no qual sua média e variância são constantes ao longo do tempo e quando o valor da covariância entre dois períodos de tempo depende apenas da distância da defasagem, ou seja:

- Média $\rightarrow E(Z_t) = \mu$
- Variância $\rightarrow Var(Z_t) = E(Z_t - \mu)^2 = \sigma^2$
- Covariância $\rightarrow \gamma_k = E[(Z_t - \mu)(Z_{t+k} - \mu)]$

onde, γ_k é a covariância ou autocovariância na defasagem k (covariância entre os valores de Z_t e Z_{t+k}).

Em suma, quando uma série temporal é estacionária sua média, variância e covariância permanecem constantes, independente do período que foram analisadas, ou seja, elas não variam no decorrer do tempo. Assim, as flutuações da série estarão em torno da sua média (medidas pela variância) mantendo, assim, uma amplitude mais ou menos constante (GUJARATI, 2000).

Um processo de grande importância das séries temporais estocásticas discretas é a presença de ruído branco. Uma sequência $\{a_t\}$ é definida como um ruído branco, se cada valor apresentar média zero e variância constante, além de ser independente de qualquer realização da própria série, ou seja, autocorrelação igual a zero (BUENO, 2008).

Tomando-se uma sequência $\{a_t\}_{t=-\infty}^{\infty}$ de variáveis aleatórias que:

- $E(a_t) = 0; \forall t$
- $E(a_t^2) = \sigma^2; \forall t$
- $E(a_t, a_{t-j}) = 0; \forall j \neq 0$

diz-se que, quando as condições acima descritas são satisfeitas, um processo é um ruído branco, ou seja, $RB(0, \sigma^2)$.

Os modelos autorregressivos integrados e de médias móveis (ARIMA) são lineares e univariados, onde seus valores correntes são relacionados apenas com seus valores passados e/ou com os valores correntes e passados dos seus erros. Estes modelos podem ser autorregressivos (AR), de médias móveis (MA), autorregressivos e de médias móveis (ARMA), ou ainda processos integrados mistos (ARIMA), dependendo do comportamento da série analisada.

Segundo Morettin e Toloí (2004), um processo linear geral, pressupõe que uma série temporal é gerada a partir de um filtro linear (ou sistema linear), cuja entrada é um ruído branco conforme ilustra Figura 2.

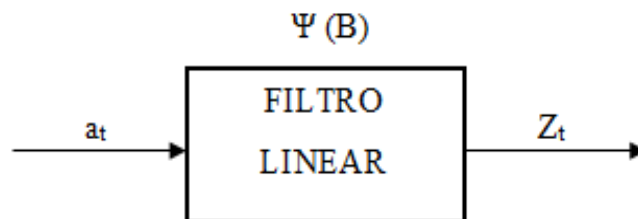


Figura 2- Filtro linear com entrada a_t , saída Z_t
 Fonte: Morettin & Toloí (2004, p.111).

formalmente, tem-se que

$$Z_t = \mu + a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \dots = \mu + \psi(B)a_t \quad (1)$$

em que

$$\psi(B) = 1 + \psi_1 B + \psi_2 B^2 + \dots \quad (2)$$

Sendo a equação (2) denominada de função de transferência do filtro e μ é um parâmetro determinado pelo nível da série.

Sabendo que Z_t descrito pela equação 1, é um processo linear em que,

- $E(a_t) = 0, \forall t$;
- $\text{Var}(a_t) = \sigma_a^2, \forall t$;
- $E(a_t, a_s) = 0, s \neq t$.

assim, chamando $\tilde{Z}_t = Z_t - \mu$, tem-se que:

$$\tilde{Z}_t = \psi(B)a_t \quad (3)$$

Com isso, se a sequência de pesos $\{\psi_j, j \geq 1\}$ for finita ou infinita e convergente, o filtro é estável (somável) e Z_t é estacionária. Neste caso, μ será a média do processo, caso contrário, Z_t é não estacionária e μ não tem significado específico.

Tem-se então que a média de Z_t é dada conforme a equação (4) descrita abaixo:

$$E(Z_t) = \mu + E(a_t + \sum_{j=1}^{\infty} \psi_j a_{t-j}) \quad (4)$$

Como $E(a_t) = 0$, para todo o t , então tem-se $E(Z_t) = \mu$ se a série convergir $\sum_{j=1}^{\infty} \psi_j$. Lembrando que a variância de Z_t é dada por:

$$\text{Var}(Z_t) = \sigma_a^2 \sum_{j=0}^{\infty} \psi_j^2 \quad \dots \quad (5)$$

Com $\psi_0 = 1$, a função de autocovariância γ_j de Z_t é dada por:

$$\gamma_j = \sigma_a^2 \sum_{i=0}^{\infty} \psi_i \psi_{i+j} \quad (6)$$

Se $\sum_{i=0}^{\infty} \psi_i \psi_{i+j}$ existir, então para todo $j = 0, \pm 1, \pm 2, \dots$. Verifica-se que, quando $j = 0$ então, $\gamma_0 = \text{Var}(Z_t)$. Dessa forma, tem-se que a média e a variância da série Z_t são constantes e a covariância só vai depender de j , portanto Z_t é uma série estacionária (MORETTIN e TOLOI, 2004).

A série \tilde{Z}_t pode ser escrita de forma alternativa, considerando-se a soma ponderada dos valores passados $\tilde{Z}_{t-1}, \tilde{Z}_{t-2}, \dots$ e também um ruído a_t :

$$\tilde{Z}_t = \pi_1 \tilde{Z}_{t-1} + \pi_2 \tilde{Z}_{t-2} + \dots + a_t = \sum_{j=1}^{\infty} \pi_j \tilde{Z}_{t-j} + a_t \quad (7)$$

assim, verifica-se que

$$a_t = (1 - \sum_{j=1}^{\infty} \pi_j B^j) \tilde{Z}_t = \pi(B) \tilde{Z}_t \quad (8)$$

onde $\pi(B)$ é o operador, descrito por

$$\pi(B) = 1 - \pi_1 B - \pi_2 B^2 - \dots \quad (9)$$

das equações (4) e (9) tem-se que

$$a_t = \pi(B) \psi(B) a_t \quad (10)$$

de modo que

$$\pi(B) = \psi^{-1}(B) \quad (11)$$

A equação descrita acima pode ser utilizada para obter os pesos π_j em função dos pesos de ψ_j e vice-versa

2.4.1 Modelos autorregressivos (AR)

O modelo autorregressivo descreve que o valor previsto de Z no período t é uma proporção de seu valor no período anterior ($t-1$), acrescido de um choque aleatório no período t , ou seja, um modelo autorregressivo depende de seus valores passados e dos erros aleatórios (MORETTIN e TOLOI, 2004).

Se $\pi_j = 0, j > p$, então tem-se um modelo autorregressivo de ordem p , o qual é denotado por AR(p), onde p indica a ordem do modelo, isto é, o número de defasagens. Na equação abaixo troca-se o nome dos pesos π_j para ϕ_j :

$$\tilde{Z}_t = \phi_1 \tilde{Z}_{t-1} + \phi_2 \tilde{Z}_{t-2} + \dots + \phi_p \tilde{Z}_{t-p} + a_t \quad (12)$$

Definindo-se o operador autorregressivo estacionário de ordem p , tem-se que:

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p \quad (13)$$

com isso, pode-se concluir que

$$\phi(B) \tilde{Z}_t = a_t \quad (14)$$

onde, p é o número de defasagens de Z_t e a_t são os resíduos não correlacionados com média zero e variância σ^2 (ruído branco).

Os modelos de ordem 1 e 2 podem ser descritos nas equações (15) e (16):

$$\text{AR}(1) \rightarrow \tilde{Z}_t = \phi_1 \tilde{Z}_{t-1} + a_t \quad (15)$$

$$\text{AR}(2) \rightarrow \tilde{Z}_t = \phi_1 \tilde{Z}_{t-1} + \phi_2 \tilde{Z}_{t-2} + a_t \quad (16)$$

Lembrando-se que $\tilde{Z}_t = Z_t - \mu$, em outras palavras, a expressão corresponde aos valores de Z_t subtraídos da média do processo.

Considerando-se que $\tilde{Z}_t = \psi(B)a_t = 11(1 + \phi_1 B + \phi_2 B^2 + \dots)a_t$, tem-se a seguinte equação:

$$\psi(B) = \sum_{j=0}^{\infty} \phi^j B^j = [\phi(B)]^{-1} = (1 - \phi B)^{-1} \quad (17)$$

Segundo a proposição descrita em Morettin e Tolo (2004) o processo será estacionário se $\psi(B)$ convergir para $|B| \leq 1$. Assim deve-se ter $|\phi| < 1$. A condição de estacionariedade do AR(p) estabelece que todas as raízes da equação devem cair fora do círculo unitário. Tendo a raiz da equação $\phi(B) = 1 - \phi B = 0$ e $B = \phi^{-1}$, assim diz-se que a raiz de $\phi(B) = 0$ e deve cair fora do círculo unitário. Em termos gerais, sendo o processo de ordem finita, será sempre inversível.

Portanto, as autocovariâncias não dependem do instante t , e sim da ordem p . Como $|\phi| < 1$ sendo a condição de estacionariedade, quanto maior o valor de p , ou seja, quanto maior a distância entre as observações, menor a sua autocovariância.

A identificação do modelo será avaliada por meio das funções de autocorrelação e autocorrelação parcial, que indicam a ordem do modelo, conforme a Figura 3:

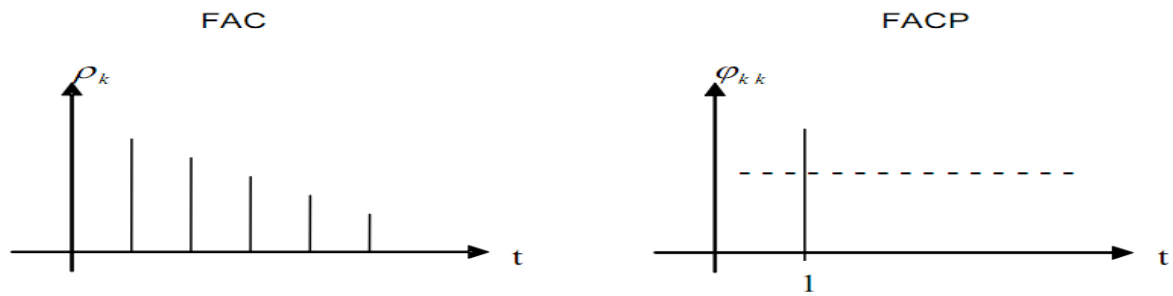


Figura 3 - Comportamento das funções de autocorrelação (FAC) e autocorrelação parcial (FACP) de um modelo AR(1).

Fonte: Moretin e Toloí (2004).

A Figura 3 mostra que a função de autocorrelação (FAC) decai exponencialmente (alternando ou não de sinal) e a autocorrelação parcial (FACP) apresenta um corte rápido no *lag* significativo, indicando a ordem p do modelo.

2.4.2 Modelos de médias móveis (MA)

Segundo o enfoque de Moretin e Toloí (2004) a média móvel para um período escolhido consiste numa série de médias aritméticas dos períodos passados. Nesses modelos a variável dependente Z_t é escrita como função linear de um número finito de defasagens dos erros aleatórios não correlacionados. Supondo que $\psi_j = 0, j > q$, tem-se então um processo de médias móveis de ordem q .

$$Z_t = \mu + a_1 + \theta_1 a_{t-1} + \dots + \theta_q a_{t-q} \quad (18)$$

lembrando-se que $\tilde{Z}_t = Z_t - \mu$, tem-se

$$\tilde{Z}_t = (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q) a_t = \theta(B) a_t \quad (19)$$

sendo que $\theta(B)$ é o operador de médias móveis de ordem q

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_p B^p \quad (20)$$

assim, podemos descrever a equação (21)

$$\tilde{Z}_t = \theta(B)a_t \quad (21)$$

Onde, q é o número de defasagens das médias móveis e a_t são os resíduos não correlacionados com média zero e variância σ^2 (ruído branco).

Os modelos de ordem um (1) e dois (2) podem ser descritos nas equações abaixo:

$$\text{MA(1)} \rightarrow \tilde{Z}_t = a_t + \theta_1 a_{t-1} \quad (22)$$

$$\text{MA(2)} \rightarrow \tilde{Z}_t = a_t + \theta_1 a_{t-1} + \theta_2 a_{t-2} \quad (23)$$

De acordo com Morettin e Toloi (2004) demonstram a proposição que evidencia um processo linear estacionário se $\theta(B) = 1 - \theta B$, tendo $\psi(B) = 1 - \theta B$ que é finito, logo:

$$a_t = [\theta(B)]^{-1} \tilde{Z}_t = \frac{1}{1 - \theta B} \tilde{Z}_t = (1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q) \tilde{Z}_t \quad (24)$$

dessa forma, pode-se obter forma invertível denotada por

$$\tilde{Z}_t = -\theta\tilde{Z}_t - \theta^2\tilde{Z}_t - \dots - \theta^q\tilde{Z}_q + a_t \quad (25)$$

Portanto, se $|\theta| < 1$, ou seja, a série $\pi(B) = \theta^{-1}(B)$ acima, vai convergir para $|B| \leq 1$. Sendo que, isto é equivalente que os zeros da equação $\theta(B) = 1 - \theta(B) = 0$ estão fora do círculo unitário.

A identificação do modelo é determinada por meio das funções de autocorrelação e autocorrelação parcial, sendo que a função de autocorrelação fornece a ordem do modelo, conforme a Figura 4:

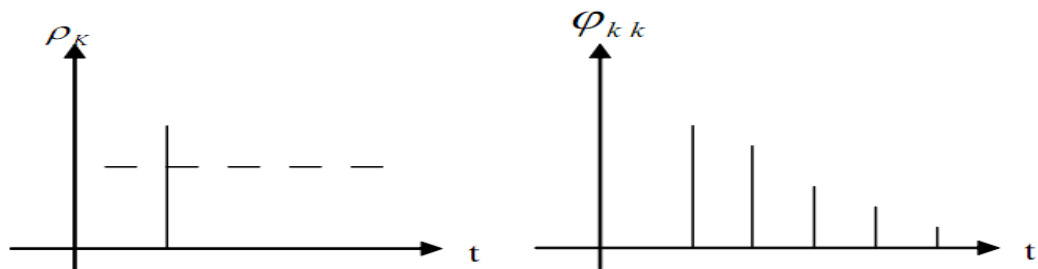


Figura 4 - Comportamento da função de autocorrelação (FAC) e função de autocorrelação parcial (FACP) de um modelo MA (1).
Fonte: Moretin e Toloí (2004).

Como mostra a Figura 4 a autocorrelação apresentará um corte rápido no *lag* significativo, indicando a ordem do modelo, e a autocorrelação parcial decairá exponencialmente.

2.4.3 Modelos autorregressivos e de médias móveis (ARMA)

De acordo com Gujarati (2000), um processo ARMA, é um modelo autorregressivo e de médias móveis. De modo geral, num processo ARMA (p,q) , haverá p termos autorregressivos e q termos de médias móveis. Dessa forma, o modelo se apresenta de forma

mais complexa, sendo explicado pelo termo autorregressivo, pelo erro estocástico e pela média móvel do termo em relação ao erro. Um modelo ARMA é descrito pela equação (26):

$$\tilde{Z}_t = \phi_1 \tilde{Z}_{t-1} + \dots + \phi_p \tilde{Z}_{t-p} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q} \quad (26)$$

Então, utilizando-se os operadores, pode-se reescrever a equação (26) de forma resumida:

$$\phi(B)\tilde{Z}_t = \theta(B)a_t \quad (27)$$

Onde, p é o número de defasagens da parte autoregressiva do modelo e q representa o número de defasagens das médias móveis e a_t é o resíduo ruído branco.

Os modelos de ordem um (1) e dois (2) estão descritos nas equações (28) e (29) respectivamente:

$$\text{ARMA}(1) \rightarrow \tilde{Z}_t = \phi_1 \tilde{Z}_{t-1} + a_t - \theta_1 a_{t-1} \quad (28)$$

$$\text{ARMA}(2) \rightarrow \tilde{Z}_t = \phi_1 \tilde{Z}_{t-1} + \phi_2 \tilde{Z}_{t-2} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} \quad (29)$$

Em suma, conforme Morettin e Toloi (2004), a condição de estacionariedade para um processo ARMA (p,q) é a mesma de um processo AR(p) e a condição de invertibilidade também é a mesma para um processo MA(q), pois utilizando-se o operador de defasagem, a equação (26), pode ser reescrita na seguinte forma:

$$\tilde{Z}_t = \psi(B)a_t = \theta(B)\theta^{-1}(B)a_t \quad (30)$$

2.4.4 Processo estocástico não estacionário – modelos autorregressivos integrado e de médias móveis (ARIMA)

No estudo de séries temporais frequentemente encontram-se séries não estacionárias, entretanto, a metodologia de *Box & Jenkins* aplica-se a um caso específico de séries não estacionárias: as séries que se tornam estacionárias após a aplicação de diferenças. O número de diferenças necessárias atribuídas a uma série é denominado de ordem de integração (VASCONCELLOS e ALVES, 2002).

O modelo aplicado a séries não estacionárias homogêneas, (ou seja, apresenta tendência estocástica, em contraposição a tendência determinística, que é expressa como função do tempo) pode ser genericamente representado por W_t , e se W_t tornar-se estacionária após a aplicação de d diferenças e a série estimada for representada por uma modelo ARMA(p, q), diz-se que Z_t é descrita por um modelo ARIMA(p, d, q) e matematicamente é denotado por:

$$W_t = \phi_1 W_{t-1} + \dots + \phi_p W_{t-p} + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q} \quad (31)$$

onde, $W_t = \Delta^d Z_t$ e utilizando-se o operador de defasagem, tem-se

$$(1 - \phi_1 B - \dots - \phi_p B^p) W_t = (1 - \theta_1 B - \dots - \theta_q B^q) a_t \quad (32)$$

$$W_t = (1 - B)^d Z_t \quad (33)$$

ou

$$(1 - B)^d \phi(B) Z_t = \theta(B) a_t \quad (34)$$

Nesse caso, $(1 - B)^d \phi(B) = 0$, apresenta d raízes sobre o círculo unitário, ou seja, d raízes unitárias e p raízes não contidas no círculo unitário.

Em relação ao estudo dos modelos não estacionários existem outros tipos de modelos como o ARI(p, d) ou ARIMA($p, d, 0$) e o IMA(d, q) ou ARIMA($0, d, q$) que são diferentes versões dos modelos AR e MA.

Um caso particular de processo não estacionário homogêneo é o ARIMA($0,1,0$), conhecido como *random walk* ou passeio aleatório, denotado por:

$$Z_t = Z_{t-1} + a_t \rightarrow \Delta Z_t = a_t \quad (35)$$

A série Z_t é integrada de ordem 1 se sua primeira diferença é um ARMA($0,0$), isto é, um ruído branco. A caracterização e a denominação de passeio aleatório decorrem deste fato, a trajetória da série do instante $t-1$ para t é totalmente aleatória (VASCONCELLOS e ALVES, 2002).

Tomando-se o modelo *random walk* e acrescentando-se uma constante na equação (35), tem-se um processo denominado *random walk* com *drift* denotado por:

$$Z_t = \mu + Z_{t-1} + a_t \quad (36)$$

2.5 Procedimentos para a análise de séries temporais

Na literatura existem inúmeros procedimentos para a análise de uma série temporal. Ao iniciar um estudo sobre séries temporais, torna-se necessário verificar a estacionariedade da mesma, onde a análise gráfica é relevante, porém uma alternativa comumente utilizada são os testes estatísticos, sendo que são utilizados para verificar a autocorrelação e a estacionariedade, que auxiliam na determinação de um resíduo com característica de ruído branco e na validação do modelo analisado.

2.5.1 Análise gráfica

Numa análise gráfica, as séries em estudo são plotadas em relação ao tempo, com o intuito de verificar se as mesmas apresentam algum tipo de componente como: tendência, sazonalidade ou ciclo e se essas séries distribuem-se de forma aleatória em torno da média. O gráfico da série indicará, no máximo, se ela é ou não estacionária (VASCONCELLOS e ALVES, 2002).

2.5.2 Teste de *Ljung-Box*, função de autocorrelação (FAC) e função de autocorrelação parcial (FACP)

Segundo Vasconcellos e Alves (2002), a referente análise consiste em determinar quais os filtros AR, I e MA compõem o processo gerador da série e também as suas respectivas ordens e caracteriza-se pela utilização da função da autocorrelação (FAC) e da função de autocorrelação parcial (FACP).

O coeficiente de autocorrelação ou correlação serial de ordem k , isto é, autocorrelação entre Z_t e Z_{t-k} é definida por:

$$\rho_k = \frac{\text{Cov}(Z_t, Z_{t-k})}{V(Z_t)} = \frac{\gamma_k}{\gamma_0} \quad (37)$$

A sequência de pares (k, ρ_k) , $k = 1, 2, \dots$, é denominada função de autocorrelação.

Como a covariância e a variância têm a mesma unidade de medida então ρ_k trata-se de um número puro e seu gráfico em relação à k corresponde ao correlograma populacional. Entretanto, sabe-se que normalmente não se trabalha com dados populacionais, mas sim, com a possibilidade de se extrair uma amostra e a partir da mesma, calcular as covariâncias e a variâncias amostrais. Assim, o coeficiente ρ_k envolve parâmetros geralmente desconhecidos.

Na prática, é necessário trabalhar com o coeficiente de autocorrelação amostral r_k com k defasagens, expresso por:

$$r_k = \frac{\sum_{t=k+1}^n (Z_t - \bar{Z})(Z_{t-k} - \bar{Z})}{\sum_{t=1}^n (Z_t - \bar{Z})^2} = \frac{\hat{\gamma}_k}{\hat{\gamma}_0} \quad (38)$$

onde n é o número de observações da série Z_t .

A análise estatística da FAC amostral demanda do conhecimento da distribuição de r_k . Se $\rho_k = 0$ para $k \neq 0$, a variância de r_k é aproximadamente igual a $1/n$. Se n for grande, a distribuição r_k é aproximadamente normal:

$$r_k \rightarrow N(0; 1/n) \quad (39)$$

A escolha do número de defasagens procede de um método empírico, onde genericamente se calcula a função de autocorrelação até um terço ou um quarto da amostra. Posteriormente, para analisar a significância estatística de ρ_k utiliza-se o seu erro-padrão (VASCONCELLOS e ALVES, 2002).

Bartlett (1946), obteve o seguinte resultado para a variância r_k quando $\rho_k = 0$ para $k > q$:

$$V(r_k) = \frac{1}{n} (1 + 2\rho_1^2 + \dots + 2\rho_q^2) \quad (40)$$

com $k = q + 1, q + 2, \dots$

Entretanto, pode-se testar se a significância estatística dos k primeiros coeficientes de autocorrelação são conjuntamente iguais à zero. Nesse caso, usa-se estatística de *Ljung - Box* que é definida por:

$$Q(K) = n(n+2) \sum_{k=1}^K \frac{r_k^2}{n-k} \quad (41)$$

onde, $Q(K)$ tem distribuição qui-quadrado (χ^2) com k graus de liberdade.

