

**UNIVERSIDADE FEDERAL DE SANTA MARIA
CENTRO DE CIÊNCIAS NATURAIS E EXATAS
CURSO DE ESPECIALIZAÇÃO EM ESTATÍSTICA E MODELAGEM
QUANTITATIVA**

**ANÁLISE DO VOLUME DE VENDAS DE ÓLEO
DIESEL E GASOLINA NO RS, 2000 A 2010**

MONOGRAFIA DE ESPECIALIZAÇÃO

Silvana Gonçalves de Almeida

Santa Maria, RS, Brasil

2010

ANÁLISE DO VOLUME DE VENDAS DE ÓLEO DIESEL E GASOLINA NO RS, 2000 A 2010

por

Silvana Gonçalves de Almeida

Monografia apresentada ao Curso de Especialização em Estatística e Modelagem Quantitativa, da Universidade Federal de Santa Maria (UFSM, RS), como requisito parcial para obtenção do grau de **Especialista em Estatística e Modelagem Quantitativa.**

Orientadora: Prof. Dr. Adriano Mendonza Souza
Co-Orientadora: Prof. Dra. Roselaine Ruviaro Zanini

Santa Maria, RS, Brasil

2010

**Universidade Federal de Santa Maria
Centro de Ciências Naturais e Exatas
Curso de Especialização em Estatística e Modelagem Quantitativa**

A Comissão Examinadora, abaixo assinada,
aprova a Monografia de Especialização

**ANÁLISE DO VOLUME DE VENDAS DE ÓLEO
DIESEL E GASOLINA NO RS, 2000 A 2010**

elaborada por
Silvana Gonçalves de Almeida

como requisito parcial para obtenção do grau de
Especialista em Estatística e Modelagem Quantitativa

COMISSÃO EXAMINADORA:

Adriano Mendonza Souza, Dr. (UFSM)
(Orientador)

Roselaine Ruviaro Zanini, Dra. (UFSM)
(Presidente/Co-Orientadora)

Dênis Rasquin Rabenschlag, Dr. (UFSM)

Anaelena Bragança de Moraes, Dra. (UFSM)

Santa Maria, 26 de julho de 2010

AGRADECIMENTOS

Agradeço,

A todos os professores do Curso de Especialização em Estatística e Modelagem Quantitativa, especialmente ao professor Adriano Mendonza Souza, pela oportunidade, paciência e exímio trabalho de orientação e à professora Roselaine Ruviaro Zanini pela dedicação e paciência na co-orientação do trabalho.

Aos meus amigos, pelo incentivo e carinho.

À minha família, que sempre encorajou minha decisão de aprimorar meus conhecimentos.

RESUMO

Monografia de Especialização
Curso de Especialização em Estatística e Modelagem Quantitativa
Universidade Federal de Santa Maria

ANÁLISE DO VOLUME DE VENDAS DE ÓLEO DIESEL E GASOLINA NO RS, 2000 A 2010

Autor: Silvana Gonçalves de Almeida

Orientador: Adriano Mendonza Souza

Co-Orientadora: Roselaine Ruviaro Zanini

Data e local da defesa: Santa Maria, 26 julho de 2010.

A presente pesquisa tem como tema a abordagem de modelos de previsão aplicados ao volume de vendas de óleo diesel e gasolina, principais derivados combustíveis do petróleo no Estado do RS, com o objetivo de verificar a estabilidade do processo por meio de gráficos de controle aplicados aos resíduos oriundos da modelagem linear e com isto, detectar causas econômicas que tiveram influência no processo. Como as observações eram autocorrelacionadas foi necessário buscar um modelo matemático no qual foram obtidos resíduos independentes e normalmente distribuídos. Os modelos encontrados para as séries de vendas de barris de óleo diesel e gasolina foram, SARIMA(0,1,1)(0,1,1)₁₂ e SARIMA(0,1,2)(0,1,1)₁₂ respectivamente. Tais modelos foram capazes de modelar satisfatoriamente os dados e juntamente com o emprego das ferramentas estatísticas de controle estatístico do processo possibilitaram uma análise do comportamento da série, detectando possíveis causas econômicas que tiveram influência sobre a variação do volume de vendas de óleo diesel e gasolina no RS. A realização de previsões em curto prazo, em um mercado regional, é importante para qualquer planejamento estratégico, decisão de investimento e financiamento de produção, também como orientação nas tomadas de decisões para o setor.

Palavras-chaves: séries temporais; gráficos de controle; previsão; óleo diesel; gasolina.

ABSTRACT

Monograph of Specialization
Specialization Course in Statistics and Quantitative Modeling
Federal University of Santa Maria

Author: Silvana Gonçalves de Almeida
Advisor: Adriano Mendonza Souza
Co-advisor: Roselaine Ruviaro Zanini
Place and Date of defense: Santa Maria, July 26, 2010

This research theme is the approach of forecasting models, applied to sales of diesel and gasoline, fuel oil major byproducts in the State of RS, in order to verify process stability through control charts, applied to waste derived from models of linear modeling and with this, detecting economic causes which influenced the process. As the observations were autocorrelated, it was necessary to seek a mathematical model in which residues were obtained independent and normally distributed. The models found for the series of sales of barrels of diesel and gasoline were SARIMA (0,1,1) (0,1,1)₁₂ and SARIMA (0,1,2) (0,1,1)₁₂ respectively. Such models were able to model the data satisfactorily and with the use of statistical tools for CEP enabled an analysis of the behavior of the series, detecting possible economic causes that influence the change in volume of sales of fuels derived from petroleum. Moreover, the realization of short-term forecasts in a regional market, is important for any strategic planning, investment decision and financing of production, as well as policy guidance for the sector.

Words keys: series; control charts; prediction; diesel; gasoline.

LISTA DE FIGURAS

FIGURA 1 -	Gráfico de controle adaptado de Montgomery, 2004.....	18
FIGURA 2 -	Representação gráfica para o teste de sequência.....	19
FIGURA 3 -	Algumas sugestões para escolha do gráfico de controle.....	20
FIGURA 4 -	Série original e diferenciada dos números de barris de óleo diesel vendidos no Estado do RS no período de janeiro de 2000 a fevereiro de 2010.....	38
FIGURA 5 -	Função de Autocorrelação e Autocorrelação Parcial da série original das vendas de barris de óleo diesel.....	38
FIGURA 6-	Função de Autocorrelação e Autocorrelação Parcial para os resíduos do modelo SARIMA $(0,1,1)(0,1,1)_{12}$ encontrado para as vendas de barris de óleo diesel.....	40
FIGURA 7 -	Gráfico Q-Q <i>plot</i> dos resíduos do modelo SARIMA $(0,1,1)(0,1,1)_{12}$	41
FIGURA 8 -	Gráfico dos valores originais e valores estimados do modelo SARIMA $(0,1,1)(0,1,1)_{12}$	42
FIGURA 9 -	Gráfico de controle para os valores individuais e amplitude móvel considerando os dados originais.....	43
FIGURA 10 -	Gráfico de controle para os valores individuais e amplitude móvel para os resíduos do modelo SARIMA $(0,1,1)(0,1,1)_{12}$	44
FIGURA 11-	Gráfico de controle para os valores individuais e amplitude móvel para os resíduos do modelo SARIMA $(0,1,1)(0,1,1)_{12}$ com a exclusão dos pontos amostrais fora de controle.....	46
FIGURA 12 -	Série original e diferenciada dos números de barris de gasolina vendidos no Estado do RS no período de janeiro de 2000 a fevereiro de 2010.....	48
FIGURA 13 -	Função de Autocorrelação e Autocorrelação Parcial da série original das vendas de barris de gasolina.....	48
FIGURA 14-	Função de Autocorrelação e Autocorrelação Parcial para os resíduos do modelo SARIMA $(0,1,2)(0,1,1)_{12}$ encontrado para as vendas de barris de gasolina.....	51
FIGURA 15 -	Gráfico Q-Q <i>plot</i> dos resíduos do modelo SARIMA $(0,1,2)(0,1,1)_{12}$	51
FIGURA 16 -	Gráfico dos valores originais e valores estimados do modelo SARIMA $(0,1,2)(0,1,1)_{12}$	52
FIGURA 17 -	Gráfico de controle para os valores individuais e amplitude móvel considerando os dados originais.....	53
FIGURA 18 -	Gráfico de controle para os valores individuais e amplitude móvel para os resíduos do modelo SARIMA $(0,1,2)(0,1,1)_{12}$	55
FIGURA 19 -	Gráfico de controle para os valores individuais e amplitude móvel para os resíduos do modelo SARIMA $(0,1,2)(0,1,1)_{12}$ com a exclusão dos pontos amostrais fora de controle.....	57

LISTA DE TABELAS

TABELA 1 - Estatística descritiva para o volume de vendas, de todos os derivados combustíveis de Petróleo no Estado do Rio Grande do Sul.....	36
TABELA 2 - Estimação dos parâmetros para os modelos SARIMA concorrentes da variável venda de barris de óleo diesel no período de janeiro de 2000 a fevereiro de 2010.....	39
TABELA 3 - Período, valores originais para os meses de março e abril de 2010 e os valores previstos do modelo SARIMA(0,1,1)(0,1,1) ₁₂	42
TABELA 4 - Estimação dos parâmetros dos modelos SARIMA e ARFIMA concorrentes da variável venda de barris de gasolina no período de janeiro de 2000 a fevereiro de 2010.....	50
TABELA 5 - Período, valores originais para os meses de março e abril de 2010 e os valores previstos do modelo SARIMA(0,1,2)(0,1,1) ₁₂	53

LISTA DE ABREVIATURAS E SIGLAS

ADF	Teste de raiz unitária de Dickey e Fuller Aumentado
AIC	Critério de Informação Akaike
AR	Autorregressivo
ARIMA	Autorregressivo Integrado de Média Móvel
ARMA	Autorregressivo de Média Móvel
CEP	Controle Estatístico de Processo
DF	Teste de raiz unitária de Dickey e Fuller
FAC	Função de Autocorrelação
FACP	Função de Autocorrelação Parcial
IC	Intervalo de Confiança
KPSS	Teste de raiz unitária proposto pelos autores Kwiatkowski, Phillips, Schmidt e Shin
LC	Linha Central
LIC	Limite Inferior de Controle
LSC	Limite Superior de Controle
MA	Média Móvel
OPEP	Organização dos Países Produtores de Petróleo

SUMÁRIO

1. INTRODUÇÃO	12
1.1 Justificativa	14
1.2 Objetivos	14
1.2.1 Objetivo geral.....	14
1.2.2 Objetivos específicos.....	15
1.3 Delimitação do estudo	15
2. REVISÃO DA LITERATURA	16
2.1 Gráficos de controle	16
2.1.1 Gráfico de controle para medidas individuais e amplitude móvel.....	20
2.2 Análise de Séries Temporais	22
2.2.1 Modelo auto-regressivo: AR(p).....	23
2.2.2 Modelo de médias móveis: MA(q).....	24
2.2.3 Modelo auto-regressivo de médias móveis: ARMA(p,q).....	24
2.2.4 Modelo auto-regressivo integrado de médias móveis: ARIMA(p,d,q).....	25
2.2.5 Modelo sazonal auto-regressivo integrado de médias móveis: SARIMA(p,d,q)(P,D,Q) _s	26
2.2.6 Modelo auto-regressivo fracionário integrado de médias móveis: ARFIMA(p,d,q).....	26
2.3 Testes de diagnóstico	28
2.3.1 Testes de raiz unitária.....	28
2.3.2 Testes de normalidade.....	30
2.3.3 Testes para autocorrelação.....	32
2.3.4 Critério penalizador.....	33
3. METODOLOGIA	34
3.1 Etapas metodológicas	34
4. RESULTADOS E DISCUSSÕES	36
4.1 Análise Descritiva	36

4.2 Identificação do modelo e estimação dos parâmetros para as vendas de óleo diesel.....	37
4.3 Análise residual para o modelo SARIMA (0,1,1)(0,1,1)₁₂.....	40
4.4 Previsão.....	40
4.5 Aplicação de gráficos de controle no resíduo do modelo SARIMA (0,1,1)(0,1,1)₁₂ para analisar a estabilidade do processo.....	43
4.6 Estimação do modelo para as vendas de gasolina.....	47
4.7 Análise residual para o modelo SARIMA (0,1,2)(0,1,1)₁₂.....	50
4.8 Previsão.....	52
4.9 Aplicação de gráficos de controle no resíduo do modelo SARIMA (0,1,2)(0,1,1)₁₂ para analisar a estabilidade do processo.....	53
4.10 Comentários gerais das análises.....	58
5. CONCLUSÃO.....	60
6. REFERÊNCIAS.....	62

1. INTRODUÇÃO

A partir de meados do século XIX, o petróleo começou a ser utilizado em maior escala, nos Estados Unidos, como substituto do óleo de baleia na iluminação e do carvão mineral na produção do vapor. Na década de 90 o óleo diesel e a gasolina passaram a ser utilizados como os principais combustíveis para os meios de transporte, o que gerou um aumento acelerado do consumo e conseqüentemente da produção de petróleo tornando-se essa a principal fonte da matriz energética mundial.

O petróleo é conhecido à milhões de anos nas regiões onde o mesmo aflorava espontaneamente para a superfície e pelas propriedades que os povos do antigo Oriente Médio faziam uso, como lubrificante, combustíveis de diversas aplicações e até mesmo para aplicações medicinais (WIKIPÉDIA, 2010).

O petróleo traz riqueza a quem o produz, e supre as necessidades de energia e matéria prima de quem o consome. Sua importância estratégica faz dele uma fonte potencial de conflitos, uma das razões pela qual o Oriente Médio é uma das regiões do mundo que mais despertam interesse e preocupação global. Hoje, o petróleo é responsável por quase a metade de toda a energia gerada no mundo. Sem ele as usinas termelétricas parariam de funcionar, deixando cidades às escuras. Veículos terrestres, navios e aviões ficariam parados, indústrias não produziriam nada e os habitantes dos países frios congelariam no inverno, sem o combustível da calefação doméstica.

O Óleo diesel e a gasolina no Brasil são componentes essenciais do processo produtivo, sendo os combustíveis mais usados no país. Ressalta-se que o óleo diesel representa o insumo básico do transporte rodoviário, que é responsável pela maior parte do transporte de carga e de passageiros por meio coletivos (MOREIRA, 1996).

Embora existam fontes renováveis de energia, como a energia solar, energia eólica, hidráulica e outras, que podem substituir em grande parte a produção de petróleo, as mesmas não são auto-suficientes a ponto de substituí-lo totalmente.

Conforme o Informe Executivo do Sistema FARSUL (POA, 14 de dezembro de 2009), em relação ao perfil de vendas existe destaque para o óleo diesel e a gasolina, produtos de maior valor agregado e que representam cerca de 52,3% da comercialização de derivados. Por

meio dos dados coletados, na Agência Nacional do Petróleo, Gás Natural e Bicomcombustíveis (ANP), no período de janeiro de 2000 a fevereiro de 2010 para o estado do Rio Grande do Sul, pode-se constatar que dos derivados combustíveis do petróleo, o óleo diesel e a gasolina apresentaram maior volume de vendas, isto se deve ao fato, de serem utilizados para gerar energia e movimentar máquinas e motores de grande porte, tais como: automóveis de passeio, furgões, ônibus, caminhões, pequenas embarcações marítimas, máquinas de grande porte, locomotivas, navios.

Como o petróleo é uma das principais *commodities* minerais produzidas pelo Brasil e além de ser a principal fonte de renda de muitos países, optou-se em trabalhar com as vendas dos derivados combustíveis do petróleo no Estado do Rio Grande do Sul, uma vez que a demanda contínua e crescente de energia de baixo custo e a disponibilidade de recursos de hidrocarbonetos colocam ainda o petróleo como uma importante fonte não-renovável da matriz energética mundial para as próximas décadas do século XXI.

Este estudo possui um caráter relevante no setor, por ser uma forma de fazer previsão do volume de vendas de barris de óleo diesel e gasolina, para o RS, e também detectar possíveis pontos anômalos do volume de vendas e suas implicações econômicas. Os modelos de previsão são ferramentas de grande importância a partir dos quais os riscos gerados pela incerteza são minimizados, proporcionando uma maior credibilidade nas tomadas de decisões. Além dos modelos de previsão, as ferramentas do controle estatístico do processo, mostraram-se eficazes, possibilitando uma melhoria no controle das previsões e também como forma de detectar possíveis pontos onde acarretou maior ou menor volume de vendas no setor.

Segundo Montgomery (2004), os gráficos estatísticos de controle são utilizados em qualquer tipo de processo, para alcançar um estado de controle estatístico e assim monitorá-lo. Existem no mínimo cinco razões que garantem a importância dos gráficos de controle. Entre elas destacam-se a técnica comprovada para a melhoria da produtividade, a eficácia na prevenção de defeitos, evitam ajustes desnecessários no processo, fornecem informações de diagnóstico e capacidade do processo.

Além disso, o uso da metodologia de séries temporais, segundo Alwan e Roberts (1998), ajuda a detectar a existência de variações e a obter resultados mais concretos das causas especiais de variação, por meio do gráfico de controle aplicado nos resíduos do modelo matemático.

Conforme Mourão (2009), a previsão da demanda é um componente primordial do planejamento estratégico da produção, de vendas e de finanças de qualquer empresa, onde os

gestores se baseiam para desenvolver os planos de capacidade, de fluxo de caixa, de produção e estoques, de mão-de-obra, de compras e vendas.

