Universidade de Lisboa
Faculdade de Ciências
Departamento de Estatística e Investigação Operacional

Modelos de previsão aplicados à optimização da gestão das actividades de um Call Center

Marta Alexandra Lourenço Machado

Relatório de Estágio

Mestrado em Estatística

2012

Universidade de Lisboa
Faculdade de Ciências
Departamento de Estatística e Investigação Operacional

Modelos de previsão aplicados à optimização da gestão das actividades de um Call Center

Marta Alexandra Lourenço Machado

Relatório de Estágio orientado pelo Prof. Doutor Kamil Feridun Turkman e supervisionado pelo Doutor Nuno Santos

Mestrado em Estatística

2012

Agradecimentos

Em primeiro lugar gostaria de agradecer ao meu orientador, ao Prof. Feridun Turkman por toda a atenção prestada ao longo deste trabalho, por toda a paciência que teve comigo, por estar sempre disponível em receber-me para tirar as minhas dúvidas e por transmitir-me conhecimentos sobre esta área.

Em segundo lugar gostaria de agradecer ao meu supervisor de estágio, Nuno Santos e a toda a sua equipa, nomeadamente à Odília, à Luísa, ao Pedro Santos, ao Pedro Constantino e ao Nuno Rocha por todo o apoio prestado e por me terem recebido tão bem no vosso local de trabalho. Foram 9 meses de estágio que tiveram muito significado para mim, que me deixaram muitas saudades e posso dizer que foi um prazer ter trabalhado convosco.

Também queria agradecer à Faculdade de Ciências da Universidade de Lisboa e à Empresa de Telecomunicações Vodafone Portugal por me terem dado a oportunidade de realizar este estágio.

Agradeço de igual modo à Prof.ª Teresa Alpuim por me disponibilizado alguns apontamentos relativos à sua cadeira de séries temporais, à Prof.ª Helena Mourino por algumas referências bibliográficas cedidas, à Profª. Patrícia Bermudez por algumas explicações relacionadas com a linguagem R, ao Prof. António José Rodrigues por alguns esclarecimentos relacionados sobre séries temporais e redes neurais e por fim à minha coordenada de mestrado, a Profª. Helena Iglésias que sempre me ajudou nas questões burocráticas acerca do mesmo e esteve sempre presente e disposta em me ajudar.

Queria agradecer também ao meu Prof. de Matemática do Secundário, Miguel Padilha e ao meu explicador dessa altura, David da Graça por me terem cativado o gosto pela matemática e deste modo eu ter seguido esta área da qual gosto muito.

Agradeço ao meu namorado e amigo Rui Martins por todo o apoio dado nesta etapa da minha vida, pela força e confiança que sempre me deste e peço-te desculpa pelas minhas ausências devido à realização este trabalho.

Por último mas não menos importante, queria agradecer profundamente aos meus pais por toda a educação que me deram e por estarem sempre presentes na minha vida. Obrigada por me terem tornado na pessoa que sou hoje e por sempre confiarem em mim e nas minhas capacidades. Também deixo aqui um beijinho muito especial ao meu irmão Miguel do qual amo muito.

A todos os meus amigos, especialmente à minha colega Soraia Pereira que me acompanhou mais nesta etapa, e a todas as pessoas de que alguma forma tiveram envolvidas neste projecto, o meu muito obrigada por tudo.

Resumo

Este estágio teve como principal objectivo encontrar uma alternativa ao método de previsão utilizado pela empresa Vodafone no que diz respeito às suas linhas de apoio aos seus clientes.

Os dados referem-se ao número de chamadas atendidas diariamente de diferentes linhas de apoio e portanto estas linhas podem ser tratadas como séries temporais.

Pretende-se encontrar quais os melhores modelos que se ajustam às series temporais diárias observadas para as várias linhas de apoio a clientes da referida operadora móvel e depois usaremos estes modelos para prever os valores futuros para estas séries temporais. Um dos objectivos é avaliar a qualidade de ajustamento bem como o poder preditivo destes modelos.

O método de previsão sugerido é a clássica Metodologia de Box-Jenkins que se baseia nos modelos de séries temporais lineares. Grande parte das nossas séries temporais, apresenta uma forte componente sazonal diária e portanto os modelos que iremos aplicar são os chamados modelos lineares multiplicativos não estacionários sazonais que designam-se por SARIMA$(p,d,q)x(P,D,Q)_s$.

Ao fazermos uma análise cuidadosa destas séries temporais, por vezes estas mostra-nos algumas não linearidades e desta maneira devemos aplicar os modelos GARCH para os resíduos obtidos dos modelos lineares SARIMA, para explicar o elevado grau de volatilidade presente em algumas séries temporais.

No final deste relatório, apresento detalhadamente um caso prático referente a uma linha de atendimento que se encontra no activo na empresa. Tendo em conta o modelo que ajustei para esta linha, pretende-se prever os valores desta para os meses de Novembro e Dezembro de 2011 e para o mês de Janeiro de 2012 e comparar esses valores com os valores reais dessa linha e também pelo método utilizado na Vodafone, para depois retirarmos conclusões acerca dos mesmos. Posteriormente, realizamos o mesmo procedimento mas agora para os meses de Junho, Julho e Agosto de 2012. Uma vez que até a este momento só faço o retrato de uma linha, também irei aplicar resumidamente estes métodos a outras linhas que faltam considerar, para depois tirarmos uma conclusão final acerca dos métodos apresentados.

Os resultados vão-nos mostrar claramente a superioridade do método Box-Jenkins sobretudo para períodos curtos de previsão em relação ao método aplicado pela Vodafone.

Palavras - Chave:

Previsão, Série Temporal, Metodologia Box-Jenkins, SARIMA, GARCH

Abstract

The main objective of this training report is to find an alternative method of prediction to the methods currently in use by the Vodafone company for customer helpline telephone traffic.

Data correspond to the daily number of telephone calls received on several different helplines and therefore are treated as time series.

We hope to find models which fit best to the observed daily time series of several different customer helpline traffic, and then use these models for prediction of future values of these time series. The objective is to assess the quality of fit as well as the predictive power of these models.

The method of prediction which we suggest is the classical Box and Jenkins Methodology based on linear time series models. Almost all the observed time series show strong weekly seasonal components and therefore the models which we employ are the seasonal non-stationary linear multiplicative models designated by SARIMA(p,d,q)X(P,D,Q)s.

The careful analysis of the time series shows that there is non-linearity and we employ GARCH models to the residuals obtained from the linear SARIMA models to explain the high degree of volatility in some of the time series.

At the end of this report a case study will be presented referring to a hotline which is in active in the company. Based on the adjusted model, the predicted values of number of calls for this line in the months of November, December 2011 and January 2012 were compared with the real values (observed) and also with the method used in Vodafone. Afterwards, the same methodology was applied for the months of June, July and August 2012 for this line, as well as in other lines and conclusions were drawn about the presented methods.

The results will clearly show the superiority of the Box and Jenkins method particularly for short term predictions over the methods employed by the Vodafone.

Key Words:

Predictions, Time Series, Box–Jenkins Methodology, SARIMA, GARCH

Conteúdo

Introdução

Nos dias de hoje as previsões desempenham um papel extremamente importante na sociedade em que vivemos e que atinge as mais diversas áreas nomeadamente nas vidas de empresas, bancos, estudos clínicos, etc.

As previsões podem ser designadas como as “estimativas do futuro” e por vezes há uma grande necessidade e utilidade de antever o futuro que se avizinha de forma a prevenir eventuais imprevistos, complicações e custos desnecessários que possam surgir nas situações do nosso dia-a-dia. Mas ao prevê-las, há que correr riscos por vezes e haverá certamente a ocorrência de erros uma vez que o futuro é sempre algo incerto e nada é tido como garantido certamente.

Como o leitor já deverá ter reparado, as previsões serão a parte central deste trabalho.

Assim, sem perder mais tempo, passo a descrever detalhadamente o que se passará nos capítulos seguintes:

No primeiro capítulo, irei abordar principalmente a natureza dos dados, explicitando alguns conceitos que o leitor possa não conhecer muito bem ou desconhecer e assim ficar mais esclarecido sobre o assunto. Também vou referir os objectivos deste trabalho e os objectivos da empresa sobre a realização deste projecto.

No capítulo seguinte, irei enunciar algumas noções básicas no domínio das séries temporais de modo a percebemos o mecanismo teórico que vem por detrás de todos os sistemas de previsão.

No terceiro capítulo vou pronunciar sobre os diversos métodos de previsão que existem actualmente e quais as metodologias novas que irei aplicar relativamente ao cálculo de previsões.

No último capítulo terei a componente prática, onde irei considerar um conjunto de dados e onde vou aplicar um ou dois métodos de previsão alternativos ao que se encontra implementado na empresa. No final de tudo, discutirei os resultados e retirarei as conclusões finais relativo a toda a análise em estudo.

Nos anexos, encontram-se outros conjuntos de dados em que se tira conclusões análogas ao do caso prático demonstrado no último capítulo deste trabalho.

Motivação

Este trabalho foi o resultado final de um estágio de 9 meses na empresa de Telecomunicações Vodafone Portugal. Durante este tempo, estive inserida num departamento que tem o nome de Grupo Operações. Este grupo tem como principais funções gerir as linhas de apoio cujos serviços a empresa fornece aos seus clientes; um exemplo de uma linha poderá ser a 16912, esta linha é designada por apoio aos clientes da referida operadora; dá assistência às designadas equipas de "Call Center", monitoriza os assistentes envolvidos na prestação de tais serviços e têm de mandar regularmente previsões de três em três meses sobre cada uma das linhas para estas empresas de "Outsourcing" (significa sub-contratação de serviços fora da empresa – caso o leitor desconheça este termo).