Se $Q(K) > \chi^2_k$, rejeita-se a hipótese de que os K primeiros coeficientes de autocorrelação são nulos. Segundo Gujarati (2000) essa estatística é mais poderosa para amostras pequenas.

De acordo com Vasconcellos e Alves (2002), o coeficiente de autocorrelação parcial de ordem k , usualmente representada por ϕ_{kk} , expressa a correlação entre Z_t e Z_{t-k} depois que a influência de $Z_{t-1}, Z_{t-2}, \dots, Z_{t-k+1}$ sobre Z_t foi descontada. O coeficiente ϕ_{jj} , $j = 1, 2, \dots$, é dado pelo último coeficiente, β_{jj} de cada uma das autoregressões descritas a seguir nas equações (42) e (43):

$$Z_t = \beta_{11}Z_{t-1} + a_t \rightarrow \phi_{11} = \beta_{11} \quad (42)$$

$$Z_t = \beta_{11}Z_{t-1} + \beta_{22}Z_{t-2} + a_t \rightarrow \phi_{22} = \beta_{22} \quad (43)$$

dessa forma, a equação geral é dada por

$$Z_t = \beta_{k1}Z_{t-1} + \beta_{k2}Z_{t-2} + \dots + \beta_{kk}Z_{t-k} + a_t \rightarrow \phi_{kk} = \beta_{kk} \quad (44)$$

Os valores de ϕ_{jj} podem ser obtidos a partir da solução do sistema de equações de *Yule-Walker* para os sucessivos valores de j :

$$\rho_1 = \phi_1 + \phi_2\rho_1 + \dots + \phi_k\rho_{k-1} \quad (45)$$

$$\rho_2 = \phi_1\rho_1 + \phi_2 + \dots + \phi_k\rho_{k-2} \quad \rightarrow \quad \phi_{jj} = \phi_j, j = 1, 2, \dots \quad (46)$$

dessa forma, a equação geral é dada por

$$\rho_k = \phi_1\rho_{k-1} + \phi_2\rho_{k-2} + \dots + \phi_k \quad (47)$$

A sequência de pares (j, ϕ_{jj}) constitui uma função de autocorrelação parcial.

Se o processo for autorregressivo de ordem p , a variância de $\hat{\phi}_{jj}$ para $j > p$ é aproximadamente igual a $1/n$ e para um valor de n moderado, ϕ_{jj} tem distribuição aproximadamente normal. Logo:

$$\hat{\phi}_{jj} \rightarrow N(0; 1/n), \text{ para } j > p \quad (48)$$

2.5.3 Teste *Dickey-Fuller*

De acordo com Bueno (2008), a verificação visual de uma série, dificilmente permite distinguí-la como sendo de tendência estocástica ou tendência determinística, com isso, pode-se cometer equívocos em relação a esta inspeção. Então, foram desenvolvidos testes para

verificação da existência de raízes unitárias. O primeiro teste foi mencionado por *Dickey e Fuller* (1979).

Considere o modelo da equação (49):

$$Z_t = \rho Z_{t-1} + a_t \quad (49)$$

O processo inicial é estimar esse modelo e utilizar um teste convencional “t” de *Student* sobre ρ , tendo como hipótese nula $H_0: \rho = 1$. De maneira geral, os pacotes computacionais reportam os testes nos coeficientes contra a hipótese nula de serem iguais a zero. Assim, o teste poderia ser alterado, subtraindo-se Z_{t-1} em ambos os lados da equação (49), obtendo-se:

$$Z_t - Z_{t-1} = \rho Z_{t-1} - Z_{t-1} + a_t \quad (50)$$

então, obtém-se que

$$\Delta Z_t = (\rho - 1)Z_{t-1} + a_t \quad (51)$$

em que, define-se $\delta = (\rho - 1)$.

Assim, $H_0: \rho = 1$ é equivalente a $H_0: \delta = 0$, com isso, gera-se um problema, pois sob a hipótese nula, a distribuição da estatística do teste deixa de ser igual à distribuição “t” de *Student*, pois Z_t não é estacionário. A distribuição do teste pode ser encontrada utilizando-se o teorema central do limite e pode ser visualizada por meio de experimentos de Monte Carlo, denominada de estatística τ (*tau*) (BUENO, 2008).

Dessa forma, o teste *Dickey-Fuller* (DF) tem as seguintes hipóteses a serem analisadas, na hipótese nula a série é não estacionária e em contrapartida, a hipótese alternativa diz que a série é estacionária.

$H_0: \delta = 0$ ou $\rho = 1$ (A série não é estacionária).

$H_1: \delta \neq 0$ ou $\rho < 1$ (A série é estacionária).

Em suma, o teste é simples de ser aplicado, basta tomar as primeiras diferenças de Z_t e fazer a regressão em relação à Z_{t-1} , após verifica-se se coeficiente angular é zero ou não. Se o valor do coeficiente for igual à zero, então existe a presença de uma raiz unitária, ou seja, a série não é estacionária. Em contrapartida, se o coeficiente resultar em um valor negativo, a série é estacionária.

2.5.4 Teste *Dickey-Fuller* aumentado

Segundo Bueno (2008), o problema do teste *Dickey-Fuller* (1979) é considerar o erro como um ruído branco. Entretanto, normalmente, o erro é estimado de um processo estacionário qualquer. Assim, esse problema pode causar distorções no poder do teste. Para solucionar o problema, estima-se o modelo com variáveis autorregressivas, com o intuito de corrigir o desvio do valor correto da estatística, isto é, encontram-se os desvios de Z_t em relação a sua média, para deslocar a distribuição de δ em direção à zero, caso a hipótese nula seja verdadeira.

Na prática, é necessário introduzir tantas variáveis autorregressivas quantas forem necessárias, para que o teste de resíduos não rejeite a hipótese de que se trata de um ruído branco. Com isso, o teste pode ser feito utilizando-se os mesmo valores críticos encontrados por *Dickey-Fuller*, desde que se realize a correção do modelo, de maneira a considerar as demais variáveis defasadas, que em nada modificam os valores críticos do teste e a interpretação do modelo. Com isso, essa variação do teste é denominada de teste de *Dickey-Fuller* aumentado (ADF), sendo que, a sua principal característica é a de considerar a presença de autocorrelação entre os resíduos. As hipóteses deste teste são as mesmas do teste *Dickey-Fuller* e consiste em estimar a regressão abaixo:

$$\Delta^d Z_t = \mu + \beta_1 Z_t + \delta Z_{t-1} + \sum_{i=1}^m \alpha_i \Delta Z_{t-i} + a_t \quad (52)$$

De maneira análoga ao teste DF, o teste ADF possui as seguintes hipóteses a serem analisadas, diz-se então que na hipótese nula a série é não estacionária e em contrapartida, a hipótese alternativa diz que a série é estacionária. Assim, as hipóteses são descritas na forma:

$H_0: \delta = 0$ ou $\rho = 1$ (A série não é estacionária)

$H_1: \delta \neq 0$ ou $\rho < 1$ (A série é estacionária).

2.5.5 Teste de *Kwiatkowski-Phillips-Schmidt-Shin* (KPSS)

Segundo Bueno (2008), existem testes alternativos para a verificação da presença de raiz unitária já que o teste ADF e DF são de baixo poder, ou seja, há uma elevada probabilidade de se cometer o Erro Tipo II (não rejeitar a hipótese nula quando ela é falsa). Para tentar aumentar o poder dos testes de estacionariedade surge então o KPSS, desenvolvido por *Kwiatkowski, Phillips, Schmidt e Shin* em 1992, tendo como hipótese nula a estacionariedade da série e a hipótese alternativa a não estacionariedade, a ideia dos autores é de utilizar este teste como um complemento ao teste ADF, DF ou de *Phillips-Perron*.

Para entender a mecânica desse teste assumamos que o processo gerador de dados é descrito pela equação (53):

$$Z_t = \varphi_t + a_t \quad (53)$$

Sendo que, φ_t é um passeio aleatório sem deslocamento, $\varphi_t = \varphi_{t-1} + a_t$ sabendo-se que a_t é independente e identicamente distribuído, ou seja, $a_t \sim i.i.d. (0, \sigma^2)$ e a_t é um processo estacionário. Neste caso, a hipótese a ser testada é de que:

$H_0: \sigma^2 = 0$ (A série é estacionária)

$H_1: \sigma^2 > 0$ (A série não é estacionária).

Conforme Bueno (2008), pode-se acrescentar uma tendência determinística no modelo da seguinte maneira:

$$Z_t = \mu + \beta_t + \varphi_t + a_t \quad (54)$$

Então, nesse caso usa-se o procedimento:

1. Estima-se a série em análise contra as variáveis determinísticas:

$$Z_t = \mu + \beta_t + a_t \quad (55)$$

2. Calculam-se os resíduos da regressão acima:

$$a_t = \mu + \hat{\beta}_t + \hat{\varphi}_t \quad (56)$$

3. Faz-se a soma parcial dos resíduos:

$$S_t = \sum_{j=1}^t \hat{a}_j \quad (57)$$

4. Então se utiliza o multiplicador de *Lagrange*;

$$KPSS = \sum_{t=1}^T \frac{S_t^2}{T^2 \hat{\sigma}^2} \quad (58)$$

Onde $\hat{\sigma}^2$ é a variância de longo prazo. Dessa forma, se a série Z_t for estacionária, então a soma parcial dos resíduos será não estacionária e o numerador do teste KPSS será um estimador da variância da soma parcial dos resíduos. Entretanto, se Z_t for não estacionária o numerador crescerá sem limites.

Os testes descritos anteriormente são utilizados de forma conjunta, onde primeiramente calcula-se o teste de *Dickey-Fuller* aumentado, e após o teste KPSS, já que o

segundo é mais consistente que o primeiro. Assim, analisada a etapa de verificação da estacionariedade da série e posteriormente realizadas as diferenças para torná-la estacionária, caso necessário, então por meio do correlograma pode-se determinar os indicativos do modelo que melhor se ajusta aos dados podendo ser autorregressivo, de médias móveis, ou ambos.

2.6 Critérios de informação

O critério de informação é uma maneira de encontrar o número ideal de parâmetros de um modelo. O objetivo do critério de informação é minimizar uma função baseado nos resíduos e penalizada pelo número de regressores incluídos no modelo (BUENO, 2004).

Na prática, é comum dois ou mais modelos gerarem resíduos cujos testes indicam ser ruído branco. Assim, a escolha do melhor modelo será o modelo mais parcimonioso, sendo que o somatório dos resíduos sejam os menores possíveis. O modelo mais parcimonioso deverá gerar menor imprecisão de estimativas, justamente por apresentar um menor número de parâmetros que um modelo com maior número de parâmetros.

Segundo Bueno (2008), as estatísticas descritivas tratam desse problema, mas não são utilizadas para determinar o melhor modelo. O critério de informação *Akaike*, denotado por *AIC* é descrito pela seguinte fórmula (59):

$$AIC = \ln \sigma^2(T) + c_T \varphi(T) \quad (59)$$

onde

T é o número de observações;

$\sigma^2(T) = T^{-1} \sum_{t=1}^T \hat{a}_t^2$ é a variância estimada dos resíduos;

c_T é o número de parâmetros estimados;

$\varphi(T)$ é a ordem do processo, que penaliza a falta de parcimônia.

A estatística de *Akaike*, também pode ser descrita pela equação (60):

$$AIC(p, q) = \ln \hat{\sigma}^2 + n \frac{2}{T} \quad (60)$$

A estatística de Schwarz, denotada por *BIC* (*Bayesian Information Criterion*) e descrita pela fórmula:

$$BIC(p, q) = \ln \hat{\sigma}^2 + n \frac{\ln T}{T} \quad (61)$$

onde

$n = p + q$, se o modelo não apresenta constante,

$n = p + q + 1$, se há constante no modelo e

T é o número de observações.

Sendo que:

$$\sigma^2 = \frac{\sum_{t=1}^T \hat{a}_t^2}{T} \quad (62)$$

De acordo com Bueno (2008), quanto maior for o número de parâmetros estimados no mesmo período da amostra, menor será o erro estimado, assim deseja-se obter o menor valor de *AIC* ou *BIC* possível. O critério *BIC* é consistente assintoticamente, tendendo a determinar um modelo mais parcimonioso que o *AIC*, enquanto o critério *AIC* funciona melhor em pequenas amostras, não opondo-se seja viesado para escolher modelos sobreparametrizados.

2.7 Diagnóstico de resíduo

Conforme Bueno (2008), estimado o modelo, deve-se verificar o ajuste dos resíduos. A pressuposição inicial de um modelo temporal é assumir que os erros sejam um ruído branco. Isso significa verificar se a FAC e FACP dos resíduos estimados revelam-se sem qualquer memória. Se a hipótese nula é rejeitada, implica que há informação ainda não captada pelo modelo e com isso obteremos previsões pobres.

Quando se verifica que o modelo estimado não produziu ruído branco, então se descarta esse modelo e testam-se outras possibilidades, até encontrar um modelo cujos resíduos se comporte como um ruído branco. Em contrapartida, se é aceita a hipótese nula de não autocorrelação dos resíduos na FAC, FACP e *Ljung-Box*, os resíduos comportam-se como um ruído branco.

O teste de *Ljung-Box* ou *Q* segue uma distribuição χ^2 com k graus de liberdade. Conforme equação (41) definida anteriormente. Sendo que as hipóteses são definidas como:

H_0 : os resíduos são ruído branco

H_1 : os resíduos não são ruído branco

De acordo com Gujarati (2000), caso seja rejeitada a hipótese nula, é aconselhável rejeitar o modelo, pois apresentará uma estrutura de correlação serial significativa nos resíduos.

2.8 Gráficos estatísticos de controle

Um método eficiente para monitorar, controlar e tentar reduzir a variabilidade de um processo produtivo consiste na determinação de gráficos estatísticos de controle, utilizados como uma ferramenta do Controle Estatístico de Processos (CEP) e como parte integrante das técnicas que buscam a qualidade de produtos e/ou serviços.

Conforme Montgomery (1997), os gráficos de controle revelam quando determinado processo sofre modificações e necessita de alguma intervenção. Esses gráficos foram desenvolvidos, inicialmente, pelo Dr. *Walter Shewhart*, com o objetivo de fazer distinção entre variabilidade controlável e não controlável, devido às causas que afetam o desempenho do processo, na qual chamou de causas comuns ou devido ao acaso e causas especiais ou assinaláveis.

O Controle Estatístico do Processo é constituído de ferramentas para monitorar o estado do processo e distinguir entre causas comuns e causas especiais de variabilidade. A

primeira causa é oriunda da variabilidade natural do processo, e a variação que ocorre, apresenta um comportamento aleatório indicando que o processo está sob controle. A segunda revela a formação de padrões especiais, apontando que algo de incomum está influenciando o processo, devendo ser identificada e removida para que todo o processo não seja afetado (SOUZA, SAMOBYL, MALAVÉ, 2004).

O paradigma habitual é que um processo industrial seja analisado no tempo (séries temporais), mas atualmente as ferramentas do CEP já se propagaram para processos administrativos e de serviços.

O gráfico consiste na plotagem de três linhas e os pontos que representam as médias de pequenas amostras, cada um de tamanho n , são de mensurações periódicas de alguma característica importante de um processo ou porcentagem de peças defeituosas ou número de defeitos. As três linhas representam dois limites de controle, um superior (LSC) e outro inferior (LIC), e uma linha central, a qual é a média da variável ou o alvo da característica. Tradicionalmente, as linhas de controle ficam numa distancia de três desvios padrões da média ou do valor alvo do processo. O uso de três é um pouco arbitrário, mas na prática funciona bem, na maioria dos casos. Os limites definem uma área razoavelmente grande, para evitar alarmes falsos (MONTGOMERY, 1997).

A Figura 5 descreve a representação genérica de um gráfico de controle:

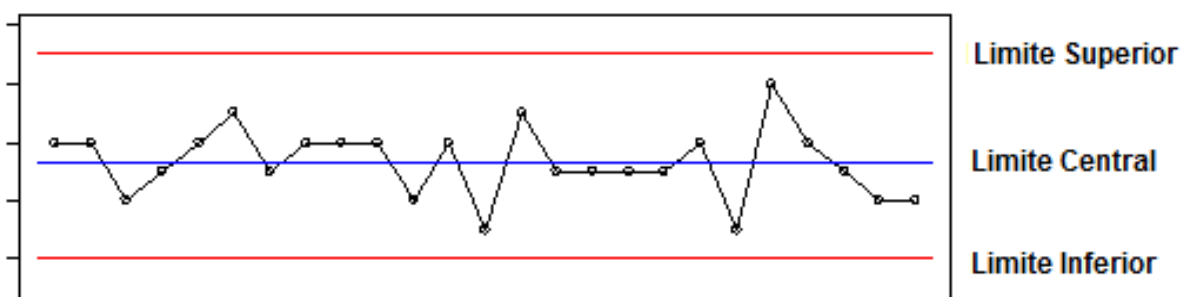


Figura 5 – Representação genérica de gráficos de controle.

Para que os resultados dos gráficos de controle sejam válidos, duas pressuposições devem ser satisfeitas: as observações devem ser independentes e identicamente distribuídas (*i.i.d.*) e, ainda, as observações devem seguir alguma distribuição de probabilidades específica, tais como a Binomial, *Poisson* ou Normal. Estas suposições são à base da teoria

dos gráficos de controle, dos limites de controle calculados e as regras para identificação de padrões não aleatórios (MONTGOMERY, 1997).

De acordo com Montgomery (1997), nas situações que não ocorra a independência entre as observações, ou seja, na presença de autocorrelação, utilizam-se outros procedimentos, como, por exemplo, o ajuste de um modelo de séries temporais do tipo ARIMA e, posteriormente, aplica-se os gráficos de controle aos resíduos do modelo.

Os gráficos de controle auxiliam na melhoria contínua da qualidade dos produtos, podendo ser utilizados para identificar a ocorrência de não conformidades em tempo real e mostrando, quando necessário, a necessidade da existência de ações interventivas e corretivas no processo que está sendo monitorado.

O desempenho de um gráfico de controle pode ser avaliado pelo ARL (*Average Run Length*), que é um parâmetro relacionado com a distribuição do tempo necessário para o gráfico de controle emitir um sinal de “fora de controle”, ou seja, expressa a sensibilidade do gráfico no sentido de detectar desvios na característica que está sendo monitorada. Para cada amostra coletada, um ponto é plotado no gráfico para monitorar as variações nas características de um produto ou serviço (PEREIRA e REQUEIJO, 2008).

O número de amostras desde o recomeço do processo até o momento no qual um sinal fora de controle é emitido, excluindo-se esse ponto, é denominado de RL (*Run Length*) e a média desse número de amostras é o ARL.

Os gráficos de controle têm forte correspondência com os testes de hipóteses. Nesse caso, a hipótese nula seria a de que o processo está sob controle. Um ponto amostral registrado fora dos limites de controle estabelecidos indica que a hipótese nula seria rejeitada.

Nessa analogia, podem-se destacar os erros Tipo I (α) e Tipo II (β), o primeiro ocorre quando se rejeita a hipótese de que o processo está sob controle e na realidade ele está (alarme falso); o segundo ocorre quando se aceita a hipótese de que o processo está sob controle e na realidade ele não está (PEREIRA e REQUEIJO, 2008).

O ARL, para os gráficos de *Shewhart*, quando o processo está estável, pode ser obtido por:

$$ARL = \frac{1}{\alpha} \quad (63)$$

onde; α é a probabilidade de que um ponto amostral exceda os limites de controle.

Para uma situação em que foi utilizado 3σ , o valor correspondente seria $\alpha = 0,0027$.

Para um processo sob controle (ARL_{SC}), tem-se:

$$ARL_{SC} = \frac{1}{0,0027} = 370 \quad (64)$$

Isso indica que, mesmo que um processo esteja sob controle, ocorrerá um sinal de alerta, em média, a cada 370 amostras. Isso equivale a dizer que $ARL_{FC} = 370$ é o número esperado de pontos registrados até que ocorra um alarme falso no processo.

Considerando um processo fora de controle (ARL_{FC}), pode-se calcular por meio da equação (65):

$$ARL_{FC} = \frac{1}{1 - \beta} \quad (65)$$

Nesse caso, o valor de ARL_{FC} é o número médio de observações que devem ser retiradas até a detecção de um verdadeiro deslocamento no processo, uma vez que ele já tenha ocorrido.

Para evitar um grande número de alarmes falsos, recomenda-se um valor grande para o ARL_{FC} , quando o processo estiver estável e, pequeno, em caso contrário.

O desvio padrão utilizado é o desvio padrão das médias (erro padrão). Teoricamente, o desvio padrão da população dividido pela raiz quadrada do tamanho da amostra σ / \sqrt{n} . Em termos estatísticos, os dois limites de controle definem um intervalo de confiança com nível de confiança de 99,73%, o número significa que um alarme falso pode ocorrer uma vez em 370 subgrupos.

2.9 Gráficos de controle com dados autocorrelacionados

A maioria dos gráficos de controle multivariados são utilizados em processos onde as variáveis são independentes. Entretanto, essa situação de independência nem sempre é possível de ser obtida e conseqüentemente, podem conduzir a decisões completamente errôneas se não forem tomadas medidas adequadas.

Conforme Pereira e Requeijo (2008), a presença de autocorrelação em um processo produtivo reflete significativamente nos gráficos de controle. Caso se verifique a existência numa ou mais variáveis, deve-se proceder à estimação dos resíduos correspondentes às p variáveis e aplicar gráficos de controle multivariados nos resíduos estimados e não nas p variáveis originais. Assim, se existindo autocorrelação, for assumida a condição de independência entre as variáveis, os parâmetros do processo e os limites de controle serão estimados incorretamente.

Dessa forma, quando não se considera a autocorrelação existente (mesmo que moderada), o risco de α aumenta de forma relevante e em conseqüência desse aumento, o ARL_{FC} decresce, acarretando na ocorrência de um grande número de falsos alarmes. Assim, a decisão de averiguar a presença de causas especiais, a análise errônea dos gráficos de controle e a aplicação de medidas corretivas, são possivelmente descartadas, pois o processo é considerado instável.

Em suma, caso se verifique a existência da autocorrelação numa ou mais variáveis, deve-se determinar os resíduos correspondentes às p variáveis autocorrelacionadas e aplicar os gráficos de controle multivariados aos resíduos estimados e não nas p variáveis originais do processo produtivo.

2.10 Gráficos de controle multivariados T^2 de Hotelling

O gráfico de controle multivariado T^2 de Hotelling apresenta duas abordagens distintas. Pode-se obter o gráfico de controle multivariado quando os dados são agrupados em amostras ($n > 1$) ou quando os dados são observações individuais ($n = 1$). Entretanto, no presente estudo abordam-se apenas os dados baseados em observações individuais.