1.1 Justificativa

Nesta pesquisa, optou-se por trabalhar com o volume de vendas dos derivados combustíveis do petróleo no Estado do Rio Grande do Sul, uma vez que a demanda contínua e crescente de energia de baixo custo e a disponibilidade de recursos de hidrocarbonetos colocam ainda o petróleo como uma importante fonte não-renovável da matriz energética mundial para as próximas décadas do século XXI.

O óleo diesel e a gasolina, combustíveis derivados do petróleo, passaram, e continuam passando, por mudanças, tais como estratégias de vendas e precificações, inseridas pelo Governo Federal ou pelo próprio mercado. Com isto, a importância de um combustível em relação a outro mudou ao longo dos tempos, o que ocasionou diversos processos de substituição entre estes devido à falta de um planejamento estratégico bem definido.

Dentro deste contexto, salienta-se a importância em estudar e analisar o comportamento da demanda de combustíveis, no qual tende a ser uma ferramenta de orientação de política para o setor, bem como de apoio para a tomada de decisão.

Cabe mencionar que além de estudar o funcionamento do mercado de combustíveis, a estimação, bem como a previsão do volume de vendas dos derivados do petróleo é justificada pela necessidade em avaliar e planejar a infra-estrutura, iniciando pela produção até a distribuição, visando garantir e atender a demanda por combustível.

As previsões do volume de vendas de óleo diesel e gasolina são fundamentais para garantir de forma satisfatória o abastecimento do mercado realizado pelas distribuidoras autorizadas para tal fim. São feitas com base no histórico de vendas das distribuidoras e nas suas promessas de compras futuras. Ao analisar o volume de vendas de óleo diesel e gasolina pode-se perceber se o mesmo encontra-se constante ao longo do tempo ou existiram períodos anômalos nas vendas e o que influenciou esse fato economicamente.

1.2 Objetivos

1.2.1 Objetivo geral

Modelar processos autocorrelacionados, referentes ao volume de vendas de barris de óleo diesel e gasolina no RS.

1.2.2 Objetivos específicos

- Realizar a modelagem linear para tratar a autocorrelação serial;
- Realizar a previsão com a finalidade de perceber o comportamento do volume de vendas nos próximos períodos;
- Comparar os efeitos da correlação, utilizando gráficos de controle nos dados originais e nos resíduos sem a presença de autocorrelação serial.

1.3 Delimitações do estudo

A delimitação da pesquisa está na utilização da metodologia de séries temporais em séries de vendas de combustíveis derivados do petróleo que juntamente com a metodologia de gráficos de controle pretende-se analisar o volume dos derivados combustíveis ao longo do período em estudo. Optou-se em trabalhar com os derivados combustíveis do petróleo no Rio Grande do Sul, que apresentaram o maior volume de vendas no período de janeiro de 2000 a fevereiro de 2010, ou seja, o óleo diesel e a gasolina, excluindo do estudo os demais combustíveis. Os modelos matemáticos utilizados para modelar a série em estudos se restringem a modelos Lineares (ARIMA).

Portanto esse primeiro capítulo teve enfoque no tema da pesquisa, justificativa, importância do trabalho, objetivos e delimitação da pesquisa.

2. REVISÃO DA LITERATURA

Este capítulo apresenta a pesquisa bibliográfica realizada para a elaboração deste trabalho. O embasamento teórico primeiramente apresentará a metodologia dos gráficos de controle, na sequência, abordam-se os modelos de previsão, os modelos auto-regressivos integrados de médias móveis (ARIMA), modelo sazonal auto-regressivo integrado de médias móveis (SARIMA) e modelo auto-regressivo fracionário de médias móveis (ARFIMA), seguidos dos testes estatísticos utilizados.

2.1 Gráficos de controle

De acordo com Montgomery (2004) controle estatístico do processo é o ramo do controle de qualidade que consiste na coleta, análise e interpretação de dados para utilização nas atividades de melhoria e controle da qualidade de produtos e serviços. Não existe uma única definição para qualidade; para Deming (2000), qualidade significa atender e, se possível, exceder as expectativas do consumidor. Com isso, pode-se dizer que os gráficos de controle auxiliam na melhoria contínua da qualidade do produto, podendo ser utilizados para reduzir não-conformidades em tempo real, mostrando quando necessário, a necessidade da existência de ações no processo que está sendo monitorado.

Segundo Montgomery e Runger (2003) o objetivo da aplicação dos gráficos de controle é identificar quando os processos analisados sofrem alterações e quando essas mudanças podem afetar a qualidade do serviço oferecido, além de verificar se os processos em análise são estatisticamente estáveis e se tendem a permanecerem assim.

Em qualquer processo, existe uma determinada quantidade de variabilidade inerente ou natural. Para que se possa controlar a qualidade de um produto é necessário ter habilidade para se medir as variações que ocorrem no processo. Segundo Siqueira (1997) existe seis fatores que contribuem para essas variações, são elas: máquinas, métodos, materiais, meio ambiente, mão-de-obra e medidas. De forma sucinta, estes fatores são exemplificados a seguir:

- Máquinas: este fator de variação inclui o desgaste de ferramentas, o ajuste e as variações das máquinas, as flutuações elétricas, hidráulicas e pneumáticas, etc;
- Métodos: as alterações dos parâmetros dos processos ou na tecnologia utilizada podem provocar variações nos produtos produzidos;
- Materiais: uma vez que variações ocorrem em produtos acabados, elas também ocorrem em matérias-primas. Variações em características tais como resistência à tração, ductibilidade, limite de escoamento, porosidade, composição química, contribuindo para as variações do produto final;
- Meio ambiente: temperatura, umidade, luminosidade e radiação podem contribuir para variações no processo e, conseqüentemente, no produto final;
- Mão-de-obra: o treinamento do operador, a forma como o operador executa uma operação, suas condições físicas e emocionais, podem contribuir para a variação de sua *performance* e, conseqüentemente, do produto final;
- Medidas: as falhas nos equipamentos de inspeção, a utilização inadequada desses equipamentos ou a aplicação incorreta de padrões de qualidade, podem contribuir para variações no produto final.

Quando esses fatores de variação estão presentes nos processos, de forma normal ou esperada, dizemos que um padrão de causas comuns ou causas aleatórias está se desenvolvendo. Causas comuns de variação são inevitáveis e são difíceis de serem identificadas, pois no processo são de pequena importância. Logo quando apenas causas comuns estão presentes no processo, se diz que o mesmo encontra-se sob controle. Entretanto, as causas de variação de grande significância, e facilmente identificáveis, são as causas especiais de variação.

Para identificar se as variações observadas no processo são decorrentes de causas comuns de variação, ou de causas especiais, usa-se os gráficos de controle que são uma ferramenta bastante útil, auxiliando na sua identificação.

De acordo com Costa, Epprecht e Carpinetti (2005), a representação de um gráfico de controle é identificada por três linhas horizontais: a linha central (LC) representa a média dos valores encontrados da característica investigada, e as demais linhas representam simetricamente em relação à central os valores referentes a três desvios-padrão (3σ), sendo estes, o limite superior de controle (LSC) e o limite inferior de controle (LIC). Os pontos (amostrais) assinalados no gráfico representam as médias ou medidas de cada observação (por determinado tempo) retirados ao longo do tempo.

Na Figura 1 tem-se um gráfico de controle que mostra quando o processo está sob controle e quando está fora de controle. As conclusões são extraídas a partir da interpretação dos limites de controle e dos testes de sequência, denominados de *runs tests*.

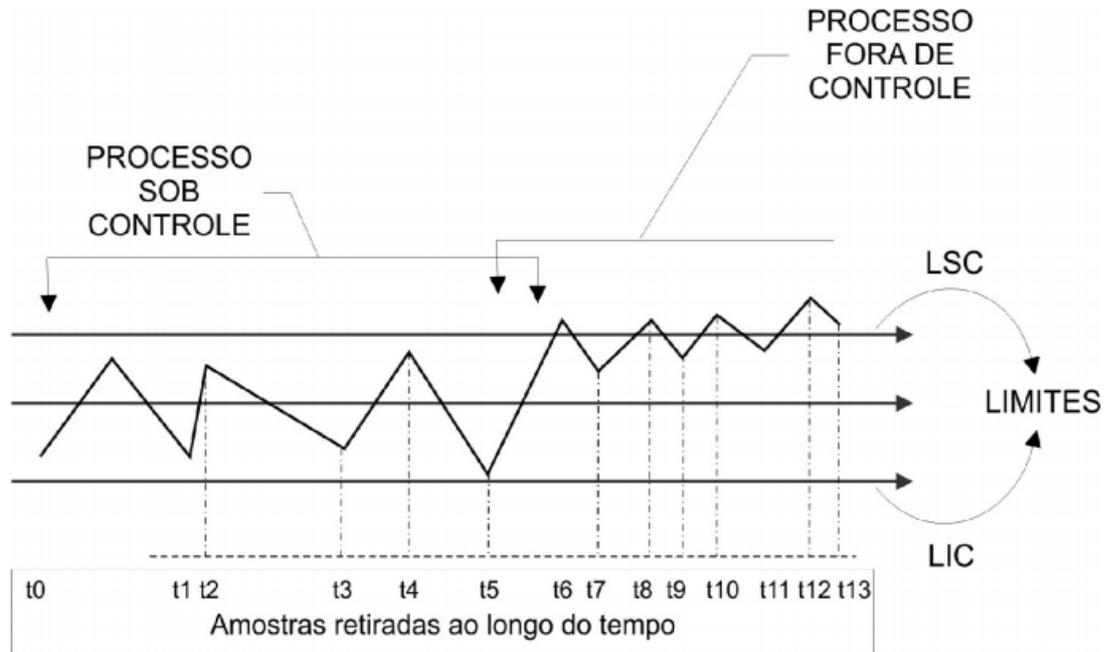


Figura 1- Gráfico de controle adaptado de Montgomery, 2004.

Pode-se dizer, baseado somente na análise visual, que o processo está sob controle quando a maioria dos pontos amostrais no gráfico está próximo da linha central. De modo geral, um gráfico de controle pode indicar uma condição fora de controle quando um ou mais pontos amostrais se localizam além dos limites de controle, ou quando exibem algum padrão de comportamento não-aleatório.

Podem-se cometer dois tipos de erro ao analisar um processo:

- Erro tipo I: concluir que o processo está fora de controle, quando realmente ele está sob controle.
- Erro tipo II: concluir que o processo está sob controle, quando está fora de controle.

Segundo Montgomery (2004), um processo pode ser considerado fora de controle mesmo que os pontos amostrais estejam dentro dos limites de controle. Com a finalidade de auxiliar na decisão do processo aplicam-se os *runs tests* ou também conhecidos como teste de sequência, que identificam os pontos amostrais fora de controle. Nos gráfico de controle de Shewhart são denominados de regras sensibilizantes, que respondem mais rapidamente a uma causa especial.

A Figura 2 demonstra a representação gráfica, com limites de um sigma (1σ), dois sigma (2σ) e três sigma (3σ), chamados de limites de ação, para a interpretação dos testes de sequências. Quando um ponto situa-se fora desses limites, procura-se uma causa especial de modo a tomar uma ação corretiva, caso seja necessário.

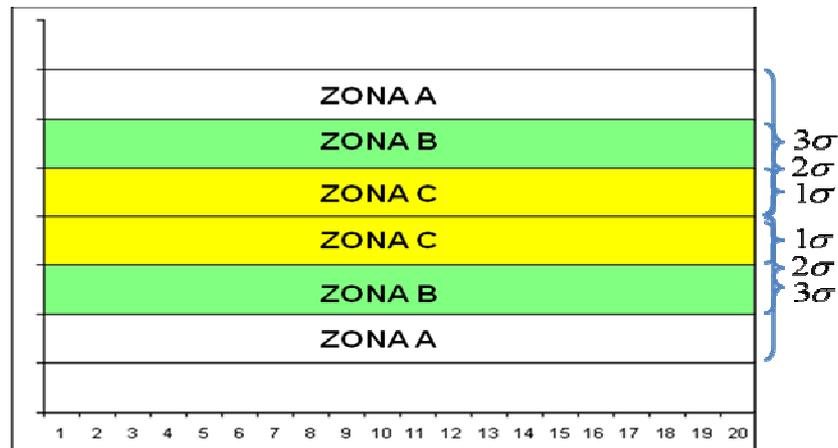


Figura 2 – Representação gráfica para o teste de sequência.

Pode-se perceber na Figura 2, que os limites dividem o gráfico de controle em três zonas, A, B e C, para cada lado da linha central. Com isso aplica-se um conjunto de regras de decisão que possuem a finalidade de detectar padrões não-aleatórios, os quais podem indicar uma condição fora de controle.

O teste de sequência ou *runs tests* levam em conta os seguintes critérios:

- Um ou mais pontos fora dos limites de controle;
- Dois ou três pontos consecutivos fora dos limites de alerta dois-sigma (2σ);
- Quatro ou cinco pontos consecutivos além dos limites um-sigma (1σ);
- Uma sequência de oito pontos consecutivos de um mesmo lado da linha central;
- Seis pontos em uma sequência sempre crescente ou decrescente;
- Quinze pontos em sequência na zona C (acima ou abaixo);
- Quatorze pontos em sequência alternados a acima ou abaixo;
- Oito pontos em sequência de ambos os lados da linha central com nenhum na zona C;
- Um padrão não aleatório dos dados;
- Um ou mais pontos perto dos limites de controle.

Conclui-se que o processo está fora de controle se um ou mais dos critérios apresentados forem verificados.

Segundo Costa, Epprecht e Carpinetti (2004) para utilizar gráficos de controle de Shewhart, é necessário que as observações da característica de qualidade de interesse sejam independentes e normalmente distribuídas.

Existem duas categorias de classificação para os gráficos de controle: gráficos de controle para variáveis e gráficos de controle para atributos, cabe salientar que os gráficos de controle para variáveis são mais sensíveis do que os gráficos para atributos, eles podem alertar para problemas antes que uma real rejeição ocorra.

Segundo Siqueira (1997) o fluxograma encontrado na Figura 3, auxilia na escolha do gráfico de controle.

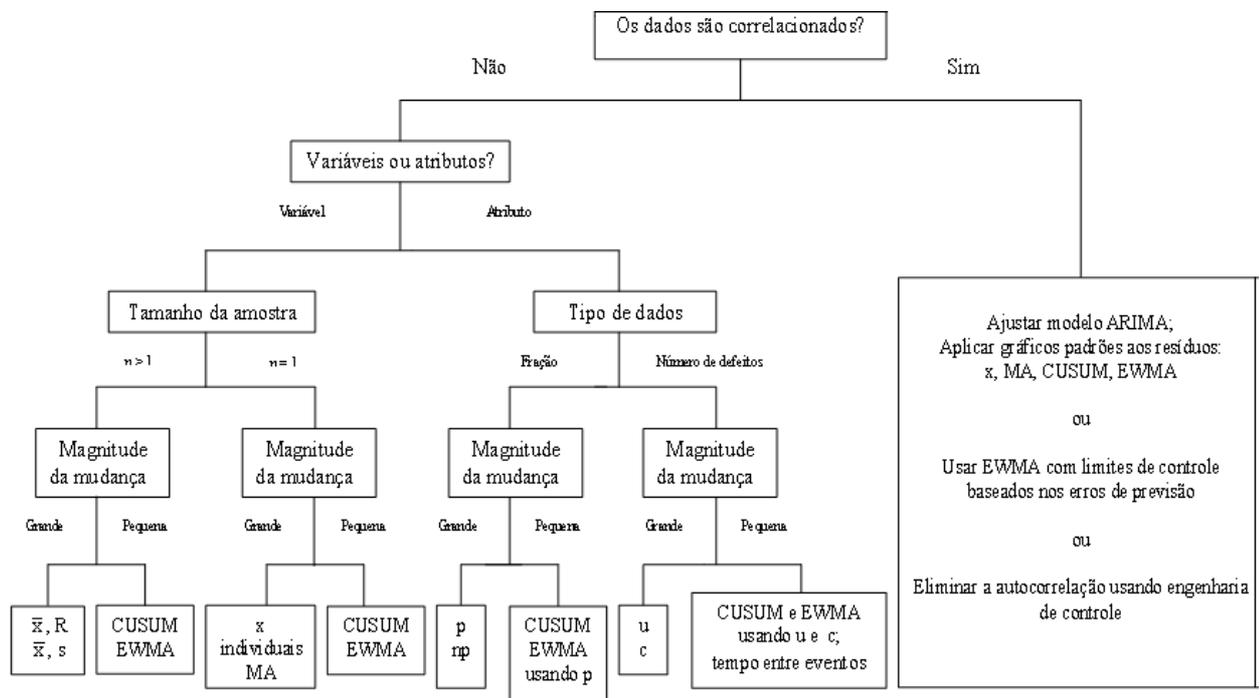


Figura 3 – Algumas sugestões para escolha do gráfico de controle.

Descreve-se a seguir, os gráficos de controle de Shewhart para medidas individuais e amplitude móvel, os quais foram empregados neste trabalho.

2.1.1 Gráficos de controle para medidas individuais e amplitude móvel

Neste caso, o tamanho da amostra para monitoramento do processo é $n=1$. Existem muitas situações em que os gráficos de controle para medidas individuais são úteis, entre elas destacam-se: inspeção e medição automática em toda a unidade produzida; taxa de produção muito lenta, sendo inconveniente acumular amostras $n>1$; várias medidas são tomadas em

uma mesma unidade do produto e nos resíduos de modelos matemáticos ajustados aos dados que possuíam autocorrelação.

Como, os gráficos de controle serão aplicados aos resíduos de modelos matemáticos estimados, as suposições necessárias para aplicá-los, segundo Costa, Epprecht e Carpinetti (2004) normalidade e independência, são plenamente satisfeitos.