Dito isto, o meu estágio teve como ponto fulcral as previsões e poderemos ver a seguir as metas a atingir no final deste relatório.

Objectivos do Estágio

- Apresentar propostas de melhoria relativamente ao modelo de previsões implementado na empresa;
- Implementação de um novo método de previsão (caso seja possível);
- Comparar as previsões obtidas com o novo modelo com os valores reais das linhas e com as previsões fornecidas pelo actual modelo de previsão da empresa;
- Verificar se houve melhorias significativas para algumas linhas;
- Possibilidade de implementação deste novo método no futuro.

A seguir, irei anunciar os objectivos essenciais do ponto de vista da empresa relativamente ao envio destes valores para as equipas de "Call Center".

Objectivos da Empresa

- Garantir o dimensionamento para atingir certos níveis de serviço de atendimento;
- Garantir a produtividade dos assistentes.

Para elucidar mais um pouco o leitor, relativamente ao primeiro ponto dos objectivos descritos acima, o que acontece é que cada linha de apoio tem um certo nível de serviço de atendimento (normalmente expressa-se em percentagem) e dever-se-á tentar manter esse valor sempre que seja possível. Assim, ao assegurar estas referências, estamos a garantir um nível de qualidade de serviço e de atendimento aos clientes. Relativo ao segundo ponto referido, há por vezes certos períodos em que há um maior aumento ou diminuição no número de chamadas; por exemplo, aos feriados há poucas chamadas e na altura do Natal há um aumento

significativo do seu número, e deste modo, há que se colocar menos ou mais assistentes nos “Call Center” consoante a procura nessas épocas.

Já por diversas vezes, já mencionei o nome de “Call Center” e caso o leitor não conheça muito bem o seu conceito, transmitirei a seguir algumas noções para que fique bem assente.

“Call Center”

A designação de “Call Center” ou alternativamente pode-se chamar como “Contact Center”, sendo que em português tomam a designação de “Centro de Chamadas” ou “Centro de Contactos”, respectivamente; estes centros têm como principal objectivo de garantir a comunicação fácil, rápida e de forma directa entre as empresas e os seus clientes, funcionando como um intercâmbio entre ambos os lados. Têm ao seu dispor pessoal especializado nas mais diversas áreas abrangentes, chegando a funcionar 24 horas por dia dependendo dos serviços disponibilizados. Têm como principal função de resolver variadíssimas situações entre elas questões de ordem técnica, resolução de problemas, também realizam-se estudos de mercado, vendas de produtos e equipamentos entre outras coisas.

Podemos encontrar diversas definições sobre o que é um “Call Center”. Apenas irei enunciar algumas encontradas nos motores de busca na internet, para título de exemplo.

“Serviço que centraliza o atendimento de chamadas telefónicas, possibilitando o apoio ao cliente, a realização de estudos de mercado, vendas e outros serviços.”

Dicionário de Língua Portuguesa - Porto Editora

“Uma central de atendimento (ou call centre, ou ainda call center) é composta por estruturas físicas e de pessoal, que têm por objetivo centralizar o recebimento de ligações telefônicas, distribuindo-as automaticamente aos atendentes e possibilitando o atendimento aos usuários finais, realização de pesquisas de mercado por telefone, vendas, retenção e outros serviços por telefone, Web, Chat ou e-mail.”

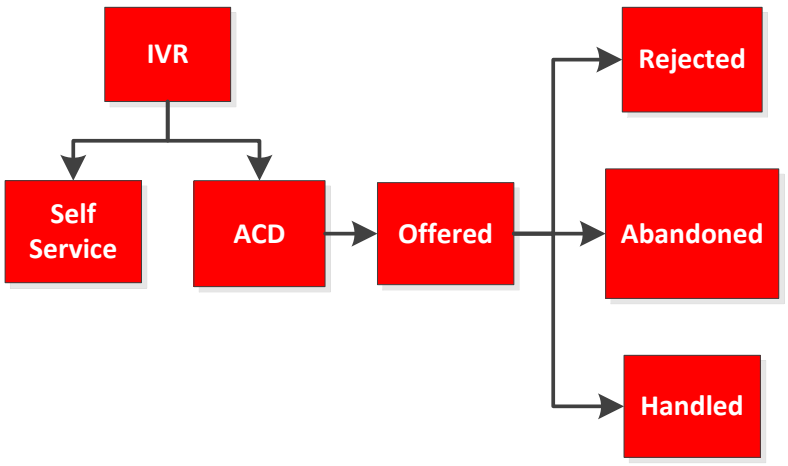
Wikipédia

Como título de curiosidade, vou enumerar algumas conhecidas empresas de “Call Center” a operar cá em Portugal relacionadas com o sector das Telecomunicações, como a Teleperformance Portugal, Randstad, RHmais e Contact.

Caso não saiba, existe uma associação portuguesa de “Contact Centers”. Para mais informações, pode consultar o respectivo site: www.apcontactcenters.com.

Distribuição das Chamadas

O esquema representado a seguir dá-nos uma visão muito simples de como é o caminho de uma chamada desde do momento em que “entra” num sistema de um “Call Center”.



Vamos imaginar que quando um cliente liga para um determinado serviço, a sua chamada vai ser logo atendida automaticamente por um atendedor. A este acontecimento dá-se o nome de IVR (atendimento automático).
Caso o cliente consiga resolver o seu problema dentro do IVR sem solicitar a ajuda de nenhum assistente, este serviço designa-se por "Self Service", caso contrário a sua chamada entra num sistema de distribuição automática de chamadas, designado por ACD.
Assim, quando a chamada entra no sistema de atendimento, ela é designada por oferecida ("offered") e depois pode acontecer três possibilidades à chamada:

- Pode ser rejeitada ("Rejected"). Por exemplo, se uma linha só funcionar em dias úteis e o cliente ligar no fim-de-semana.
- Caso os clientes desligam a chamada a meio do processo de espera, a chamada considera-se abandonada ("Abandoned"). Caso não o façam, a chamada pode ser tratada/resolvida com a ajuda de um assistente personalizado ("Handled").

Dados

Relativamente aos meus conjuntos de dados, só me vai importar para a minha análise as chamadas atendidas que são dadas pela seguinte expressão:

Chamadas Atendidas = Offered – Rejected

Base de Dados

Quanto ao sítio aonde tenho acesso aos dados, eles encontram-se disponíveis no programa Business Objects XI fornecido pela empresa através da aplicação Citrix. Convém referir como nota que esta base de dados só armazena os dados mais recentes. Como no meu caso tive a necessidade de precisar de dados mais antigos, foi-me fornecido um ficheiro excel com respeito a previsões de anos anteriores para cada uma das linhas em estudo (não tendo de facto os valores reais das linhas).

Introdução às Séries Temporais

Uma série temporal, por vezes também denominada por sucessão cronológica, é um conjunto de observações feitas sequencialmente ao longo do tempo e apresenta normalmente as seguintes notações: $\{X_t\ , t \in T\}$ ou $\{X(t)\ , t \in T\}$, em que os valores de X são chamados de estados e T é o conjunto de índices.
Uma característica muito importante neste tipo de dados é que as observações vizinhas são dependentes e um dos grandes interesses deste estudo é analisar e modelar essa dependência.

Pode-se classificar este tipo de séries em dois tipos:

Univariada:

$$\{X_t\ ,\ \ t \in T\}$$

Exemplo: Seja X_t a quantidade diária de leite produzida (em litros) durante o mês de Março de 2012.

Multivariada:

$$\{X_1(t), \dots, X_k(t),\ \ t \in T\}$$

Exemplo: Designamos por $X_1(t)$ as vendas semanais de uma revista e por $X_2(t)$ os custos associados à publicidade da mesma durante o ano de 2011.

E o conjunto T pode ser:

Discreto:

$$T = \{0,1,2, \dots, N\}$$

Exemplo: Número de nascimentos mensais ocorridos no Hospital do Barlavento Algarvio entre os anos de 2000 e 2011.

Contínuo:

$$T = [0, +\infty)$$

Exemplo: Registo da temperatura de uma arca frigorífica durante uma semana em que $T = [0,24]$ se a unidade de tempo é em horas.

A variável de interesse, X, também pode ser considerada discreta ou contínua. A maior parte das vezes, utiliza-se mais a discreta. Neste trabalho como estamos perante um conjunto de dados que dizem respeito ao número de chamadas telefónicas atendidas a nossa variável de interesse será discreta.

Uma série temporal pode conter duas componentes:

Determinística: Quando os valores da série podem ser escritos através de uma função matemática.

> **Exemplo:** Podemos modelar a tendência através de uma função de senos e cosenos.

Estocástica: Caso a série envolva termos aleatórios.

> **Exemplo:** O modelo final de uma série temporal pode depender do passado dos erros da própria série.

Uma outra curiosidade relativamente à serie temporal, ela própria, por vezes, pode não ter os instantes de observação igualmente espaçados.

> **Exemplo:** Imaginemos que estamos perante um ensaio clínico em que nos interessa medir o crescimento de uma planta nas primeiras semanas e que depois, quando esta atingir uma determinada altura, só nos vai interessar tirar as medidas de 3 em 3 semanas.

Deste modo, ficamos com uma sucessão cronológica em que temos diferentes instantes de observação mas neste projecto só nos vamos focalizar em séries sem valores omissos.