Conforme Pereira e Requeijo (2008), em muitas situações o controle de um processo é determinado pela avaliação conjunta das variáveis, que por fatores distintos são correlacionadas. Quando as características do processo são correlacionadas, elas devem ser examinadas conjuntamente. Uma forma simples de monitorar essas variáveis é ignorar a correlação e utilizar múltiplos gráficos univariados, ou seja, um gráfico para cada característica. Entretanto, este método pode ser insatisfatório, pois não leva em consideração a correlação existente entre as variáveis, não sinalizando corretamente quando o processo está fora de controle. Assim, a solução mais adequada para contornar esses inconvenientes consiste em utilizar um gráfico de controle que contemplem estatísticas multivariadas para monitorar todas as características conjuntamente.

Nesse contexto, define-se por controle estatístico multivariado o conjunto de métodos e técnicas baseada nos gráficos de controle, que permitem controlar e monitorar simultaneamente várias características de qualidade de um processo produtivo, sendo que a estabilidade é obtida quando as médias, as variâncias e as covariâncias das variáveis envolvidas no processo permanecem estáveis ao longo do tempo.

A estatística T^2 de Hotelling desempenha um importante papel no controle de qualidade multivariado, e seu gráfico apresenta os mesmos procedimentos dos gráficos desenvolvidos por *Shewhart* para avaliar a estabilidade de um sistema. Segundo Ryan (1989) os gráficos multivariados são mais sensíveis na detecção de pontos fora de controle, quando as variáveis estiverem correlacionadas, pois o uso de gráficos univariados não contemplam a correlação existente entre as variáveis.

De acordo com Montgomery (1997), quando se deseja controlar uma característica de qualidade por meio de gráficos univariados, a probabilidade de ocorrer falsos alarmes é 0,27%. Entretanto, se almejarmos controlar duas variáveis simultaneamente, a probabilidade de um ponto situar-se fora dos intervalos definidos pelos limites de controle é igual a 0,000729% [(0,0027)(0,0027) = 0,00000729]. Em contrapartida, a probabilidade simultânea de um ponto estar dentro dos limites de controle é aproximadamente de 99,46% [(1 – 0,0027) (1 – 0,0027) = 0,994607].

O risco do tipo I, conjunto, para p variáveis independentes, em que α_i é o risco do tipo I para a variável i , é expresso por:

$$\alpha = P(\text{erro tipo I}) = 1 - \prod_{i=1}^p (1 - \alpha_i) \quad (66)$$

Assim, a probabilidade de todos os pontos das p variáveis pertencerem de forma simultânea aos intervalos definidos pelos seus limites de controle é definida por:

$$P(\text{ } p \text{ pontos entre os limites de controle}) = \prod_{i=1}^p (1 - \alpha_i) \quad (67)$$

Então, se as p variáveis analisadas não são independentes, isso implica que normalmente, as equações (66) e (67) não serão consideradas válidas, acarretando assim, a dificuldade em detectar essas distorções. Nesse contexto, uma forma de evitar distorções indesejáveis, consiste em controlar as variáveis conjuntamente, quer sejam independentes e/ou correlacionadas.

No caso multivariado, calcula-se uma medida para verificar a que distância um novo conjunto de pontos está próximo do alvo desejado, e esta distância é calculada com base na estrutura de covariância dos dados. Geometricamente, o gráfico T^2 de *Hotelling* é proporcional à distância quadrática de uma observação multivariada, onde pontos equidistantes cercam o alvo e, quanto maior o valor da estatística T^2 , maior será a distância da observação em relação ao alvo.

Segundo Pereira e Requeijo (2008), os estudos da abordagem de estatísticas multivariadas abrangem o controle da média de diversas variáveis e o controle das suas variâncias e covariâncias. Com isso, a interpretação dos gráficos de controle multivariados refletem em uma maior dificuldade do que a interpretação dos gráficos univariados, pois não só é necessário detectar as possíveis variáveis fora de controle, como também investigar se essa situação ocorre devido a uma alteração da correlação entre as diversas variáveis.

Os gráficos de controle multivariados baseiam-se no pressuposto de que a distribuição conjunta das p variáveis é uma distribuição normal multivariada $N_p(\mu, \Sigma)$, com vetor média $\mu^t = (\mu_1, \mu_2, \dots, \mu_p)$ e matriz das variâncias-covariâncias Σ . Assume-se também, que as variáveis X_1, X_2, \dots, X_p são independentes e identicamente distribuídas (*i.i.d.*) conforme as distribuições normais.

Para a construção do procedimento de *Hotelling*, é necessário que as p variáveis correlacionadas sejam oriundas de medidas realizadas simultaneamente, e sigam uma distribuição normal multivariada, com vetor média μ e matriz de variância-covariância Σ .

Se os parâmetros populacionais são desconhecidos, seus valores podem ser estimados utilizando-se a informação amostral. A média populacional é estimada a partir da média amostral denotada por \bar{X} e a matriz de variância-covariância populacional é estimada por meio de S , quando o processo for considerado estável, ou seja, sob controle estatístico.

Os gráficos de controle multivariados são utilizados para o controle das médias e das variâncias-covariâncias das várias características da qualidade. O controle estatístico multivariado é utilizado para a detecção de alterações pequenas ou moderadas do vetor média do processo.

Conforme Pereira e Requeijo (2008), para o controle estatístico multivariado, consideram-se três situações, uma para a Fase I e duas para a Fase II. Essa maneira criteriosa de abordar os gráficos multivariados depende da distribuição seguida pela estatística que controla as três situações. Na Fase I, do controle estatístico multivariado (CEPM), para o controle do vetor média, o processo é controlado por um gráfico T^2 de Hotelling, utilizando-se no mínimo de 100 observações para cada uma das características da qualidade. Após a verificação da estabilidade do processo, estimam-se o vetor média e a matriz das variâncias - covariâncias do processo.

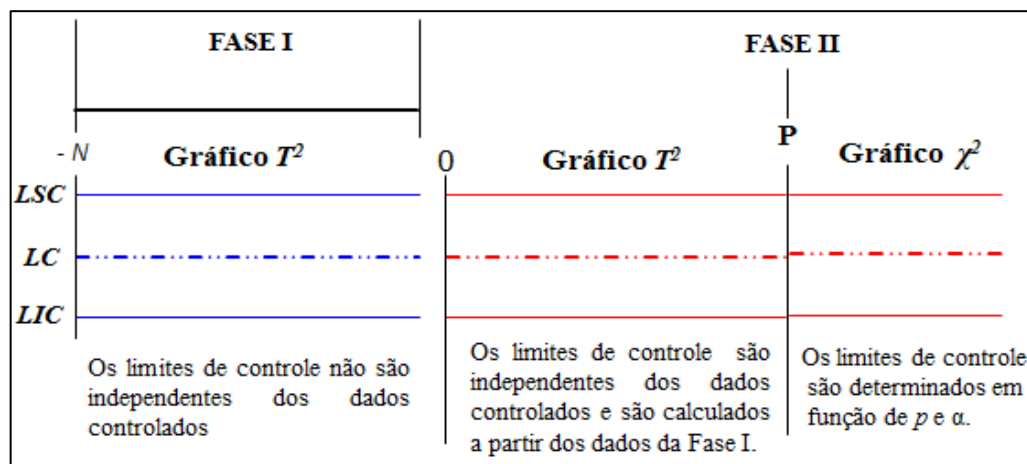


Figura 6 – Fase I e Fase II do controle estatístico do processo multivariado.

Fonte: PEREIRA E REQUEIJO (2008, p. 560).

Na Fase II do CEP multivariado, existem duas possibilidades a serem analisadas. O momento imediatamente a seguir da Fase I (que corresponde ao ponto P) e outro posterior a Fase I. Para o período compreendido entre a Fase I e o ponto P , controla-se o processo por

meio de um gráfico T^2 , onde os limites de controle são calculados com base nos parâmetros estimados na Fase I. Para o período que inicia no ponto P , considerando-se que o processo encontra-se estável, o controle é feito por meio de um gráfico χ^2 , cujos limites dependem do número de p variáveis e do nível de significância α . Assim, pode-se efetuar o monitoramento do processo propriamente dito, ou seja, monitorar as características da qualidade no futuro e em tempo real.

2.10.1 Gráficos de controle multivariados baseados em observações individuais

De acordo com Pereira e Requeijo (2008), quando não é possível definir subgrupos racionais dentro de cada amostra, o processo é então monitorado por meio de amostras de tamanho unitário ($n = 1$). Assim, consideram-se as p características da qualidade, com m observações individuais.

- **Fase I – Gráfico T^2**

Assume-se que as p variáveis X_1, X_2, \dots, X_p são independentes e identicamente distribuídas conforme distribuições normais multivariada $N_p(\mu, \Sigma)$, com vetor média $\mu^t = (\mu_1, \mu_2, \dots, \mu_p)$ e matriz das variâncias-covariâncias Σ . Como as m observações individuais de cada variável definem-se para cada variável j , a média \bar{X}_j e a variância S_j^2 e para as características j e h a covariância S_{jh}^2 , então:

$$\bar{X}_j = \frac{1}{m} \sum_{k=1}^m X_{jk} \quad (68)$$

$$S_{jj} = S_j^2 = \frac{1}{m-1} \sum_{k=1}^m (X_{jk} - \bar{X}_j)^2 \quad (69)$$

$$S_{jh} = \frac{1}{m-1} \sum_{k=1}^m (X_{jk} - \bar{X}_j) (X_{hk} - \bar{X}_h) \quad (70)$$

Sendo que, as médias \bar{X}_j ($j = 1, 2, \dots, p$) são os elementos do vetor das média \bar{X} e as variâncias S_j^2 e as covariâncias S_{jh} são os elementos da matriz das variâncias-covariâncias S .

O vetor média \bar{X} é definido por:

$$\bar{X} = (\bar{X}_1, \bar{X}_2, \dots, \bar{X}_p)^t \quad (71)$$

onde, o vetor X_k , composto pelas p características da qualidade no instante k é dado por:

$$\bar{X}_k = (X_{1k}, X_{2k}, \dots, X_{pk}) \quad (72)$$

e a matriz das variâncias-covariâncias S é definida por;

$$S = \begin{pmatrix} S_{11} & S_{12} & S_{13} & \dots & S_{1p} \\ S_{21} & S_{21} & S_{23} & \dots & S_{2p} \\ \cdot & \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \cdot & \dots & \cdot \\ S_{p1} & S_{p2} & S_{p3} & \dots & S_{pp} \end{pmatrix} \quad (73)$$

Segundo Montgomery (1997), a estatística T^2 apresenta duas fases distintas para a avaliação do processo. A primeira (Fase I) é utilizada para verificar se o processo se encontra sob controle estatístico e estimar os parâmetros \bar{X} e S com os dados disponíveis, pressupondo-se que o processo já se encontra estável. Após a verificação da estabilidade do processo é possível estabelecer os limites de controle.

Na Fase I do processo, é aplicado o gráfico T^2 de Hotelling, no instante k onde a estatística a ser considerada é dada por:

$$T^2 = (X_k - \bar{X})^t S^{-1} (X_k - \bar{X}) \quad (74)$$

onde, \bar{X} e S representam respectivamente as estimativas do para o vetor das médias e a matriz das variâncias-covariância do processo.

Os limites de controle da Fase I, do gráfico T^2 são definidos pelas equações abaixo descritas:

$$LSC_{T^2} = \frac{(m-1)^2}{m} \beta_{\left(\alpha; \frac{p}{2}, (m-p-1)/2\right)} \quad (75)$$

$$LIC_{T^2} = 0 \quad (76)$$

onde, m é o número de observações individuais que estão sendo analisadas e p é o número de variáveis envolvidas no processo.

Na equação (75), a expressão $\beta_{(\alpha; p/2, (m-p-1)/2)}$, representa o percentil à direita para uma probabilidade α , da distribuição *Beta* com parâmetros $p/2$ e $(m-p-1)/2$. Para facilitar o cálculo da estatística *Beta*, Manson Tracy e Toung (1995, 1997), sugerem a determinação do valor de $\beta_{\alpha; p/2, (m-p-1)/2}$, recorrendo a distribuição de *Fisher*, pela seguinte equação (PEREIRA e REQUEIJO, 2008).

$$\beta_{\alpha; \frac{p}{2}, (m-p-1)/2} = \frac{\frac{p}{m-p-1} \cdot F_{\alpha; p, (m-p-1)}}{1 + \frac{p}{m-p-1} \cdot F_{\alpha; p, (m-p-1)}} \quad (77)$$

Para detectar uma causa especial de variação compara-se o valor da estatística T^2 com o LSC da Fase I, ou seja, se $T^2 > LSC_{T^2}$, então o processo é considerado instável. Quando forem detectadas causas especiais de variação, os valores das observações individuais das p características referentes aos pontos fora dos limites de controle serão excluídos do gráfico T^2 , recalculando-se as médias \bar{X}_j , as variâncias S_j^2 e as covariâncias S_{jh} determinando-se também um novo vetor de média \bar{X} e uma nova matriz das covariâncias S . A estatística T^2 é novamente recalculada com base nos novos valores de \bar{X} e S (PEREIRA e REQUEIJO, 2008).

Considera-se que um processo está estável, quando todos os pontos estão inclusos dentro do intervalo definido pelos limites de controle. Assim, é possível estimar os parâmetros do processo, sendo \bar{X} o estimador de μ e S o estimador de Σ ($\hat{\mu} = \bar{X}$ e $\hat{\Sigma} = S$).

- **Fase II – Gráfico T^2**

Depois de verificado que o processo encontra-se estável na Fase I, pode-se então, partir para a Fase II do processo, que se refere à monitoração do processo propriamente dito, ou seja, ao controle a ser efetuado sobre os dados, no futuro, em tempo real. Nessa fase deve-se considerar numa primeira aproximação que os parâmetros do processo são estimativas determinadas na Fase I e os novos limites de controle são determinados em função das estimativas dos parâmetros da Fase I.

A hipótese assumida é a de que a média do processo é igual ao valor central de especificação. Assim, a estatística T^2 de Hotelling possui apenas o limite superior especificado, pois ela sempre fornece valores positivos. A estatística da Fase II é calculada pela equação (74).

Na Fase II de avaliação, os novos limites são estabelecidos para monitorar observações futuras e são definidos pelas equações (78) e (79):

$$LSC_{T^2} = \frac{p(m+1)(m-1)}{m(m-p)} F_{\frac{\alpha}{2}; p, m-p} \quad (78)$$

$$LIC_{T^2} = 0 \quad (79)$$

onde, $F_{\alpha/2; p, m-p}$ representa o percentil da distribuição de *Fisher* com p e $m - p$ graus de liberdade.

Detecta-se uma causa especial de variação, quando $T^2 > LSC_T^2$ nessas condições, a variável ou as variáveis responsáveis pela instabilidade devem ser identificadas e posteriormente, tomar as devidas medidas corretivas no processo. Nessa fase torna-se necessário verificar se a existência de causas especiais não será devida a uma alteração da correlação entre as características da qualidade em estudo (PEREIRA e REQUEIJO, 2008).

Como pode ser observado nas equações (76) e (79), o LIC_T^2 é definido como zero, a explicação para isso decorre que, qualquer mudança na média levará a um aumento na estatística T^2 e assim o LIC_T^2 pode ser ignorado.

- **Fase II – Gráfico χ^2**

Conforme Pereira e Requeijo (2008), quando o processo permanece estável por um longo período de tempo, considera-se que os parâmetros μ e Σ são conhecidos e estáveis. Assim, a estatística χ^2 usada no gráfico de controle no instante k é definida por:

$$(\chi^2)_k = (X_k - \mu)^t \Sigma^{-1} (X_k - \mu) \quad (80)$$

Normalmente, $\hat{\mu} = \bar{X}$ e $\hat{\Sigma} = S$, sendo que \bar{X} e S são determinados a partir de todas as observações individuais para o gráfico T^2 na Fase I e Fase II. Entretanto, os limites de controle do gráfico χ^2 são definidos em função do número de variáveis p e do nível de significância α . Assim, detecta-se uma causa especial de variação quando $(\chi^2)_k > LSC_{\chi^2}$.

Quando as observações são individuais, a detecção de uma causa especial de variação por meio do gráfico χ^2 consiste na identificação da(s) variável(eis) responsável(eis) pela situação de instabilidade do processo, bem como da análise de eventuais alterações das correlações entre as variáveis (PEREIRA e REQUEIJO, 2008).

O número de observações necessárias para elaborar um gráfico de controle multivariado T^2 obedece duas propostas levando em consideração o número mínimo de amostras ou observações individuais.

A primeira regra foi desenvolvida por Lowry e Montgomery (1995), que determinaram o valor mínimo do número de amostras m . Quando na Fase II, comparam-se os

limites de controle por meio da distribuição do χ^2 e da distribuição *de Fisher*. Na Tabela 1, descrevem-se alguns desses valores de m :

Tabela 1 – Valor mínimo de m

| Número de variáveis p | Dimensão da amostra n | | | |
|-------------------------|-------------------------|-----|----|----|
| | 1 | 3 | 5 | 10 |
| 2 | 100 | 75 | 50 | 20 |
| 3 | 150 | 75 | 50 | 25 |
| 4 | 150 | 75 | 50 | 25 |
| 10 | 250 | 150 | 75 | 50 |

Fonte: Pereira e Requeijo (2008, p.569).

Entretanto, a proposta de Lowry e Montgomery (1995) é extremamente conservadora, sugerindo um valor sutil para o número de amostras m a ser determinado na Fase I do CEP multivariado. Assim, como uma alternativa mais fidedigna para a determinação do menor valor de m , deve-se explorar sobre a semelhança entre os limites de controle, a partir dos parâmetros estimados e calculados em relação aos verdadeiros valores dos parâmetros (PEREIRA e REQUEIJO, 2008).

Para que os valores desses limites de controle sejam suficientemente análogos, o número de amostras m deverá pertencer ao seguinte intervalo:

$$\frac{800p}{3(n-1)} \leq m \leq \frac{400p}{n-1} \quad (81)$$

Assim, levando-se em consideração os argumentos mencionados, propõem-se o seguinte procedimento para implantação dos gráficos de controle multivariados:

- Coletar pelo menos 25 a 30 amostras (m) de dimensão igual ou superior a 4, no caso da utilização de observações individuais, considera-se no mínimo $m = 100$;
- Construir o respectivo gráfico de T^2 , para a Fase I, verificar se o processo encontra-se sob controle estatístico e estimar os parâmetros das p variáveis;
- Estabelecer os limites de controle do gráfico *de* T^2 para a Fase II, com base nas m amostras ou observações individuais utilizadas na Fase I;

- Construir o gráfico T^2 para a Fase II e monitorar o processo;
- Quando o número de amostras m coletadas para a Fase I e Fase II pertencer ao intervalo definido pela expressão (81), pode-se começar a controlar o processo por meio do gráfico χ^2 e para observações individuais deve-se considerar para m o intervalo $180 p \leq m \leq 300 p$;
- Determinar os limites de controle do gráfico χ^2 e monitorar o processo.

2.11 Interpretação dos gráficos de controle T^2 de Hotelling

Quando se utilizam gráficos multivariados, uma situação fora de controle é detectada quando um ponto se encontra fora do intervalo definido pelos limites de controle. Nestas condições, pode-se concluir que existe uma possível causa especial associada a uma ou mais variáveis ou a alteração da correlação entre duas ou mais variáveis. A identificação da variável (eis) responsável (eis) pela instabilidade do processo é considerada um imperativo do controle estatístico multivariado, pois somente dessa forma, será possível solucionar convenientemente os problemas relativos à instabilidade do processo e tomar as ações corretivas adequadas (PEREIRA e REQUEIJO, 2008).

Na literatura encontram-se vários métodos para a identificação dessas possíveis variáveis e/ou correlações, tais como: gráficos univariados das p variáveis com limites baseados nas desigualdades de *Bonferroni*; decomposição de T^2 em componentes principais (CP); decomposição de T^2 em componentes ortogonais; estatística d e o método do Ponto crítico, $C_{R,\alpha}$.

Aborda-se nessa pesquisa, apenas a decomposição de T^2 em componentes principais. Dessa forma, propicia-se uma ferramenta que incorpore as informações dos resíduos estimados oriundos dos modelos ARIMA, nos quais o gráfico multivariado T^2 é aplicado, onde a ponderação dos autovalores permite identificar as variáveis que expliquem o máximo possível da variação existente nos dados responsáveis pelas possíveis causas de instabilidade no processo.

O método de análise de componentes principais (ACP) foi desenvolvido por *Pearson* (1901) e posteriormente por *Hotelling* (1933) e é considerada uma técnica matemática da análise multivariada que possibilita investigações com grande número de dados, bem como a

identificação das variáveis responsáveis pelas variações entre os resultados, sem perda significativa de informações.

Essa técnica apresenta como principal objetivo a redução de um conjunto de dados, principalmente quando esses dados são constituídos de um grande número de variáveis inter-relacionadas.

De acordo com Souza e Vicini (2007), ao transformar um conjunto de variáveis originais, em um novo conjunto de variáveis, que mantém ao máximo, a variabilidade do conjunto de dados, reflete numa redução de informações. Esse novo conjunto de variáveis modificadas é denominado de Componentes Principais (CP), sendo que são independentes e não correlacionadas, facilitando a análise do conjunto, principalmente quando muitas variáveis devem ser analisadas.

Segundo Reis (2001), partindo-se de um vetor de variáveis aleatórias, com vetor média μ e matriz de variância-covariância Σ , procura-se encontrar um novo conjunto de variáveis, as quais são não correlacionadas entre si, sendo que as variâncias de cada nova variável são descritas em ordem decrescente de seus valores, ou seja, da mais explicativa para a menos explicativa.

Assim, para determinar as CP's deve-se calcular o vetor de média μ e da matriz de variância-covariância Σ . Após, determinar o valor dos autovalores e autovetores, e descrever as combinações lineares que formarão um novo conjunto de variáveis não correlacionadas.

Considerar os valores das CP's ao invés dos valores originais implica em perda de informações, refletidas pelo grau de explicação ou variância explicada por cada componente, entretanto, existe um ganho na análise devido a esta redução. Considerando-se que X_1, X_2, \dots, X_n são as variáveis; os r_{ij} correspondem à carga fatorial ou a correlação entre as CP's e também, as previsões dos modelos utilizados. Os a_{ij} correspondem à contribuição de cada um dos modelos na formação da CP's, ou seja, os pesos que cada modelo exerce na formação da CP. Os λ_j são os autovalores referentes às componentes principais, que representam à variação correspondente a explicação de cada componente, sendo que são selecionados habitualmente, os autovalores maiores que 1, ou seja, $\lambda_j > 1$.

Cada componente principal é representada pela equação descrita abaixo:

$$CP_j = \lambda_j = a_{1j}X_1 + a_{2j}X_2 + a_{3j}X_3 + \dots + a_{ij}X_i \quad (82)$$

onde, o vetor de constantes deve manter a condição de normalidade.

O uso da ACP, inicialmente, apresentou o propósito de reduzir a dimensionalidade de conjuntos de dados multivariados, sem ocorrer perda significativa de informações. Posteriormente, a ACP tornou-se útil na identificação das possíveis variáveis fora de controle, mostrando-se mais eficiente do que o gráfico T^2 para detectar pequenas mudanças do valor alvo do processo (WOODALL e NCUMBE, 1985).