Os gráficos de controle para medidas individuais e amplitude móvel, possuem habilidade de detectar grandes deslocamentos na média do processo.

A linha média para o gráfico de medidas individuais é a média geral μ_0 do processo, e os limites de controle distam $3\sigma_0$ da linha média, tal que μ_0 é dado por;

$$\mu_0 = \frac{x_1 + x_2 + \dots + x_m}{m} \quad (1)$$

Se x_i é o valor da i -ésima medida da característica x , então a amplitude móvel associada a ela será dada por:

$$MR_i = \text{máx} \{x_i, x_{i-1}\} - \text{mín} \{x_i, x_{i-1}\} \quad (2)$$

Onde MR_i é a amplitude móvel para duas observações consecutivas.

$$\overline{MR} = \frac{MR_1 + MR_2 + \dots + MR_m}{m} \quad (3)$$

Onde \overline{MR} é a média das amplitudes móveis.

Logo, temos os seguintes limites de controle para o gráfico de medidas individuais:

$$\begin{aligned} LSC &= \mu_0 + 3 \frac{\overline{MR}}{d_2} \\ LC &= \mu_0 \\ LIC &= \mu_0 - 3 \frac{\overline{MR}}{d_2} \end{aligned} \quad (4)$$

Para o gráfico de amplitude móvel (\overline{MR}), os limites de controle são dados pelas expressões:

$$\begin{aligned} LSC &= D_4 \overline{MR} \\ LC &= \overline{MR} \\ LIC &= D_3 \overline{MR} \end{aligned} \quad (5)$$

As constantes d_2 , D_4 e D_3 , utilizadas para calcular os limites de controle para o gráfico de medidas individuais e amplitude móvel, encontram-se tabuladas em Montgomery (2004).

Até há pouco tempo, os gráficos de controle eram utilizados exclusivamente para o monitoramento de processos industriais; porém hoje não estão mais limitados ao “chão de fábrica”. Segundo Costa, Epprecht e Carpinetti (2005) a eficácia de um gráfico de controle é medida pela rapidez com que esse dispositivo detecta alterações no processo, seja ele para monitorar processos que produzem regularmente certa porcentagem de itens defeituosos, pesquisa de satisfação e como é o caso neste trabalho, detectar períodos com variação expressiva no volume de vendas de óleo diesel e gasolina.

2.2 Análise de Séries Temporais

Segundo Morettin (2006), uma série temporal é uma sequência de dados obtidos em intervalos regulares de tempo durante um período específico. O qual pode ser obtido por meio de observações periódicas do evento de interesse.

Conforme Morettin e Tolo (2004), uma das suposições mais frequentes que se faz a respeito de uma série temporal é que ela seja estacionária, ou seja, desenvolve-se no tempo, aleatoriamente ao redor de uma média constante, refletindo alguma forma de equilíbrio estável. Todavia, a maior parte das séries que encontra-se, na prática apresenta alguma forma de não-estacionariedade, geralmente justificada devido à presença de quatro componentes:

- Tendência: sentido de deslocamento da série ao longo do tempo;
- Sazonalidade: movimento ondulatório de curta duração, em geral, inferior a um ano; geralmente associada a mudanças climáticas.
- Ciclo: movimento ondulatório que ao longo de vários anos tende a ser periódico;
- Ruído aleatório: compreende a variabilidade intrínseca aos dados e não pode ser modelado.

A análise de séries temporais, segundo o enfoque de Box e Jenkins (1970) tem como objetivo principal a realização de previsão. Essa metodologia permite que valores futuros de uma série sejam previstos tomando por base apenas seus valores presentes e passados. Isso, geralmente é feito, explorando a correlação temporal que existe entre os valores exibidos pela série.

A relação temporal considerada segundo Ehlers (2007) é representada formalmente por um conjunto de processos estocásticos genericamente denominados modelos ARIMA, que por envolverem apenas uma série de tempo, eles são classificados como modelos univariados.

Os modelos ARIMA resultam da combinação de três componentes também denominados “filtros”: o componente auto-regressivo (AR), o filtro de integração (I) e o componente de médias móveis (MA).

2.2.1 Modelo auto-regressivo: AR(p)

De acordo com esse modelo, Y_t é descrito apenas por seus valores passados e pelo ruído branco ε_t . Um modelo AR, mais simples, é aquele em que Y_t depende somente de Y_{t-1} e de ε_t . Diz-se, nesse caso, que o modelo é auto-regressivo de ordem 1, representado por AR(1). A representação algébrica desse modelo é a seguinte; supondo que Y_t seja a série original menos sua média, logo, a média de Y_t é igual a zero.

$$Y_t = \phi Y_{t-1} + \varepsilon_t \quad (6)$$

Onde ϕ é um parâmetro e $E(\varepsilon_t) = 0$; $E(\varepsilon_t^2) = \sigma^2$; $E(\varepsilon_t \varepsilon_s) = 0$ para $t \neq s$. Por se tratar de um modelo fracamente estacionário, a variância de $Y_t(\gamma_0)$ deve ser constante e as autocovariâncias (γ_k) devem ser independentes de t.

O modelo auto-regressivo genérico, representado por AR(p) pressupõe que seja o resultado da soma ponderada de seus p valores passados, além do ruído branco ε_t :

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \varepsilon_t \quad (7)$$

Utilizando o operador de defasagem, $B y_t = Y_{t-1}$; $B^2 Y_t = Y_{t-2}$; $B^p Y_t = Y_{t-p}$, o modelo pode ser reescrito como:

$$\begin{aligned} Y_t &= \phi_1 B y_t + \phi_2 B^2 y_t + \dots + \phi_p B^p y_t + \varepsilon_t \\ (1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p) Y_t &= \varepsilon_t \\ \phi(B) Y_t &= \varepsilon_t \end{aligned} \quad (8)$$

As variâncias e as autocovariâncias são, agora, dadas respectivamente por:

$$\begin{aligned} \gamma_0 &= \phi_1 \gamma_1 + \phi_2 \gamma_2 + \dots + \phi_p \gamma_p + \sigma_\varepsilon^2 \\ &\dots \end{aligned} \quad (9)$$

$$\gamma_k = \phi_1 \gamma_{k-1} + \phi_2 \gamma_{k-2} + \dots + \phi_p \gamma_{k-p}$$

As condições de estacionariedade do AR(p) estabelecem que todas as p raízes da equação característica $\phi(B) = 0$ caem fora do círculo unitário.

2.2.2 Modelo de médias móveis: MA(q)

O modelo de médias móveis de primeira ordem MA (1) é dado por:

$$Y_t = \varepsilon_t - \theta \varepsilon_{t-1} \quad (10)$$

Onde θ é o parâmetro de médias móveis.

O modelo de médias móveis genérico envolve q valores defasados de ε e é indicado por MA(q) cuja equação é dada por:

$$Y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q} \quad (11)$$

A série Y_t resulta, portanto, da soma ponderada do valor corrente e dos q valores passados de ε . Como não se impõe a restrição de que a soma dos “pesos” dos q + 1 ruídos brancos seja igual à unidade, a expressão “média móvel” não traduz com precisão a natureza do processo gerador de Y_t .

Fazendo uso do operador de defasagem, o modelo MA(q) passa a ser representado por:

$$\begin{aligned} Y_t &= (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q) \varepsilon_t \\ Y_t &= \theta(B) \varepsilon_t \end{aligned} \quad (12)$$

Os modelos AR e MA são facilmente identificáveis por meio das funções de autocorrelações e autocorrelações parciais, que apresentam características definidas conforme mostram Morettin e Tolo (2004), bem como as características de estacionariedade e invertibilidade dos modelos.

2.2.3 Modelo auto-regressivo de médias móveis: ARMA (p,q)

Conforme indica o próprio nome, esse modelo é uma combinação dos dois anteriores, tal que, Y_t é descrito por seus valores passados e pelos choques aleatórios corrente e passados.

Um caso particular é o modelo ARMA (1,1), dado pela equação (13) sendo a especificação mais simples que um processo dessa natureza pode apresentar:

$$Y_t = \theta Y_{t-1} + \varepsilon_t - \theta \varepsilon_{t-1} \quad (13)$$

A especificação genérica de um modelo ARMA admite um componente auto-regressivo de ordem p e um componente de médias móveis de ordem q ; ARMA (p,q) . A equação é dada por:

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q} \quad (14)$$

Usando o operador de defasagem B , tem-se:

$$\begin{aligned} (1 - \phi_1 B - \dots - \phi_p B^p) Y_t &= (1 - \theta_1 B - \dots - \theta_q B^q) \varepsilon_t \\ \phi(B) Y_t &= \theta(B) \varepsilon_t \end{aligned} \quad (15)$$

As condições de estacionariedade e de invertibilidade de um modelo ARMA (p,q) requerem, respectivamente, que todas as p raízes de $\phi(B) = 0$ e todas as q raízes de $\theta(B) = 0$ caiam fora do círculo unitário.

2.2.4 Modelo auto – regressivo integrado de médias móveis: ARIMA (p,d,q)

A metodologia de Box e Jenkins aplica-se a um caso específico de séries temporais não estacionárias; séries que se tornam estacionárias após a aplicação de diferenças. Chama-se ordem de integração o número de diferenças necessárias para tornar uma série estacionária, o qual é o procedimento correto a ser utilizado. Pois na análise de séries temporais, procura-se sempre por um conjunto de observações que sejam estáveis ao longo do tempo, garantindo as propriedades de um processo estacionário.

O modelo aplicado a séries não estacionárias homogêneas pode ser genericamente formulado da seguinte maneira: se Y_t tornar-se estacionária após a aplicação de d diferenças e a série resultante for representada por um modelo ARMA (p,q) , diz-se que Y_t é descrito por um modelo ARIMA (p,d,q) representado por:

$$W_t = \phi_1 W_{t-1} + \phi_2 W_{t-2} + \dots + \phi_p W_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q} \quad (16)$$

Onde $W_t = \Delta^d Y_t$

Utilizando o operador de defasagem tem-se a seguinte expressão:

$$W_t = (1 - B)^d Y_t \quad (17)$$

Segundo Morettin e Tolo (2004), a construção e aplicação de modelos ARIMA seguem um ciclo iterativo da metodologia Box e Jenkins. O primeiro passo consiste em descobrir qual entre as várias versões do modelo ARIMA descreve o comportamento da série, essa etapa é denominada identificação. O segundo passo consiste em estimar os parâmetros do modelo identificado, essa etapa é chamada estimação. Segue-se a ela a etapa de verificação, que consiste em avaliar se o processo de estimação foi bem sucedido. Se a conclusão for negativa,

repete-se o procedimento desde a etapa de identificação com o objetivo de encontrar uma especificação de modelo adequado. Quando se obtém um modelo satisfatório, passa-se para a última etapa da metodologia, qual seja, a previsão, que se constitui no objetivo primordial do enfoque de Box e Jenkins.

2.2.5 Modelo sazonal auto-regressivo integrado de médias móveis: SARIMA(p,d,q)(P,D,Q)_s

Os modelos ARIMA vistos anteriormente exploram a correlação de valores de Y_t observados em instantes de tempo consecutivos.

Quando a periodicidade da série é inferior a um ano (séries mensais e trimestrais, por exemplo), outro tipo de correlação serial passa a ter importância: a correlação entre os instantes de tempo distantes entre si por s ou múltiplos de s , onde s é o número de observações contidas em um ano ($s=12$ para dados mensais e $s=4$ para dados trimestrais).

Para acomodar esse tipo de série, surgem os modelos ARIMA sazonais, também conhecidos por SARIMA, os quais são, na verdade, extensões dos modelos ARIMA.

Esse modelo aplica-se à série não estacionária, mas que, após a aplicação de D diferenças “sazonais”, são descritas por um processo sazonal estacionário ARMA (P,Q)_s:

$$\Phi(B^s)(1 - B^s)^D Y_t = \Theta(B^s)\varepsilon_t \quad (18)$$

D é denominado ordem de integração sazonal.

2.2.6 Modelos auto-regressivos fracionários integrados de médias móveis: ARFIMA(p,d,q)

Os modelos ARFIMA (p,d,q) são responsáveis por capturar e modelar processos com longa dependência serial nos dados, onde esta longa dependência serial costuma-se chamar de memória longa. É um processo estacionário em que a função de autocorrelação decresce hiperbolicamente para zero, isto é, $\rho_j \sim C_j^{-\alpha}, j \rightarrow \infty$ onde $C > 0$ e $0 < \alpha < 1$.

Estes modelos são capazes de descrever as dinâmicas de memória curta e longa de processos fracionários, onde d deve explicar a estrutura de correlação de ordens altas, enquanto que ϕ e θ explicam a estrutura de correlação de ordens baixas.

Outras características dos modelos de memória longa segundo Morettin (2008), são que estas séries apresentam persistência nas autocorrelações amostrais, isto é, dependência significativa entre as observações por um longo intervalo de tempo, e que esse tipo de série é

que sua função densidade espectral é não limitada na frequência zero, o que equivale a dizer que sua função de auto-correlação não é absolutamente somável.

Os modelos ARIMA (p,d,q), introduzidos por Box e Jenkins (1970), incluem o parâmetro d, um inteiro que estabelece o nível de diferenciações necessárias para tornar uma série temporal estacionária de 2ª ordem. Estes modelos são adequados para a modelagem do comportamento de séries temporais em curto prazo. A partir dos anos 80, Granger & Joyeux e ainda Hosking (1981) propõem uma generalização desta modelagem em relação ao parâmetro d, podendo este assumir não só valores inteiros, mas também representar graus de diferenciação fracionários.

Modelos com esta propriedade permitem estudar séries caracterizadas por longas dependências temporais. Estes modelos intitulam-se ARFIMA (p,d,q), onde F significa "fractional".

O processo Y_t é um ARFIMA (p,d,q) se este é a solução da equação de diferenças:

$$\phi(B)(1-B)^d Y_t = \theta(B)a_t \quad (19)$$

Onde: $\phi(B)$ e $\theta(B)$ representam os parâmetros

$$\phi(Z) = 1 - \phi_1 Z - \dots - \phi_p Z^p \text{ e } \theta(Z) = 1 - \theta_1 Z - \dots - \theta_q Z^q \text{ no operador retardo } B : B_j Y_t = Y_{t-j}$$

O termo $(1-B)^d$ é definido pela expansão binomial $(1-B)^d = \sum_{k=0}^{\infty} \binom{d}{k} (-1)^k B^k$.

{at} é um processo ruído branco onde $E(at) = 0$, σ_a^2 .

Segundo Morettin (2008), a razão da escolha dessa família de processos, para fins de modelagem das séries com o comportamento de memória longa, é que o efeito do parâmetro d em observações distantes decai hiperbolicamente conforme a distância aumenta, enquanto os parâmetros ϕ e θ decaem exponencialmente. Então, d deve ser escolhido com o objetivo de explicar a estrutura de correlação de ordens da série, enquanto os parâmetros ϕ e θ explicam a estrutura de correlação de ordens baixas.

Se os polinômios $\phi(Z)$ e $\theta(Z)$ têm suas raízes fora do círculo unitário, e não possuem raízes comuns, o processo $(1-B)^d Y_t$ é estacionário (de 2ª ordem) e invertível.

Quando $d = 0$, Y_t é um modelo auto-regressivo de médias móveis, ARMA (p,q).

Quando $d \neq 1$ e é não-inteiro, a função de autocorrelação $\rho(k)$ tem um decaimento hiperbólico, $\rho(k) \sim e|k|^{2d-1}$, $|k| \rightarrow \infty$. As autocorrelações originadas de um modelo ARMA (p,q)

têm um decaimento exponencial $\rho_k \sim a^k$, $0 < a < 1$ (Box & Jenkins, 1976). Tem-se então, no caso do ARFIMA, um processo de "longa dependência", "longa persistência" ou "long memory", se $0 < d < 0,5$. No caso de $-0,5 < d < 0$, o processo é de dependência intermediária ou "intermediate memory". Neste caso, a função de autocorrelação exibirá dependências negativas entre observações mais distantes. No domínio da frequência, a característica fracionária de "d" é detectada pelo comportamento da função espectral que tende ao infinito, quando a frequência se aproxima de zero.

Em seguida, seguem-se os testes de diagnósticos com o objetivo de verificar diversos pressupostos que devem ser satisfeitos.

2.3 Testes de diagnósticos

No processo de modelagem de séries temporais e na aplicação do controle estatístico do processo diversos pressupostos devem ser considerados, os quais foram mencionados anteriormente e que devem ser satisfeitos. Nesta seção serão apresentados os testes utilizados no presente trabalho, para a verificação de estacionariedade das séries, da autocorrelação, da normalidade e o critério penalizador.

Na seção 2.3.1, apresentam-se os testes de raiz unitária, Dickey e Fuller e KPSS, para analisar a estacionariedade; na seção 2.3.2, os testes de normalidade, Lilliefors e gráfico Q-Q Plot; na seção 2.3.3, encontra-se a estatística de Ljung–Box para detectar autocorrelação e, na seção 2.3.4, o critério para escolha do melhor modelo.

2.3.1 Testes de raiz unitária (Teste de Dickey e Fuller)

O teste de raiz unitária Dickey e Fuller, segundo Morettin (2006) estima a seguinte auto-regressão:

$$\nabla Y_t = \gamma Y_{t-1} + \varepsilon_t \quad (20)$$

Onde $\nabla = (Y_t - Y_{t-1})$, ou seja, é o operador diferença e $\gamma = \rho - 1$. Nesse caso, a hipótese nula, H_0 é de que existe pelo menos uma raiz unitária, logo a série não é estacionária e $\gamma = 0$. Por sua vez, a hipótese alternativa H_1 é que a série seja estacionária, nesse caso não há nenhuma raiz unitária e conseqüentemente $\gamma < 0$.