Decomposição da Série

Uma sucessão cronológica decompõe-se normalmente pelos seguintes elementos:

Tendência - T_t : Indica a variação "em média" ao longo do tempo, isto é, há uma mudança ao nível da série.

Sazonalidade - S_t : Refere-se a efeitos relacionados com as variações periódicas (diário, semanal, mensal, anual, etc.)

Ciclos - C_t : Reporta-se a variações na série que apesar de periódicas não têm qualquer periodicidade conhecida.

Componente Aleatória (Ruído) - ε_t : É tudo o que não se consegue definir ou modelar na série e tomam a designação de erros.

Uma vez que já sabemos como se decompõe usualmente uma série temporal, agora já estamos nas condições de especificar a equação geral destas séries:

Modelo Aditivo:

$$X_t = T_t + S_t + C_t + \varepsilon_t$$

Aplica-se caso a variação periódica aumente com a média da série.

Modelo Multiplicativo:

$$X_t = T_t \times S_t \times C_t \times \varepsilon_t$$

Utiliza-se caso o efeito sazonal seja directamente proporcional ao nível médio da série.

Modelo Misto:

$$X_t = (T_t + C_t) \times S_t + \varepsilon_t \quad ou \quad X_t = T_t \times S_t \times C_t + \varepsilon_t$$

Modelos

Nomeadamente ao que se refere que tipos de modelos a utilizar no estudo das séries temporais, estes podem-se classificar em dois tipos:

Modelos Paramétricos: Tem um número finito de parâmetros.

Esta classe diz respeito essencialmente à análise no domínio do tempo e esta última atribui um papel importante às funções de autocovariância e autocorrelação. Dos modelos mais utilizados encontram-se os modelos lineares, nomeadamente os autoregressivos e de médias móveis (ARMA), os autoregressivos integrados e de médias móveis (ARIMA) e modelos não lineares (ARCH / GARCH).

Modelos Não-Paramétricos: Tem um número infinito de parâmetros.

Este tipo de modelos incide mais sobre a análise no domínio da frequência e um dos casos muito utilizados na prática designa-se por análise espectral.
A vantagem deste tipo de modelo comparativamente à outra classe de modelos considerada é que aqui não há problema quanto à correlação serial dos dados, pois na análise espectral as componentes são ortogonais.

Assim, relativamente a este trabalho, a modelação de séries temporais irá incidir mais sobre os modelos paramétricos estocásticos discretos.

A seguir, passo a enumerar os principais objectivos que se pretende ao realizar uma análise a uma série temporal:

i) Entender o mecanismo gerador da série;
ii) Descrever o comportamento da série (tendências, variações sazonais, valores atípicos, ciclos, etc.)
iii) Encontrar periodicidades (através do uso da análise espectral)
iv) Tentar obter explicações para determinadas perturbações na série (através do uso da análise de intervenção ou da inclusão de variáveis explicativas)
v) Tentar descrever a trajectória da série.
vi) Predizer os valores futuros da série para curto, médio ou longo prazo e deste modo poder tomar decisões antecipadamente.

Assim, para o nosso estudo consideremos que temos uma série temporal univariada discreta que designaremos por X_t e que as observações são equiespaçadas no tempo de tal modo que representamos elas por $X_1, X_2, \dots, X_N$.

Em particular, há uma classe de processos que nos vão interessar, os chamados processos estacionários.

O estudo deste tipo de processos pode-se fazer no domínio do tempo ou no domínio da frequência. No caso do domínio da frequência, será importante o uso do periodograma e da densidade espectral que veremos mais à frente; já no caso do domínio do tempo as funções de autocovariância e autocorrelação são as que se destacam mais nesta área e abordaremos estes conceitos já a seguir.

Um processo X_t diz-se estacionário quando o comportamento da série não se altera com o passar do tempo, ou seja, as características de X_{t+k}, $\forall k \in \mathbb{Z}$ são as mesmas que X_t.

Estes processos caracterizam somente pelos momentos de primeira e de segunda ordem, sendo eles:

Função Média:

$$E(X_t) = \mu$$

Função de Autocovariância (FACV):

$\gamma_k = \gamma(t, t+k) = E\{[X_t - E(X_t)][X_{t+k} - E(X_{t+k})]\} = E\{(X_t - \mu)(X_{t+k} - \mu)\} = E(X_t X_{t+k}) - E(X_t)E(X_{t+k}) = E(X_t X_{t+k}) - \mu^2$, $\forall k \in \mathbb{Z}$

Caso se tenha $k = 0$, temos que:

$$\gamma_0 = \gamma(t, t) = E(X_t^2) - E(X_t)^2 = Var(X_t)$$

onde $Var(X_t)$ é a **Função Variância** do processo X_t e denota-se usualmente por uma constante, σ^2.

A função de Autocovariância mede o grau de intensidade com que covariam os pares de valores do processo separados por um intervalo, que designamos por "lag" de amplitude k. Assim, esta função fornece a forma de dependência temporal do processo X_t mas como depende de X, ela não representa a força dessa dependência e por isso é comum ser substituída pela:

Função de Autocorrelação (FAC):

$$\rho_k = \frac{\gamma_k}{\gamma_0}$$

Esta função mede a correlação entre as variáveis X_t e X_{t+k} e tem em conta o efeito das variáveis intermédias $X_{t+1}, X_{t+2}, \dots, X_{t+k-1}$.

Função de Autocorrelação Parcial (FACP):

Ao contrário da função anterior, esta elimina o efeito das variáveis intermédias.

Para chegarmos à sua expressão, uma das maneiras é supormos que ajustamos uma regressão linear múltipla de X_{t+k} sobre $X_{t+k-1}, X_{t+k-2}, \dots, X_{t+1}, X_t$, isto é:

$$X_{t+k} = \phi_{k1}X_{t+k-1} + \phi_{k2}X_{t+k-2} + \ \dots + \phi_{kk}X_t + \ \varepsilon_{t+k}$$

onde ϕ_{kj} , $j = 1,2, \dots, k$ são os coeficientes de regressão e ε_{t+k} é o erro não correlacionado com X_{t+k-j} para $j \geq 1$. Multiplicando ambos os membros da expressão anterior por X_{t+k-j} , $j = 1,2, \dots, k$, tomando os valores esperados e dividindo por γ_0 , obtemos o sistema seguinte:

$$\rho_j = \phi_{k1}\rho_{j-1} + \phi_{k2}\rho_{j-2} + \ \dots + \phi_{kk}\rho_{j-k} \ , \qquad j = 1,2, \dots, k$$

e ao resolvemos em ordem aos coeficientes ϕ_{kj} , $j = 1,2, \dots, k$ pela regra de Cramer, obtemos a expressão da função de autocorrelação parcial que é dada por:

$$\phi_{kk} = \frac{\begin{vmatrix} 1 & \rho_1 & \dots & \rho_{k-2} & \rho_1 \\ \rho_1 & 1 & \dots & \rho_{k-3} & \rho_2 \\ \dots & \dots & \dots & \dots & \dots \\ \rho_{k-1} & \rho_{k-2} & \dots & \rho_1 & \rho_k \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{k-2} \\ \dots & \dots & \dots & \dots & \dots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \dots & 1 \end{vmatrix}}$$

Para $k = 1,2,3$ temos que:

$$\phi_{11} = \rho_1$$

$$\phi_{22} = \frac{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & \rho_2 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}}$$

$$\phi_{33} = \frac{\begin{vmatrix} 1 & \rho_1 & \rho_1 \\ \rho_1 & 1 & \rho_2 \\ \rho_2 & \rho_1 & \rho_3 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{vmatrix}}$$

Estas duas últimas funções referidas, FAC e FACP, desempenham um papel extremamente fundamental visto que ambas as funções identificam univocamente qual o modelo a ser ajustado ao conjunto de dados dado.

Características da FACV e da FAC:

i) $\gamma_0 = \sigma^2 \;\; ; \;\; \rho_0 = 1$
ii) $-\gamma_0 \leq \gamma_k \leq \gamma_0 \;\; ; \;\; -1 \leq \rho_k \leq 1$
iii) $\gamma_k = \gamma_{-k} \;\; ; \;\; \rho_k = \rho_{-k}$ *(funções pares)*
iv) $São\ funções\ semidefinidas\ positivas$

Uma vez que muitas séries apresentam tendências, instabilidades e oscilações no tempo, podemos recorrer a transformações que estabilizam a média e/ou a variância da série, que desde modo convertem a série não estacionária em estacionária.

Não Estacionaridade na Variância

Para estabilizar a variância da série, normalmente costuma-se aplicar a Transformação de Box-Cox que representa-se por:

$$T(X_t) = X_t^{(\lambda)} = \begin{cases} \dfrac{X_t^{(\lambda)} - 1}{\lambda}, & \lambda > 0 \\ ln(X_t), & \lambda = 0 \end{cases}$$

Temos que ter em conta que esta transformação só se pode aplicar para valores das séries positivas ($X_t > 0$) e que deve ser sempre aplicada antes que se faça qualquer transformação na média da série.

Não Estacionaridade na Média

Para os casos em que as médias dos processos não são constantes, podemos realizar diferenciações na série de modo a torná-la estacionária.
Conhecemos dois tipos de diferenciação, sendo eles os seguintes:

i) **Operador Diferenciação Simples** de potência d:

$$\nabla^d = (1 - B)^d$$

que elimina a tendência da série , em que d é um número inteiro que diz respeito à ordem da diferenciação e B é um operador de atraso tal que $B^k X_t = X_{t-k}$, $k \in \mathbb{Z}^+$.

ii) **Operador Diferenciação Sazonal** de potência D:

$$\nabla_S^D = (1 - B^S)^D$$

que elimina os movimentos periódicos da série, em que D é um número inteiro que indica a ordem da diferenciação sazonal.