Assim, a técnica de ACP é utilizada para identificar as possíveis variáveis que poderão afetar o processo e detectar em que momento isso ocorreu, bem como na busca de causas especiais responsáveis pela instabilidade de um processo produtivo, quando estas forem constatadas fora de controle, pelo uso de gráficos T^2 e por meio dos coeficientes de correlação entre as componentes e as variáveis originais.

Quando uma mudança ocorre em uma componente principal, geralmente não se conhece qual variável está afetando a combinação linear, quando isso ocorre, deve-se recorrer ao cálculo das correlações entre a componente e as variáveis, para determinar qual a variável que apresenta a maior influência na CP, determinando assim, a variável que possivelmente estará afetando essa componente.

2.12 Comentários gerais do capítulo

Neste capítulo inicialmente, abordou-se as características da empresa em estudo e o processo de lingotamento contínuo. A avaliação e análise desse processo são de extrema relevância para o controle da fabricação dos lingotes de aço.

Em seguida, descreveu-se a correlação existente entre as variáveis e a autocorrelação nas observações. Após foi apresentada a análise exploratória dos dados, para verificar o comportamento das variáveis analisadas, bem como uma descrição teórica sobre séries temporais e a metodologia de *Box-Jenkins*. Utilizaram-se os modelos ARIMA, pois possibilitam modelar a média condicional e eliminar a autocorrelação existente nos dados.

Na sequência, foram descritos os testes de estacionariedade, os critérios de ajuste e o diagnóstico dos resíduos.

A seguir, foram apresentadas as técnicas de gráficos de controle quando os dados encontram-se autocorrelacionados e a estatística T^2 de Hotelling, ferramentas utilizadas para

classificar o processo quanto a sua estabilidade. A estatística T^2 foi aplicada nos resíduos estimados dos modelos ARIMA com o intuito de identificar o conjunto de variáveis geradoras da instabilidade no sistema, pois quando isso ocorre, medidas de intervenção no sistema devem ser adotadas. Também foi mencionado o número mínimo de observações, bem com a interpretação do gráfico de controle multivariado T^2 de *Hotelling*.

Por fim, foram descritas a metodologia de decomposição de T^2 em componentes principais (CP).

A metodologia empregada neste estudo é o assunto a ser abordado no próximo capítulo.

3 METODOLOGIA

No capítulo 2, apresentaram-se a revisão de literatura, abordando os temas: modelos lineares ARIMA e gráficos de controle multivariados T^2 de Hotelling e análise de componentes principais. Assim, por meio destas técnicas foi desenvolvida a metodologia com o intuito de determinar os objetivos propostos nesse estudo.

A metodologia será utilizada para avaliar a estabilidade do processo produtivo de lingotamento contínuo usado na fabricação de tarugos de aço, da seção 240 milímetros.

3.1 Caracterização do estudo

A pesquisa é caracterizada como um estudo de caso, em uma empresa do setor siderúrgico, fabricante de aços especiais, sendo aplicada a dados reais.

A amostra considerada neste estudo foi de 228 medições, sendo analisadas um conjunto de 8 variáveis para a temperatura do aço no distribuidor ($T_2, T_3, T_4, T_5, T_6, T_7, T_8, T_9$), em graus Célsius, e um conjunto de 9 variáveis para os pesos do distribuidor ($PD_1, PD_2, PD_3, PD_4, PD_5, PD_6, PD_7, PD_8, PD_9$), em toneladas. O período de coleta foi compreendido entre setembro de 2009 a dezembro de 2010.

3.2 Coleta dos dados

A coleta dos dados efetuou-se junto à aciaria, disponibilizada pela empresa em planilha eletrônica, com intuito de viabilizar a importação dos dados para os programas estatísticos utilizados na análise.

Os dados provenientes da máquina de lingotamento contínuo foram obtidos em intervalos de aproximadamente 1 hora. As variáveis foram selecionadas em conjunto com a equipe de apoio e a equipe da seção de lingotamento contínuo, onde se optou pelo estudo do

Aço DIN20MnCr5 devido ao fato, do mesmo apresentar um maior número de trincas nos tarugos da seção 240 mm.

Os programas computacionais utilizados para a análise das variáveis foram: *Statistica* versão 9.1, *Eviews* 6.0 e planilha eletrônica do tipo *Excel*.

3.3 Característica e ajuste dos dados

O presente estudo aborda os dados baseados no processo de produção de tarugos da seção 240 mm, pois a mesma representa 1/3 da produção total da empresa, sendo que o restante pertence à seção 155 mm. A escolha da seção de 240 mm é justificada devido ao fato de apresentar maior número de defeitos no produto final, embora ambas as seções apresentem sucateamento.

A determinação do período relativo à coleta dos dados baseou-se, em consonância com a equipe de apoio responsável pelas referidas informações, sendo que neste período todos os tarugos avaliados pela linha de inspeção apresentaram um número significativo de defeitos, os quais foram classificados posteriormente como sucatas. As variáveis analisadas foram:

- temperatura do aço no distribuidor (°C): nesta etapa, a leitura dos dados efetuou-se a partir da segunda temperatura, sendo que a primeira leitura é descartada, devido a grande variabilidade inerente ao processo, pois esta primeira temperatura é que recebe o material da etapa anterior e, segundo o corpo técnico, não ocasiona maiores problemas nas etapas posteriores. Nesse caso, o sistema descarta a primeira leitura e atribui o valor zero nos registros das medições, pois o aço fundido, teoricamente, não poderia apresentar um valor discrepante para a primeira leitura da temperatura.
- peso do distribuidor (t): inicialmente a sucata é transportada para o forno panela onde o aço é fundido. Após a fundição, o aço escorre para o distribuidor sendo que, nesta etapa, podem-se efetuar as medições, tanto de velocidade, como de temperatura do aço fundido. Para o desenvolvimento do estudo utilizou-se 9 leituras, distribuídas igualmente ao longo das 228 observações.

Segundo a equipe de apoio da empresa, as irregularidades de maior relevância no processo produtivo dos tarugos de aço são as trincas superficiais, as trincas internas e as macroinclusões. Quando detectadas essas irregularidades, seja na produção ou laminação, que

dependendo da sua dimensão, os tarugos retornam para a fase inicial como sucata, e serão reprocessados e transformados em novos tarugos.

Salienta-se que, para evitar a ocorrência desses defeitos, é necessário o estudo da combinação de diversas variáveis integrantes do processo como, por exemplo: composição química, temperatura, refrigeração, entre outros, inclusive, as questões operacionais e a manutenção dos equipamentos utilizados.

A presente pesquisa foi desenvolvida com o propósito de investigar e detectar as possíveis variáveis causadoras dessas irregularidades nos tarugos de aço.

Apresentam-se a seguir, as etapas metodológicas utilizadas nas séries das variáveis em estudo.

3.4 Etapas da metodologia

Para atingir os objetivos propostos, foram estabelecidos os seguintes passos:

1º) Coleta dos dados na indústria siderúrgica no processo produtivo de tarugos de aço da seção 240 mm, no período de setembro de 2009 a dezembro de 2010, disponibilizados em planilha eletrônica;

2º) Investigação do comportamento das variáveis originais por meio de estatísticas descritivas, incluindo-se também, a verificação da existência de correlação cruzada entre as variáveis. Caso a correlação for significativa, Jacson e Morris (1957) sugerem que elas não devem ser analisadas separadamente, pois o efeito conjunto das variáveis pode distorcer os resultados da análise univariada.

A influência no comportamento de uma variável sobre as demais pode revelar informações importantes sobre o comportamento conjunto, gerando grandes vantagens nas informações devido à relação existente entre elas, justificando assim, a utilização de técnicas multivariadas (TRACY, YOUNG, MANSON, 1997).

3º) Verificação da existência de autocorrelação nas variáveis em estudo. Se este efeito estiver presente, utiliza-se a metodologia de *Box-Jenkins* (ARIMA) para estudar e estimar o comportamento médio do processo de produção do aço nas diferentes etapas do lingotamento, sendo que essa etapa foi analisada por meio da FAC e FACP. Depois de verificada a

existência de autocorrelação nas variáveis, procedeu-se a modelagem ARIMA, com o intuito de eliminar esta autocorrelação, e obter os resíduos estimados oriundos dos modelos matemáticos, que posteriormente, foram utilizados na construção dos gráficos de controle multivariados. Nesta etapa de modelagem a estacionariedade foi avaliada pelos testes de raiz unitária *Dickey-Fuller* Aumentado e o teste KPSS, sendo que, se após sucessivas diferenciações a série se tornar estacionária, então os modelos ARIMA são aplicáveis.

4º) Com o intuito de encontrar o melhor modelo matemático que represente o processo gerador de cada série, foram estimados os diversos modelos ARIMA denominados de concorrentes. A escolha do modelo mais adequado para cada série foi verificado, por meio da presença de ruído branco nos resíduos estimados, isto é, obtendo-se uma série de resíduos aleatórios, independentes e identicamente distribuídos (*i.i.d.*) com média zero e variância constante e não correlacionados. Assim, para auxiliar na decisão da escolha do melhor modelo estimado, utilizaram-se os critérios penalizadores *Akaike* (AIC) e *Schwarz* (BIC). Os modelos que apresentaram os menores valores desses critérios, seguindo o princípio da parcimônia e a significância dos parâmetros estimados foram os modelos selecionados. Ressalta-se que, para todas as etapas descritas anteriormente, foi utilizado o nível de significância $\alpha = 5\%$ e o programa computacional *Eviews* 6.0.

5º) Construção dos gráficos de controle multivariados T^2 de *Hotelling* a partir dos resíduos estimados dos modelos ARIMA, com o intuito de avaliar a estabilidade do sistema produtivo e identificar as possíveis variáveis causadoras da instabilidade do processo. Os gráficos de controle multivariados dividem-se em duas fases distintas, a Fase I e Fase II e estas, foram estudadas e verificadas na pesquisa. Para o desenvolvimento do algoritmo do método estatístico multivariado T^2 de *Hotelling*, foram realizadas as seguintes etapas descritas abaixo:

1. Fase I do gráfico de controle T^2 de *Hotelling*:
 - a) Após determinados os resíduos estimados dos modelos ARIMA para cada variável em estudo, primeiramente, foi calculado o vetor média \bar{X} para todos os resíduos estimados;
 - b) Em seguida, determinou-se a matriz de variância-covariância (S) e sua matriz inversa (S^{-1}) dos resíduos estimados;
 - c) Em sequência, foi determinada a matriz $(X_k - \bar{X})$, a sua transposta $(X_k - \bar{X})^t$ e após foi calculado o valor da estatística $T^2 = (X_k - \bar{X})^t S^{-1} (X_k - \bar{X})$, por meio da

multiplicação das respectivas matrizes descritas anteriormente. Ressalta-se que, para a determinação do valor da estatística T^2 , foram utilizados os valores $\alpha_i = 0,27\%$, o valor de $\alpha_{cal} = 2,14\%$ e o valor de $\beta_{cal} = 0,0785$. Posteriormente, foram calculados os limites de controle LSC e LIC por meio das equações (75) e (76), e assim, foi plotado o gráfico de controle multivariado T^2 de Hotelling, para verificar se o processo produtivo encontrava-se sob controle estatístico. O programa computacional usado foi à planilha eletrônica.

Na Fase I, de um processo multivariado é verificada a estabilidade do processo por meio do gráfico T^2 de Hotelling, ou seja, se existem apenas causas comuns de variação, bem como a estimação dos limites de controle e os parâmetros do processo, como o vetor média e a matriz de variância-covariância. Assim, estando o processo sob controle estatístico, parte-se para o monitoramento do processo, em contrapartida, havendo um sinal de instabilidade faz-se uma investigação das causas geradoras desta instabilidade, sendo que, as devidas correções devem ser realizadas no processo para estabilizá-lo, ou seja, se encontre sob controle estatístico (SOUZA e RIGÃO, 2005).

2. Fase II do gráfico de controle T^2 de Hotelling:

- a) Depois de realizada a Fase I do gráfico de controle T^2 , foram estimados os parâmetros do processo: o vetor média e a matriz de variância-covariância;
- b) Em sequencia, foi determinada a matriz $(X_k - \bar{X})$, a sua transposta $(X_k - \bar{X})^t$ e após foi calculado o valor da estatística $T^2 = (X_k - \bar{X})^t S^{-1} (X_k - \bar{X})$, por meio da multiplicação das matrizes. Ressalta-se que, para a determinação do valor da estatística T^2 , foram utilizados os valores $\alpha_i = 0,27\%$ e o valor de $\alpha_{cal} = 2,14\%$. O programa computacional usado foi a planilha eletrônica do tipo *Excel*;
- c) E por fim, foram calculados os novos limites de controle LSC e LIC pelas equações (78) e (79), para a Fase II do processo e assim, foi plotado o gráfico de controle multivariado T^2 , com intuito de monitorar o processo produtivo em estudo.

Ao detectar um ou mais pontos fora dos limites de controle, torna-se necessário identificar as possíveis variáveis causadoras da instabilidade, sendo uma informação de grande relevância em um processo, pois consiste num subsídio adicional para os operadores diagnosticarem a real causa da instabilidade.

6º) Foi realizada a decomposição da estatística T^2 em componentes principais, pois quando uma situação fora de controle for detectada, de acordo com Pereira e Requeijo (2008), essa

situação deve ser investigada criteriosamente, pois sugere a existência de uma causa especial de variação associada a uma ou mais variáveis ou devido a alteração da correlação existente entre duas ou mais variáveis. Com isso, no presente estudo, foi utilizado o método da decomposição de T^2 em CP para identificar as possíveis variáveis e/ou correlações existentes no processo, com o intuito de sugerir as ações corretivas adequadas para o processo em estudo. O programa computacional utilizado nessa etapa foi o *Statistica 9.1*.

Em uma análise com gráficos de controle que envolve dados multivariados e correlacionados, o acompanhamento das causas dessa instabilidade pode ter início a partir da redução substancial do número de variáveis. Dessa forma, torna-se possível descrever com maior precisão o valor de um conjunto de dados, por meio de um subconjunto de variáveis independentes, com uma pequena perda de informação (HENNING et al., 2010).

Para as componentes consideradas fora de controle, Souza, Samohyl e Malavé (2004) sugerem que as mesmas sejam submetidas ao estudo das correlações, ou seja, deve-se proceder a comparação entre os resultados das correlações dos resíduos estimados, com as correlações das componentes instáveis, com o intuito de avaliar e confirmar as possíveis variáveis causadoras da instabilidade nesse processo.

No capítulo 4, apresenta-se a análise e a discussão dos resultados obtidos nessa pesquisa.

4 ANÁLISE E DISCUSSÃO DOS RESULTADOS

Nesse capítulo são apresentados os resultados referentes à aplicação da metodologia desenvolvida no capítulo 3 e as técnicas estatísticas explicitadas no capítulo 2.

4.1 Análise exploratória dos dados

Com o objetivo de realizar um estudo sobre o comportamento das variáveis temperaturas do aço no distribuidor e pesos do distribuidor no processo de lingotamento contínuo, inicialmente, foi realizada uma investigação das séries por meio da análise descritiva de cada uma das variáveis.

Na Tabela 2 são apresentadas as principais estatísticas descritivas das temperaturas do aço no distribuidor.

Tabela 2 – Análise descritiva das variáveis originais das temperaturas do aço no distribuidor em graus centígrados (°C), (n = 228).

| Temperaturas do aço no distribuidor (°C) | Média | Desvio Padrão | C.V.P (%) | Assimetria |
|--|-----------|---------------|-----------|------------|
| Temperatura2 (T ₂) *1.456,000-1.563,000 | 1.535,794 | 13,359 | 0,869 | -2,165 |
| Temperatura3 (T ₃) *1.446,000-1.563,000 | 1.535,873 | 15,425 | 1,004 | -2,096 |
| Temperatura4 (T ₄) *1.455,000-1.563,000 | 1.536,632 | 13,424 | 0,874 | -1,496 |
| Temperatura5 (T ₅) *1.464,000-1.563,000 | 1.536,268 | 13,659 | 0,889 | -1,089 |
| Temperatura6 (T ₆) *1.438,000-1.563,000 | 1.533,882 | 15,771 | 1,028 | -1,970 |
| Temperatura7 (T ₇) *1.424,000-1.563,000 | 1.531,890 | 16,426 | 1,072 | -2,450 |
| Temperatura8 (T ₈) *1.423,000-1.562,000 | 1.531,535 | 17,259 | 1,127 | -2,341 |
| Temperatura9 (T ₉) *1.456,000-1.563,000 | 1.536,820 | 14,399 | 0,937 | -1,802 |

*valores mínimo-máximo. C.V.P é o coeficiente de variação de *Pearson*, expresso em porcentagem.

Observa-se pela Tabela 2 que as médias e os desvios padrões apresentaram-se aproximadamente constantes para todas as temperaturas. Os coeficientes de variação mostram-se baixos, ou seja, o grau de concentração dos dados em torno da média é relativamente alto, tendo assim uma boa representatividade para as médias das variáveis. Em relação à assimetria, devido ao fato de todos os coeficientes serem negativos, as variáveis apresentam uma distribuição assimétrica à esquerda (ou negativa), ou seja, as médias encontram-se levemente deslocadas para a esquerda.

Na Tabela 3 são apresentadas as principais estatísticas descritivas dos pesos do distribuidor. Com base nos resultados da Tabela 3, observa-se que as médias das variáveis PD₂ à PD₇ mantiveram-se praticamente constantes, ocorrendo um leve decaimento nos valores das médias de PD₈ e PD₉ em relação às demais variáveis.

Tabela 3 – Análise descritiva das variáveis originais dos pesos do distribuidor em toneladas (t), (n = 228).

| Peso do Distribuidor (t) | Média | Desvio Padrão | C.V.P (%) | Assimetria |
|---|--------------|----------------------|------------------|-------------------|
| Peso Distribuidor1 (PD ₁) *4,600-20,700 | 14,724 | 2,193 | 14,894 | -2,532 |
| Peso Distribuidor2 (PD ₂) *13,300-19,200 | 15,316 | 0,823 | 5,373 | +1,819 |
| Peso Distribuidor3 (PD ₃) *13,300-19,200 | 15,288 | 0,801 | 5,239 | +1,840 |
| Peso Distribuidor4 (PD ₄) *13,300-19,200 | 15,268 | 0,786 | 5,141 | +1,872 |
| Peso Distribuidor5 (PD ₅) *7,200-19,200 | 15,259 | 0,939 | 6,154 | -1,682 |
| Peso Distribuidor6 (PD ₆) *13,000-19,200 | 15,294 | 0,833 | 5,446 | +1,530 |
| Peso Distribuidor7 (PD ₇) *7,900-19,200 | 14,876 | 1,234 | 8,295 | -0,897 |
| Peso Distribuidor8 (PD ₈) *4,300-19,300 | 13,858 | 2,821 | 20,356 | -1,718 |
| Peso Distribuidor (PD ₉) *2,200-19,800 | 12,789 | 4,293 | 33,567 | -1,111 |

*valores mínimo-máximo. C.V.P é o coeficiente de variação de *Pearson*, expresso em porcentagem.

O desvio padrão das variáveis PD₁, PD₈ e principalmente da PD₉, mostram-se mais elevados que as demais variáveis. Em relação aos coeficientes de variação, a maioria das variáveis apresentou baixa dispersão em relação à média, com exceção das variáveis PD₁ e PD₈ que apresentaram um médio valor de dispersão, e a variável PD₉ exibiu um valor elevado de dispersão, superior a 30%, indicando uma baixa representatividade da média para esta variável.

Quanto à assimetria PD₁, PD₇, PD₈ e PD₉, apresentam uma distribuição assimétrica à esquerda e as demais variáveis descrevem uma distribuição assimétrica à direita (positiva), pois seus coeficientes são positivos.

4.2 Análise da correlação e autocorrelação nas variáveis

Na Tabela 4 são mostradas as correlações existentes entre as temperaturas do aço no distribuidor, sendo que as mesmas foram significativas. As temperaturas que se apresentaram mais correlacionadas umas com as outras foram: T₂, T₃, T₄, T₅ e T₆, com as temperaturas T₃, T₄, T₅, T₆ e T₇, justificando assim, que um procedimento multivariado pode ser aplicado.

Tabela 4 – Análise da correlação existente entre as variáveis das temperaturas do aço no distribuidor (n = 228).

| Variáveis | T ₂ | T ₃ | T ₄ | T ₅ | T ₆ | T ₇ | T ₈ | T ₉ |
|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|
| T ₂ | 1,00 | | | | | | | |
| T ₃ | 0,83 | 1,00 | | | | | | |
| T ₄ | 0,64 | 0,69 | 1,00 | | | | | |
| T ₅ | 0,57 | 0,57 | 0,76 | 1,00 | | | | |
| T ₆ | 0,48 | 0,49 | 0,63 | 0,84 | 1,00 | | | |
| T ₇ | 0,46 | 0,51 | 0,63 | 0,64 | 0,68 | 1,00 | | |
| T ₈ | 0,35 | 0,38 | 0,41 | 0,39 | 0,41 | 0,55 | 1,00 | |
| T ₉ | 0,31 | 0,31 | 0,39 | 0,39 | 0,38 | 0,40 | 0,43 | 1,00 |

Na Tabela 5 são descritas as correlações existentes entre os pesos do distribuidor, sendo que as mesmas apresentaram-se significativas, por meio dos valores das correlações da respectiva matriz. As variáveis dos pesos do distribuidor que se apresentaram mais

correlacionadas umas com as outras foram: PD₂, PD₃, PD₄, e PD₅ com os pesos PD₃, PD₄, PD₅ e PD₆.

Tabela 5 – Análise da correlação existente entre as variáveis dos pesos do distribuidor (n = 228).

| Variáveis | PD ₁ | PD ₂ | PD ₃ | PD ₄ | PD ₅ | PD ₆ | PD ₇ | PD ₈ | PD ₉ |
|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| PD ₁ | 1,00 | | | | | | | | |
| PD ₂ | 0,27 | 1,00 | | | | | | | |
| PD ₃ | 0,28 | 0,93 | 1,00 | | | | | | |
| PD ₄ | 0,30 | 0,78 | 0,85 | 1,00 | | | | | |
| PD ₅ | 0,18 | 0,61 | 0,65 | 0,75 | 1,00 | | | | |
| PD ₆ | 0,24 | 0,68 | 0,68 | 0,74 | 0,69 | 1,00 | | | |
| PD ₇ | 0,11 | 0,49 | 0,48 | 0,54 | 0,45 | 0,64 | 1,00 | | |
| PD ₈ | - 0,04 | 0,02 | 0,05 | 0,07 | 0,03 | 0,13 | 0,30 | 1,00 | |
| PD ₉ | -0,12 | -0,08 | -0,07 | -0,08 | -0,03 | -0,08 | 0,00 | 0,59 | 1,00 |

4.3 Modelagem de *Box-Jenkins* das variáveis

A primeira etapa realizada para a modelagem das séries da temperatura do aço no distribuidor e o peso do distribuidor foi à verificação da estacionariedade das séries que serão incluídas no modelo. Essa etapa consiste em verificar se as séries são produzidas por um processo estocástico estacionário, ou seja, se a média e a variância são constante ao longo do tempo (SILVA e MAIA, 2004). Para tanto, inicialmente, foi realizada a inspeção visual das séries ao longo do tempo das variáveis temperaturas do aço no distribuidor e pesos do distribuidor e posteriormente, os correlogramas FAC e FACP. Depois de verificada a análise observou-se que, aparentemente, as séries não eram estacionárias, ou seja, não foi possível verificar com precisão, um indicativo de que as séries encontravam-se estacionárias em nível.