Alternativamente, o teste de DF pode ser conduzido utilizando-se a seguinte expressão:

$$Y_t = \rho Y_{t-1} + \varepsilon_t \quad (21)$$

Nesse caso, testa-se a hipótese nula de que $\rho = 1$, isto é, tenha uma raiz unitária contra a hipótese alternativa de que não tenha raiz unitária, ou seja, $\rho < 1$.

A equação (21) pode ser ampliada para incorporar a presença ou não de intercepto e/ou tendência, ou seja, o procedimento Dickey e Fuller permite que se teste a existência ou não de raiz unitária naqueles casos em que seja necessária a introdução ou não de constante e/ou tendência. O teste para ambos os casos é conduzido de maneira semelhante ao apresentado acima. A diferença é que no caso em que seja necessária à introdução de uma constante, a auto-regressão a ser utilizada é a seguinte:

$$\nabla Y_t = \alpha + \gamma X_{t-1} + \varepsilon_t \quad (22)$$

Onde α é o intercepto. Caso seja necessária a presença da tendência e do intercepto, a equação a ser utilizada é a seguinte:

$$\nabla Y_t = \alpha + \beta t + \gamma X_{t-1} + \varepsilon_t \quad (23)$$

Sendo que, o termo t representa a tendência linear.

O teste Dickey e Fuller parte da pressuposição de que os termos de erros nas equações acima são identicamente e independentemente distribuídos (i.i.d.), isto é não apresentam autocorrelação. Em função disso, o teste Dickey e Fuller foi ampliado de forma a incorporar defasagens em relação a variável que está sendo analisada. Ao proceder dessa maneira obtém-se o teste Dickey e Fuller Aumentado (ADF), o qual adiciona a equação (23) a própria variável defasada e diferenciada, assumindo a seguinte forma:

$$\nabla Y_t = \alpha + \beta t + \gamma X_{t-1} \sum \rho_{j+1} + \nabla Y_{t-j} + \varepsilon_t \quad (24)$$

A principal vantagem do teste ADF em relação ao DF é que ao introduzir um número suficiente de defasagens, garante-se que os resíduos não apresentem autocorrelação. Em linhas gerais, para se determinar o número ideal de defasagens utiliza-se algum critério de informação, como por exemplo, o Critério de Akaike (AIC) discutido no tópico 2.3.4.

Teste Kwiatkowski, Phillips, Schmidt e Shin (KPSS)

O teste KPSS desenvolvido pelos autores Kwiatkowski, Phillips, Schmidt e Shin em 1992 tem como objetivo distinguir a raiz unitária de séries cujos dados não são suficientemente conclusivos. A idéia do teste é testar a variância de passeio aleatório, caso essa variância for nula, então o processo é estacionário (BUENO, 2008).

As hipóteses testadas são:

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

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

Para a realização do teste KPSS é aplicada a estatística:

$$KPSS = \sum_{t=1}^T \frac{S_t^2}{T^2 \hat{\nu}^2} \quad (25)$$

Onde $\hat{\nu}^2$ representa a variância de longo prazo.

Regra de decisão: se a estatística calculada for menor que os valores críticos, não se rejeita a hipótese nula, H_0 , conclui-se que a série é estacionária.

Segundo Bem (1998) se a série analisada for não estacionária, adota-se na forma de primeira ou segunda diferença para torná-la estacionária.

Os testes empregados para identificar a estacionariedade são aplicados nos dados originais e caso a série analisada for não estacionária aplica-se diferenças e os testes novamente devem ser utilizados, com a finalidade de testar se a estacionariedade na série diferenciada.

2.3.2 Testes de normalidade

Vários testes estatísticos têm sido usados para avaliar a normalidade dos resíduos do modelo estimado, no presente trabalho, destacam-se: Teste de Lilliefors e gráfico Q-Q plot.

Teste de Lilliefors.

O teste foi introduzido por Lilliefors em 1967, é utilizado nos casos em que a média e a variância não são previamente especificadas, mas sim, estimadas dos dados amostrais.

Sejam X_1, X_2, \dots, X_n uma amostra aleatória de uma população X com distribuição desconhecida. Pretende-se testar se X tem distribuição $N(\mu, \sigma^2)$ sem especificar μ e σ , isto é, para algum μ e algum σ . As hipóteses a testar são:

Este teste padroniza os dados originais, usando estimativas de μ e σ :

$$Z_i = \frac{X_i - \bar{X}}{S} \quad (i = 1; \dots; n) \quad (26)$$

Logo, as hipóteses a serem testadas são as seguintes:

H_0 : X segue uma distribuição normal com média 0 e variância 1.

H_1 : X não segue uma distribuição normal com média 0 e variância 1.

Sendo F_0 a função de distribuição $N(0,1)$, a estatística do teste de Lilliefors é dada por:

$$D_n^* = \sup_{-\infty < z < +\infty} |F_n(z) - F_0(z)| \quad (27)$$

Onde F_n é a função de distribuição da amostra depois de padronizar, i.e., definida para as variáveis aleatórias Z_1, Z_2, \dots, Z_n . Mais uma vez, o valor observado da estatística teste D_n^* , d_n^* , é obtido substituindo na estatística a função distribuição da amostra padronizada pela função distribuição empírica. A hipótese H_0 é rejeitada, para um nível de significância α , se o valor observado d_n^* for superior ou igual ao ponto crítico $D_{n,\alpha}^*$, com $D_{n,\alpha}$ tal que;

$$P(D_n^* \geq D_{n,\alpha}^* / H_0 \text{ é verdadeira}) = \alpha$$

Os pontos críticos são consultados na tabela elaborada por Lilliefors (anexo)

Normal Q-Q plot

O pressuposto de normalidade pode ser testado usando um gráfico de probabilidade normal (*Normal Probability Plot*). Que representa o quartil de probabilidade esperado se a distribuição fosse normal em função dos resíduos (Normal Q-Q Plot). Para produzir estes gráficos, padroniza-se os resíduos de forma a terem um desvio padrão unitário:

$$d'_i = \frac{d_i - 0}{S} \quad (28)$$

Onde:

$$S^2 = \frac{SSE}{n - k - 1} = \frac{\sum_{i=1}^n d_i^2}{n - k - 1}$$

E ordenam-se por ordem crescente. No gráfico Q-Q plot os quartis de probabilidade esperados, ou seja, os z_i tais que $P(Z < z_i) = \frac{i - 0,5}{n}$; estes valores são representados no eixo das ordenadas e no eixo das abscissas representam-se os

resíduos padronizados. Se no gráfico, os erros possuírem distribuição normal, todos os pontos dos gráficos devem estar posicionados mais ou menos sobre uma reta (diagonal).

2.3.3 Teste para autocorrelação

Existe autocorrelação residual quando os erros são correlacionados com os valores anteriores ou posteriores na série. Este problema também é chamado de correlação serial, e ocorre, principalmente, em aplicações envolvendo séries temporais. A presença de autocorrelação significa que as estimativas dos parâmetros obtidos pelo método mínimos quadrados ordinários não são eficientes, isto é, não apresentam variância mínima.

Segundo Ehlers (2007), pode-se testar se um grupo de autocorrelações é significativamente diferente de zero utilizando-se as chamadas estatísticas Q, ao invés de verificar as autocorrelações de forma individual.

Ljung e Box (1978) propuseram o seguinte teste na qual a estatística é denotada pela expressão:

$$Q(K) = n(n+2) \sum_{k=1}^h \frac{\hat{r}_k^2}{(n-k)} \quad (29)$$

Em que n representa o número de observações; \hat{r}_k^2 é a autocorrelação no lag k e h representa o número de lags que estão sendo testados, seguindo uma distribuição Qui-quadrado (χ^2) com p graus de liberdade.

As hipóteses a serem testadas são:

H_0 : Os dados são ruído branco;

H_1 : Os dados não são ruído branco.

Caso a estatística calculada exceda o valor crítico da tabela de χ^2 para determinado nível de significância, pode-se rejeitar a hipótese nula de que todos os \hat{r}_k são iguais a zero. Isso implica em aceitar a hipótese alternativa de que pelo menos uma autocorrelação (\hat{r}_k) é estatisticamente diferente de zero.

2.3.4 Critério penalizador AIC

Com o objetivo de escolher o melhor modelo matemático, levando em conta a parcimônia do modelo, aplicou-se o critério penalizador *AIC* (*Akaike Info Criterion*). Como vários modelos podem ser julgados adequados em termos do comportamento dos resíduos, uma forma de distinguir qual é o modelo mais adequado entre os modelos candidatos é utilizar alguns critérios penalizadores. O critério de informação *AIC* é calculado com base na variância estimada e também leva em consideração o número de parâmetros estimados.

Pode-se definir algebricamente esse critério como:

$$AIC = -2(\ln l/T) + 2(K/T) \quad (30)$$

Onde T é o tamanho da amostra; l é o valor da função log-verossimilhança e K é o número de parâmetros estimados.

Como regra de decisão, escolhe-se o modelo que minimiza o critério penalizador, isto é, o melhor modelo encontrado é aquele que apresentar o menor valor de *AIC*, para que se possam realizar previsões em curto prazo e, além disso, aplicar gráficos de controle no modelo mais adequado.

Mais informações sobre o assunto pesquisar em Morettin (2006).

Portanto, esse segundo capítulo expõe a revisão de literatura necessária para dar suporte ao desenvolvimento desta pesquisa.

3 METODOLOGIA

Neste capítulo são apresentadas as informações sobre os métodos utilizados na pesquisa, englobando: o método da pesquisa, a coleta de dados, as etapas metodológicas e os programas computacionais utilizados para desempenhar, da melhor forma, o seu desenvolvimento.

O presente trabalho é uma pesquisa quantitativa, pois, segundo Malhotra (2001), procura quantificar os dados e aplica alguma forma de análise estatística.

A base de dados utilizada neste trabalho foi obtida junto ao site da ANP - Agência Nacional do Petróleo, Gás Natural e Biocombustíveis (<http://www.anp.gov.br/>). As observações referem-se ao volume de vendas dos combustíveis derivados do petróleo, realizadas pelas distribuidoras de derivados de petróleo, em barris de óleo diesel e gasolina no Estado do Rio Grande do Sul, no período de janeiro de 2000 a fevereiro de 2010. Os registros das vendas foram analisados mensalmente totalizando em 122 observações.

3.1 Etapas metodológicas

As etapas metodológicas para o cumprimento dos objetivos perfazem os seguintes passos:

1) aplicação da estatística descritiva para verificar o comportamento das variáveis em estudo, e selecionar os derivados combustíveis de petróleo de maior volume de venda;

2) encontrar um modelo matemático que capte bem o comportamento dos dados da série analisada, utilizando a metodologia de Box e Jenkins, modelo SARIMA e ARFIMA;

3) analisar os resíduos do modelo encontrado, com o objetivo de verificar se são independentes e normalmente distribuídos e com isso satisfazer as pressuposições da aplicação de gráficos de controle;

4) encontrado o modelo que eliminou a autocorrelação serial, realizar previsões do volume de vendas de óleo diesel e gasolina devido à necessidade de avaliar e planejar a infraestrutura, iniciando desde a produção até a distribuição, visando garantir e atender a demanda por combustíveis;

5) aplicar gráficos de controle nos dados originais e nos resíduos do modelo encontrado a fim de analisar o comportamento do volume de vendas de óleo diesel e gasolina e detectar possíveis pontos com volumes de vendas que não apresentem a mesma regularidade identificada no processo ;

6) os procedimentos utilizados para verificar se a série de tempo em estudo é estacionária, foram à análise do correlograma, os testes baseados em funções de autocorrelação, assim como os testes desenvolvidos por Dickey e Fuller e KPSS;

7) para a escolha do melhor modelo foi verificado um conjunto de observações, assim como, estimados os parâmetros do modelo identificado juntamente com a significância estatística (p-valor), logo foram analisados os critérios de ajuste para a escolha do melhor modelo estimado e finalizando com a verificação do ruído branco.

8) o critério estatístico utilizado para justificar a escolha de um modelo de previsão foi o *Akaike Information Criteria (AIC)*;

9) na verificação dos resíduos foi analisado se os mesmos apresentam média zero, variância constante, distribuição normal e independência. Para isso, foi utilizado o teste de Ljung e Box para verificar a inexistência de autocorrelação entre os resíduos, bem como o teste de Lilliefors para testar a normalidade dos mesmos no modelo encontrado;

10) para identificar se o processo estava fora de controle utilizou-se o teste de sequência com a finalidade de detectar padrões não aleatórios, que indicam condições fora de controle;

Os resultados das análises estatísticas foram obtidos utilizando-se os softwares Statistica versão 7.0; EViews versão 7.0 e planilha eletrônica.

Para que os objetivos deste trabalho pudessem ser alcançados, uma revisão da literatura foi realizada, apresentada no capítulo anterior a fim de mostrar a importância do CEP. Além disso, foi necessário estudar as teorias de gráficos de controle, os modelos SARIMA, ARFIMA de séries temporais.

4 RESULTADOS E DISCUSSÕES

Este capítulo apresenta a análise e discussão dos resultados mediante a aplicação da metodologia proposta para os conjuntos de dados reais referentes ao volume de vendas de barris dos combustíveis derivados do petróleo no período de 2000 a 2010 com observações mensais. Primeiramente foi realizada uma análise descritiva do volume de vendas dos combustíveis derivados do petróleo, para o RS, posteriormente realizou-se a modelagem para tratar a autocorrelação e com isso obter resíduos identicamente independentes que satisfaçam as condições e pressuposições para a aplicação de gráficos de controle e por meio deste, detectar possíveis pontos anômalos na economia.

4.1 Análise descritiva

Com o objetivo de verificar quais dos derivados combustíveis de petróleo apresentaram maior volume de vendas realizou-se uma análise descritiva, de todos os derivados combustíveis do petróleo vendidos no RS e os resultados encontram-se na Tabela 1. Entre todos os combustíveis derivados do petróleo selecionaram-se os que apresentaram maior média, isto é, o óleo diesel e a gasolina. O fato deve-se à grande utilidade dos dois derivados do petróleo e cabe salientar que gasolina é um dos produtos de maior importância do petróleo.

Tabela 1 – Estatística descritiva para o volume de vendas, em barris, de todos os derivados combustíveis de Petróleo no Estado do Rio Grande do Sul.

Variáveis	Média	Desvio - Padrão	C.V (%)	Assimetria	Curtose	Mínimo	Máximo
Álcool hidratado	66478	26548,34	39,94	1,60	1,85	36625	153509
Gasolina C	856730	86796,41	10,13	1,61	5,04	692839	1290918
Gasolina de aviação	2574	1963,38	76,28	1,23	0,82	439	8622
GLP	305467	31628,48	10,35	-0,09	-0,31	238407	390544
Óleo combustível	163742	62085,04	37,92	0,87	0,84	69967	364903
Óleo diesel	1396003	149604,92	10,72	0,39	-0,38	1116061	1787042
Querosene /aviação	60843	9230,62	15,17	0,48	-0,18	43767	84887
Querosene iluminante	3867	2702,09	69,88	0,87	-0,42	907	11246

C.V.: coeficiente de variação de Pearson

Optou-se em trabalhar com os dois derivados de petróleo que apresentaram um grande volume de vendas em barris; isto é uma maior média que são o óleo diesel e gasolina C para conhecer o comportamento dessas variáveis no decorrer do período analisado.

Pela análise descritiva das variáveis em estudo, verificou-se, por meio do coeficiente de variação, que as médias das variáveis são representativas. Isso não significa que todos os valores da amostra estejam dentro dos limites de controle desejados, uma vez que o coeficiente de variação é uma medida de dispersão utilizada quando se deseja comparar a variação amostral. Dentro do conjunto de dados existe uma pequena variabilidade, pois o coeficiente de variação encontrado foi em torno de 10% para as duas variáveis.

Ao analisar o valor da assimetria e da curtose para os dados das vendas de gasolina, observou-se o seguinte: como a assimetria é diferente de zero e a curtose é maior de três, há um indicativo de que a série exibe sinais de volatilidade. Neste caso a curtose excede três e a distribuição é dita leptocúrtica. Já para as vendas de óleo diesel a assimetria é próxima de zero e a curtose é inferior a três e, por esses resultados a variável não apresenta uma distribuição normal. Uma vez que a distribuição é inteiramente descrita por seus parâmetros de média e desvio padrão e sua importância deve-se ao fato de um grande número de fenômenos aleatórios serem aproximados por meio da distribuição normal. Desse caso, por meio da Tabela 1, percebe-se que o volume de vendas de óleo diesel e gasolina não seguem uma distribuição normal, no processo de vendas de barris de óleo diesel e gasolina trabalhar com uma distribuição normal significa possuir igual dispersão dos dados.

Definidas as variáveis que se vai trabalhar, o próximo passo é identificar um modelo matemático que capte bem o comportamento da série no decorrer do período analisado.