Usualmente, estes operadores devem ser usados com moderação porque ao efectuar-se cada diferenciação, está-se a perder informações significativas. Por norma não é necessário mais do que uma ou duas diferenciações para a série ficar estável em relação à média e/ou variância.

Modelos Lineares

Neste capítulo iremos enunciar os modelos que estão na base da construção do modelo final ajustado aos dados que se pretende obter.
Estes modelos estão divididos em duas categorias: estacionários e não estacionários. Comecemos por ver os processos estacionários e posteriormente os processos não estacionários.

Processos Estacionários

Ruído Branco

Um processo Z_t designa-se por ruído branco se:

$E(Z_t) = \mu_Z$ (normalmente denota-se por $\mu_Z = 0$)
$Var(Z_t) = \sigma_Z^2$

$$\gamma_k = \begin{cases} \sigma_Z^2 \ , \ k = 0 \\ 0 \ \ , \ k \pm 1, k \pm 2, \ldots \end{cases} \qquad \rho_k = \begin{cases} 1 \ , \ k = 0 \\ 0 \ , \ k \pm 1, k \pm 2, \ldots \end{cases}$$

Se as v.a.'s Z_t são não correlacionadas, isto é, $cov(Z_t, Z_s) = 0, t \neq s$, dizemos que Z_t é um processo de ruído branco discreto.

Se considerarmos que as v.a.'s Z_t forem independentes, elas também serão não correlacionadas. E portanto estamos perante uma sequência de v.a.'s independentes e identicamente distribuídas (i.i.d.) que têm a designação de um processo puramente aleatório.

Processo autoregressivo de ordem p – AR (p)

Um processo X_t diz-se autoregressivo de ordem p se satisfaz a seguinte equação:

$$X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} - \ \ldots \ - \ \phi_p X_{t-p} = \varepsilon_t \quad \Leftrightarrow \quad \phi_p(B) X_t = \varepsilon_t$$

onde $\{\varepsilon_t\} \sim RB(0, \sigma^2)$, $\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \ \ldots - \phi_p B^p$ é o polinómio autoregressivo de ordem p e tem-se que $|\phi_i| < 1 \ \ \forall i = 1,2, \ldots, p$.

Estes modelos para serem estacionários significam que as raízes de $\phi_p(B)$ têm que ficar fora do círculo unitário e podemos realçar o facto que estes processos são sempre invertíveis.

Processo de médias móveis de ordem q – MA (q)
Um processo X_t diz-se de médias móveis de ordem q se X_t for invertível e se satisfaz a seguinte equação:

$$X_t = \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \ \dots \ + \ \theta_q\varepsilon_{t-q} \quad \Leftrightarrow \quad X_t = \theta_q(B)\varepsilon_t$$

onde $\{\varepsilon_t\}\sim RB(0,\sigma^2)$, $\theta_q(B) = 1 + \theta_1 B + \theta_2 B^2 + \ \dots \ + \theta_q B^q$ é o polinómio de médias móveis de ordem q e tem-se que $|\theta_i| < 1 \ \ \forall i = 1,2,\dots,q$.

Para a condição de invertibilidade ser verificada basta que as raízes de $\theta_q(B)$ fiquem fora do círculo unitário. Também é importante referir que estes processos são sempre estacionários.

Processo autoregressivo sazonal de ordem P – AR(P)s
Um processo X_t diz-se autoregressivo sazonal de ordem P se X_t satisfaz a seguinte equação:

$$X_t - \Phi_1 X_{t-S} - \Phi_2 X_{t-2*S} - \ \dots \ - \ \Phi_p X_{t-P*S} = \varepsilon_t \quad \Leftrightarrow \quad \Phi_P(B^S)X_t = \varepsilon_t$$

onde $\{\varepsilon_t\}\sim RB(0,\sigma^2)$, $\Phi_P(B^S) = 1 - \Phi_1 B^S - \Phi_2 B^{2*S} - \ \dots - \Phi_P B^{P*S}$ é o polinómio autoregressivo sazonal de ordem P em B^Se tem-se que $|\Phi_i| < 1 \ \ \forall i = 1,2,\dots,P$.

Processo de médias móveis sazonal de ordem Q – MA(Q)s
Um processo X_t diz-se de médias móveis sazonal de ordem Q se X_t satisfaz a seguinte equação:

$$X_t = \varepsilon_t + \Theta_1\varepsilon_{t-S} + \Theta_2\varepsilon_{t-2*S} + \ \dots \ + \ \Theta_Q\varepsilon_{t-Q*S} \quad \Leftrightarrow \quad X_t = \Theta_Q(B^S)\varepsilon_t$$

onde $\{\varepsilon_t\}\sim RB(0,\sigma^2)$, $\Theta_Q(B^S) = 1 + \Theta_1 B^S + \Theta_2 B^{2*S} + \ \dots \ + \Theta_Q B^{Q*S}$ é o polinómio de médias móveis sazonal de ordem Q em B^S e tem-se que $|\Theta_i| < 1 \ \ \forall i = 1,2,\dots,Q$.

Processo autoregressivo e de médias móveis de ordem p+q – ARMA(p,q)
Um processo X_t diz-se autoregressivo e de médias móveis de ordem p+q se X_t satisfaz a seguinte equação:

$$X_t - \phi_1 X_{t-1} - \ \dots \ - \phi_p X_{t-p} = \varepsilon_t + \theta_1\varepsilon_{t-1} + \ \dots \ + \ \theta_q\varepsilon_{t-q}$$

$$\Leftrightarrow \phi_p(B)X_t = \theta_q(B)\varepsilon_t$$

onde $\{\varepsilon_t\}\sim RB(0,\sigma^2)$, $\phi_p(B)$ é o polinómio autoregressivo de ordem p, $\theta_q(B)$ é o polinómio de médias móveis de ordem q e tem-se que $|\phi_i| < 1 \ \ \forall i = 1,2,\dots,p$ e $|\theta_i| < 1 \ \ \forall i = 1,2,\dots,q$.

Caso este processo seja estacionário é necessário que as raízes de $\phi_p(B)$ estejam fora do círculo unitário e caso o processo seja invertível é preciso que as raízes de $\theta_q(B)$ estejam fora do círculo unitário.

Processo autoregressivo e de médias móveis sazonal de ordem P+Q – ARMA(P,Q)s

Um processo X_t diz-se autoregressivo e de médias móveis sazonal de ordem P+Q se X_t satisfaz a seguinte equação:

$$X_t - \Phi_1 X_{t-S} - \ \dots \ - \Phi_P X_{t-P*S} = \varepsilon_t + \Theta_1 \varepsilon_{t-S} + \ \dots \ + \Theta_Q \varepsilon_{t-Q*S}$$

$$\Leftrightarrow \Phi_P(B^S) X_t = \Theta_Q(B^S)\varepsilon_t$$

onde $\{\varepsilon_t\}\sim RB(0,\sigma^2)$, $\Phi_P(B^S)$ é o polinómio autoregressivo sazonal de ordem P em B^S , $\Theta_Q(B^S)$ é o polinómio de médias móveis sazonal de ordem Q em B^S e tem-se que $|\Phi_i| < 1 \ \ \forall i = 1,2,\dots,P$ e $|\Theta_i| < 1 \ \ \forall i = 1,2,\dots,Q$.

Modelo Multiplicativo

Processo autoregressivo e de médias móveis sazonal - SARMA(p,q)x(P,Q)s

Um processo X_t diz-se autoregressivo e de médias móveis sazonal se X_t satisfaz a seguinte equação:

$$\phi_p(B)\Phi_P(B^S)X_t = \theta_q(B)\Theta_Q(B^S)\varepsilon_t$$

onde $\{\varepsilon_t\}\sim RB(0,\sigma^2)$, $\phi_p(B)$ é o polinómio autoregressivo de ordem p, $\Phi_P(B^S)$ é o polinómio autoregressivo sazonal de ordem P em B^S , $\theta_q(B)$ o polinómio de médias móveis de ordem q, $\Theta_Q(B^S)$ é o polinómio de médias móveis sazonal de ordem Q em B^S e tem-se que $|\phi_i| < 1 \ \ \forall i = 1,2,\dots,p$, $|\Phi_i| < 1 \ \ \forall i = 1,2,\dots,P$, $|\theta_i| < 1 \ \ \forall i = 1,2,\dots,q$ e $|\Theta_i| < 1 \ \ \forall i = 1,2,\dots,Q$.

Processos Não Estacionários

Processo autoregressivo integrado de médias móveis – ARIMA (p,d,q)

Um processo X_t diz-se autoregressivo integrado de médias móveis se X_t satisfaz a seguinte equação:

$$\phi_p(B)(1-B)^d X_t = \theta_q(B)\varepsilon_t$$

onde $\{\varepsilon_t\}\sim RB(0,\sigma^2)$, $\phi_p(B)$ é o polinómio autoregressivo de ordem p, $\theta_q(B)$ é o polinómio de médias móveis de ordem q, d é a ordem da diferenciação simples e tem-se que $|\phi_i| < 1 \ \ \forall i = 1,2,\dots,p$, $|\theta_i| < 1 \ \ \forall i = 1,2,\dots,q$ e $d \in \mathbb{Z}^+$.