Para confirmação da não estacionariedade das séries, aplicaram-se os testes ADF e o teste KPSS. Os resultados obtidos nos testes de raiz unitária ADF e KPSS para as séries em nível e para as séries em primeira diferença da variável temperatura do aço no distribuidor são apresentados na Tabela 6.

No teste de raiz unitária ADF, na Tabela 6, para as temperaturas do aço, observa-se que a hipótese nula foi rejeitada, sendo que as séries não revelaram a presença de raiz unitária, ou seja, são estacionárias, pois o *p-valor* apresentou um valor inferior ao nível de

significância $\alpha = 5\%$ para todas as variáveis analisadas, tanto em nível e em primeira diferença.

Na Tabela 6, encontram-se também, os resultados do teste KPSS em nível e em primeira diferença das séries analisadas.

Tabela 6 - Resultados dos testes de raiz unitária (ADF e KPSS) para a série em nível e primeira diferença das temperaturas do aço no distribuidor (n = 228)

| Ordem de Integração da Série | Teste ADF | Teste KPSS |
|-------------------------------------|---|---|
| | $H_0: \delta = 0$ ou $\rho = 1$ $H_1: \delta \neq 0$ ou $\rho < 1$ | $H_0: \delta \neq 0$ ou $\rho < 1$ $H_1: \delta = 0$ ou $\rho = 1$ |
| Temperatura2 (T₂) | | |
| Série em nível | $\tau = -9,7421$ (p < 0,05) | LM = 0,7412 (p < 0,05) |
| Série em 1ª diferença | $\tau = -15,4409$ (p < 0,05) | LM = 0,0752 (p > 0,05) |
| Temperatura3 (T₃) | | |
| Série em nível | $\tau = -10,9770$ (p < 0,05) | LM = 0,6348 (p < 0,05) |
| Série em 1ª diferença | $\tau = -12,5519$ (p < 0,05) | LM = 0,0572 (p > 0,05) |
| Temperatura4 (T₄) | | |
| Série em nível | $\tau = -7,1804$ (p < 0,05) | LM = 1,0043 (p < 0,05) |
| Série em 1ª diferença | $\tau = -15,7398$ (p < 0,05) | LM = 0,0867 (p > 0,05) |
| Temperatura5 (T₅) | | |
| Série em nível | $\tau = -10,6970$ (p < 0,05) | LM = 1,3008 (p < 0,05) |
| Série em 1ª diferença | $\tau = -15,7370$ (p < 0,05) | LM = 0,2680 (p > 0,05) |
| Temperatura6 (T₆) | | |
| Série em nível | $\tau = -7,5259$ (p < 0,05) | LM = 1,5571 (p < 0,05) |
| Série em 1ª diferença | $\tau = -15,6641$ (p < 0,05) | LM = 0,1822 (p > 0,05) |
| Temperatura7 (T₇) | | |
| Série em nível | $\tau = -10,8595$ (p < 0,05) | LM = 1,4579 (p < 0,05) |
| Série em 1ª diferença | $\tau = -18,0583$ (p < 0,05) | LM = 0,1988 (p > 0,05) |
| Temperatura8 (T₈) | | |
| Série em nível | $\tau = -9,8818$ (p < 0,05) | LM = 0,7814 (p < 0,05) |
| Série em 1ª diferença | $\tau = -12,3597$ (p < 0,05) | LM = 0,1930 (p > 0,05) |
| Temperatura9 (T₉) | | |
| Série em nível | $\tau = -8,8031$ (p < 0,05) | LM = 0,7113 (p < 0,05) |
| Série em 1ª diferença | $\tau = -10,4292$ (p < 0,05) | LM = 0,1728 (p > 0,05) |

τ estatística “tau”, LM: estatística qui-quadrado. Hipótese nula do teste ADF: série não estacionária e do teste KPSS: série estacionária. ADF: Teste *Augmented Dickey-Fuller* e KPSS: Teste *Kwiatkowski-Phillips-Schmidt-Shin*.

Os resultados obtidos no teste KPSS, não mostraram-se em consonância com os resultados do teste ADF, em nível, pois o *p-valor* mostrou-se menor que o nível de significância adotado. Assim, tornou-se necessária aplicação de uma diferença nas séries originais para verificar tal estacionariedade. Após, observou-se que a série tornou-se estacionária tanto em nível, quanto em primeira diferença, em ambos os testes.

Na Tabela 7, estão descritos os resultados dos testes ADF e KPSS para os pesos do distribuidor. Por meio dela, é possível verificar que o *p-valor* apresentou-se significativo em nível, não sendo necessário aplicar diferenças para torná-las estacionárias.

Tabela 7 – Resultado dos testes de raiz unitária (ADF e KPSS) para a série em nível dos pesos do distribuidor (n = 228)

| Ordem de Integração da Série | Teste ADF H ₀ : $\delta = 0$ ou $\rho = 1$ H ₁ : $\delta \neq 0$ ou $\rho < 1$ | Teste KPSS H ₀ : $\delta \neq 0$ ou $\rho < 1$ H ₁ : $\delta = 0$ ou $\rho = 1$ |
|-------------------------------------|---|--|
| | Peso1 (PD₁) | |
| Série em nível | $\tau = -16,3749$ (p < 0,05) | LM = 0,1038 (p > 0,05) |
| | Peso2 (PD₂) | |
| Série em nível | $\tau = -5,6831$ (p < 0,05) | LM = 0,1474 (p > 0,05) |
| | Peso3 (PD₃) | |
| Série em nível | $\tau = -5,7605$ (p < 0,05) | LM = 0,1283 (p > 0,05) |
| | Peso4 (PD₄) | |
| Série em nível | $\tau = -5,9647$ (p < 0,05) | LM = 0,1612 (p > 0,05) |
| | Peso5 (PD₅) | |
| Série em nível | $\tau = -7,0173$ (p < 0,05) | LM = 0,2705 (p > 0,05) |
| | Peso6 (PD₆) | |
| Série em nível | $\tau = -6,5550$ (p < 0,05) | LM = 0,2173 (p > 0,05) |
| | Peso7 (PD₇) | |
| Série em nível | $\tau = -5,7442$ (p < 0,05) | LM = 0,2694 (p > 0,05) |
| | Peso8 (PD₈) | |
| Série em nível | $\tau = -15,3865$ (p < 0,05) | LM = 0,0811 (p > 0,05) |
| | Peso9 (PD₉) | |
| Série em nível | $\tau = -6,4023$ (p < 0,05) | LM = 0,1571 (p > 0,05) |

τ estatística “tau”, LM: estatística qui-quadrado. Hipótese nula do teste ADF: série não estacionária de do teste KPSS série estacionária. ADF: Teste *Augmented Dickey-Fuller* e KPSS: Teste *Kwiatkowski-Phillips-Schmidt-Shin*.

Depois de verificada a estacionariedade das variáveis temperaturas por meio dos testes ADF e KPSS em primeira diferença e do peso do distribuidor em nível, iniciou-se então, o processo de modelagem, com o intuito de buscar um modelo matemático que descrevesse o comportamento das séries (temperatura e peso do distribuidor) e estimar os resíduos que proporcionassem os menores erros possíveis, bem como tratar a correlação serial existentes nas variáveis.

Na Tabela 8 descrevem-se os modelos ARIMA selecionados e seus respectivos modelos matemáticos, da variável temperatura do aço no distribuidor, sendo que os modelos escolhidos foram, os modelos ARIMA(0,1,2) para as variáveis T_2 e T_9 e os modelos ARIMA(1,1,1) para as variáveis T_3 , T_4 , T_5 , T_6 , T_7 e T_8 .

Tabela 8 – Modelos ARIMA selecionados para as variáveis temperaturas do aço no distribuidor em graus centígrados ($^{\circ}\text{C}$), ($n = 228$).

| Temperaturas do aço no distribuidor | Modelos ARIMA | Modelos Matemáticos |
|--|----------------------|---|
| Temperatura2 | ARIMA(0,1,2) | $\Delta Z_t = -0,6749a_{t-1} - 0,2915a_{t-2}$ |
| Temperatura3 | ARIMA(1,1,1) | $\Delta Z_t = +0,2698Z_{t-1} - 0,9801a_{t-1}$ |
| Temperatura4 | ARIMA(1,1,1) | $\Delta Z_t = +0,2909Z_{t-1} - 0,9700a_{t-1}$ |
| Temperatura5 | ARIMA(1,1,1) | $\Delta Z_t = +0,2339Z_{t-1} - 0,9654a_{t-1}$ |
| Temperatura6 | ARIMA(1,1,1) | $\Delta Z_t = +0,1828Z_{t-1} - 0,9676a_{t-1}$ |
| Temperatura7 | ARIMA(1,1,1) | $\Delta Z_t = +0,1820Z_{t-1} - 0,9530a_{t-1}$ |
| Temperatura8 | ARIMA(1,1,1) | $\Delta Z_t = +0,3586Z_{t-1} - 0,9800a_{t-1}$ |
| Temperatura9 | ARIMA(0,1,2) | $\Delta Z_t = -0,4972a_{t-1} - 0,4757a_{t-2}$ |

Modelo ARIMA (p,d,q) correspondem aos Modelos Autorregressivos Integrados de Médias Móveis, onde p representa a parte autoregressiva, d o número de diferenças aplicadas na série e q a parte de médias móveis.

Na Tabela 9 são apresentados os modelos ARIMA selecionados para as variáveis pesos do distribuidor, bem como as suas equações matemáticas.

Observando-se os resultados, verifica-se que, para as variáveis PD_1 , PD_5 , PD_6 , PD_7 e PD_8 foram escolhidos os modelos ARIMA(1,0,1), para PD_2 , PD_3 e PD_4 os modelos ARIMA(2,0,0) e para PD_9 o ARIMA(3,0,3).

Tabela 9 – Modelos ARIMA selecionados para as variáveis pesos do distribuidor em toneladas (t), (n = 228).

| Peso do Distribuidor | Modelos ARIMA | Modelos Matemáticos |
|-----------------------------|----------------------|---|
| Peso Distribuidor1 | ARIMA(1,0,1) | $Z_t = 0,9999Z_{t-1} - 0,9922a_{t-1}$ |
| Peso Distribuidor2 | ARIMA(2,0,0) | $Z_t = 15,3282 + 0,3150Z_{t-1} + 0,3088Z_{t-2}$ |
| Peso Distribuidor3 | ARIMA(2,0,0) | $Z_t = 15,3039 + 0,3772Z_{t-1} + 0,2514Z_{t-2}$ |
| Peso Distribuidor4 | ARIMA(2,0,0) | $Z_t = 15,2779 + 0,3501Z_{t-1} + 0,2529Z_{t-2}$ |
| Peso Distribuidor5 | ARIMA(1,0,1) | $Z_t = 15,2741 + 0,8855Z_{t-1} - 0,6992a_{t-1}$ |
| Peso Distribuidor6 | ARIMA(1,0,1) | $Z_t = 15,3179 + 0,9169Z_{t-1} - 0,7142a_{t-1}$ |
| Peso Distribuidor7 | ARIMA(1,0,1) | $Z_t = 14,9322 + 0,8923Z_{t-1} - 0,6997a_{t-1}$ |
| Peso Distribuidor8 | ARIMA(1,0,1) | $Z_t = 13,8756 - 0,9549Z_{t-1} + 0,9958a_{t-1}$ |
| Peso Distribuidor9 | ARIMA(3,0,3) | $Z_t = 12,8004 - 0,8830Z_{t-1} - 0,8988Z_{t-2} - 0,8796Z_{t-3} + 0,6375a_{t-1} + 0,6249a_{t-2} + 0,6360a_{t-3}$ |

Modelo ARIMA (p,d,q) correspondem aos Modelos Autorregressivos Integrados de Médias Móveis, onde p representa a parte autoregressiva, d o número de diferenças aplicadas na série e q a parte de médias móveis.

Na modelagem da variável peso do distribuidor, constatou-se uma discrepância entre os modelos selecionados, principalmente para a variável PD₉, pois apresentou 3 parâmetros autorregressivos e 3 de médias móveis, sendo que a maioria das variáveis exibiram 1 parâmetro autorregressivo e 1 de médias móveis.

Após a determinação dos melhores modelos matemáticos que representassem a processo gerador das séries em estudo, foram utilizados os resíduos estimados oriundos destes modelos, com o propósito de desenvolver o algoritmo dos gráficos de controle multivariados T^2 de Hotelling.

Salienta-se que, a aplicação da metodologia de *Box-Jenkins*, foi de extrema necessidade e relevância, pois de acordo com Pereira e Requeijo (2008), a presença de autocorrelação em um processo produtivo reflete significativamente nos gráficos de controle. Caso se verifique a existência de autocorrelação em uma ou mais variáveis, deve-se proceder à estimação dos resíduos correspondentes as variáveis e aplicar gráficos de controle multivariados nos resíduos estimados e não nas variáveis originais. Assim, se for assumida a independência entre as variáveis e existir autocorrelação significativa, os parâmetros do processo e os limites de controle serão estimados incorretamente.

4.4 Gráfico de controle multivariado T^2 de Hotelling

Na Figura 7 são apresentados os resultados referentes ao gráfico de controle multivariado T^2 dos resíduos originais estimados, oriundos dos modelos ARIMA das variáveis temperaturas do aço no distribuidor da Fase I do CEPM.

Com base na análise do gráfico T^2 dos resíduos originais estimados, verifica-se que o processo referente as 8 medições das temperaturas em estudo, não se encontra estável. Na Fase I, foram detectados 21 pontos fora dos limites de controle, sendo LIC = 0 e LSC = 17,58. Os respectivos pontos fora de controle detectados foram: 1, 3, 8, 9, 12, 38, 56, 83, 97, 99, 107, 142, 158, 171, 180, 181, 198, 202, 206, 210 e 221, ou seja, aproximadamente 9,3% dos pontos encontraram-se fora dos limites de controle. Porém, não é presumível identificar qual ou quais são as possíveis variáveis causadoras desta instabilidade, assim uma investigação mais minuciosa foi realizada, por meio da decomposição de T^2 em componentes principais, que será descrita posteriormente.

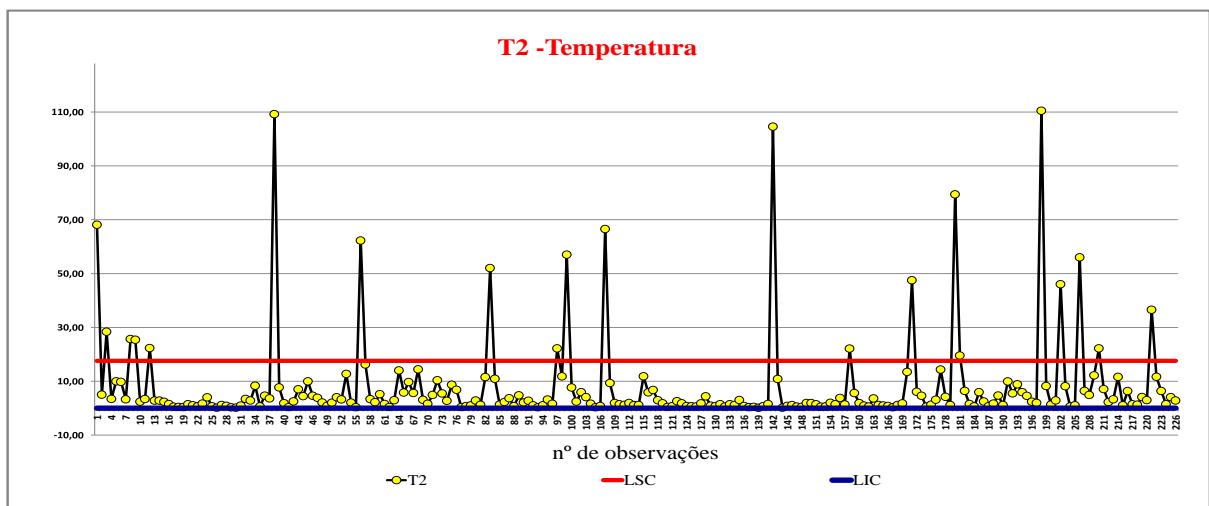


Figura 7 – Gráfico de controle multivariado T2 dos resíduos originais estimados dos modelos ARIMA das variáveis temperaturas do aço no distribuidor, na Fase I do CEPM.

Assim, como o processo não se encontrou sob controle estatístico para as temperaturas na Fase I, foi necessário intervir no processo, como o intuito de estabilizá-lo. Para tanto, optou-se por 2 alternativas: a primeira optou-se por substituir os pontos que encontravam-se fora de controle pela média dos resíduos das temperatura e a segunda foi deletar os pontos que não estavam dentro dos limites de controle estimados.

Na Figura 8 são descritos os resultados do gráfico T^2 dos resíduos substituídos pelas médias dos resíduos das temperaturas do aço no distribuidor.

Observa-se que o gráfico T^2 assinala uma instabilidade no processo, mesmo após a primeira intervenção no processo, pois foram encontrados 31 pontos fora dos limites de controle (LIC = 0 e LSC = 17,58), isto é, os pontos: 5, 6, 39, 45, 53, 57, 64, 68, 72, 82, 84, 98, 100, 108, 115, 117, 143, 170, 172, 177, 185, 189, 191, 193, 194, 199, 203, 207, 209, 214 e 222, ou seja, aproximadamente 13,7% dos pontos encontraram-se fora dos limites de controle.

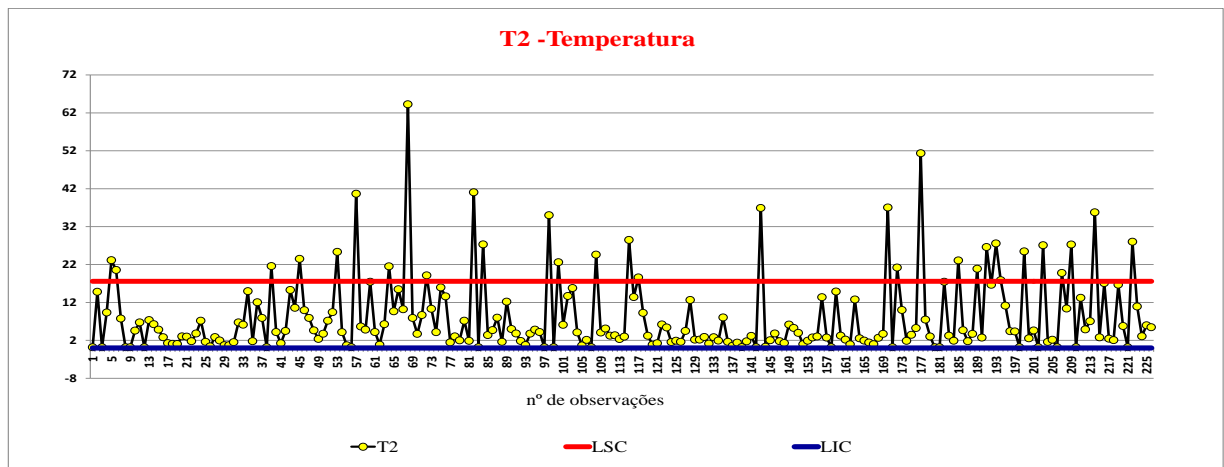


Figura 8 – Gráfico de controle multivariado T^2 dos resíduos substituídos pelas médias dos resíduos das variáveis temperaturas do aço no distribuidor, na Fase I do CEP.

Neste contexto, pode-se intuir que, quando se substitui os pontos que se encontram fora dos limites de controle, pelas respectivas médias dos resíduos, refletiu em uma mudança no processo. Assim, verifica-se que, ao substituir os pontos pela média dos resíduos, não foi uma alternativa coerente, pois houve um aumento do número de pontos fora de controle, isto é, de 21 para 31 pontos. Uma explicação coerente para este aumento é o fato de que a média neste caso, não seria um representante adequado para o conjunto de dados em estudo, sendo que os resíduos apresentam valores discrepantes.

Na Figura 9 são mostrados os resultados do gráfico T^2 dos resíduos excluídos que se encontravam fora dos limites de controle na Fase I do CEP. Neste caso, observa-se que o processo ainda continuou instável, pois foram detectados 28 pontos fora dos limites de controle, sendo os pontos: 3, 4, 33, 39, 47, 50, 57, 61, 75, 76, 89, 90, 97, 104, 131, 157, 158, 163, 169, 173, 175, 177, 182, 185, 188, 190, 194 e 201, isto é aproximadamente 13,5% dos pontos encontraram-se fora dos limites de controle.

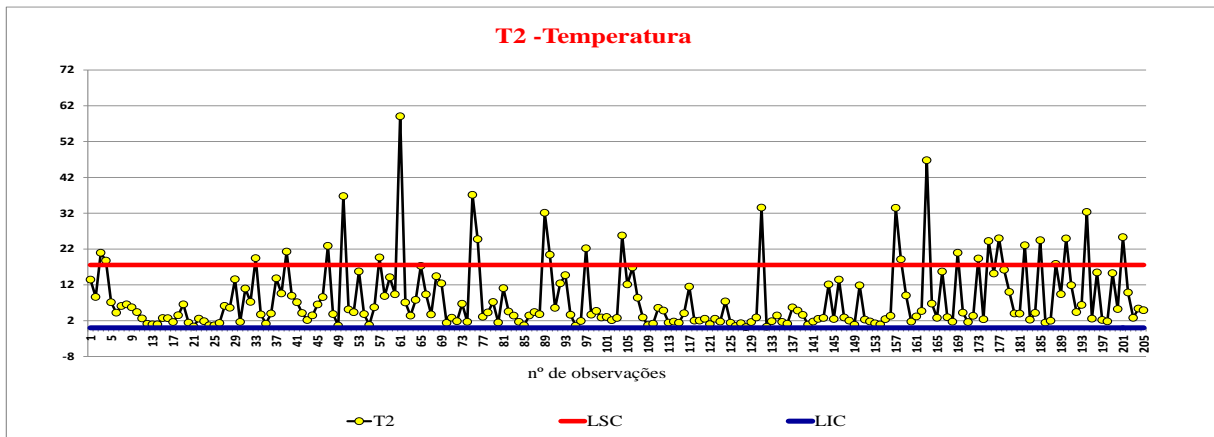


Figura 9 – Gráfico de controle multivariado T^2 dos resíduos excluídos fora dos limites de controle, das variáveis temperaturas do aço no distribuidor, na Fase I do CEPM.

Dessa forma, observa-se que mediante a redução do número de resíduos e consequentemente, a estimação dos novos limites de controle: $LIC = 0$ e $LSC = 17,54$, não foi possível tornar o processo sob controle estatístico. Entretanto, com base na análise univariada, que sugere que os pontos que se encontram fora dos intervalos dos limites de controle estimados sejam excluídos do processo, optou-se então, pela escolha do gráfico T^2 onde foi efetuada a exclusão dos pontos fora de controle e procedeu-se para a Fase II.