4.2 Identificação do modelo e estimação dos parâmetros para o volume de venda de óleo diesel

Para identificar o modelo apropriado que representa a série de vendas (em barris) de óleo diesel, em uma série temporal, o primeiro passo que se deve realizar é a inspeção gráfica, pois a mesma mostra o comportamento da série analisada. Na Figura 4 apresenta-se o gráfico das vendas de barris de óleo diesel no Rio Grande do Sul representado pela linha contínua, em que se observa que a mesma não se apresenta estacionária, isto é, seus valores não variam em torno de uma média constante, o que se pode constatar também pelo teste de Dickey-Fuller, que verifica a existência (ou não) de raiz unitária na série. O teste apresentou o seguinte resultado: para a série original o *p-valor* encontrado foi de 0,5679, o qual excede o nível de significância de 5%, isto é, neste caso aceita-se a hipótese nula, ou seja, a série não é estacionária.

Para a série com uma diferença o p -valor encontrado foi de 0,0001, o qual é inferior ao nível de significância de 5%, isto é rejeita-se a hipótese nula, concluindo-se que a série diferenciada apresenta-se estacionária. Para comprovar a estacionariedade da série aplicou-se o teste KPSS, pois o ADF, quando as raízes são próximas de um, pode apontar resultados errôneos. Quanto ao teste KPSS, constatou-se que a série com uma diferença é estacionária, pois a estatística calculada é de 0,1090, que é menor que o valor crítico de 0,4630, o que indica que não há rejeição da hipótese nula de estacionariedade. Assim, os testes de Dickey-Fuller e KPSS asseguraram que a série com uma diferença é estacionária. Portanto sugere a necessidade de que a mesma seja diferenciada para se estabelecer um modelo ARIMA, de forma que os parâmetros estimados sejam estáveis e representativos da série em estudo.

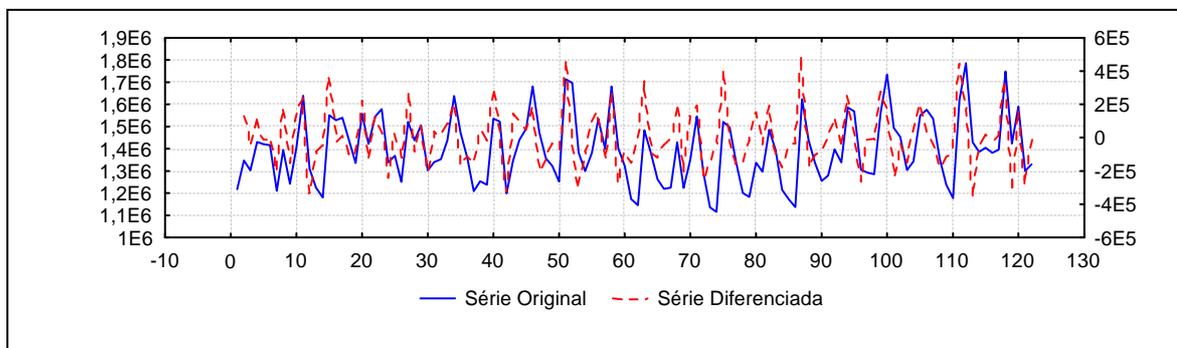


Figura 4: Série original e diferenciada do número de barris de óleo diesel vendidos no Estado do Rio Grande do Sul no período de janeiro de 2000 a fevereiro de 2010.

A próxima etapa consistiu em analisar as funções de autocorrelação (FAC) e autocorrelação parcial (FACP) da série original das vendas de barris de óleo diesel. Por meio do comportamento dessas funções pode-se ter uma idéia do tipo de modelo a ser utilizado. As figuras 5a e 5b, apresentam, respectivamente, a FAC e a FACP da série de vendas de barris de óleo diesel.

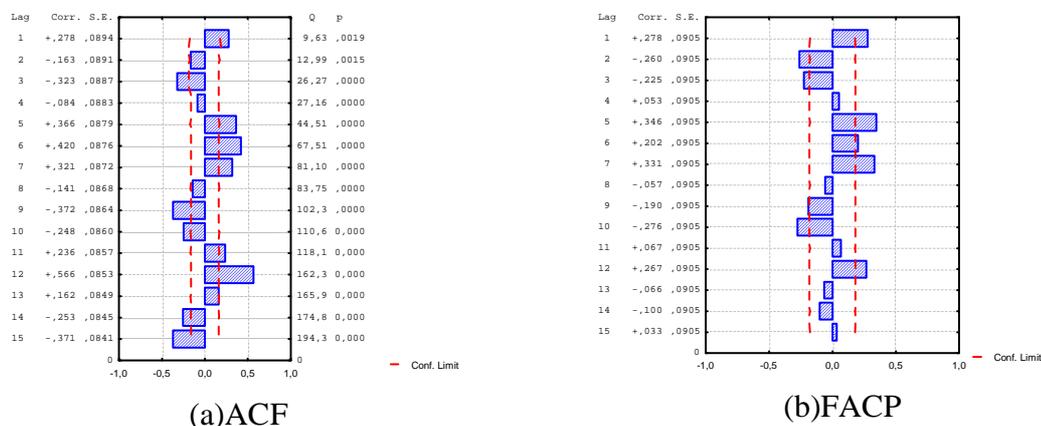


Figura 5 – (a) Função de Autocorrelação e (b) Autocorrelação Parcial da série original das vendas de barris de óleo diesel.

Analisando-se a Figura 5 pode-se observar as estatísticas Ljung-Box, que evidenciam a presença de autocorrelação para os primeiros 15 *lags*, pois as autocorrelações são maiores que o intervalo de confiança ao nível de significância de 5%. Analisando a função de autocorrelação pode-se ter uma idéia do tipo de modelo a ser utilizado, neste caso a ACF apresenta uma alternância o que sugere um modelo sazonal. Após a identificação de prováveis modelos segue-se para a etapa de estimação dos respectivos parâmetros. Posteriormente para sanar o efeito da autocorrelação serial presente na série original, a qual se removida por meio dos modelos de previsão ou modelos matemáticos, serão aplicadas técnicas do controle estatístico do processo.

Analisando-se os resultados apresentados na Tabela 2, é possível observar os modelos significativos encontrados, uma vez que todos apresentam coeficientes com *p-valor* menor que 5% e respeitam as condições de estacionariedade e invertibilidade, já que os valores estimados, em módulo, são menores que a unidade. A importância, da série do volume de vendas de óleo diesel apresentar-se estacionária, deve-se ao fato da média e variância manter-se constante ao longo do período em análise e com isso pode-se aplicar a modelagem proposta.

Tabela 2 – Estimação dos parâmetros para os modelos SARIMA concorrentes da variável venda de barris de óleo diesel no período de janeiro de 2000 a fevereiro de 2010.

Modelos	Parâmetros	AIC
SARIMA (0,1,1)(0,1,1) ₁₂	q(1)=0,814 Qs(1)=0,561	23,236
SARIMA (0,1,1)(1,1,0) ₁₂	q(1)=0,805 Ps(1)= -0,451	23,286
SARIMA (2,1,0)(0,1,1) ₁₂	p(1)= -0,760 p(2)= -0,394 Qs(1)= 0,529	23,342
SARIMA (0,1,1)(2,1,0) ₁₂	q(1)= 0,808 Ps(1)= -0,541 Ps(2)= -0,223	23,265

AIC: critério de informação Akaike

Mediante os modelos concorrentes obtidos, o melhor modelo ajustado para as vendas de barris de óleo diesel foi o SARIMA (0,1,1)(0,1,1)₁₂, o qual foi determinado pelo critério penalizador AIC.

4.3 Análise residual para o modelo SARIMA (0,1,1)(0,1,1)₁₂

Tendo encontrado o modelo matemático que tratou a autocorrelação serial da série é necessário analisar os resíduos do modelo a fim de verificar se os mesmos são independentes e normalmente distribuídos, de forma a cumprir as pressuposições básicas para o uso dos gráficos de controle.

Analisando a FAC e FACP residual do modelo SARIMA (0,1,1)(0,1,1)₁₂, na Figura 6, observou-se que nenhum *lag* apresentou autocorrelação acima dos limites de confiança de 95%. Dessa forma, para confirmar que os resíduos não são autocorrelacionados, utilizou-se o teste de Ljung-Box Q*. A estatística de teste tabelada $Q(15-1)=23,7$ é maior para as 15 primeiras autocorrelações. Assim, a hipótese nula de que essas 15 autocorrelações apresentam valor zero não é rejeitada, e os resíduos podem ser considerados “ruído branco”.

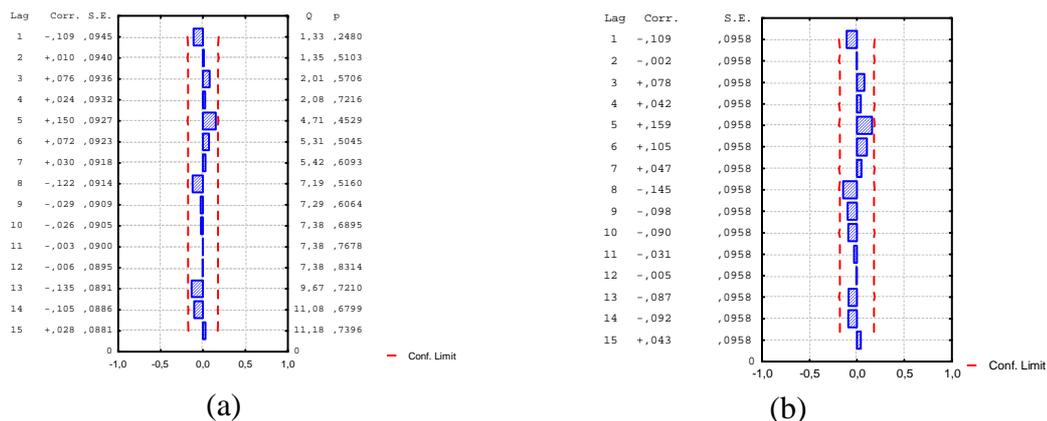


Figura 6 – (a) Função de Autocorrelação e (b) Autocorrelação Parcial para os resíduos do modelo SARIMA (0,1,1)(0,1,1)₁₂ identificado para as vendas de barris de óleo diesel.

Para utilizar gráficos de controle é necessário que as observações da característica de qualidade sejam independentes e normalmente distribuídas. Observou-se que os resíduos do modelo não são autocorrelacionados, ou seja, eles são independentes. Analisando-se o gráfico Q-Q plot exibido na Figura 7, o qual compara os quartis teóricos da distribuição normal com os quartis dos resíduos do modelo, verifica-se que existe uma relação linear entre estes quartis, o que indica que os resíduos do modelo encontrado seguem uma distribuição normal.

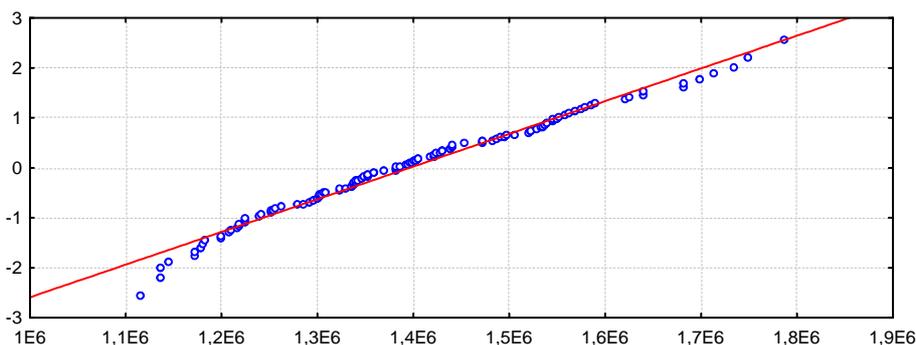


Figura 7 – Gráfico Q-Q plot dos resíduos do modelo SARIMA (0,1,1)(0,1,1)₁₂.

Para testar a normalidade dos resíduos do modelo encontrado, também se utilizou o teste de Lilliefors, pois a média e o desvio padrão da distribuição esperada não são conhecidos. Para o teste encontrou-se um *p-valor* superior a 0,20, então se aceita a hipótese de que a distribuição dos resíduos do modelo SARIMA (0,1,1)(0,1,1)₁₂ é normal.

Portanto, os resíduos do modelo SARIMA (0,1,1)(0,1,1)₁₂ são independentes e normalmente distribuídos, satisfazendo as condições para o uso de gráficos de controle. No presente trabalho, os gráficos de controle serão utilizados acrescidos de outros modelos adequados à transformação de observações autocorrelacionadas em observações que sejam independentes e normalmente distribuídas. Com isso, justifica-se o uso de gráfico de controle nos resíduos do modelo, com o propósito de detectar possíveis pontos anômalos no volume de vendas de óleo diesel e gasolina RS.

4.4 Previsão

No sentido de auxiliar nas tomadas de decisões e como forma de perceber como o processo de demanda de óleo diesel irá se comportar nos próximos meses tem-se a previsão do volume de vendas do derivado do petróleo.

A previsão é dada por meio do modelo SARIMA (0,1,1)(0,1,1)₁₂, que foi estimado com 122 observações, sendo assim, tem-se os valores futuros para as observações 123 a 128 e pode-se comparar o valor previsto com o valor real para os meses de março e abril de 2010.

Na Figura 8 observa-se os valores originais, e os valores estimados por meio do modelo SARIMA (0,1,1)(0,1,1)₁₂.

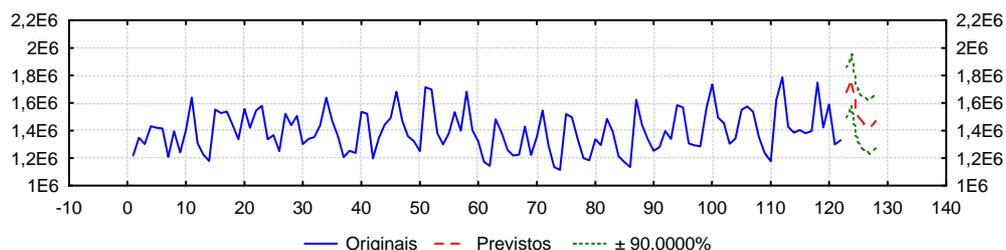


Figura 8 – Gráfico dos valores originais e valores estimados do modelo SARIMA (0,1,1)(0,1,1)₁₂.

Na Tabela 3 apresentam-se os valores previstos seis passos à frente, o período, os valores originais para março e abril de 2010 e os intervalos de confiança.

Tabela 3 – Período, valores originais para os meses de março e abril de 2010 e os valores previstos do modelo SARIMA (0,1,1)(0,1,1)₁₂.

Período	Originais	Previstos	Limite inferior	Limite Superior	Erro Padrão
mar/10	1.924.094	1.678.025	1.496.110	1.859.941	109639,10
abr/10	1.803.190	1.765.794	1.580.750	1.950.838	111525,00
mai/10	1.550.581	1.510.005	1.321.884	1.698.126	113379,40
jun/10	1.503.954	1.449.293	1.258.144	1.640.441	115204,10
jul/10	1.504.583	1.425.326	1.231.197	1.619.455	117000,2
ago/10		1.468.880	1.271.816	1.665.945	118769,2

Ao analisar os valores previstos em relação aos originais para os meses de março a julho de 2010, pode-se notar que o valor real para o mês de março não se encontrou entre os limites de previsão, isto se deve a alguma causa especial que se tratará mais adiante. Com relação aos demais valores, observa-se que os mesmos encontraram-se entre os limites de previsão, o que indica que o modelo encontrado captou bem o comportamento da série. O valor real para o mês de agosto não estava disponível no site da agência nacional do petróleo de modo a fazer uma comparação com o valor previsto. Salienta-se, que as previsões são úteis, como forma de visualizar o comportamento do volume de vendas de óleo diesel no RS e assim perceber como o mercado dessa *commodity* ira se comportar nos próximos períodos.

4.5 Aplicação de gráficos de controle no resíduo do modelo SARIMA (0,1,1)(0,1,1)₁₂ para analisar a estabilidade do processo

Com o objetivo de perceber como se comporta a demanda de umas das principais *commodity* ao longo de dez anos e com isso detectar possíveis pontos anômalos, faz-se o uso de gráficos de controle para medidas individuais e amplitude móvel.

Para dar início a esta etapa, apresentam-se os gráficos de controle para medidas individuais e amplitude móvel, na Figura 9, usando os dados originais com base nas amostras coletadas.

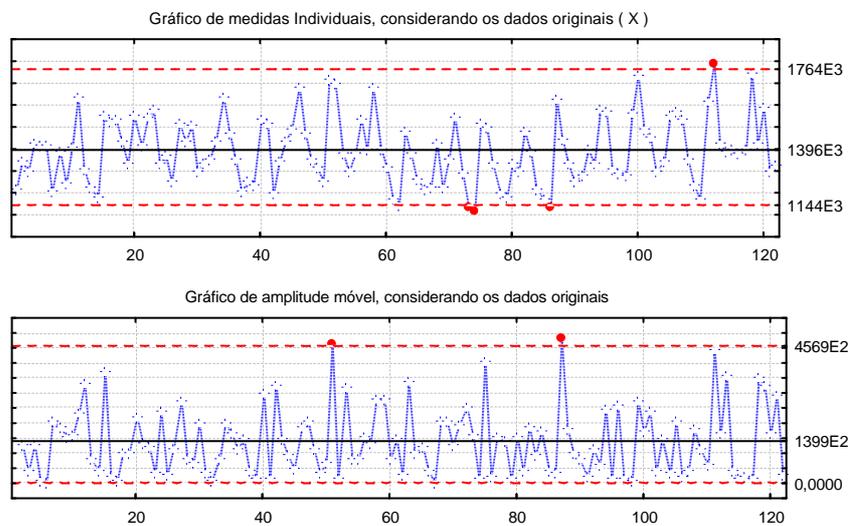


Figura 9 – Gráficos de controle para os valores individuais e amplitude móvel considerando os dados originais.