Processo autoregressivo integrado de médias móveis sazonal – SARIMA (p,d,q)x(P,D,Q)s

Um processo X_t diz-se autoregressivo integrado de médias móveis sazonal se X_t satisfaz a seguinte equação:

$$\phi_p(B)\Phi_{\mathrm{P}}(\mathrm{B}^{\mathrm{S}})(1-B)^d(1-B^S)^D X_t = \theta_q(B)\Theta_Q(B^S)\varepsilon_t$$

onde $\{\varepsilon_t\}\sim RB(0,\sigma^2)$, $\phi_p(B)$ é o polinómio autoregressivo de ordem p, $\theta_q(B)$ é o polinómio de médias móveis de ordem q, d é a ordem da diferenciação simples, $\Phi_P(B^S)$ é o polinómio auto regressivo sazonal de grau P em B^S, $\Theta_Q(B^S)$ é o polinómio de médias móveis sazonal de grau Q em B^S, D é a ordem da diferenciação sazonal e tem-se que $|\phi_i| < 1 \ \ \forall i = 1,2,\dots,p$, $|\Phi_i| < 1 \ \ \forall i = 1,2,\dots,P$, $|\theta_i| < 1 \ \ \forall i = 1,2,\dots,q$, $|\Theta_i| < 1 \ \ \forall i = 1,2,\dots,Q$, $d \in \mathbb{Z}^+$ e $D \in \mathbb{Z}^+$

Convém realçar que estes processos, quando se efectua uma ou duas diferenciações (não sendo necessário mais) sejam elas simples ou sazonais tornam-se processos estacionários pertencendo à classe dos modelos lineares ARMA.

Pressupostos do Modelo Linear

Resíduos:

- São variáveis aleatórias com média zero e de variância constante - hipótese de homocedasticidade;
- São independentes e não correlacionados;
- Seguem uma distribuição normal, isto é, $\hat{\varepsilon}_t\sim N(0,\sigma^2)$.

Quando estamos a visualizar os gráficos das funções de autocorrelação dos resíduos em valores absolutos e dos resíduos ao quadrado, caso se notem elevadas correlações significativas nesses gráficos, isso demonstra-nos que o modelo linear não é o mais indicado para descrever os nossos dados mas sim um modelo não linear.

Modelos Não Lineares

Por vezes determinadas séries apresentam uma variância não constante ao longo do tempo e portanto os modelos lineares não são os modelos mais indicados para a modelação desses dados. Por isso, tiveram que ser introduzidos alguns modelos que tivessem em conta essa condição. A essa classe dá-se o nome de Modelos Não Lineares e veremos já de seguida alguns exemplos, nomeadamente os modelos ARCH e GARCH dos quais fazem parte, cujo objectivo será modelar essa variância condicional.

Modelos ARCH

Esta classe de modelos foi introduzida por Engle em 1982 com o propósito de estimar a variância da série.
Os erros ε_t, não são correlacionados serialmente mas a variância condicional, por vezes chamada de volatilidade, depende dos erros passados que provêem de uma função quadrática.

Processo autoregressivo com heterocedasticidade condicional de ordem r – ARCH (r)

Um processo ε_t diz-se autoregressivo com heterocedasticidade condicional de ordem r se ε_t é dado pela seguinte expressão:

$$\varepsilon_t = a_t\sqrt{h_t} \;\; com \;\; h_t = \alpha_0 + \alpha_1\varepsilon_{t-1}^2 + \; \dots + \alpha_r\varepsilon_{t-r}^2$$

em que $\{a_t\}$ é uma sequência de variáveis aleatórias i.i.d. com média zero e variância um, em que se tem $\alpha_0 > 0$ e $\alpha_i \geq 0 \; \forall i = 1,2, \dots, r$.

Alguns autores sugerem na prática que $\{a_t\}$ tenha uma distribuição aproximadamente normal isto é, $\{a_t\} \sim N(0,1)$ ou em alternativa $\{a_t\}$ tenha uma distribuição t-student com v graus de liberdade e neste caso teremos que $a_t \sim t_v$.

Consideremos agora o caso particular em que r=1,

$$\varepsilon_t = a_t\sqrt{\alpha_0 + \alpha_1\varepsilon_{t-1}^2}$$

com $\alpha_0 > 0, \alpha_1 \geq 0$ e a_t é um processo de ruído branco.

De seguida, vamos ver algumas propriedades estatísticas deste modelo:

i) $E(\varepsilon_t) = 0$
ii) $Var(\varepsilon_t) = \alpha_0 + \alpha_1\varepsilon_{t-1}^2$

iii) Se ε_t for um processo estacionário de 2ª ordem, então $Var(\varepsilon_t) = \frac{\alpha_0}{1-\alpha_1}$

iv) $Cov(\varepsilon_{t+k}, \varepsilon_t) = 0 \ \ \forall k \geq 1$

Relativamente à curtose, este processo apresenta caudas mais pesadas do que uma distribuição normal, sendo que:

$K = 3\frac{1-\alpha_1^2}{1-3\alpha_1^2} > 3$ e diz-se que a função é leptocúrtica neste caso.

A título de curiosidade, estes modelos são muito utilizados no que diz respeito a dados económicos, uma vez que estes apresentam uma grande volatilidade em certos períodos de tempo. Como consequência há um reflexo nos dados, ao terem distribuições com caudas mais pesadas.

Modelos GARCH

Em 1986, Bollerslev fez uma generalização do modelo ARCH, tendo dado o nome GARCH ("ARCH generalizado") e deste modo, um modelo GARCH pode ser utilizado para descrever a volatilidade com menos parâmetros do que um modelo ARCH, visto ser mais parcimonioso.
A diferença para o modelo ARCH reside no facto em que no GARCH a variância condicional depende também da variância condicional passada.

Processo autoregressivo com heterocedasticidade condicional generalizado – GARCH (r,s)

Um processo ε_t diz-se autoregressivo com heterocedasticidade condicional generalizado se ε_t é dado pela seguinte expressão:

$$\varepsilon_t = a_t\sqrt{h_t} \ \ com \ \ h_t = \alpha_0 + \sum_{i=1}^{r} \alpha_i \, \varepsilon_{t-i}^2 + \sum_{j=1}^{s} \beta_j \, h_{t-j}$$

em que $\{a_t\}$ é uma sequência de variáveis aleatórias i.i.d. com média zero e variância um, em que se tem $\alpha_0 > 0$, $\alpha_i \geq 0 \ \ \forall i > 0$, $\beta_j \geq 0 \ \ \forall j > 0$, $\sum_{i=1}^{q}(\alpha_i + \beta_i) < 1$ e $q = \max(r, s)$.

Como no caso do modelo ARCH, usualmente supõem que $\{a_t\}$ tem uma distribuição normal ou então uma distribuição t-student.

Consideremos agora um caso muito utilizado na prática, um GARCH(1,1) que é dado pela seguinte expressão:

$$\varepsilon_t = a_t\sqrt{\alpha_0 + \alpha_1\varepsilon_{t-1}^2 + \beta_1 h_{t-1}}$$

Com $\alpha_0 > 0, \alpha_1 \geq 0, \beta_1 \geq 0, \alpha_1 + \beta_1 < 1$ e temos em particular que $E(\varepsilon_t^2) = \frac{\alpha_0}{1-(\alpha_1+\beta_1)}$.

Através do cálculo da curtose, reparamos que o modelo apresenta caudas mais pesadas do que o modelo gaussiano uma vez que apresenta um valor de K superior a 3 (como no caso do modelo ARCH) como poderemos ver a seguir:
$K = 3\frac{1-(\alpha_1+\beta_1)^2}{1-(\alpha_1+\beta_1)^2-2\alpha_1^2}$.

Construção do Modelo

Sabemos que um dos grandes objectivos na análise de séries temporais será a construção de um modelo óptimo que seja capaz de captar de forma satisfatória o comportamento da série.

Dito isto, um bom modelo deve ter em conta as seguintes características:

- Deve obedecer ao **princípio da parcimónia**, isto é, deve conter o menor número de parâmetros possível.
- Deve ser relativamente simples e flexível para poder se adaptar ao futuro (incerto) e facilitar o seu manuseamento.
- O modelo em causa é probabilístico, devido à incerteza do presente/futuro.
- Deve realizar boas previsões.

Existem três etapas que se devem adoptar para a escolha de um modelo óptimo segundo a Metodologia de Box-Jenkins:

Identificação $\Rightarrow$ Estimação $\Rightarrow$ Avaliação do diagnóstico

Identificação: Neste passo pretende estacionalizar a série, caso esta não seja. Realizamos esta operação através da aplicação das transformações que já foram discutidas anteriormente, de modo a poder-se estabilizar a variância, eliminar a tendência e/ou a sazonalidade da série caso exista. Depois de estacionarizada a série, o passo a seguir é ajustar um modelo ARMA aos dados e para isso necessitamos de analisar as FAC e FACP estimadas que são dadas pelo próprio software que estamos a trabalhar.

Estimação: Nesta segunda etapa, uma vez que já temos o modelo identificado, só nos falta procedermos à estimação dos parâmetros do referido modelo. Aqui pouco ou nada podemos fazer, uma vez que é o próprio pacote estatístico que faz o seu cálculo e que depois nos fornece os valores obtidos para tais parâmetros.