A Fase II do CEPM refere-se ao monitoramento do processo propriamente dito, ou seja, ao controle a ser efetuado sobre os dados, no futuro, em tempo real. Dessa forma, considera-se que os parâmetros do processo são as estimativas determinadas na Fase I e os limites de controle são determinados em função dessas estimativas.

Na Figura 10 são apresentados os resultados do gráfico T^2 dos resíduos excluídos que estavam fora dos limites de controle para as variáveis temperaturas do aço no distribuidor, na Fase II do CEPM.

O gráfico T^2 identifica claramente a instabilidade no processo, pois revela a existência de 25 causas especiais de variação, pois os respectivos valores calculados em alguns pontos da estatística T^2 foram superiores aos novos limites de controle estimados, $LIC = 0$ e $LSC = 19,28$, indicando assim, que o processo não se encontra sob controle estatístico. Os pontos detectados foram: 3, 33, 39, 47, 50, 57, 61, 75, 76, 89, 90, 97, 104, 131, 157, 163, 169, 173, 175, 177, 182, 185, 190, 194 e 201, isto é, aproximadamente 11% dos pontos localizaram-se fora dos limites de controle.

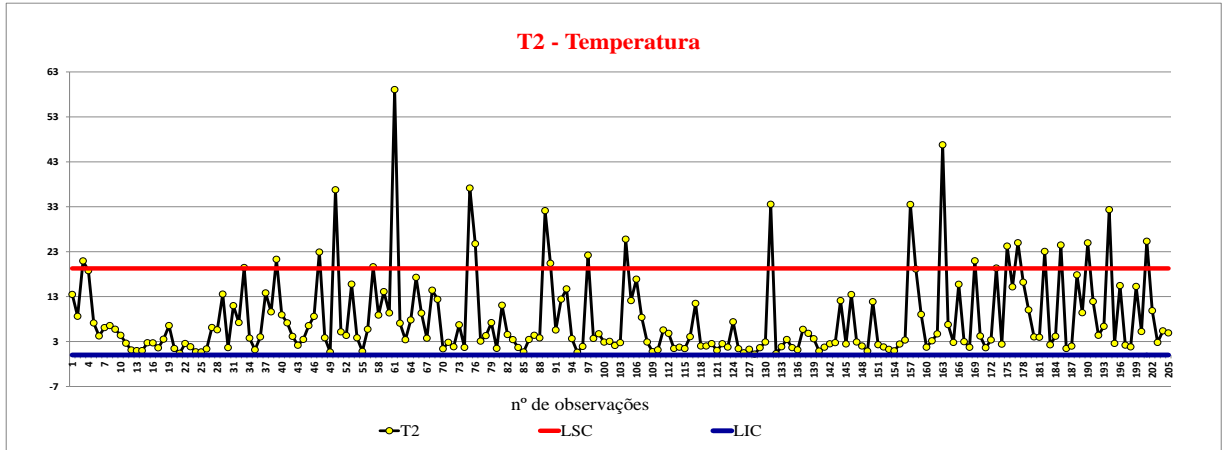


Figura 10 – Gráfico de controle multivariado T^2 dos resíduos excluídos fora dos limites de controle, das variáveis temperaturas do aço no distribuidor, na Fase II do CEP.

Na Figura 11 são apresentados os resultados referentes ao gráfico de controle multivariado T^2 dos resíduos originais estimados, oriundos dos modelos ARIMA das variáveis pesos do distribuidor, na Fase I do CEP.

Analisando-se a Figura 11 que corresponde aos resíduos originais estimados, verificar-se que o processo relativo as 9 medidas do peso do distribuidor, não se apresentou estável, pois na Fase I, foram detectados 24 pontos fora dos limites de controle estimados, sendo $LIC = 0$ e $LSC = 18,71$. Os respectivos pontos fora de controle detectados foram: 14, 36, 46, 51, 52, 53, 57, 62, 70, 71, 76, 85, 105, 111, 114, 130, 148, 178, 179, 184, 187, 189, 205 e 213, ou seja, aproximadamente 10,7% dos pontos localizaram-se fora dos limites de controle.

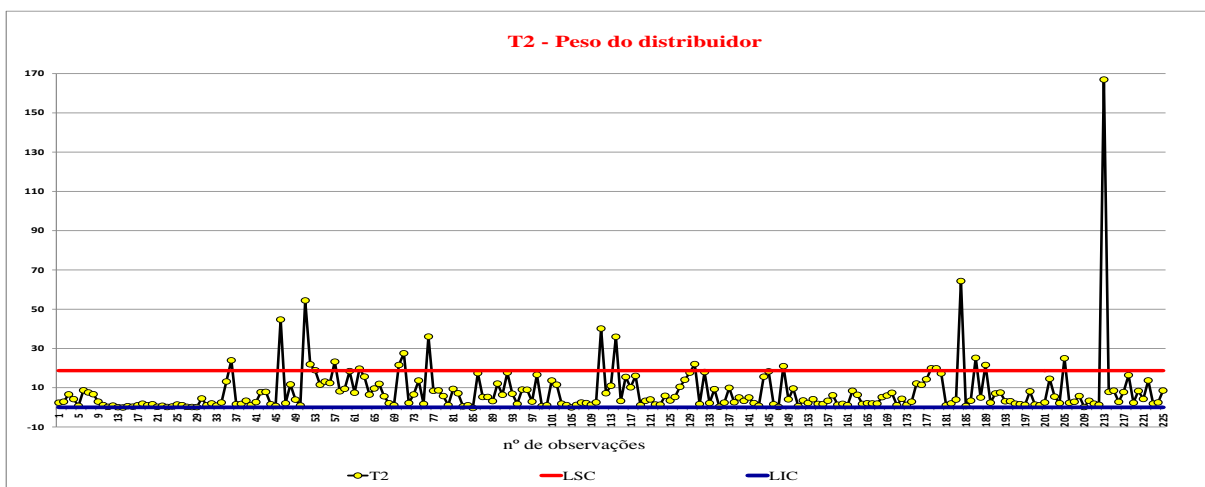


Figura 11 – Gráfico de controle multivariado T^2 dos resíduos originais estimados dos modelos ARIMA das variáveis pesos do distribuidor, na Fase I do CEP.

De maneira análoga, as variáveis temperatura, como o processo não se encontrou sob controle estatístico para os pesos do distribuidor na Fase I, interviu-se no processo, por meio de 2 alternativas,: substituir os pontos que encontravam-se fora de controle pela média dos resíduos dos pesos do distribuidor e deletar os pontos que não estavam dentro dos limites de controle estimados.

Na Figura 12 são descritos os resultados do gráfico T^2 dos resíduos substituídos pelas médias dos resíduos dos pesos do distribuidor, na Fase I do CEPM.

Com base na análise da Figura 12, é possível verificar que o processo ainda se encontrou instável, mesmo após a primeira intervenção no processo, pois foram detectados 29 pontos fora dos limites de controle (LIC = 0 e LSC = 18,71), isto é, os pontos: 35, 56, 60, 63, 66, 69, 74, 86, 92, 98, 101,113, 116, 117,118, 127, 128, 129, 132, 144,145, 175, 176, 177, 180, 202, 214, 218 e 222, ou seja, aproximadamente 12,8% dos pontos encontraram-se fora dos limites de controle.

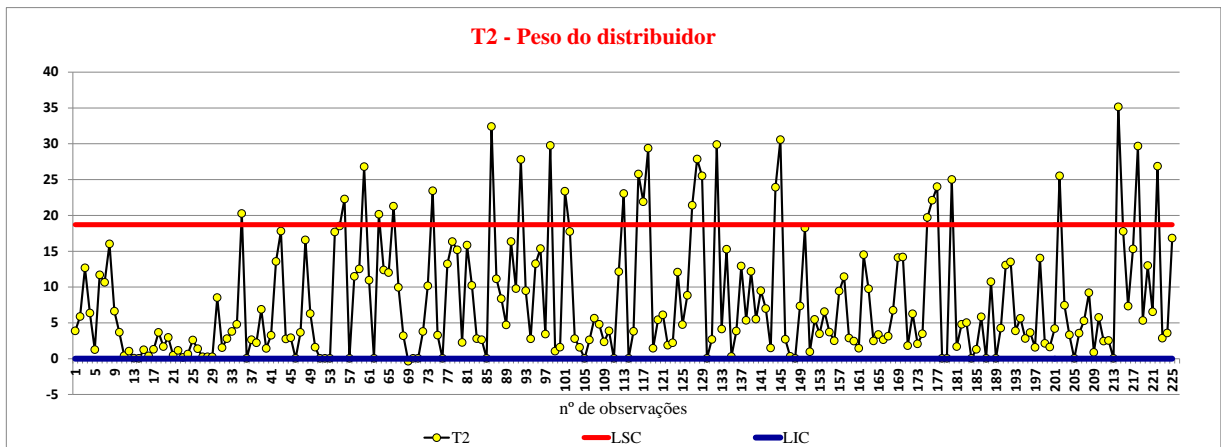


Figura 12 – Gráfico de controle multivariado T^2 dos resíduos substituídos pelas médias dos resíduos das variáveis pesos do distribuidor, na Fase I do CEPM.

Assim, verifica-se que ao substituir os pontos pela média dos resíduos, não foi uma boa alternativa, pois refletiu em um aumento do número de pontos fora de controle, de 24 para 29 pontos. Uma explicação lógica para este aumento é o fato de que a média neste caso, não seria um representante apropriado para o conjunto de dados em estudo, sendo que os resíduos apresentam valores discrepantes.

Na Figura 13 são mostrados os resultados do gráfico T^2 dos resíduos excluídos que encontravam fora dos limites de controle na Fase I do CEPM das variáveis pesos do

distribuidor. Observa-se que o processo ainda prossegue instável, pois foram detectados 26 pontos fora dos limites de controle, sendo os pontos: 50, 53, 58, 61, 64, 74, 80, 86, 89, 99, 101, 102, 103, 112, 113, 114, 116, 128, 129, 159, 160, 161, 180, 190, 194 e 198, isto é, aproximadamente 12,8% dos pontos localizaram-se fora dos limites de controle.

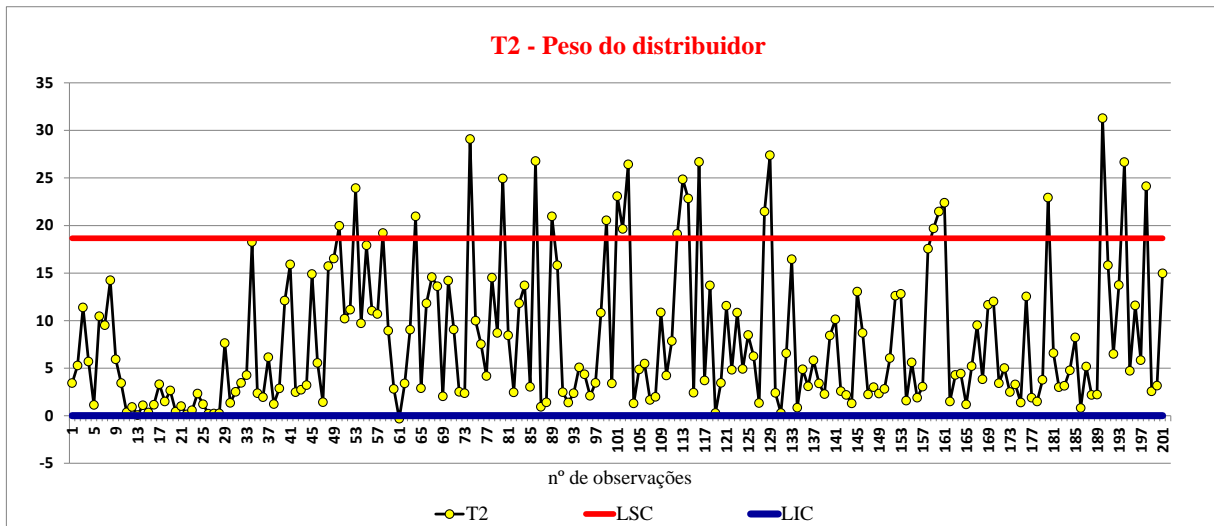


Figura 13 – Gráfico de controle multivariado T^2 dos resíduos excluídos fora dos limites de controle das variáveis pesos do distribuidor, na Fase I do CEPM.

Portanto, com a redução dos resíduos e a estimação dos novos limites de controle: LIC = 0 e LSC = 18,65, não foi possível tornar o processo sob controle estatístico. Entretanto, com base na análise univariada, que sugere que os pontos que se encontram fora dos intervalos dos limites de controle estimados sejam excluídos do processo, então, optou-se pela escolha do Gráfico T^2 onde foram excluídos dos pontos amostrais fora de controle para proceder para a Fase II.

Na Figura 14 são apresentados os resultados do T^2 dos resíduos excluídos que estavam fora dos limites de controle para os pesos do distribuidor, na Fase II do CEPM.

A análise da Figura 14 revela a existência de 26 causas especiais de variação, sendo que alguns pontos obtiveram os valores da estatística T^2 superiores aos novos limites de controle estimados, LIC = 0 e LSC = 20,68; cujos pontos são: 50, 53, 58, 61, 64, 74, 80, 86, 89, 99, 101, 102, 103, 112, 113, 114, 116, 128, 129, 159, 160, 161, 180, 190, 194 e 198, isto é, aproximadamente 12,8% dos pontos localizaram-se fora dos limites de controle.

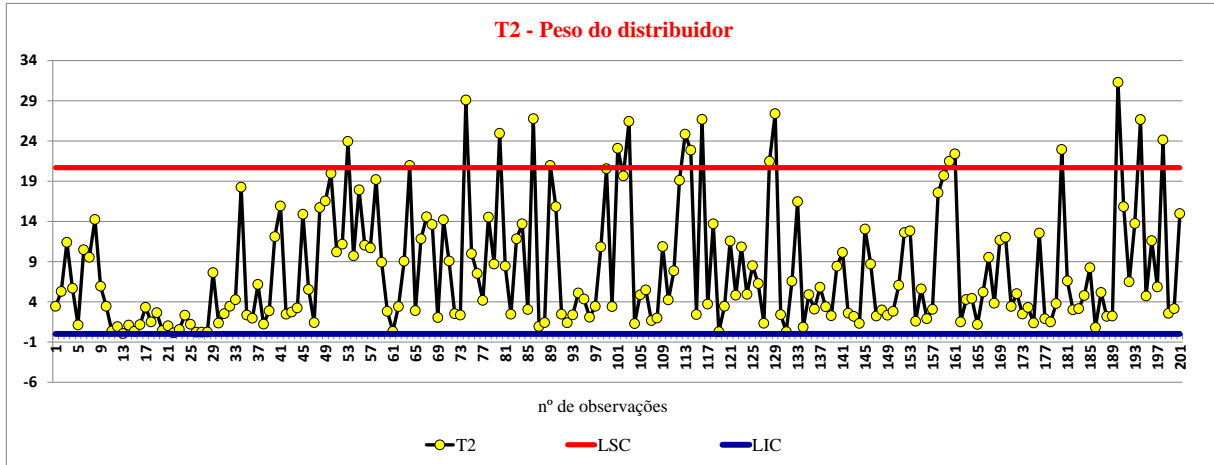


Figura 14 – Gráfico de controle multivariado T^2 dos resíduos excluídos fora dos limites de controle das variáveis pesos do distribuidor, na Fase II do CEP.

Com isso, verifica-se que, para estabilizar o processo, tanto para as temperaturas quanto para os pesos do distribuidor, medidas corretivas deveriam ser aplicadas no processo no momento em que ocorreu tal instabilidade, pois assim, o processo poderia possivelmente, tornar-se estável.

Ressalta-se que, a Fase II que envolve a estatística χ^2 não foi possível ser determinada, pois para o desenvolvimento desta fase, o processo deve permanecer estável por um longo período de tempo e o número mínimo de observações m devem estar compreendido no intervalo $180 p \leq m \leq 300 p$, o que não ocorreu no presente estudo.

4.5 Decomposição dos resíduos em componentes principais

Após constatar que o gráfico de controle multivariado T^2 registrou pontos fora dos limites estimados, um diagnóstico auxiliar foi realizado por meio da decomposição em CP, com o objetivo de eliminar a correlação existente entre as variáveis, reduzir o número de variáveis a serem analisadas, além de detectar as possíveis variáveis causadoras da instabilidade do processo.

Na Tabela 10 são apresentados os resultados da decomposição dos resíduos em componentes principais da temperatura do aço no distribuidor, com os seus respectivos

autovalores, o percentual da variância explicada, os autovalores acumulados e o percentual da variância explicada acumulada.

Essa decomposição revelou que apenas 2 componentes (fatores), são responsáveis por 78,18% da variabilidade dos dados.

Na Tabela 10, pode-se observar que apenas 2 componentes apresentaram autovalor maior que 1, isto é, $\lambda_1 = 5,17$ e $\lambda_2 = 1,08$, cujas variâncias explicadas correspondem, respectivamente a 64,65% e 13,53% em relação a cada componente, respectivamente.

Tabela 10 – Autovalores e variância explicada de cada componente dos resíduos da variável temperatura do aço no distribuidor

| Número de Componentes | Autovalores | % da variância explicada | Autovalores acumulados | % da variância explicada acumulada |
|------------------------------|--------------------|---------------------------------|-------------------------------|---|
| 1 | 5,17 | 64,65 | 5,17 | 64,65 |
| 2 | 1,08 | 13,53 | 6,25 | 78,18 |
| 3 | 0,62 | 7,69 | 6,87 | 85,87 |
| 4 | 0,46 | 5,81 | 7,33 | 91,68 |
| 5 | 0,29 | 3,57 | 7,62 | 95,25 |
| 6 | 0,19 | 2,46 | 7,81 | 97,71 |
| 7 | 0,11 | 1,35 | 7,92 | 99,06 |
| 8 | 0,08 | 0,94 | 8,00 | 100,00 |

Na Tabela 11 são descritos os resultados da decomposição dos resíduos em componentes principais do peso do distribuidor, com os seus respectivos autovalores, o percentual da variância explicada, os autovalores acumulados e o percentual da variância explicada acumulada.

Tabela 11 – Autovalores e variância explicada de cada componente dos resíduos da variável peso do distribuidor.

| Número de Componentes | Autovalores | % da variância explicada | Autovalores acumulados | % da variância explicada acumulada |
|------------------------------|--------------------|---------------------------------|-------------------------------|---|
| 1 | 4,36 | 48,51 | 4,36 | 48,51 |
| 2 | 1,61 | 17,84 | 5,97 | 66,35 |
| 3 | 0,91 | 10,11 | 6,88 | 76,46 |
| 4 | 0,82 | 9,19 | 7,70 | 85,65 |
| 5 | 0,50 | 5,48 | 8,20 | 91,13 |
| 6 | 0,36 | 3,99 | 8,56 | 95,12 |
| 7 | 0,24 | 2,72 | 8,80 | 97,84 |
| 8 | 0,13 | 1,41 | 8,93 | 99,25 |
| 9 | 0,07 | 0,75 | 9,00 | 100,00 |

De maneira análoga, o gráfico T^2 para as 9 variáveis originais peso do distribuidor também apresentou pontos fora de controle. Observando-se a Tabela 11, verifica-se que foram selecionadas as 2 primeiras componentes, sendo $\lambda_1 = 4,36$ e $\lambda_2 = 1,61$, cujas variâncias explicadas correspondem, respectivamente a 48,51% e 17,84%, sendo que a variância total explicada por essas 2 componentes é aproximadamente de 66,35%.

Na Tabela 12 são mostrados os resultados das cargas fatoriais de cada um dos fatores (CP) para as variáveis temperaturas e pesos do distribuidor.

Tabela 12 – Carga fatorial de cada componente dos resíduos das variáveis temperaturas do aço no distribuidor e peso do distribuidor

| Temperatura | | | Peso do distribuidor | | |
|--------------------|---------|--------|-----------------------------|---------|--------|
| Variável | Fator 1 | Fator2 | Variável | Fator 1 | Fator2 |
| Resíduo T_2 | -0,766 | 0,312 | Resíduo PD_1 | -0,375 | -0,092 |
| Resíduo T_3 | -0,853 | 0,338 | Resíduo PD_2 | -0,861 | -0,139 |
| Resíduo T_4 | -0,911 | 0,176 | Resíduo PD_3 | -0,882 | -0,109 |
| Resíduo T_5 | -0,911 | 0,121 | Resíduo PD_4 | -0,899 | -0,012 |
| Resíduo T_6 | -0,899 | 0,082 | Resíduo PD_5 | -0,888 | 0,017 |
| Resíduo T_7 | -0,817 | -0,146 | Resíduo PD_6 | -0,861 | 0,086 |
| Resíduo T_8 | -0,633 | -0,615 | Resíduo PD_7 | -0,596 | 0,192 |
| Resíduo T_9 | -0,564 | -0,647 | Resíduo PD_8 | -0,087 | 0,892 |
| | | | Resíduo PD_9 | 0,024 | 0,852 |

Uma análise feita em relação às cargas fatoriais (ou correlações) de cada componente para a temperatura do aço no distribuidor revela que T_4 , T_5 , T_6 , T_3 , T_7 e T_2 apresentam forte correlação com a primeira componente e T_8 e T_9 uma moderada correlação com a segunda componente. Observa-se que os valores superiores a 0,7, são os que possuem uma maior contribuição na formação da CP e, portanto, são mais representativos.

Assim, com base nos resultados da Tabela 12, as temperaturas T_4 e T_5 , subsequentemente às variáveis T_6 , T_3 , T_7 e T_2 , nesta ordem, são as possíveis variáveis causadoras da instabilidade no processo, ou seja, a temperatura T_4 deveria ser a primeira variável a ser analisada e ajustada, posteriormente a T_5 , a T_6 e assim sucessivamente.

De maneira análoga, na Tabela 12 visualizam-se as ponderações de cada resíduo estimado relativo às variáveis pesos do distribuidor, as quais irão compor a combinação linear, sendo que as variáveis PD_4 , PD_5 , PD_3 , PD_6 e PD_2 apresentam uma forte correlação com

a primeira componente e PD_8 e PD_9 expõem uma forte correlação com a segunda componente. Com isso, as possíveis variáveis PD_4 , PD_5 , PD_3 , PD_6 e PD_2 , nesta ordem, deveriam ser analisadas e ajustadas constantemente, pois o que se procura e espera, em longo prazo, é a estabilidade do sistema, sendo que a variabilidade pode diminuir gradativamente, à medida que os ajustes e correções forem sendo efetuados.

Assim, analisando-se os resultados obtidos por meio da comparação das respectivas correlações dos resíduos estimados e das componentes instáveis, obteve-se, então, a confirmação de que as supostas variáveis citadas anteriormente, realmente são as possíveis variáveis causadoras da instabilidade no sistema.

Em estudo realizado por Casarin (2012), com as mesmas variáveis, porém com técnicas estatísticas diferentes, os resultados corroboram com os obtidos no presente estudo. Entretanto, para a temperatura do aço no distribuidor apenas uma componente mostrou-se significativa, mas constataram-se praticamente, as mesmas temperaturas como as possíveis candidatas responsáveis pela instabilidade no sistema, sendo T_4 e T_5 as mais expressivas, seguidas de T_6 , T_7 , T_3 e T_2 , contrariando apenas na ordem das correlações de T_7 e T_3 .