Analisando a Figura 9, pode-se verificar que o gráfico de medidas individuais e da amplitude móvel apresentou seis pontos fora dos limites de controle, ocasionando uma instabilidade da variável volume de vendas, em barris de óleo diesel.

Os seguintes limites de controle para o volume de vendas de barris de óleo diesel foram estimados:

$$\text{Gráfico de medidas individuais} \begin{cases} LSC = 1764 \cdot 10^3 \\ LC = 1396 \cdot 10^3 \\ LIC = 1144 \cdot 10^3 \end{cases}$$

$$\text{Gráfico de amplitude móvel} \begin{cases} LSC = 4569 \cdot 10^2 \\ LC = 1399 \cdot 10^2 \\ LIC = 0,00 \end{cases}$$

Aplicando os *runs tests*, detectaram-se três conjuntos de três pontos localizados na zona A ou fora dela, correspondendo às observações 50, 51 e 52; 72, 73 e 74; 111, 112 e 113, e um conjunto de cinco pontos amostrais localizados na zona B ou fora dela, correspondendo às observações 65, 66, 67, 68 e 69.

Na série de vendas de barris de óleo diesel, essas observações referem-se aos seguintes períodos: março de 2004, janeiro e fevereiro de 2006, janeiro e março de 2007 e abril de 2009. Cabe salientar que o volume de barris vendido em cada período é respectivamente 1.714.207, 1.135.610, 1.116.061, 1.171.896, 1.624.543 e 1.787.042. Em relação à média anual de cada período pode-se dizer que a demanda para março de 2004 encontrou-se superior a média anual (1.449.179); para janeiro e fevereiro de 2006, o volume de venda de barris encontrou-se inferior a venda anual (1.309.923); em janeiro de 2007 as vendas encontraram-se inferiores ao período (1.370.409), porém em março do mesmo ano, encontraram-se superiores ao período e em abril de 2009 a demanda de óleo diesel encontrou-se superior a média anual do período (1.465.366).

Uma estratégia frequentemente utilizada consiste em ajustar as observações da característica de qualidade a um modelo de previsão apropriado e monitorar o processo com cartas de controle para os resíduos identicamente independentes, resultantes conforme realizado no item 4.5, para o modelo encontrado da demanda de óleo diesel.

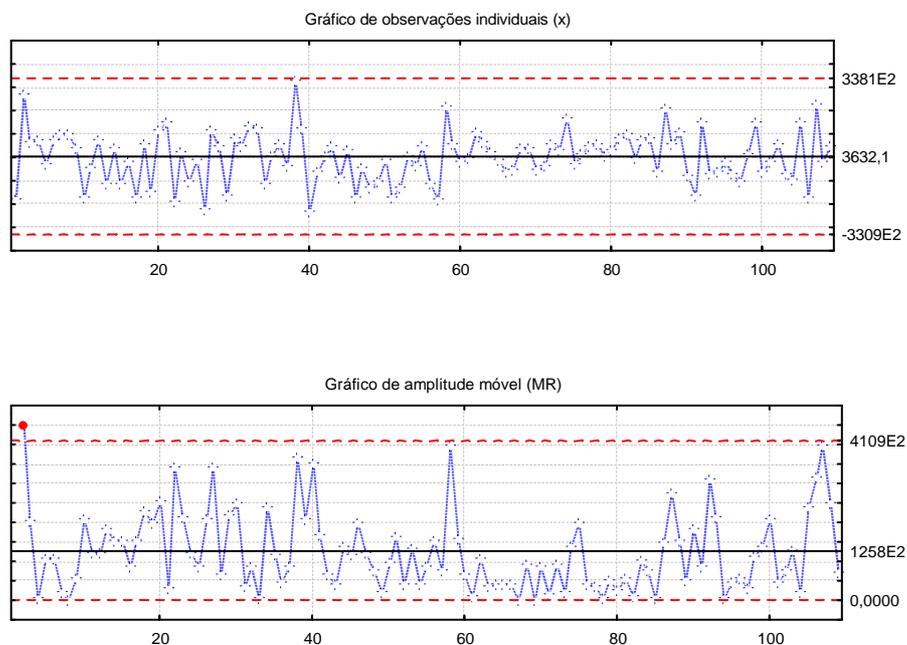


Figura 10 – Gráficos de controle para os valores individuais e amplitude móvel para os resíduos do modelo SARIMA $(0,1,1)(0,1,1)_{12}$.

Na Figura 10, têm-se gráficos de controle para os valores individuais e amplitude móvel para os resíduos do modelo SARIMA (0,1,1)(0,1,1)₁₂.

Estimou-se os seguintes limites de controle para os resíduos do modelo SARIMA(0,1,1)(0,1,1)₁₂:

$$\text{Gráfico de medidas individuais} \begin{cases} LSC = 3381 \cdot 10^2 \\ LC = 3632,10 \\ LIC = -3309 \cdot 10^2 \end{cases}$$

$$\text{Gráfico de amplitude móvel} \begin{cases} LSC = 4109 \cdot 10^2 \\ LC = 1258 \cdot 10^2 \\ LIC = 0,00 \end{cases}$$

De acordo com os resíduos do modelo estimado, traçou-se o gráfico de controle para medidas individuais e amplitude móvel, tendo como auxílio na decisão a aplicação dos *runs tests*. Neste caso detectou-se que quinze pontos amostrais em sequência situados na zona C, correspondendo à observação com início na amostra 59; nove pontos amostrais em sequência estão no mesmo lado da linha central, representado pelo conjunto de amostras com início nas observações 60 e 76. Observou-se também que 14 amostras alternam para cima e para baixo, correspondendo às observações com início na amostra 61 e término na amostra 74; a cada três pontos amostrais traçados, dois se encontram localizados na zona A ou fora dela, correspondendo às observações com início nas amostras 38 e 105. Além disso, observou-se que, a cada cinco pontos amostrais traçados, quatro se encontram localizados na zona B ou fora dela, correspondendo à observação com início na amostra 104.

Ao realizar outra análise, excluindo o ponto amostral fora de controle, para o gráfico da amplitude móvel, que corresponde à primeira observação, tem-se uma visão da instabilidade que está ocorrendo no processo. Consequentemente os limites de controle dos gráficos de medidas individuais e amplitude móvel foram recalculados.

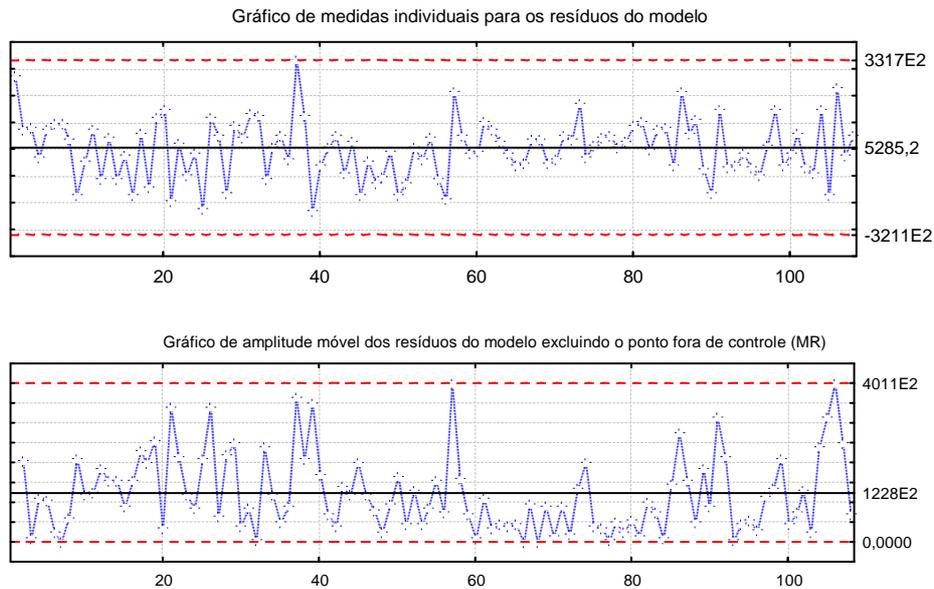


Figura 11 – Gráfico de controle para os valores individuais e amplitude móvel para os resíduos do modelo SARIMA $(0,1,1)(1,0,0)_{12}$ com a exclusão dos pontos amostrais fora de controle.

Recalculou-se os limites de controle para os resíduos do modelo SARIMA $(0,1,1)(0,1,1)_{12}$, com a exclusão do ponto amostral fora dos limites de controle para o gráfico de amplitude móvel.

$$\text{Gráfico de medidas individuais} \begin{cases} LSC = 3317 \cdot 10^2 \\ LC = 5285,20 \\ LIC = -3211 \cdot 10^2 \end{cases}$$

$$\text{Gráfico de amplitude móvel} \begin{cases} LSC = 4011 \cdot 10^2 \\ LC = 1228 \cdot 10^2 \\ LIC = 0,00 \end{cases}$$

Os gráficos da Figura 11 apresentam o monitoramento dos resíduos para o modelo SARIMA $(0,1,1)(0,1,1)_{12}$. De acordo com os resíduos do modelo estimado, traçou-se o gráfico de medidas individuais, mostrado na Figura 11, pelo qual se detecta quinze amostras em sequência situadas na zona C correspondendo à observação com início na amostra 58. No gráfico de amplitude móvel, constata-se que dois conjuntos de nove pontos amostrais em sequência estão no mesmo lado da linha central, com início nas observações 59 e 75, além de que uma sequência de quatorze pontos alterna para cima e para baixo com início na

observação 60. Também, observa-se que, a cada três pontos amostrais traçados, dois se encontram localizados na zona A ou fora dela, correspondendo às observações com início nas amostras 37 e 104, e a cada cinco pontos amostrais traçados, quatro se encontram localizados na zona B ou fora dela, correspondendo à observação com início na amostra 103.

Economistas costumam dizer que o preço de um produto é resultado da relação entre oferta e demanda, e de fato, essa é uma das raízes que esta em alta na primeira metade de 2008. A demanda por petróleo aumentou, especialmente por parte da China e da Índia, países que nos últimos anos vêm registrando uma alta taxa de crescimento econômico. Por outro lado, a cotação do petróleo também é influenciada por problemas enfrentados pelos países exportadores. A indústria petrolífera e os oleodutos na Nigéria e no Iraque, por exemplo, costumam ser alvo de ataques. Devido a esses problemas, cresce a preocupação de que a oferta não vá suprir a demanda, e o preço sobe.

Em relação ao volume de vendas de óleo diesel, que se apresentou fora de controle, principalmente, os pontos que correspondem às observações do ano de 2006, deve-se mencionar a forte influência do colapso provocado pela estiagem e pela queda de preço das *commodities* agrícolas que puxou para baixo o desempenho do comércio sulino em 2005, provocando um baixo volume de vendas de barris de óleo diesel. Durante todo o período analisado observou-se a fraca demanda de óleo diesel nos anos de 2005 e conseqüentemente em 2006, o menor volume de venda anual de barris foi registrado em primeiro lugar no ano de 2006 (1.309.923 barris) e, posteriormente, 2005 (1.311.534 barris).

A alta demanda registrada em alguns períodos deve-se ao fato de coincidir com a colheita das lavouras de soja, trigo e arroz que fazem aumentar o volume de vendas de óleo diesel devido ao fato de ser utilizado no transporte dessas *commodities*.

4.6 Estimação do modelo para as vendas de gasolina

Analogamente ao procedimento realizado para a variável venda de barris de óleo diesel procedeu-se a identificação do modelo apropriado que representa a série de vendas de barris de gasolina, sendo que o primeiro passo foi a inspeção gráfica, pois a mesma mostra o comportamento da série analisada. Na Figura 12 apresenta-se o gráfico do volume de vendas em barris de gasolina no Rio Grande do Sul na linha contínua, em que se observa que a mesma não se apresenta estacionária, isto é, seus valores não variam em torno de uma média constante, o que se pode constatar também pelo teste de Dickey-Fuller, que verifica a existência de raiz unitária na série. O teste apresentou o seguinte resultado: para a série

original o *p*-valor encontrado foi de 1,000, o qual excede o nível de significância de 5%, isto é, neste caso aceita-se a hipótese nula, ou seja, conclui-se que a série não é estacionária. Para a série com uma diferença o *p*-valor encontrado foi de 0,0470, o qual é inferior ao nível de significância de 5%, isto é, rejeita-se a hipótese nula, concluindo-se que a série diferenciada apresenta-se estacionária. Para comprovar a estacionariedade da série aplicou-se o teste KPSS, pois o ADF, quando as raízes são próximas de um, pode apontar resultados errôneos. Quanto ao teste KPSS, constatou-se que a série com uma diferença é estacionária, pois a estatística calculada é de 0,2258, que é menor que o valor crítico de 0,4630, o que indica a aceitação da hipótese nula de estacionariedade. Assim, os testes de Dickey – Fuller e KPSS asseguraram que a série de vendas de barris de gasolina com uma diferença é estacionária.

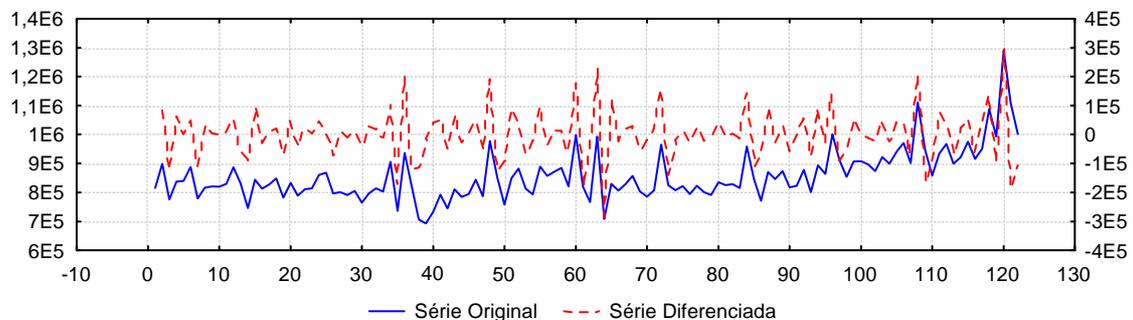


Figura 12 - Série original e diferenciada do número de barris de gasolina vendidos no Estado do Rio Grande do Sul no período de janeiro de 2000 a fevereiro de 2010.

A próxima etapa consiste em analisar as funções de autocorrelação (FAC) e autocorrelação parcial (FACP) da série original das vendas de barris de óleo diesel. Por meio do comportamento dessas funções pode-se ter uma idéia do tipo de modelo a ser utilizado. As Figuras 13a e 13b, apresentam, respectivamente, a FAC e a FACP da série de vendas de barris de gasolina no período analisado.

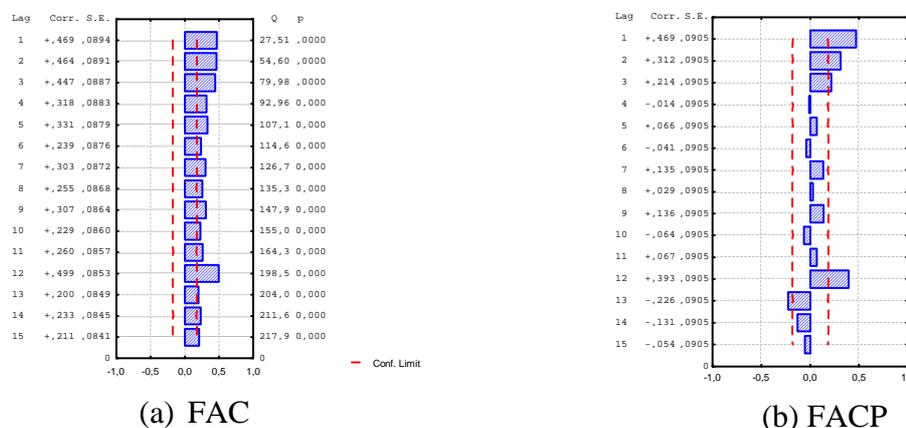


Figura 13 – (a) Função de Autocorrelação e (b) Autocorrelação Parcial da série original das vendas de barris de gasolina.

Talvez em uma análise menos criteriosa baseada somente na FAC amostral da Figura 13 pode-se observar o decaimento lento na função de autocorrelação, indicando a presença de longa dependência serial, mostrando claramente a persistência nos dados, sugerindo um modelo ARFIMA.

Portanto, além da modelagem SARIMA, aplica-se também a modelagem ARFIMA, com o propósito que encontrar o melhor modelo possível. O próximo procedimento é a identificação dos modelos.

Analisando-se os resultados apresentados a Tabela 4, é possível observar os modelos significativos encontrados, uma vez que todos apresentam coeficientes com *p-valor* menor que 5% e respeitam as condições de estacionariedade e invertibilidade, já que os valores estimados, em módulo, são menores que a unidade.

Mediante os modelos concorrentes obtidos, o melhor modelo ajustado para as vendas de barris de gasolina foi o SARIMA $(0,1,2)(0,1,1)_{12}$, o qual foi determinado pelo critério penalizador AIC.

Pode-se visualizar, que mesmo a série apresentando características de modelo ARFIMA, o que melhor captou o comportamento da série foi o modelo sazonal.

Tabela 4 – Estimação dos parâmetros para os modelos SARIMA e ARFIMA concorrentes da variável venda de barris de gasolina no período de janeiro de 2000 a fevereiro de 2010.