Avaliação do diagnóstico: Neste último passo, o modelo escolhido é analisado através de vários critérios que dão credibilidade à qualidade estatística do modelo seleccionado, entre eles encontram-se os critérios de informação, o princípio da parcimónia e a significância dos valores estimados para os parâmetros do modelo. Uma outra condição que também devemos ter em conta neste passo é sabermos se o modelo ajustado ao nosso conjunto de dados é adequado, para isso procedemos ao estudo dos correspondentes resíduos. Segundo os pressupostos conhecidos, estes devem-se comportar como um ruído branco e possuírem uma distribuição aproximadamente normal.

Identificação

Para identificarmos o modelo a escolher, teremos que olhar para os gráficos das FAC e FACP da série estacionarizada.

A seguir apresento uma pequena tabela que nos dá as características destas funções para conseguirmos identificar qual o tipo de modelo a seleccionar. Neste caso, iremos só nos concentrar nos modelos AR, MA e ARMA, sendo estes os mais fáceis de se identificar.

Modelo	FAC (ρ_k)	FACP (ϕ_{kk})
AR(p)	Decaimento exponencial e/ou sinusoidal para zero	É nula para k>p
MA(q)	É nula para k>q	Decaimento exponencial e/ou sinusoidal para zero
ARMA(p,q)	Decaimento exponencial e/ou sinusoidal para zero	Decaimento exponencial e/ou sinusoidal para zero

Uma vez que nos nossos modelos (que iremos ver mais à frente na parte prática) são essencialmente modelos multiplicativos, a sua identificação relativamente aos parâmetros faz-se da seguinte maneira: a componente não sazonal é modelada através da observação dos primeiros valores das FAC e FACP estimadas e a componente sazonal pelo comportamento destas funções sobre os múltiplos de S, ou seja, podemos tratar as FAC e FACP estimadas como sendo compostas por duas partes distintas: uma nos "lags" k=1,2,3… e outra nos "lags" k=S, 2S, 3S,....A primeira parte ajuda a identificar os parâmetros p e q e a segunda parte ajuda a identificar os parâmetros P e Q.

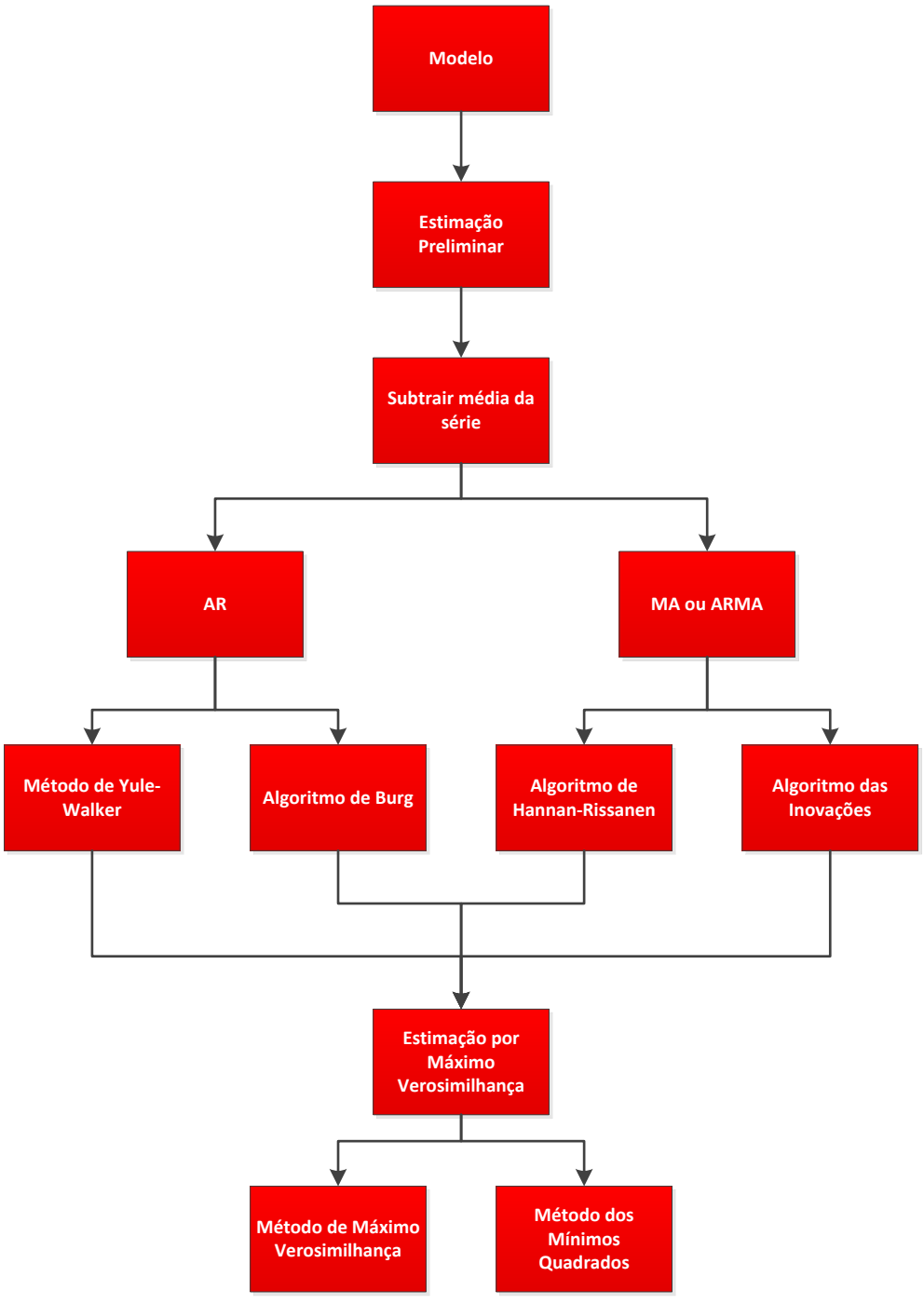
Para mais informações referentes a esta parte, aconselharia o leitor a consultar os livros que se encontram na bibliografia, nomeadamente o primeiro e o terceiro livro mencionados.

Estimação

Neste capítulo não irei dar muito enfâse nem em entrar em grandes desenvolvidos relativamente à estimação dos modelos, uma vez que esta é uma parte muito teórica e ligeiramente complexa e que vem logo incluída dentro dos packages estatísticos.

Apresento a seguir o diagrama aonde aborda todas as etapas da estimação dos modelos.

Estimação dos Modelos



Para mais informações acerca deste capítulo, o leitor por favor consulte algum dos três primeiros livros referenciados na bibliografia.

Critérios de Informação

Perante uma série temporal, podem existir vários modelos que se ajustam à série em causa de uma maneira aceitável. A questão que se coloca é: De entre dos vários modelos propostos, qual se ajustará melhor? Para responder a esta questão, foram introduzidos os chamados Critérios de Informação para medir a qualidade de ajustamento entre os modelos.

Estes critérios têm por ideia base em comparar modelos construídos com base na maximização do logaritmo da função de verosimilhança, penalizando os modelos com mais parâmetros. Assim, quanto menor for o valor do critério de informação de um modelo, melhor será o modelo.

Os critérios de informação mais conhecidos são:

Critério de Informação Akaike (AIC) que foi proposto por Akaike (1974) que é dado pela seguinte expressão:

$$AIC = -2\ln(L) + 2p$$

em que L representa a função de máxima verosimilhança e p é o número de parâmetros a serem estimados do modelo respectivo.

Critério de Informação Akaike Corrigido (AICC) que foi proposto por Bozdogan (1987) que se considera ser uma versão melhorada relativamente ao critério anterior.

$$AICC = -2\ln(L) + 2p + 2\frac{p(p+1)}{n-p-1}$$

Critério de Informação Bayesiano (BIC) que foi desenvolvido por Schwarz (1978) e é semelhante ao AIC, mas coloca uma penalização maior pela inclusão de coeficientes adicionais a serem estimados:

$$BIC = -2\ln(L) + 2p\ ln(n)$$

em que n representa o número total de observações do modelo que foi ajustado.

Convém referir como nota que só se pode comparar modelos se forem aplicadas as mesmas transformações em eles próprios.

Previsão

A previsão dos valores futuros de uma série temporal representa um dos grandes objectivos da sua análise, sendo que o melhor critério para escolher um modelo de previsão tem que ter por base a sua capacidade preditiva, ou seja, quão perto estão as previsões dos valores posteriormente observados.

Suponha que observamos uma série até ao instante t e queremos prever o valor da série para o instante $t+h$.
Denotemos por $\hat{X}_t(h)$ os valores a prever da série e por X_{t+h} os valores reais da sucessão, em que t é a origem da previsão e h é o horizonte de previsão. Dito isto, $\hat{X}_t(h)$ é a previsão de X_{t+h}.
Sabemos que $\hat{X}_t(h)$ é uma variável aleatória conhecida que depende apenas da história do processo até ao instante t.
Representamos por $e_t(h)$ o erro de previsão que é dado pela diferença entre X_{t+h} e $\hat{X}_t(h)$.

Sabemos que ao fazer previsões futuras de uma série temporal, elas não são certas porque o futuro envolve incertezas e por isso devemos contar com erros que se cometem ao fazer este tipo de análises, sendo que o nosso objectivo é precisamente de minimizá-los ao máximo possível e por isso foram introduzidas várias medidas para esse propósito.

Medidas de Desempenho

Servem para medir a precisão e a eficácia dos resultados obtidos através dos diferentes métodos de previsão.
Baseiam-se no cálculo do erro médio de previsão de cada um dos períodos, independentemente do sentido (positivo ou negativo) da flutuação.