Para o peso do distribuidor, também houve consonância nos resultados obtidos, pois nos dois estudos foram encontradas duas CP's significativas, porém para a CP1, destaca-se uma pequena inversão na ordem dos valores das correlações, sendo encontrados por Casarin (2012), PD_4 , PD_3 , PD_2 , PD_6 e PD_5 e para a CP2 as variáveis foram às mesmas PD_8 e PD_9 .

Assim, quando as variáveis do processo encontram-se correlacionadas, a modificação ou a alteração em uma das variáveis, pode induzir a uma alteração nas demais variáveis, refletindo assim, na instabilidade em todo processo.

4.6 Comentários gerais do capítulo

Os gráficos multivariados são considerados mais complexos em sua implantação comparando-se com os gráficos de controle univariados introduzidos por *Shewhart*, entretanto a crescente busca pela qualidade de processos e serviços e pela minimização de perdas exigiu a utilização de técnicas cada vez mais complexas, modernas e eficazes.

Por meio da aplicação das técnicas multivariadas, tornou-se possível encontrar as possíveis variáveis causadoras da instabilidade no sistema produtivo, sendo que medidas

corretivas fossem adotadas nas variáveis fora de controle, sem a escolha aleatória na sequência, nas quais deveriam ser ajustadas.

A metodologia proposta somente poderá ser aplicada quando dispõem-se de um conjunto de dados com no mínimo três variáveis correlacionadas e não autocorrelacionadas, pois a análise univariada poderia ser deficiente e não captar possíveis distorções no processo.

A análise de componentes principais é mais eficiente quanto maior for o coeficiente de correlação entre as variáveis envolvidas no processo, reduzindo-se o número de variáveis em componentes e, facilitando assim, a identificação da variável ou conjunto de variáveis responsáveis pela instabilidade do sistema.

No próximo capítulo serão descritas as considerações finais desse trabalho, além de sugestões para futuras pesquisas.

5 CONSIDERAÇÕES FINAIS E RECOMENDAÇÕES

O presente capítulo apresenta uma síntese das principais conclusões que foram sendo inferidas ao longo dos capítulos que compõem o decorrente estudo, tanto em nível de investigação teórica e científica, como a nível de aplicação prática. Serão descritas de forma sucinta algumas ações a serem desenvolvidas em consequência dos resultados obtidos e recomendações que podem sustentar o desenvolvimento de futuros trabalhos científicos.

5.1 Considerações Finais

A metodologia de investigação descrita no presente estudo procurou desenvolver modelos de controle estatístico de processos, incidindo em temáticas como a estatística aplicada ao controle da qualidade, a existência de autocorrelação e o estudo multivariado.

O objetivo deste trabalho foi desenvolver uma metodologia auxiliar para a avaliação de um processo produtivo de lingotamento contínuo na presença de correlação cruzada por meio dos gráficos de controle multivariados T^2 de Hotelling.

Ressalta-se que, a aplicação da metodologia de *Box-Jenkins*, foi relevante e indispensável, pois a presença de autocorrelação em um processo produtivo reflete expressivamente nos gráficos de controle. Caso exista autocorrelação nas variáveis, deve-se proceder à estimação dos resíduos oriundos das variáveis originais e aplicar gráficos de controle multivariados nos resíduos estimados e não nas variáveis originais. Caso contrário, se for assumida a independência entre as variáveis e haver autocorrelação significativa, os parâmetros do processo e os limites de controle serão estimados incorretamente.

A aplicação dos modelos lineares autorregressivos propiciou modelar e estimar os resíduos dos melhores modelos para a aplicação dos gráficos multivariados, pois a correlação serial nas variáveis foi eliminada. Em decorrência deste fato, houve um ganho na interpretação dos gráficos de controle, sendo que foi possível identificar os pontos fora de controle, isentando-se assim, da ocorrência de falsos alarmes em função da autocorrelação existente nos dados.

A verificação da correlação existente entre as variáveis apresentou-se significativa, sendo este um pressuposto fundamental para a aplicação dos gráficos multivariados.

Com a utilização dos gráficos de controle multivariados T^2 de Hotelling foi possível verificar a presença de causas especiais no processo produtivo de lingotamento contínuo, isto é, o processo encontrou-se instável, pois vários pontos se localizaram fora dos limites de controle estimados para cada fase dos gráficos multivariados.

Nesse contexto, detectados os pontos fora de controle do gráfico T^2 , uma investigação auxiliar foi realizada por meio da decomposição em CP, sendo que o objetivo principal da determinação dessas componentes foi eliminar a correlação existente entre as temperaturas de aço no distribuidor e entre pesos do distribuidor, reduzir o número destas variáveis, além de, detectar as possíveis variáveis causadoras da instabilidade do processo.

Posteriormente, as componentes consideradas fora de controle foram submetidas a um estudo comparativo de suas correlações com as correlações dos resíduos estimados, com o intuito de avaliar e confirmar as variáveis que foram as prováveis causadoras da instabilidade nesse processo. Analisando-se os resultados obtidos por meio destas correlações, obteve-se uma confirmação das possíveis variáveis responsáveis pela instabilidade do sistema.

Assim, com a aplicação das técnicas propostas no presente estudo, foi possível constatar que as variáveis temperaturas de aço no distribuidor e pesos do distribuidor, do processo de lingotamento contínuo de fabricação de tarugos de aço da seção 240 mm não se encontram sob controle estatístico, ou seja, o processo encontra-se instável. Com isso, constata-se com os resultados obtidos que o objetivo proposto foi alcançado, pois foi possível avaliar a estabilidade do processo em estudo.

Em decorrência destes resultados, o estudo pode servir de apoio para as equipes envolvidas no processo de lingotamento contínuo, com o intuito de atuar nos pontos críticos do processo, bem como fornecer informações auxiliares para que medidas corretivas possam ser tomadas, contribuindo, assim na redução de custos e na qualidade dos seus produtos e serviços.

5.2 Recomendações para estudos futuros

Como futuras sugestões com base no tema desenvolvido e com o propósito de suprir certas lacunas em relação a avaliação de processos produtivos, sugere-se:

- utilizar modelos que envolvam o estudo da variância e/ou volatilidade do processo, como os modelos: ARCH, GARCH, EGARCH, dentre outros;
- utilizar o vetor autorregressivo – VAR/ VAR-VEC, como alternativa ao gráfico T^2 ;
- utilizar a metodologia de redes neurais para processo lineares e não lineares;
- realizar um estudo comparativo dos resíduos estimados detectados fora de controle no gráfico T^2 por meio da decomposição ortogonal e/ou estatística d ;
- aplicar técnicas estatísticas que permitam detectar pequenas ou moderadas alterações dos parâmetros de processo, como os gráficos de controle CUSUM e EWMA para dados independentes;
- aplicar técnicas estatísticas que permitam detectar pequenas ou moderadas alterações dos parâmetros de processo, como os gráficos de controle MCEWMA e EWMAST para dados autocorrelacionados.

REFERÊNCIAS BIBLIOGRÁFICAS

BARTLETT, M. S. On the Theoretical Specification and Sampling Properties of Autocorrelated Times Series. **Supplement to the Journal of the Royal Statistical Society** v. 8, n. 1. Wiley, 1946.

BOX, G. E. O.; JENKINS, G.M. **Time series Analysis: Forecasting and Control**. San Francisco: Holden-Day, 1970.

BUENO, R. L. S. **Econometria de Séries Temporais**. São Paulo: Cengage Learning, 2008.

CASARIN, V. A. **Avaliação da estabilidade do processo de lingotamento contínuo por meio de gráficos de controle com variáveis dependentes**. 2012. 232 f. Tese (Doutorado em Engenharia de Minas, Metalúrgica e de Materiais) – Universidade Federal do Rio Grande do Sul, Porto Alegre, 2012a.

CASARIN, V. A.; SOUZA, A. M.; MENEZES, R; SPIM, J.A. Continuous casting process stability evaluated by means of residuals control charts in the presence of cross-correlation and autocorrelation. **International Journal of Academic Research**. v. 4, n. 3. May, 2012b.

DICKEY, D. A.; FULLER, W. A. Distribution of the estimators for autoregressive times series with a unit root. **Journal of the American Statistical Association** **74**, 1979.

GARCIA, A.; SPIM, J.A.; SANTOS, C.A.; CHEUNG, N. **Lingotamento contínuo de aços**. São Paulo: Associação Brasileira de Metalurgia e Materiais, 2006.

GUJARATI, D. N. **Econometria Básica**. 3. ed. São Paulo: Makron Books, 2000.

JACKSON, J.E., MORRIS, R.H. Application of multivariate quality control to pothographic processing. **Journal of the American Statistical Association**. V.52, n.2, 1957.

KWIATKOWSKI, D.; PHILLIPS, P. C. B.; SCHMIDT, Peter; SHIN, Yongcheol. Testing the null hypothesis of stationarity against the alternative of a unit root. **Journal of Econometrics**, v. 54, 1992.

KOURTI, T.; MACGREGOR, J. F. Multivariate SPC methods for process and product monitoring. **Journal of Quality Technology**, v.28, n.4, 1996.

HENNING, E. et al. Um estudo comparativo entre o desempenho de gráficos de controle estatístico multivariados com a aplicação da análise de componentes principais. In: SIMPÓSIO BRASILEIRO DE PESQUISA OPERACIONAL – Saúde humana, saúde animal e ecossistemas, 42., 2010, Bento Gonçalves. **Anais**. Bento Gonçalves: Simpósio Brasileiro de Pesquisa Operacional, 2010.

HOTTELLING, H. Analysis of a complex of Statistical variables into principal components. **The Journal of Educational Psychology**, v.24, 1933.

LOWRY, C. A.; MONTGOMERY, D. C. A review of multivariate control chart. **IIE Transactions**, v.27, 1995.

MONTGOMERY, D. C. **Introduction to statistical quality control**. 3. ed. New York : John Wiley & Sons Inc., 1997.

MORETTIN, P. A.; TOLOI, C. M. C. **Análise de Séries Temporais**. São Paulo, *Edgard Blücher*, 2004.

PEREIRA, Z. L.; REQUEIJO, J. G. **Planejamento e controlo estatístico de processos**. Faculdade de Ciências e Tecnologia – FET. Universidade Nova de Lisboa, 2008.

PEARSON, K. On lines and planes of closest fit to systems of points in space. **Philosophical Magazine**. v.2, 190. Disponível em <http://pbil.univ-lyon1.fr/R/pearson1901.pdf> Acesso em 10 set., 2012.

REIS, E. **Estatística Multivariada Aplicada**. Edições. Sílabo, 2001.

RYAN, T. P. **Statistical Methods for quality improvement**. New York: John Wiley & Sons Inc., 1989.

SHEWHART, W. A., *Economic Control of Quality of Manufactured Product*. D. Van Nostrand Company, Inc, New York, 1931.

SILVA, E. K.; MAIA, S. F. **Política monetária no Brasil (1994-2002): uma análise utilizando vetores auto regressivos**. 2004. Disponível em <[http://www.dce.sebrae.com.br/bte/bte.nsf/0/17470e185e05fa3f03256ff2004216e8/\\$FILE/NT000A6FAA.pdf](http://www.dce.sebrae.com.br/bte/bte.nsf/0/17470e185e05fa3f03256ff2004216e8/$FILE/NT000A6FAA.pdf)> Acesso em: 12 nov., 2012.

SOUZA, A. M.; SAMOHYL, R. W.;MALAVÉ, C. O. Aplicação de um modelo paramétrico multivariado para controle da temperatura de fornos de túnel. **Revista Produção**. v.14, n.2, 2004.

SOUZA, A. M.; RIGÃO, M. H. Identificação de variáveis fora de controle em processos produtivos multivariados. **Revista Produção**. v. 15, n.1. Jan/Abr., 2005.

SOUZA, A. M.; VICINI, L. Geração de subsídios para a tomada de decisão na cadeia produtiva da bovinocultura do Brasil. **GEPROS: Gestão de Produção, Operações e Sistemas**, ano 2, v.4, p.49-64, 2007.

SOUZA, F; et al. Previsão do preço da gasolina para a região sul do Brasil. **IJIE – Iberoamerican Journal of industrial Engineering**. v.3, n.1, julho, 2011.

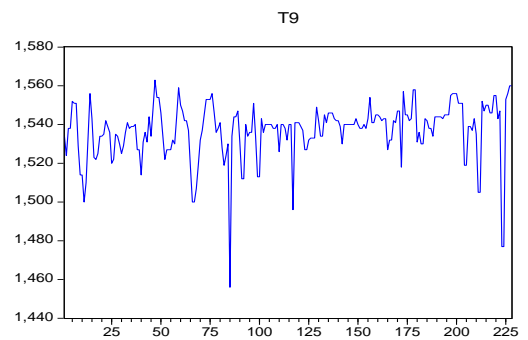
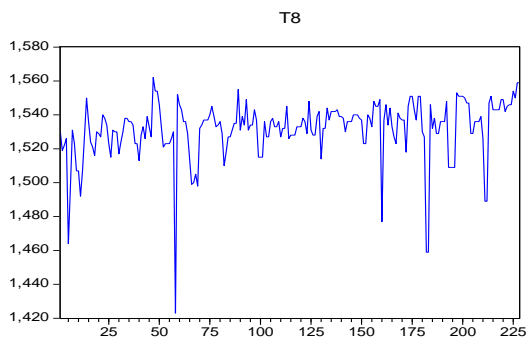
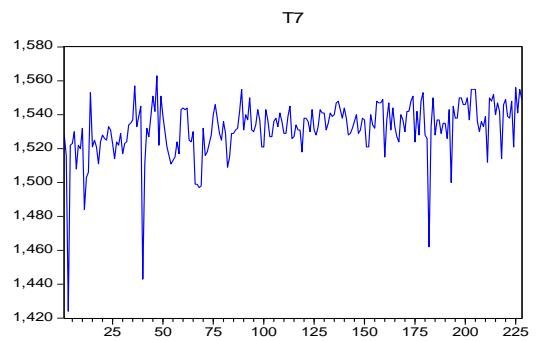
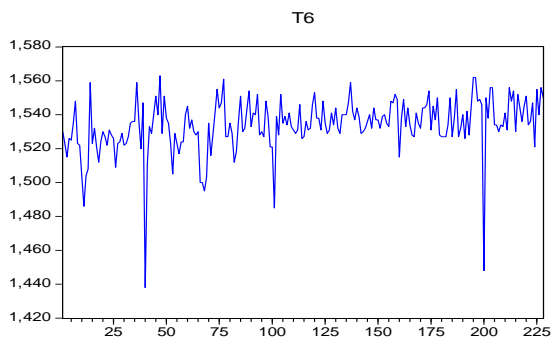
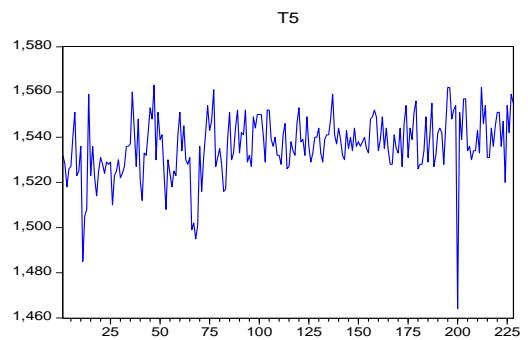
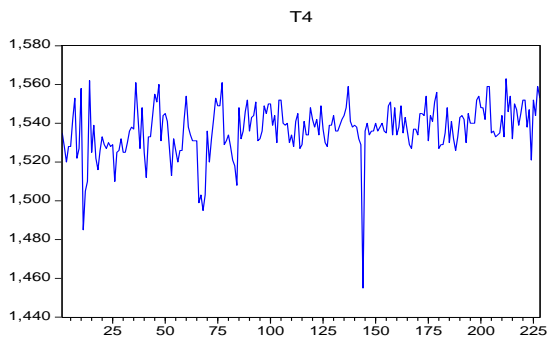
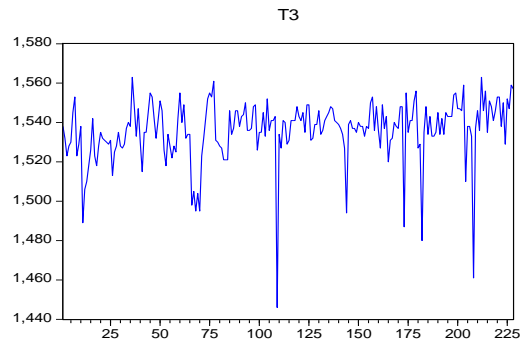
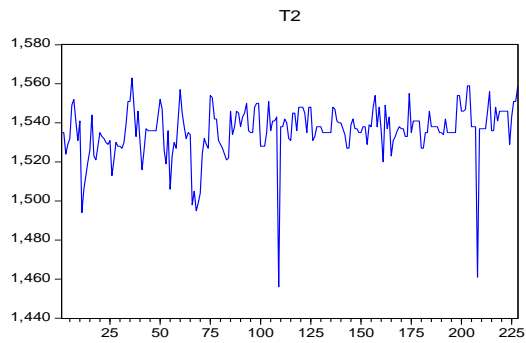
TRACY, N. D.; YOUNG, J. C.; MASON, R. L. Multivariate control charts for individual observations. **Journal of Quality Technology**, v.24. n.2, 1997.

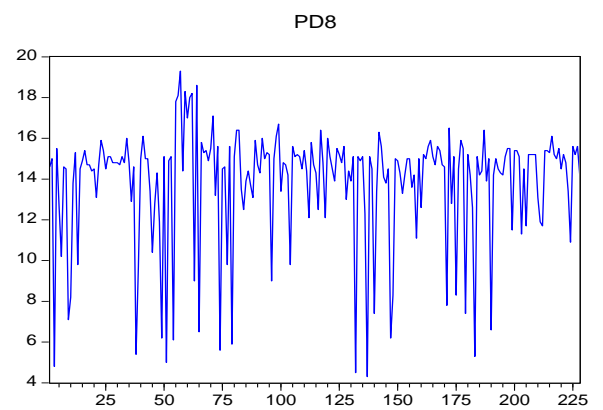
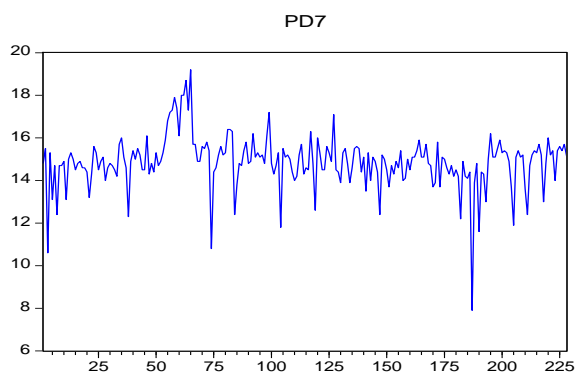
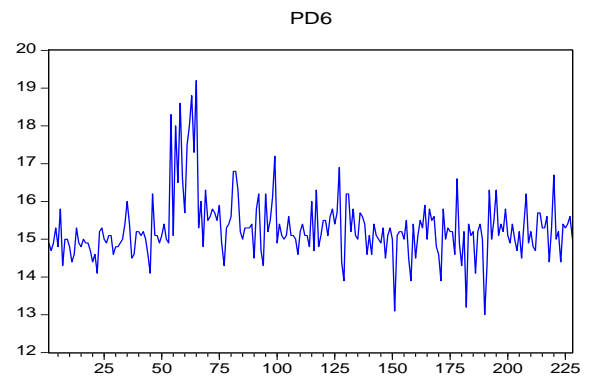
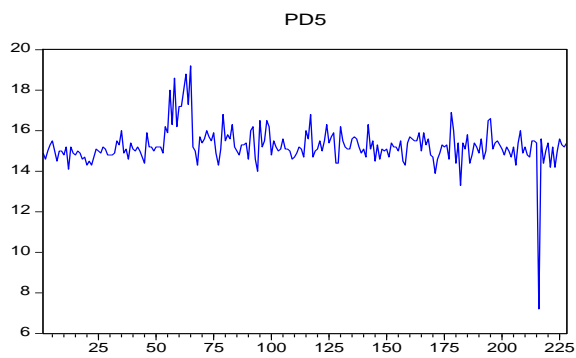
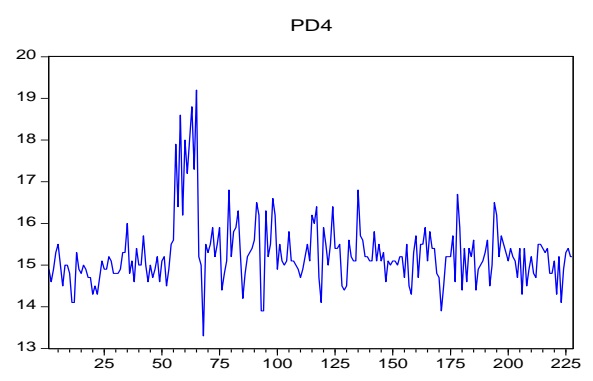
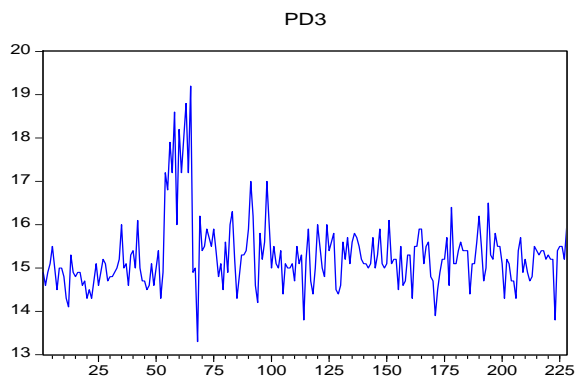
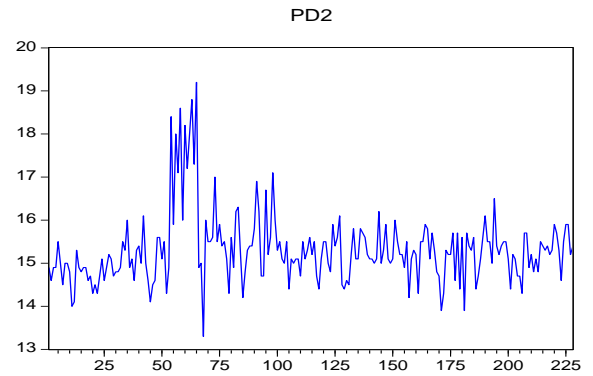
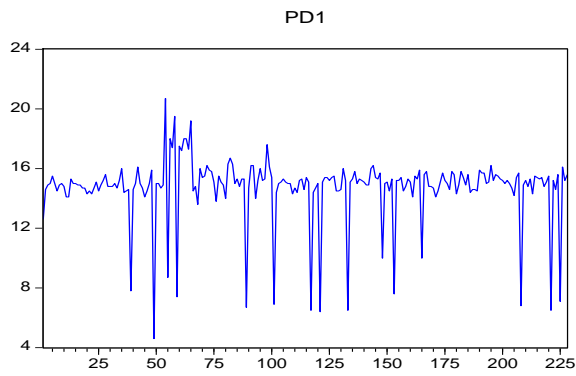
VASCONCELLOS, M. A. S.; ALVES, D. **Manual de econometria**. São Paulo. Atlas, 2002.