Modelos	Parâmetros	AIC
SARIMA (0,1,1)(1,1,0) ₁₂	q(1) = 0,773	21,852
	Ps(1) = -0,377	
SARIMA (0,1,1)(0,1,1) ₁₂	q(1) = 0,770	21,814
	Qs(1) = 0,496	
SARIMA (1,1,1)(0,1,1) ₁₂	p(1) = 0,271	21,792
	q(1) = 0,637	
	Qs(1) = 0,504	
SARIMA (0,1,2)(0,1,1) ₁₂	q(1) = 0,918	21,783
	q(2) = -0,259	
ARFIMA (1,d,0)	Qs(1) = 0,509	25,208
	d = 0,499	
ARFIMA (0,d,1)	AR-1 = -0,298	25,206
	d = 0,499	
ARFIMA (1,d,2)	MA-1 = -0,292	25,213
	d = 0,499	
	AR-1 = 0,568	
ARFIMA (3,d,1)	MA-1 = -0,929	25,183
	MA-2 = -0,343	
	d = 0,476	
	AR-1 = 0,557	
	AR-2 = 0,210	
	AR-3 = 0,219	
	MA-1 = -0,912	

AIC: Critério de Informação Akaike

4.7 Análise residual para o modelo SARIMA (0,1,2)(0,1,1)₁₂

Tendo encontrado o modelo matemático que eliminou a autocorrelação serial da série é necessário analisar os resíduos do modelo a fim de verificar se os mesmos são independentes e normalmente distribuídos, de forma a cumprir as pressuposições básicas para o uso dos gráficos de controle.

Analisando a FAC residual do modelo, na Figura 14, observou-se que o lag 13 apresentou uma autocorrelação acima dos limites de confiabilidade de 95%. Dessa forma,

para confirmar que os resíduos não são autocorrelacionados utilizou-se o teste de Ljung-Box Q^* . A estatística de teste tabelada $Q(15-2)=22,4$ é maior nas 15 primeiras autocorrelações. Assim, a hipótese nula de que essas 15 autocorrelações apresentam valor zero é aceita, e os resíduos são considerados ruído branco.

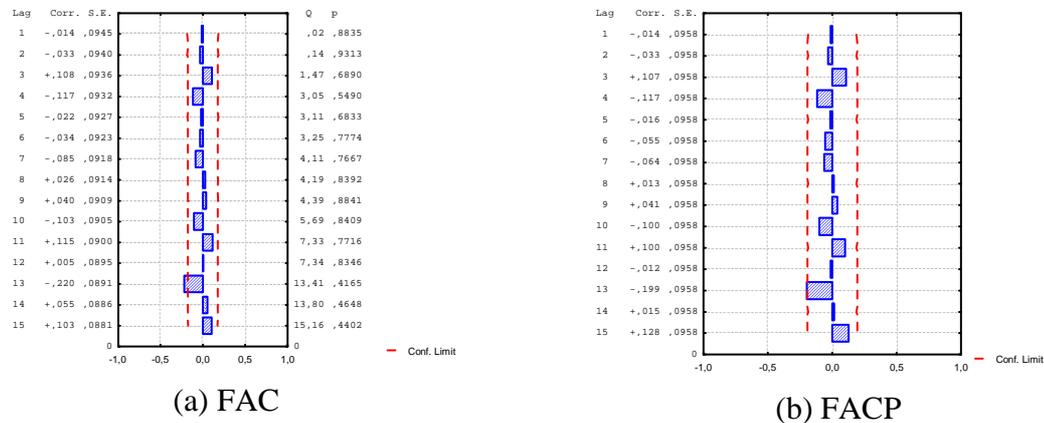


Figura 14 – (a) Função de Autocorrelação e (b) Autocorrelação Parcial para os resíduos do modelo SARIMA (0,1,2)(0,1,1)₁₂ identificado para o volume de vendas de barris de gasolina.

Para utilizar gráficos de controle é necessário que as observações da característica de qualidade sejam independentes e normalmente distribuídas. Observou-se que os resíduos do modelo não são autocorrelacionados, ou seja, eles são independentes. Analisando-se o gráfico Q-Q plot exibido na Figura 15, verifica-se que existe uma relação linear entre estes quartis, o que indica que os resíduos do modelo encontrado seguem uma distribuição normal.

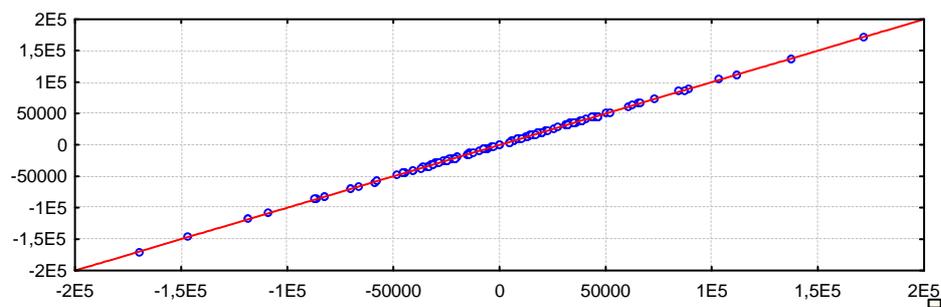


Figura 15 – Gráfico Q-Q plot dos resíduos do modelo SARIMA (0,1,1)(1,0,0)₁₂.

Para testar a normalidade dos resíduos do modelo encontrado, também se utilizou o teste de Lilliefors. Para o teste encontrou-se um p -valor superior a 0,20, aceita-se a hipótese de que a distribuição dos resíduos do modelo SARIMA (0,1,2)(0,1,1)₁₂ são normais.

Portanto os resíduos do modelo SARIMA (0,1,2)(0,1,1)₁₂ são independentes e normalmente distribuídos, satisfazendo as pressuposições para que se possa aplicar o controle estatístico do processo. Analogamente o procedimento da demanda de óleo diesel, faz-se o uso de gráficos de controle para o volume de barris de gasolina, uma vez que se realizou a

transformação de observações autocorrelacionadas (dados originais) em observações independentes e normalmente distribuídas (resíduos do modelo). Com isso, justifica-se o uso de gráfico de controle nos resíduos do modelo, com o propósito de detectar possíveis pontos anômalos na demanda de gasolina no RS.

4.8 Previsão

Encontrado o modelo que captou o comportamento da série em estudo, podem-se realizar previsões no sentido, de auxiliar, nas tomadas de decisões e como forma de perceber como o processo de vendas de gasolina irá se comportar nos próximos meses tem-se a estimativa do volume de vendas de barris de gasolina.

O modelo SARIMA $(0,1,2)(0,1,1)_{12}$, foi estimado com 122 observações, sendo assim, a previsão foi realizada para as observações 123 à 128. Pode-se comparar o valor previsto com o valor real para os meses de março e abril de 2010.

Na Figura 16 observa-se os valores originais e os valores estimados por meio do modelo SARIMA $(0,1,2)(0,1,1)_{12}$.

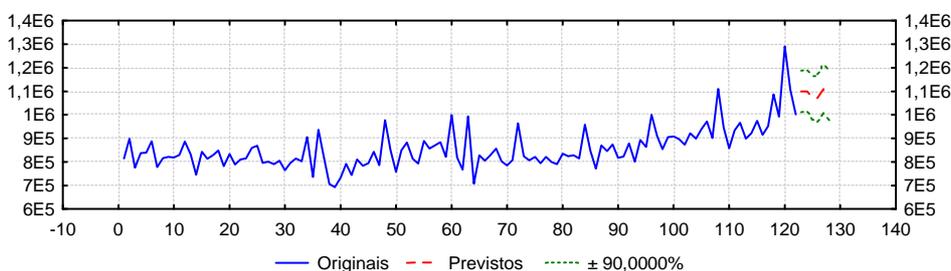


Figura 16 – Valores originais, valores estimados do modelo SARIMA $(0,1,2)(0,1,1)_{12}$.

Na Tabela 5 apresentam-se os valores previstos seis passos à frente, incluindo o período, os valores originais para março e abril de 2010 e os intervalos de confiança.

Analisando a Tabela 5, pode-se perceber que o modelo matemático estimado captou bem o comportamento da série em estudo, sendo que os valores originais de março à julho, das vendas de barris de gasolina, encontram-se entre os limites de previsão dados pelo modelo estimado.

Tabela 5 – Período, valores originais para os meses de março e abril de 2010 e os valores previstos do modelo SARIMA (0,1,2)(0,1,1)₁₂.

Período	Originais	Previstos	Limite inferior	Limite Superior	Erro Padrão
mar//10	1.155.360	1.090.027	1.011.200	1.186.163	52929,00
abr/10	1.102.028	1.100.039	1.011.916	1.188.163	53107,00
mai/10	1.039.360	1.072.701	979.620	1.165.782	56094,74
jun/10	1.057.216	1.067.825	970.037	1.165.613	58931,17
Jul/10	1.105.066	1.109.165	1.006.887	1.211.443	61637,21
ago/10		1.084.017	977.438	1.190.596	64229,35

4.9 Aplicação de gráficos de controle no resíduo do modelo SARIMA (0,1,2)(0,1,1)₁₂ para analisar a estabilidade do processo

Para dar início a esta etapa, apresenta-se os gráficos de controle para medidas individuais e amplitude móvel, na Figura 17, usando os dados originais com base nas amostras coletadas.

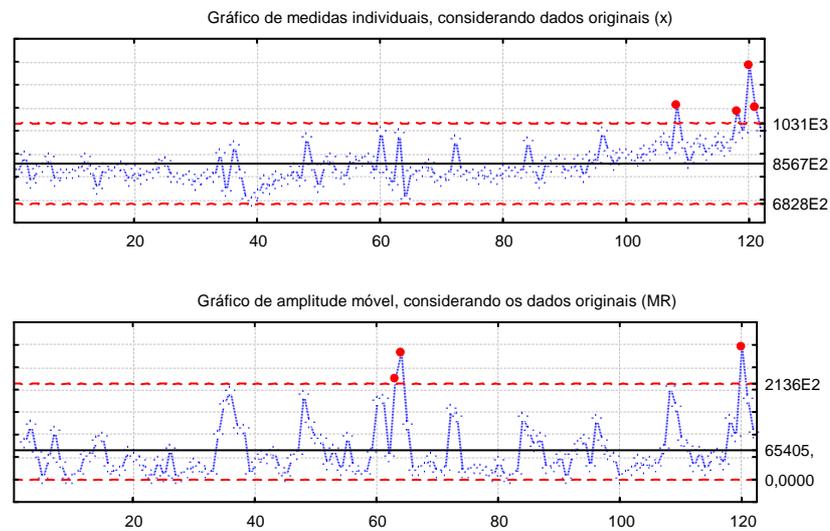


Figura 17 – Gráficos de controle para os valores individuais e amplitude móvel considerando os dados originais.

Os seguintes limites de controle para a variável venda de barris de gasolina foram estimados:

$$\text{Gráfico de medidas individuais} \begin{cases} LSC = 1031 \cdot 10^3 \\ LC = 8567 \cdot 10^2 \\ LIC = 6828 \cdot 10^2 \end{cases}$$

$$\text{Gráfico de amplitude móvel} \begin{cases} LSC = 2136 \cdot 10^2 \\ LC = 65405 \\ LIC = 0,00 \end{cases}$$

Analisando a Figura 17, pode-se verificar que o gráfico de medidas individuais e da amplitude móvel apresentou sete pontos fora dos limites de controle, ocasionando uma elevada instabilidade da variável venda de barris de gasolina.

Em relação à série de demanda de gasolina, esses pontos correspondem às seguintes observações: março e abril de 2005, dezembro de 2008, outubro e dezembro de 2009 e janeiro de 2010. Com relação à observação de 2005, esta se apresenta inferior à média anual (830.812), em dezembro de 2008 a demanda de gasolina apresentou-se superior a média registrada no ano (924.611), em outubro e de dezembro de 2009 o volume de vendas se mostrou superior a média registrada no ano (978.482) já o ponto correspondente a janeiro de 2010, não se sabe, pois não foi registrada a média do período, mais se pode dizer que a tendência é da mesma apresentar-se superior, pois nessa época do ano o aumento do volume de vendas é consideravelmente maior, pois aumenta o fluxo de veículos nas estradas devido ao período de férias.

O que influencia verdadeiramente os preços é a percepção que as pessoas têm da oferta e da demanda, isso significa que, quando dados reais sobre o petróleo são divulgados, como o balanço semanal das reservas de petróleo dos Estados Unidos, eles acabam tendo uma influência muito grande - já que poucos países são tão transparentes nessa área. Tirando isso, o mercado do petróleo não é como o de qualquer outra *commodity*, ele tem algumas características próprias. Primeiramente, existe a organização dos países produtores de petróleo (Opep), o cartel que controla 55% do petróleo exportado no mundo. A organização coordena aumentos e quedas da produção, para que todos os seus membros adotem a mesma medida. Nem sempre isso dá certo, mas certamente a existência da organização dos países produtores de petróleo (Opep) indica que o mercado de petróleo não opera livremente, como acontece com outras commodities. Em segundo lugar, o petróleo tem a característica de ser um produto não-renovável, ou seja, um dia vai acabar. Nisso, assemelha-se às *commodities* como o minério de ferro ou manganês.

Portanto, detectadas as observações que se mostraram fora do controle para o volume de vendas de gasolina, pode-se aplicar o teste de sequência para perceber como está se comportando o processo de vendas de barris de gasolina no RS.

Aplicando os *runs tests*, detectaram-se que cinco conjuntos de nove pontos amostrais em sequência estão do mesmo lado da linha central, representadas pelo conjunto de amostras com início nas observações 13, 37, 73, 74, 98, 99 e 108. Observou-se também que a cada três pontos amostrais traçados, dois se encontram localizados na zona A ou fora dela, correspondendo as observações com início nas amostras 34, 37, 59, 62, 107, 117, 119 e 120, um conjunto de sete amostras em que, a cada cinco pontos amostrais traçados, quatro se encontram localizados na zona B ou além dela, correspondendo as observações com início nas amostras 25, 37, 60, 79, 105, 111 e 116. Também se verificou que oito pontos amostrais em sequência situados além da zona C, correspondendo à observação com início na amostra 114.

Uma estratégia frequentemente utilizada consiste em ajustar as observações da característica de qualidade a um modelo de previsão apropriado e monitorar o processo com cartas de controle para os resíduos independentes resultantes. O interesse do trabalho é controlar a variabilidade do modelo matemático encontrado para as vendas de barris de gasolina no Estado do Rio Grande do Sul.

Na Figura 18, têm-se gráficos de controle para os valores individuais e amplitude móvel para os resíduos do modelo SARIMA $(0,1,2)(0,1,1)_{12}$.

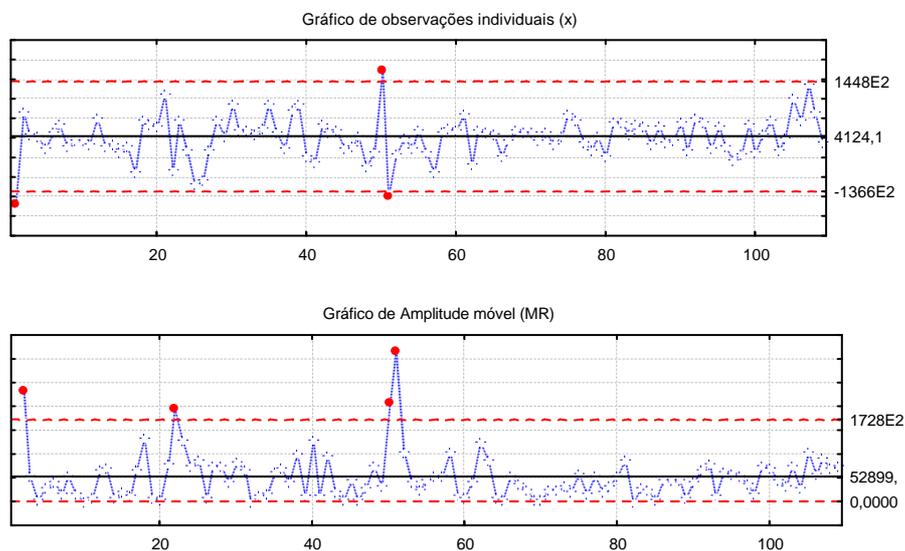


Figura 18 – Gráficos de controle para os valores individuais e amplitude móvel para os resíduos do modelo SARIMA $(0,1,2)(0,1,1)_{12}$.

Os seguintes intervalos de controle para os resíduos do modelo SARIMA $(0,1,2)(0,1,1)_{12}$, da variável vendas barris de gasolina foram estimados:

$$\begin{aligned} \text{Gráfico de medidas individuais} & \begin{cases} LSC = 1448 \cdot 10^2 \\ LC = 4124,1 \\ LIC = -1366 \cdot 10^2 \end{cases} \\ \text{Gráfico de amplitude móvel} & \begin{cases} LSC = 1728 \cdot 10^2 \\ LC = 52899 \\ LIC = 0,00 \end{cases} \end{aligned}$$

De acordo com os resíduos do modelo estimado, traçou-se o gráfico de controle para medidas individuais e amplitude móvel, tendo como auxílio na decisão a aplicação dos *runs tests*. Pelo qual se detecta que nove pontos amostrais em sequência estão no mesmo lado da linha central, representado pelo conjunto de amostras com início na observação 28. Observou-se também que em três conjuntos de dados, a cada três pontos amostrais traçados, dois se encontram localizados na zona A ou fora dela, correspondendo às observações com início nas amostras 24, 105 e 49. Além disso, nove pontos amostrais em sequência estão no mesmo lado da linha central correspondendo à observação com início na amostra 82 e 15 pontos amostrais localizam-se na zona C correspondendo a amostra com início na observação 90.