As medidas mais utilizadas na prática são:

Erro Quadrático Médio:

$$\frac{1}{n}\sum_{t=1}^{n} e_t^2\,(h)$$

Erro Absoluto Médio:

$$\frac{1}{n}\sum_{t=1}^{n} |e_t(h)|$$

Raíz do Erro Quadrático Médio:

$$\sqrt{\frac{1}{n}\sum_{t=1}^{n} e_t^2\,(h)}$$

Tendo em conta as três medidas anteriores, quanto menor for os valores destas, melhor é considerado o método de previsão.

Erro Percentual Total:

$$\frac{\sum_{t=1}^{n} e_t(h)}{\sum_{t=1}^{n} X_{t+h}} \times 100$$

Erro Médio Percentual:

$$\frac{1}{n}\sum_{t=1}^{n} \frac{e_t(h)}{X_{t+h}} \times 100$$

Erro Médio Percentual Absoluto:

$$\frac{1}{n}\sum_{t=1}^{n} \left|\frac{e_t(h)}{X_{t+h}}\right| \times 100$$

Ao aplicar as expressões anteriores, os resultados indicam a percentagem de erro de previsão para o método em análise. Quanto mais afastado de zero, maior é o erro de previsão. Quanto mais próximo de zero, menor o erro e portanto mais preciso e eficaz é o método.

Presença de Outliers

Quando estamos a analisar um conjunto de dados, notamos que por vezes alguns valores parecem não estar em conformidade com as restantes observações. A estas observações dá-se o nome de outliers, valores atípicos ou observações discordantes.

Neste trabalho utilizou-se o gráfico do Box-Plot para avaliar a existência de outliers recorrendo ao uso da ferramenta "R". No enquanto, existem outros métodos para a sua identificação, nomeadamente os testes de Dixon, Grubbs, Z-scores, os chamados modelos de discordância e pelo gráfico QQ-Plot.

Dependendo dos conjuntos de dados em causa, em alguns deles foram detectados vários outliers. O que achei importante fazer foi verificar se esses valores foram de facto verdadeiros ou são valores incorrectos que estão contidos na base de dados. Para confirmar tal facto, fui ver os relatórios diários que se fazem diariamente para cada um dos conjuntos de dados e pelo o que tive acesso, em grande parte das situações referem-se a problemas que aconteceram nesses dias sendo de facto reais e por isso preferi não retirar esses valores dos meus conjuntos de dados, apesar de saber que existe uma grande possibilidade de o modelo ajustado a esses dados não "capturar" esses valores.

Deve-se realçar o facto que os outliers podem ter influência nas FAC e FACP amostrais, no valor dos parâmetros estimados do modelo e, consequentemente, no modelo obtido e nas suas previsões futuras, e por isso torna-se muito essencial a sua análise.

Na prática, para eliminar estes outliers alguns autores sugerem substituir essas observações atípicas pela média ou mediana do referido conjunto. Tentei aplicar este conceito a alguns dos meus dados mas os resultados que obtive não foram satisfatórios.

A título de curiosidade para o leitor, Fox (1972) introduziu dois modelos estatísticos usados para a detecção destes valores:

Aditivos: Este tipo de outlier pode ser considerado como um erro grosseiro de medição e que afecta uma única observação.

Inovadores: Nesta classe de outliers verifica-se um "choque" num determinado período da série afectando as observações seguintes.

Testes para a Autocorrelação

Depois de encontrado o modelo estimado, a FAC dos resíduos dever-se-á comportar como a FAC de um processo de ruído branco, isto é, os resíduos devem ser não correlacionados, independentes e identicamente distribuídos.
Como veremos a seguir, foram introduzidos alguns testes de forma a testar esta veracidade.

Teste de Box-Pierce

Em 1970 Box e Pierce formularam um teste para testar se os primeiros k lags da FAC são iguais a zero, isto é, se os resíduos são não correlacionados.

Hipótese nula e hipótese alternativa:

$$H_0: \hat{\rho}_1(\varepsilon) = \hat{\rho}_2(\varepsilon) = \ \dots \ = \hat{\rho}_k(\varepsilon) = 0 \quad vs \quad H_1: \exists j \ tal \ que \ \hat{\rho}_j(\varepsilon) \neq 0$$

Estatística de teste:

$$Q_{BP} = N \sum_{i=1}^{k} \hat{\rho}_i^2(\varepsilon) \ \approx \ \chi_k^2$$

onde N é o número total de observações, $\hat{\rho}_i$ é a autocorrelação dos resíduos no lag i e k é o número de lags para testar.

Sob a validade da hipótese nula pode-se provar que Q_{BP} tem uma distribuição χ^2 com k graus de liberdade e assim, rejeitamos a hipótese nula para um nível de significância α quando $Q_{BP} > \chi^2_{1-\alpha\,,k}$.

Uns anos mais tarde, Ljung e Box (1978) sugeriram uma versão melhorada do teste anterior que é descrita a seguir:

Teste de Ljung-Box

Estatística de teste:

$$Q_{LB} = N(N+2) \sum_{i=1}^{k} \frac{\hat{\rho}_i^2(\varepsilon)}{N-i} \ \approx \ \chi_k^2$$

Que converge também mas muito mais rapidamente que o teste anterior, para uma distribuição χ^2 com k graus de liberdade e rejeitamos a hipótese nula do mesmo modo que o teste descrito anteriormente.

Testes para a Normalidade

Para verificar se um certo conjunto de dados tem uma determinada distribuição, podemos recorrer à construção de um histograma, de um gráfico QQ-Plot ou então recorrer a testes.

Uma vez que os resíduos devem possuir uma distribuição normal, deveremos testar com base nos testes descritos a seguir se tal requisito se verifica.

Teste Jarque-Bera

Este teste foi introduzido por Jarque e Bera (1981) e é baseado nas diferenças entre as medidas de assimetria e curtose da distribuição dos nossos dados em estudo relativamente à distribuição Normal.

Hipótese nula e hipótese alternativa:

$$H_0: Os\ dados\ seguem\ uma\ distribuição\ Normal$$
$$vs$$
$$H_1: Os\ dados\ não\ seguem\ uma\ distribuição\ Normal$$

Estatística de teste:

$$Q_{JB} = \frac{N}{6}\left(S^2 + \frac{1}{4}(K-3)^2\right) \approx \chi^2_2$$

em que N é o número total de observações, S é o coeficiente de assimetria e K representa a medida de curtose.

Sob a validade da hipótese nula, podemos provar que Q_{JB} tem uma distribuição χ^2 com 2 graus de liberdade e rejeitamos a hipótese para um nível de significância α quando $Q_{JB} > \chi^2_{1-\alpha,2}$.

Teste de Shapiro-Wilk

Este teste foi proposto por Shapiro e Wilk (1965) com o propósito de testar se um determinado conjunto de dados segue uma distribuição normal.

Estatística de Teste:

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

em que $x_{(i)}$ são os valores da amostra ordenados e para se efectuar o cálculo da constante b deve-se ter em conta a seguinte expressão:

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

em que a_{N-i+1} são constantes geradas pelas médias, variâncias e covariâncias das estatísticas de ordem de uma amostra de tamanho N de uma distribuição Normal e encontram-se tabelados. Assim, rejeitamos a hipótese nula ao nível de significância de α se $Q_W < Q_{W\ crítico}$ em que o valor de $Q_{W\ crítico}$ encontra-se também tabelado.

Teste de Kolmogorov-Smirnov

Este teste observa a máxima diferença absoluta entre a função de distribuição assumida para os dados, no nosso caso a Normal, e a função de distribuição empírica dos dados. Como critério, comparamos esta diferença com um valor crítico, para um dado nível de significância.
Convém não esquecer que ao aplicar este tipo de teste, estamos a considerar que a população que queremos provar que seja Normal, os valores da média ($\bar{x}$) e desvio-padrão (s) da população em causa sejam conhecidos.

Assim, considera-se primeiramente para a estatística de teste a expressão:

$$Q_{KS} = sup|F(x) - F_n^*(x)|$$

em que $F(x)$ representa a função de distribuição do modelo assumido na hipótese H_0 e $F_n^*(x)$ representa a função de distribuição acumulada empírica observada. Mas como $F_n^*(x)$ é uma função em escada, com saltos nos pontos da amostra observada, então devemos considerar duas novas estatísticas dadas por:

$$Q_{KS}^{+} = \sup_{x_{(i)}}\left|F\left(x_{(i)}\right) - F_n^*(x_{(i)})\right| \quad \text{e} \quad Q_{KS}^{-} = \sup_{x_{(i)}}\left|F\left(x_{(i)}\right) - F_n^*(x_{(i-1)})\right|$$

Estas duas medem as distâncias (verticalmente) entre os gráficos das duas funções, teórica e empírica, nos pontos $x_{(i-1)}$ e $x_{(i)}$.

Deste modo, a nova estatística de teste será dada por:

$$Q_{KS} = \max(Q_{KS}^{+}, Q_{KS}^{-})$$

Para o cálculo do valor observado da estatística de teste, passo a enumerar os passos que se deve seguir para achar o seu respectivo valor:

Primeiro devemos ordenar os nossos dados que estamos a considerar ($x_{(i)}$), a seguir devemos obter os valores de $F_n^*(x_{(i)})$ que são dados por $\frac{i}{N}$, $\forall i = 1, \dots, N$. Depois devemos transformar os nossos dados para ficarem padronizados através da expressão $Z_{(i)} = \frac{x_{(i)} - \bar{x}}{s}$, onde $\bar{x}$ é a média e s é o desvio padrão dos dados.
Por fim, o valor de $F\left(x_{(i)}\right)$ é encontrado na tabela da distribuição normal padrão tendo em consideração os valores padronizados.