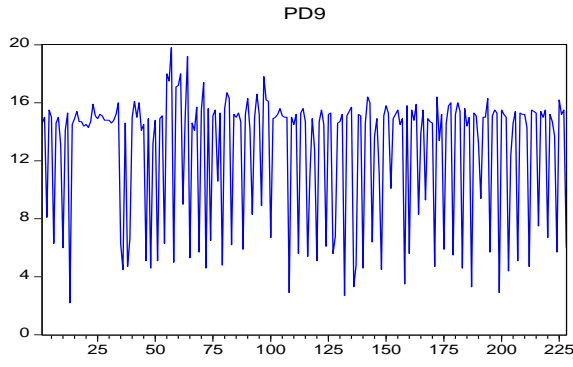
WOODALL, W. H.; NCUMBE, M. Multivariate CUSUM quality control procedure. **Technometrics**, v.27, n.3, 1985.

LISTA DE APÊNDICES

APÊNDICE A – Gráficos das variáveis originais das temperaturas do aço no distribuidor.



APÊNDICE B – Gráfico das variáveis originais dos pesos do distribuidor.



APÊNDICE C – FAC e FACP das variáveis originais das temperaturas do aço no distribuidor.

Variável T₂

| Autocorrelation | Partial Correlation | AC | PAC | Q-Stat | Prob |
|-----------------|---------------------|-----------|--------|--------|-------|
| | | 1 0.395 | 0.395 | 35.976 | 0.000 |
| | | 2 0.244 | 0.105 | 49.809 | 0.000 |
| | | 3 0.148 | 0.024 | 54.887 | 0.000 |
| | | 4 0.014 | -0.085 | 54.934 | 0.000 |
| | | 5 -0.077 | -0.093 | 56.330 | 0.000 |
| | | 6 -0.070 | -0.007 | 57.501 | 0.000 |
| | | 7 -0.084 | -0.024 | 59.173 | 0.000 |
| | | 8 -0.053 | 0.013 | 59.849 | 0.000 |
| | | 9 -0.060 | -0.036 | 60.723 | 0.000 |
| | | 10 -0.049 | -0.020 | 61.308 | 0.000 |
| | | 11 0.050 | 0.094 | 61.904 | 0.000 |
| | | 12 0.093 | 0.068 | 64.001 | 0.000 |
| | | 13 0.152 | 0.096 | 69.869 | 0.000 |
| | | 14 0.226 | 0.123 | 82.193 | 0.000 |
| | | 15 0.184 | 0.023 | 90.545 | 0.000 |

Variável T₃

| Autocorrelation | Partial Correlation | AC | PAC | Q-Stat | Prob |
|-----------------|---------------------|-----------|--------|--------|-------|
| | | 1 0.296 | 0.296 | 20.225 | 0.000 |
| | | 2 0.224 | 0.149 | 31.855 | 0.000 |
| | | 3 0.119 | 0.020 | 35.139 | 0.000 |
| | | 4 0.045 | -0.025 | 35.616 | 0.000 |
| | | 5 -0.062 | -0.100 | 36.510 | 0.000 |
| | | 6 -0.073 | -0.046 | 37.780 | 0.000 |
| | | 7 -0.074 | -0.018 | 39.064 | 0.000 |
| | | 8 -0.092 | -0.041 | 41.072 | 0.000 |
| | | 9 0.003 | 0.070 | 41.075 | 0.000 |
| | | 10 -0.031 | -0.026 | 41.307 | 0.000 |
| | | 11 0.010 | 0.013 | 41.333 | 0.000 |
| | | 12 0.065 | 0.062 | 42.359 | 0.000 |
| | | 13 0.086 | 0.045 | 44.176 | 0.000 |
| | | 14 0.140 | 0.098 | 48.981 | 0.000 |
| | | 15 0.159 | 0.080 | 55.195 | 0.000 |

Variável T₄

| Autocorrelation | Partial Correlation | AC | PAC | Q-Stat | Prob |
|-----------------|---------------------|----------|--------|--------|-------|
| | | 1 0.354 | 0.354 | 29.007 | 0.000 |
| | | 2 0.255 | 0.148 | 44.090 | 0.000 |
| | | 3 0.155 | 0.030 | 49.668 | 0.000 |
| | | 4 0.097 | 0.006 | 51.859 | 0.000 |
| | | 5 0.037 | -0.027 | 52.172 | 0.000 |
| | | 6 -0.017 | -0.047 | 52.241 | 0.000 |
| | | 7 -0.052 | -0.046 | 52.892 | 0.000 |
| | | 8 -0.013 | 0.032 | 52.932 | 0.000 |
| | | 9 -0.009 | 0.014 | 52.950 | 0.000 |
| | | 10 0.029 | 0.043 | 53.158 | 0.000 |
| | | 11 0.094 | 0.090 | 55.285 | 0.000 |
| | | 12 0.113 | 0.057 | 58.390 | 0.000 |
| | | 13 0.189 | 0.119 | 67.068 | 0.000 |
| | | 14 0.184 | 0.064 | 75.398 | 0.000 |
| | | 15 0.236 | 0.120 | 89.080 | 0.000 |

Variável T₅

| Autocorrelation | Partial Correlation | AC | PAC | Q-Stat | Prob |
|-----------------|---------------------|----------|--------|--------|-------|
| | | 1 0.319 | 0.319 | 23.527 | 0.000 |
| | | 2 0.240 | 0.154 | 36.871 | 0.000 |
| | | 3 0.127 | 0.015 | 40.629 | 0.000 |
| | | 4 0.029 | -0.053 | 40.820 | 0.000 |
| | | 5 0.025 | 0.008 | 40.964 | 0.000 |
| | | 6 -0.032 | -0.042 | 41.210 | 0.000 |
| | | 7 0.061 | 0.092 | 42.104 | 0.000 |
| | | 8 0.112 | 0.100 | 45.085 | 0.000 |
| | | 9 0.054 | -0.023 | 45.797 | 0.000 |
| | | 10 0.044 | -0.019 | 46.257 | 0.000 |
| | | 11 0.154 | 0.151 | 52.004 | 0.000 |
| | | 12 0.127 | 0.056 | 55.904 | 0.000 |
| | | 13 0.181 | 0.105 | 63.862 | 0.000 |
| | | 14 0.162 | 0.062 | 70.309 | 0.000 |
| | | 15 0.284 | 0.201 | 90.164 | 0.000 |

Variável T₆

| Autocorrelation | Partial Correlation | AC | PAC | Q-Stat | Prob |
|-----------------|---------------------|----------|--------|--------|-------|
| | | 1 0.270 | 0.270 | 16.782 | 0.000 |
| | | 2 0.229 | 0.169 | 28.970 | 0.000 |
| | | 3 0.086 | -0.012 | 30.708 | 0.000 |
| | | 4 -0.061 | -0.130 | 31.580 | 0.000 |
| | | 5 -0.024 | 0.004 | 31.715 | 0.000 |
| | | 6 -0.018 | 0.029 | 31.795 | 0.000 |
| | | 7 -0.003 | 0.013 | 31.796 | 0.000 |
| | | 8 0.051 | 0.043 | 32.414 | 0.000 |
| | | 9 0.079 | 0.057 | 33.893 | 0.000 |
| | | 10 0.061 | 0.012 | 34.762 | 0.000 |
| | | 11 0.134 | 0.096 | 39.133 | 0.000 |
| | | 12 0.114 | 0.061 | 42.278 | 0.000 |
| | | 13 0.181 | 0.126 | 50.237 | 0.000 |
| | | 14 0.167 | 0.081 | 57.105 | 0.000 |
| | | 15 0.260 | 0.186 | 73.706 | 0.000 |

Variável T₇

| Autocorrelation | Partial Correlation | AC | PAC | Q-Stat | Prob |
|-----------------|---------------------|----------|--------|--------|-------|
| | | 1 0.308 | 0.308 | 21.979 | 0.000 |
| | | 2 0.208 | 0.125 | 32.053 | 0.000 |
| | | 3 0.138 | 0.049 | 36.513 | 0.000 |
| | | 4 0.103 | 0.031 | 38.987 | 0.000 |
| | | 5 0.090 | 0.034 | 40.896 | 0.000 |
| | | 6 0.037 | -0.020 | 41.203 | 0.000 |
| | | 7 -0.027 | -0.063 | 41.379 | 0.000 |
| | | 8 0.133 | 0.163 | 45.617 | 0.000 |
| | | 9 0.080 | 0.016 | 47.131 | 0.000 |
| | | 10 0.110 | 0.058 | 50.058 | 0.000 |
| | | 11 0.121 | 0.061 | 53.614 | 0.000 |
| | | 12 0.143 | 0.074 | 58.554 | 0.000 |
| | | 13 0.198 | 0.110 | 68.135 | 0.000 |
| | | 14 0.184 | 0.067 | 76.410 | 0.000 |
| | | 15 0.236 | 0.159 | 90.111 | 0.000 |

Variável T₈

| Autocorrelation | Partial Correlation | AC | PAC | Q-Stat | Prob |
|-----------------|---------------------|-----------|--------|--------|-------|
| | | 1 0.385 | 0.385 | 34.196 | 0.000 |
| | | 2 0.139 | -0.011 | 38.656 | 0.000 |
| | | 3 0.062 | 0.014 | 39.551 | 0.000 |
| | | 4 -0.007 | -0.040 | 39.561 | 0.000 |
| | | 5 0.033 | 0.056 | 39.820 | 0.000 |
| | | 6 0.031 | 0.003 | 40.045 | 0.000 |
| | | 7 -0.030 | -0.054 | 40.252 | 0.000 |
| | | 8 -0.049 | -0.029 | 40.828 | 0.000 |
| | | 9 -0.064 | -0.034 | 41.823 | 0.000 |
| | | 10 -0.002 | 0.049 | 41.824 | 0.000 |
| | | 11 0.087 | 0.087 | 43.641 | 0.000 |
| | | 12 0.100 | 0.042 | 46.057 | 0.000 |
| | | 13 0.100 | 0.044 | 48.504 | 0.000 |
| | | 14 0.027 | -0.039 | 48.683 | 0.000 |

Variável T₉

| Autocorrelation | Partial Correlation | AC | PAC | Q-Stat | Prob |
|-----------------|---------------------|-----------|--------|--------|-------|
| | | 1 0.476 | 0.476 | 52.440 | 0.000 |
| | | 2 0.130 | -0.125 | 56.375 | 0.000 |
| | | 3 -0.003 | -0.018 | 56.378 | 0.000 |
| | | 4 -0.131 | -0.140 | 60.426 | 0.000 |
| | | 5 -0.131 | -0.004 | 64.475 | 0.000 |
| | | 6 -0.055 | 0.023 | 65.184 | 0.000 |
| | | 7 -0.020 | -0.013 | 65.279 | 0.000 |
| | | 8 -0.077 | -0.109 | 66.676 | 0.000 |
| | | 9 -0.080 | -0.019 | 68.216 | 0.000 |
| | | 10 -0.057 | -0.012 | 68.999 | 0.000 |
| | | 11 0.012 | 0.064 | 69.033 | 0.000 |
| | | 12 0.076 | 0.034 | 70.437 | 0.000 |
| | | 13 0.129 | 0.064 | 74.468 | 0.000 |
| | | 14 0.168 | 0.084 | 81.395 | 0.000 |
| | | 15 0.153 | 0.058 | 87.176 | 0.000 |

APÊNDICE D – FAC e FACP das variáveis originais dos pesos do distribuidor.

Variável PD₁

| Autocorrelation | Partial Correlation | AC | PAC | Q-Stat | Prob |
|-----------------|---------------------|-----------|--------|--------|-------|
| | | 1 -0.086 | -0.086 | 1.6966 | 0.193 |
| | | 2 -0.027 | -0.034 | 1.8617 | 0.394 |
| | | 3 -0.065 | -0.071 | 2.8405 | 0.417 |
| | | 4 0.203 | 0.192 | 12.459 | 0.014 |
| | | 5 -0.092 | -0.067 | 14.444 | 0.013 |
| | | 6 0.001 | -0.002 | 14.444 | 0.025 |
| | | 7 -0.037 | -0.019 | 14.760 | 0.039 |
| | | 8 0.003 | -0.050 | 14.763 | 0.064 |
| | | 9 -0.059 | -0.035 | 15.595 | 0.076 |
| | | 10 0.083 | 0.071 | 17.285 | 0.069 |
| | | 11 -0.043 | -0.030 | 17.714 | 0.088 |
| | | 12 0.139 | 0.149 | 22.410 | 0.033 |
| | | 13 0.029 | 0.072 | 22.622 | 0.046 |
| | | 14 -0.084 | -0.119 | 24.347 | 0.042 |
| | | 15 -0.035 | -0.007 | 24.642 | 0.055 |

Variável PD₂

| Autocorrelation | Partial Correlation | AC | PAC | Q-Stat | Prob |
|-----------------|---------------------|-----------|--------|--------|-------|
| | | 1 0.458 | 0.458 | 48.479 | 0.000 |
| | | 2 0.453 | 0.308 | 96.124 | 0.000 |
| | | 3 0.273 | -0.017 | 113.45 | 0.000 |
| | | 4 0.285 | 0.088 | 132.46 | 0.000 |
| | | 5 0.215 | 0.035 | 143.33 | 0.000 |
| | | 6 0.203 | 0.023 | 153.04 | 0.000 |
| | | 7 0.257 | 0.147 | 168.77 | 0.000 |
| | | 8 0.203 | 0.005 | 178.61 | 0.000 |
| | | 9 0.218 | 0.029 | 189.99 | 0.000 |
| | | 10 0.039 | -0.180 | 190.35 | 0.000 |
| | | 11 0.084 | 0.002 | 192.06 | 0.000 |
| | | 12 -0.006 | -0.038 | 192.07 | 0.000 |
| | | 13 -0.003 | -0.054 | 192.07 | 0.000 |
| | | 14 -0.016 | 0.008 | 192.13 | 0.000 |
| | | 15 0.052 | 0.082 | 192.81 | 0.000 |

Variável PD₃

| Autocorrelation | Partial Correlation | AC | PAC | Q-Stat | Prob |
|-----------------|---------------------|-----------|--------|--------|-------|
| | | 1 0.504 | 0.504 | 58.739 | 0.000 |
| | | 2 0.441 | 0.250 | 103.84 | 0.000 |
| | | 3 0.300 | 0.010 | 124.79 | 0.000 |
| | | 4 0.290 | 0.091 | 144.51 | 0.000 |
| | | 5 0.224 | 0.018 | 156.28 | 0.000 |
| | | 6 0.183 | -0.005 | 164.18 | 0.000 |
| | | 7 0.232 | 0.130 | 176.97 | 0.000 |
| | | 8 0.160 | -0.035 | 183.07 | 0.000 |
| | | 9 0.191 | 0.054 | 191.79 | 0.000 |
| | | 10 -0.002 | -0.211 | 191.79 | 0.000 |
| | | 11 0.046 | 0.021 | 192.30 | 0.000 |
| | | 12 -0.049 | -0.072 | 192.88 | 0.000 |
| | | 13 -0.038 | -0.027 | 193.24 | 0.000 |
| | | 14 -0.029 | 0.044 | 193.45 | 0.000 |
| | | 15 0.018 | 0.068 | 193.53 | 0.000 |

Variável PD₄

| Autocorrelation | Partial Correlation | AC | PAC | Q-Stat | Prob |
|-----------------|---------------------|-----------|--------|--------|-------|
| | | 1 0.470 | 0.470 | 51.129 | 0.000 |
| | | 2 0.418 | 0.252 | 91.587 | 0.000 |
| | | 3 0.267 | 0.002 | 108.19 | 0.000 |
| | | 4 0.254 | 0.075 | 123.34 | 0.000 |
| | | 5 0.131 | -0.062 | 127.39 | 0.000 |
| | | 6 0.101 | -0.018 | 129.79 | 0.000 |
| | | 7 0.179 | 0.166 | 137.39 | 0.000 |
| | | 8 0.102 | -0.042 | 139.88 | 0.000 |
| | | 9 0.134 | 0.042 | 144.19 | 0.000 |
| | | 10 0.021 | -0.100 | 144.30 | 0.000 |
| | | 11 -0.045 | -0.147 | 144.79 | 0.000 |
| | | 12 -0.068 | 0.006 | 145.90 | 0.000 |
| | | 13 -0.094 | -0.031 | 148.04 | 0.000 |
| | | 14 -0.030 | 0.076 | 148.26 | 0.000 |
| | | 15 0.001 | 0.088 | 148.26 | 0.000 |

Variável PD₅

| Autocorrelation | Partial Correlation | AC | PAC | Q-Stat | Prob |
|-----------------|---------------------|-----------|--------|--------|-------|
| | | 1 0.265 | 0.265 | 16.175 | 0.000 |
| | | 2 0.304 | 0.252 | 37.683 | 0.000 |
| | | 3 0.180 | 0.061 | 45.208 | 0.000 |
| | | 4 0.209 | 0.100 | 55.396 | 0.000 |
| | | 5 0.184 | 0.080 | 63.355 | 0.000 |
| | | 6 0.157 | 0.035 | 69.180 | 0.000 |
| | | 7 0.219 | 0.121 | 80.574 | 0.000 |
| | | 8 0.068 | -0.075 | 81.665 | 0.000 |
| | | 9 0.090 | -0.024 | 83.622 | 0.000 |
| | | 10 0.131 | 0.086 | 87.748 | 0.000 |
| | | 11 0.022 | -0.088 | 87.867 | 0.000 |
| | | 12 0.017 | -0.057 | 87.940 | 0.000 |
| | | 13 -0.000 | -0.014 | 87.940 | 0.000 |
| | | 14 0.015 | -0.016 | 87.994 | 0.000 |
| | | 15 0.022 | 0.036 | 88.116 | 0.000 |

Variável PD₆

| Autocorrelation | Partial Correlation | AC | PAC | Q-Stat | Prob |
|-----------------|---------------------|-----------|--------|--------|-------|
| | | 1 0.336 | 0.336 | 26.091 | 0.000 |
| | | 2 0.351 | 0.268 | 54.665 | 0.000 |
| | | 3 0.278 | 0.123 | 72.656 | 0.000 |
| | | 4 0.314 | 0.160 | 95.678 | 0.000 |
| | | 5 0.189 | -0.016 | 104.06 | 0.000 |
| | | 6 0.266 | 0.112 | 120.77 | 0.000 |
| | | 7 0.305 | 0.165 | 142.78 | 0.000 |
| | | 8 0.190 | -0.039 | 151.41 | 0.000 |
| | | 9 0.172 | -0.016 | 158.49 | 0.000 |
| | | 10 0.153 | -0.019 | 164.10 | 0.000 |
| | | 11 0.163 | 0.019 | 170.51 | 0.000 |
| | | 12 0.025 | -0.119 | 170.66 | 0.000 |
| | | 13 0.027 | -0.099 | 170.85 | 0.000 |
| | | 14 -0.024 | -0.102 | 170.99 | 0.000 |
| | | 15 0.010 | 0.002 | 171.02 | 0.000 |

Variável PD₇

| Autocorrelation | Partial Correlation | AC | PAC | Q-Stat | Prob |
|-----------------|---------------------|-----------|--------|--------|-------|
| | | 1 0.319 | 0.318 | 23.386 | 0.000 |
| | | 2 0.251 | 0.166 | 37.965 | 0.000 |
| | | 3 0.256 | 0.157 | 53.271 | 0.000 |
| | | 4 0.229 | 0.101 | 65.516 | 0.000 |
| | | 5 0.209 | 0.076 | 75.804 | 0.000 |
| | | 6 0.245 | 0.119 | 89.995 | 0.000 |
| | | 7 0.158 | -0.007 | 95.909 | 0.000 |
| | | 8 0.115 | -0.026 | 99.039 | 0.000 |
| | | 9 0.009 | -0.130 | 99.058 | 0.000 |
| | | 10 0.083 | 0.030 | 100.72 | 0.000 |
| | | 11 0.059 | -0.008 | 101.55 | 0.000 |
| | | 12 -0.011 | -0.070 | 101.58 | 0.000 |
| | | 13 -0.013 | -0.026 | 101.62 | 0.000 |
| | | 14 0.015 | 0.028 | 101.67 | 0.000 |
| | | 15 -0.015 | 0.011 | 101.73 | 0.000 |

Variável PD₈

| Autocorrelation | Partial Correlation | AC | PAC | Q-Stat | Prob |
|-----------------|---------------------|-----------|--------|--------|-------|
| | | 1 -0.025 | -0.025 | 0.1477 | 0.701 |
| | | 2 0.036 | 0.035 | 0.4480 | 0.799 |
| | | 3 0.038 | 0.040 | 0.7817 | 0.854 |
| | | 4 0.071 | 0.072 | 1.9559 | 0.744 |
| | | 5 0.033 | 0.034 | 2.2139 | 0.819 |
| | | 6 -0.105 | -0.111 | 4.8260 | 0.566 |
| | | 7 0.059 | 0.046 | 5.6494 | 0.581 |
| | | 8 -0.017 | -0.014 | 5.7164 | 0.679 |
| | | 9 0.047 | 0.047 | 6.2359 | 0.716 |
| | | 10 -0.061 | -0.049 | 7.1426 | 0.712 |
| | | 11 0.084 | 0.082 | 8.8646 | 0.634 |
| | | 12 0.008 | -0.002 | 8.8814 | 0.713 |
| | | 13 -0.080 | -0.077 | 10.438 | 0.658 |
| | | 14 -0.003 | -0.015 | 10.441 | 0.729 |
| | | 15 -0.072 | -0.066 | 11.720 | 0.700 |

Variável PD₉

| Autocorrelation | Partial Correlation | AC | PAC | Q-Stat | Prob |
|-----------------|---------------------|-----------|--------|--------|-------|
| | | 1 -0.200 | -0.200 | 9.2690 | 0.002 |
| | | 2 -0.204 | -0.255 | 18.946 | 0.000 |
| | | 3 -0.061 | -0.182 | 19.804 | 0.000 |
| | | 4 0.443 | 0.377 | 65.671 | 0.000 |
| | | 5 -0.110 | 0.055 | 68.542 | 0.000 |
| | | 6 -0.270 | -0.175 | 85.785 | 0.000 |
| | | 7 -0.027 | -0.117 | 85.954 | 0.000 |
| | | 8 0.312 | 0.073 | 109.18 | 0.000 |
| | | 9 -0.070 | 0.024 | 110.37 | 0.000 |
| | | 10 -0.123 | 0.107 | 114.03 | 0.000 |
| | | 11 -0.081 | -0.079 | 115.61 | 0.000 |
| | | 12 0.306 | 0.092 | 138.42 | 0.000 |
| | | 13 -0.061 | 0.013 | 139.33 | 0.000 |
| | | 14 -0.138 | -0.032 | 143.97 | 0.000 |
| | | 15 -0.143 | -0.146 | 148.99 | 0.000 |