Ao realizar outra análise excluindo os pontos amostrais fora de controle, tanto do gráfico para medidas individuais, os pontos 1, 50 e 51 como para a amplitude móvel, os pontos 1, 22, 50 e 51, tem-se uma visão da instabilidade que está ocorrendo no processo. E conseqüentemente os limites de controle dos gráficos, de medidas individuais e amplitude móvel, são recalculados.

Os gráficos da Figura 19 apresentam o monitoramento dos resíduos para o modelo SARIMA (0,1,2)(0,1,1)₁₂. Como auxílio na decisão tem-se a aplicação dos *runs tests*. De acordo com os resíduos do modelo estimado, traçou-se o gráfico de medidas individuais, mostrado na Figura 18, pelo qual se detecta que, a cada três pontos amostrais traçados, dois se encontram localizados na zona A ou fora dela, correspondendo às observações com início nas amostras 22 e 101. Também se observou que, a cada cinco pontos amostrais traçados, quatro se encontram localizados na zona B ou fora dela, correspondendo às observações com início na amostra 100. No gráfico, de amplitude móvel, apresentado na Figura 18, constata-se o processo encontra-se sob controle de acordo aos *runs tests*.

Gráfico para medidas individuais aplicado aos resíduos do modelo com exclusão dos pontos amostrais fora de controle

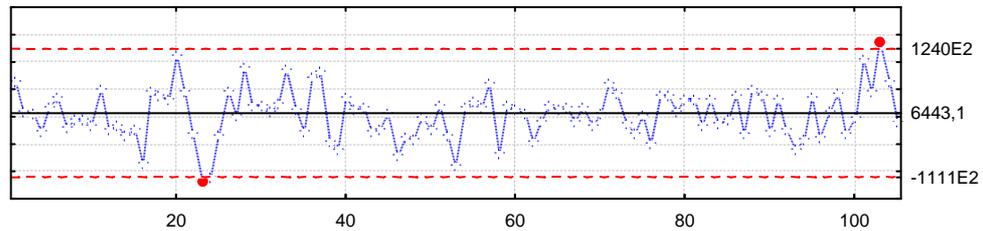


Gráfico para amplitude móvel aplicado aos resíduos do modelo com a exclusão dos pontos amostrais fora de controle

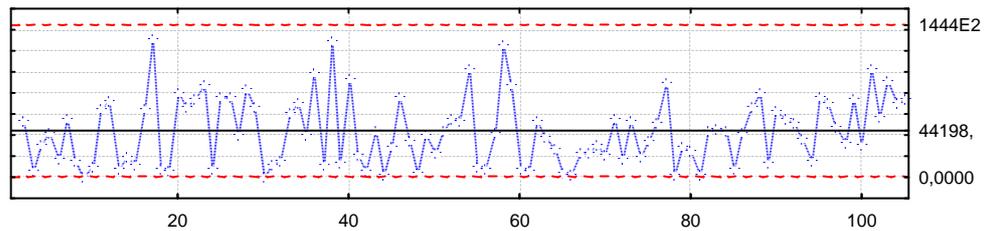


Figura 19 – Gráficos de controle para os valores individuais e amplitude móvel para os resíduos do modelo SARIMA (0,1,2)(0,1,1)₁₂ com a exclusão dos pontos amostrais fora de controle.

Recalculou-se os intervalos de controle para os resíduos do modelo SARIMA(0,1,2)(0,1,1)₁₂, com a exclusão do ponto amostral fora dos limites de controle para os gráficos de medidas individuais e amplitude móvel, tem-se os seguintes limites de controle:

$$\text{Gráfico de medidas individuais} \begin{cases} LSC = 1240 \cdot 10^2 \\ LC = 6443,1 \\ LIC = -1111 \cdot 10^2 \end{cases}$$

$$\text{Gráfico de amplitude móvel} \begin{cases} LSC = 1444 \cdot 10^2 \\ LC = 44198 \\ LIC = 0,00 \end{cases}$$

Com relação, a aplicação de gráficos de controle no processo de demanda de gasolina, pode-se dizer que o processo de vendas de barril de gasolina se mantém instável devido o petróleo ser influenciado por fatores aleatórios. Se um furacão passar pelo Golfo do México, pode interromper a produção na região. O mesmo pode acontecer no Oriente Médio se, por exemplo, um novo conflito começar por lá, e tudo isso tende a elevar os preços e consequentemente influenciar na demanda do derivado ao redor do mundo. Contudo a

metodologia foi válida uma vez que pode detectar pontos anômalos relacionados a fatores econômicos.

4.10 Comentário geral das análises

Utilizando-se a metodologia proposta, foi possível monitorar o comportamento da demanda de barris de óleo diesel no Estado do Rio Grande do Sul, utilizando o modelo SARIMA $(0,1,1)(0,1,1)_{12}$, o qual propiciou estimar os valores futuros para o período, como forma de auxiliar nas tomadas de decisões além de detectar pontos anômalos na economia.

Com relação aos resultados obtidos após a aplicação dos gráficos de controle para medidas individuais e amplitude móvel, apresentam-se as seguintes considerações:

- nos dados originais da variável, havia seis pontos amostrais fora dos limites de controle, sendo que três deles localizados no gráfico para medidas individuais e dois pontos localizados no gráfico para amplitude móvel;
- nos resíduos do modelo SARIMA $(0,1,1)(0,1,1)_{12}$, constatou-se um ponto amostral fora do limite de controle, localizado no gráfico de amplitude móvel;
- retirando-se o ponto fora de controle, no gráfico para amplitude móvel, constatou-se que nenhum ponto amostral ultrapassou os limites de controle.

Dessa forma, pode-se dizer que o modelo SARIMA $(0,1,1)(0,1,1)_{12}$ foi capaz de captar as variações que havia na série das vendas de barris de óleo diesel e detectou pontos amostrais fora dos limites de controle.

Sobre a previsão, referente ao mês de março, que ficam fora dos limites de previsão pode-se dizer que este fato deve ter ocorrido devido às causas especiais, pois a demanda por óleo diesel aumentou consideravelmente, embora a sazonalidade seja desfavorável. Segundo mencionou o diretor financeiro da Petrobras, Almir Barbassa, no site do canal rural, que a Petrobrás vendeu mais derivados, destacando o óleo diesel, no mercado o que contribuiu para o aumento dos resultados do primeiro trimestre.

Em relação aos padrões, indicados nos gráficos de controle, pode-se dizer que os mesmos referem-se às causas especiais, algumas trazem consequências em curto prazo, aumentam ou diminuem o volume de vendas de barris de derivados do petróleo, tais como taxa de importação, cotação do dólar entre outras. Como foi analisada anteriormente, a influência do colapso provocado pela estiagem e pela queda de preço das *commodities* agrícolas puxou para baixo o desempenho do comércio sulino em 2005, que trouxe consequências à curto e longo prazo, ocasionando a diminuição da demanda de óleo diesel no período. Com isso o Rio

Grande do Sul foi o estado que sofreu perdas de quase 80% na safra agrícola e o comércio varejista registrou a maior queda. As vendas caíram 2,1%, o pior desempenho no país.

Outro fator significativo, que fez aumentar o volume de vendas de óleo diesel em alguns períodos refere-se à boa colheita de soja, trigo e milho que com relação a logística do processo, faz-se necessário um aumento na demanda de óleo diesel.

Analogamente foi possível analisar o comportamento da demanda de barris de gasolina no Estado do Rio Grande do Sul, utilizando o modelo SARIMA $(0,1,2)(0,1,1)_{12}$, o qual proporcionou uma estimação para a média.

Com relação aos resultados obtidos após a aplicação dos gráficos de controle para medidas individuais e amplitude móvel, apresentam-se as seguintes considerações sobre a variável venda de barris de gasolina:

- nos dados originais da variável, havia sete pontos amostrais fora dos limites de controle, sendo que quatro pontos no gráfico para medidas individuais e três pontos no gráfico de amplitude móvel;

- nos resíduos do modelo SARIMA $(0,1,2)(0,1,1)_{12}$, constataram-se sete pontos amostrais fora dos limites de controle, sendo que três pontos localizados no gráfico para medidas individuais e quatro pontos localizados no gráfico de amplitude móvel.

Retirando-se os pontos fora de controle, no modelo SARIMA $(0,1,2)(0,1,1)_{12}$, tanto no gráfico para medidas individuais como para amplitude móvel, constataram-se dois pontos amostrais fora dos limites de controle para o gráfico de medidas individuais e para o gráfico de amplitude móvel, nenhum ponto encontro-se fora dos limites.

Dessa forma, pode-se dizer que o modelo SARIMA $(0,1,2)(0,1,1)_{12}$ captou as variações que aconteceram na série das vendas de barris de gasolina e detectou pontos amostrais fora dos limites de controle. A demanda de gasolina tem influência em diversos fatores, dentre eles, destaca-se o preço do barril de petróleo que aumentou 110% entre o início de 2007 e abril de 2008, o que elevou o preço dos transportes e dos insumos, como fertilizantes e adubos. A interrupção da produção de petróleo na Nigéria, a contínua queda do dólar ante o euro e o temor pela crescente demanda da commodity na China fizeram com que o petróleo chegasse próximo à casa dos 120 dólares em abril de 2008. Além do preço dos alimentos, o alto custo do petróleo influencia também outros setores da economia, desencadeando uma inflação geral.

5. CONCLUSÃO

Nesta pesquisa efetuou-se uma análise empírica do volume de vendas de óleo diesel e gasolina no RS, óleo diesel e gasolina, uma vez que o consumo destes combustíveis tem seguido uma trajetória crescente, fato que pode ser explicado pela expansão do agronegócio juntamente com o deslocamento de pólos de produção e comerciais para o interior do Estado.

Dada a importância e uso de diesel e gasolina, o presente estudo desenvolveu um modelo baseado na metodologia de Box-Jenkins para o estudo de séries temporais, com a finalidade de prever até o mês de agosto de 2010 a evolução do volume de vendas de diesel e gasolina pelas distribuidoras no Estado do Rio Grande do Sul. De um modo geral pode-se dizer que as previsões se confirmaram para aqueles valores que se pode comparar. Além disto, buscaram-se tratar o efeito da autocorrelação presente nos dados, pois este é um dos fatores preponderantes para a utilização do CEP como ferramenta de análise, que possibilita visualizar o comportamento da variável ao longo do tempo, permitindo detectar possíveis variações no volume de vendas dos derivados combustíveis do petróleo no período de análise sua utilização será considerada como uma ferramenta de apoio à decisão e planejamento estratégico na região.

O emprego das ferramentas estatísticas de CEP permitiu revelar o comportamento do volume de vendas de óleo diesel e gasolina, mostrando que grandes variações no processo têm forte influência econômica, fato esse observado no ano de 2005, em que a influência do colapso provocado pela estiagem e pela queda de preço das *commodities* agrícola fez baixar o desempenho do comércio sulino, o que ocasionou a diminuição do volume de vendas dos combustíveis.

É importante destacar que a demanda por óleo diesel aumentou consideravelmente, embora a sazonalidade seja desfavorável. Segundo mencionou o diretor financeiro da Petrobras, Almir Barbassa, “a Petrobrás vendeu mais derivados, destacando o óleo diesel, no mercado o que contribuiu para o aumento dos resultados do primeiro trimestre de 2010”.

Buscou-se, por meio deste estudo, evidenciar o comportamento do volume de vendas dos principais combustíveis derivados do petróleo no RS, associando essa variação a fatos econômicos que possam ter contribuído significativamente para tal. Além disso, pode-se dizer que os resultados obtidos do modelo ajustado para previsão em curto prazo conduzem para um cenário favorável quanto à participação destes tipos de combustíveis na área estadual.

De maneira geral concluiu-se, que os modelos encontrados para a série de vendas de óleo diesel, SARIMA $(0,1,1)(0,1,1)_{12}$ e para a série de vendas de gasolina SARIMA $(0,1,2)(0,1,1)_{12}$, foram capazes de explicar satisfatoriamente os dados, possibilitando uma análise de seu comportamento, de mesmo modo que tornou-se possível fornecer previsões a curto prazo, sinalizando bem o comportamento do volume de vendas de óleo diesel e gasolina. Além disso, deve-se mencionar que os grandes números de barris vendidos pelas distribuidoras, em março e abril, tiveram como influência o fato das distribuidoras adotarem políticas mais agressivas de fidelização, vendendo óleo diesel com preço referente a janeiro, sendo que este tipo de ação garantiu o aumento nas vendas de diesel, conforme publicação da revista Brasil Energia.

Deixam-se como sugestões para trabalhos futuros a utilização dos modelos autorregressivos de heterocedasticidade condicional, ARCH, para modelar a variância condicional da série de volumes de vendas de óleo diesel e gasolina e a utilização de estimadores conjuntos para parâmetros de modelo SARIMA-ARCH.

Além disso, sugere-se a utilização de gráficos de controle das somas acumulativas (CUSUM) e o gráfico de controle da média móvel ponderada exponencialmente (EWMA) pois, esses gráficos são indicados para o monitoramento de processos sujeitos a pequenas perturbações e também consegue-se maior rapidez na sinalização de pequenos desajustes.

6. REFERÊNCIAS

AGÊNCIA NACIONAL DO PETRÓLEO, GÁS NATURAL E BIOCOMBUSTÍVEIS. Disponível em <<http://www.anp.gov.br>>. Acesso em: 10 mai. 2010.

ALVES, L. Diesel de consumo brasileiro. Equipe Brasil Escola. Disponível em: <<http://www.brasilecola.com>> Acesso em: 20 jun. 2010.

ALWAN, L. C.; ROBERTS, H. V. Times series modeling for statistical process control. **Journal of Business & Economics Statistics**, v. 6, n. 1, p. 87-95, 1998.

BARBASSA, A. Entrevista disponibilizada em 21 de junho de 2010, a internet. Disponível em: <<http://www.canalrural.com.br>> Acesso em: 24 jun. 2010.

BEM, A. B. **A utilização da regressão dinâmica para prever a demanda por matrícula no ensino fundamental em Santa Catarina**. 1998. 183f. Dissertação (Mestrado em Engenharia de Produção) – Universidade Federal de Santa Catarina, Florianópolis, 1998.

BOX, G. E. P.; PIERCE, D. A.. Distribution of autocorrelations in autoregressive moving average models. **Journal of Time Series Analysis Association**, 65, 1509-1526, 1970.

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

BOX, G. E. P.; JENKINS, G. M. ; REINSEL, G. **Time Series Analysis : Forecasting and Control**. Third Edition. Englewood Cliffs: Prentice Hall, 1994.

BRESSAN, A. A.; LIMA, J. E.. Modelos de previsão de preços aplicados aos contratos futuros de boi gordo da BM&F. **Nova Economia**, 12(1), 117-140, Belo Horizonte, 2002.

BUENO, R. L. S. **Econometria de séries temporais**. São Paulo: Cengage Learning, 2008.

COSTA, A. F. B.; EPPRECHT, E. K.; CARPINETTI, L. C. R. **Controle estatístico da qualidade**. São Paulo: Atlas, 2004.

COSTA, A. F. B.; EPPRECHT, E. K.; CARPINETTI, L. C. R. **Controle estatístico da qualidade**. São Paulo: Atlas, 2005.

DEMING, W. E. Out of crisis. Cambridge: MIT Press, 2000.

EHLERS, R. S. Análise de séries temporais. In: Notas de aula de séries temporais. São Paulo: USP, 2007. Disponível em: <<http://www.icm.usp.br/~ehlers/notas/stemp.pdf>>. Acesso em: 30 mai. 2010.

GRANGER, C.; JOYEUX, R. An introduction to long memoru time series models and fractional diffetencing. **Journal of Time Series Analysis**, v. 1, 1980.

HOSKING. J. Fractional differencing. **Biometrika**, n.68, 1981.

LJUNG, G. M; BOX, G. E. P. On a measure of lack of it time series models. **Biometrika**, v. 65, n. 2, p. 297-303, 1978.

MALHOTRA, N. K. **Pesquisa de Marketing**: uma orientação aplicada. 3. ed. Porto Alegre: Bookman, 2001

MONTGOMERY, D. C. **Introdução ao controle estatístico da qualidade**. 4.ed. Rio de Janeiro: LTC, 2004.

MONTGOMERY, D. C; RUNGER, G. C. **Estatística aplicada e probabilidade para engenheiros**. 2.ed. Rio de Janeiro: LTC, 2003.

MOREIRA, A. R. B. **Modelos para a projeção do consumo nacional e regional de óleo diesel**. IPEA: Rio de Janeiro, 1996.

MORETTIN, P. A. **Econometria financeira**: um curso em séries temporais financeiras. São Paulo: ABE, 2006.

MORETTIN, P. A.; TOLOI, C. M. C. **Análise de séries temporais**. 2.ed. São Paulo: Edgard Blücher, 2004.

MOURÃO, B., SOUZA JUNIO, C., BORSATO, G., ANDRADE, G., QUELHAS, O. **Previsão de vendas em uma empresa de combustíveis**. Revista Carioca de Produção-RECAP. Rio de Janeiro, 2009.

KLIDZIO, R. **Modelos de previsão aplicados ao controle de qualidade com dados autocorrelacionados**. 2009. 153p. Dissertação (Mestrado em Engenharia de Produção) – Universidade Federal de Santa Maria, Santa Maria, 2009.

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

SIQUEIRA, L. G. P. **Controle estatístico do Processo**: série qualidade. São Paulo: Pioneira, 1997.

WIKIPÉDIA. A enciclopédia livre. Disponível em <<http://wikipedia.org/wiki/Petrobras>>. Acesso em: 25 jun. 2010.