Deste modo, agora estamos nas condições de aplicar o teste. Se Q_{KS} é maior que o valor crítico (que se encontra tabelado), então rejeitamos a hipótese de normalidade dos dados com (1-α)100% de confiança. Caso contrário, não rejeitamos a hipótese de normalidade.

Testes para efeitos ARCH

Uma das condições que os resíduos devem satisfazer é que eles não devem apresentar nenhum tipo de estrutura na sua composição, caso indiquem significa que o modelo ajustado aos dados não conseguiu captar toda a informação disponível na natureza nos dados.
Assim, queremos testar se os resíduos são de facto independentes e identicamente distribuídos mas para que tal aconteça, teremos de testar se há a existência de feitos ARCH no modelo ou não.

Teste de McLeod-Li

Este teste foi proposto por McLeod e Li (1983) para verificar se existe efeitos ARCH no modelo ajustado aos dados.

Hipótese nula e hipótese alternativa:

$$H_0: O\ modelo\ \text{é}\ linear$$
$$vs$$
$$H_1: O\ modelo\ \text{é}\ \text{não} - linear\ do\ tipo\ ARCH$$

Estatística de teste:

$$Q_{ML} = N(N+2)\sum_{i=1}^{k}\frac{\hat{\rho}_i^2(\varepsilon^2)}{N-i} \approx \chi_k^2$$

em que N representa a dimensão da nossa amostra e $\hat{\rho}_i^2$ representa a autocorrelação ao quadrado que diz respeito ao quadrado dos resíduos no lag i. Esta estatística de teste tem assimptoticamente uma distribuição qui-quadrado com k graus de liberdade e rejeitamos a hipótese para um nível de significância de α quando $Q_{ML} > \chi^2_{1-\alpha,k}$.

Análise no domínio da frequência

Embora este trabalho incida mais sobre a análise do domínio do tempo relativamente a sucessões cronológicas, não quis deixar de referir algumas poucas curiosidades sobre a temática no domínio da frequência. Desta maneira, podemos complementar a nossa análise com base nestes dois métodos de estudos.

Periodograma

O periodograma é tido com a estimativa da função de densidade espectral e desempenha um papel muito relevante no que diz respeito à detecção de ciclos numa determinada série temporal. No eixo das abcissas do gráfico considera-se as frequências de Fourier, ω_j, sendo que estas variam entre $[-\pi, \pi]$.

A expressão do periodograma é dada por:

$$I(\omega_j) = \frac{N}{2}\left[a^2(\omega_j) + b^2(\omega_j)\right]$$

em que $a(\omega_j) = \frac{2}{N}\sum_{t=1}^{N} X_t \cos(\omega_j\, t)$, $b(\omega_j) = \frac{2}{N}\sum_{t=1}^{N} X_t \sin(\omega_j t)$ e $\omega_j = \frac{2\pi j}{N}$, $j = 0,1, \dots, \left[\frac{N}{2}\right]$.

De uma forma mais abreviada, a expressão descrita acima é equivalente à seguinte:

$$I(\omega_j) = \frac{2}{N}\left|\sum_{t=1}^{N} X_t e^{-it\omega_j}\right|^2$$

Testes para a Periodicidade

Teste de Fisher

Este teste foi apresentado por Fisher (1929) com a finalidade de se detectar periodicidades numa série temporal com base nas ordenadas do periodograma.

Hipótese nula e hipótese alternativa:

$$H_0: Não\ existe\ periodicidades\ na\ série$$
$$vs$$
$$H_1: Existe\ periodicidades\ na\ série$$

Estatística de teste:

$$Q_F = \frac{\max I(\omega_j)}{\sum_{j=1}^{\left[\frac{N}{2}\right]} I(\omega_j)}$$

Sabe-se que para N ímpar a distribuição exacta de Q_F, sob H_0, é dada por $P(Q_F > b) = n(1-b)^{n-1} - \binom{n}{2}(1-2b)^{n-1} + \ \ldots + (-1)^x \binom{n}{x}(1-xb)^{n-1}$, em que $n = \left[\frac{N}{2}\right]$ e x é o maior inteiro menor que $\frac{1}{b}$, ou seja, $x = \left[\frac{1}{b}\right]$.

Assim, se o valor observado da estatística de teste, Q_F, for maior que $b(\alpha)$ então rejeitamos H_0, e portanto podemos afirmar que a série em causa apresenta uma periodicidade igual a $\frac{1}{\omega^*}$ em que ω^* representa a frequência em ciclos que corresponde ao máximo de $I(\omega_j)$.

Análise de Intervenção

Muitas vezes, quando estamos a analisar cuidadosamente a série em causa, notamos que ela mesma não se comporta sempre da mesma maneira, parece que o nível da série muda ou então algum factor influencia ou afecta ela num determinado instante de tempo. À ocorrência deste fenómeno dá-se o nome de intervenção.

Este tipo de interrupção na série, por vezes pode ser temporária ou então permanente e um dos seus principais objectivos é poder avaliar qual o seu verdadeiro impacto no comportamento da série em estudo.

Box e Tiao (1975) desenvolveram estes modelos de intervenção que são um caso particular dos modelos de função de transferência de Box e Jenkins (1970). Consideramos que a variável dependente, Y_t , é tida como estocástica e a variável independente, X_t , é tida como determinística sendo que esta toma valores zero ou um consoante a ausência ou presença de intervenção.

Assim, vamos considerar as variáveis indicadoras que indicam a sinalização da intervenção no instante de tempo T:

Função de degrau

$$X_t = S_t = \begin{cases} 0, & t < T \\ 1, & t \geq T \end{cases}$$

Este tipo de função é aplicado quando a intervenção é do tipo permanente depois do instante de tempo T.

Função de impulso

$$X_t = P_t = \begin{cases} 0, & t = T \\ 1, & t \neq T \end{cases}$$

Esta função utiliza-se quando se pensa que só afectará Y_t durante o instante de tempo T, voltando a série ao nível inicial imediatamente depois de $t = T$.

Perante uma série temporal, quando vamos estudá-la para obtermos o modelo que se ajusta melhor relativamente aos dados em causa, vamos ter que introduzir uma função na expressão do modelo para caracterizar o acontecimento de uma intervenção. A estas funções dá-se o nome de *Funções de Transferência.*

De seguida, apresentamos o modelo geral que podemos considerar com k intervenções:

$$Y_t = C + \sum_{j=1}^{k} v_j\,(B) X_t + N_t$$

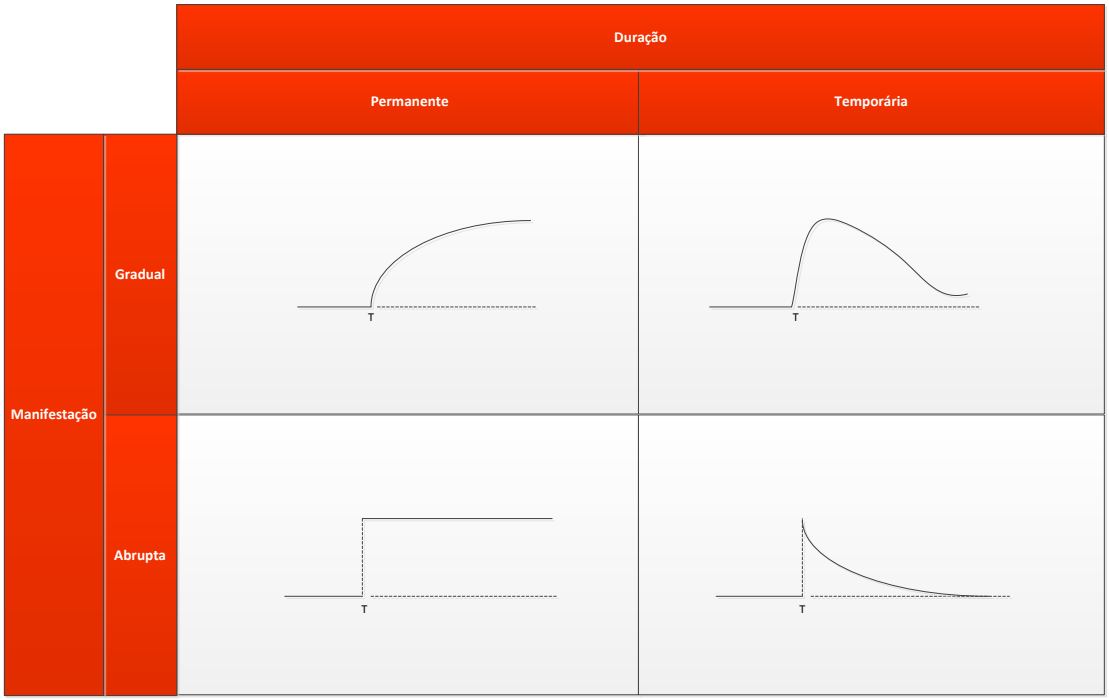
em que C é uma constante (caso seja preciso diferenciar a série, este termo desaparece), X_t é a variável de intervenção do tipo i) ou ii), $v_j(B)\ \ j = 1, \dots, k$ são as funções de transferência que têm a forma $\frac{w_j(B)B^{b_j}}{\delta_j(B)}$ onde $w_j(B) = w_{j,0} + w_{j,1}B + \ \dots \ + \ w_{j,s}B^s$ e $\delta_j(B) = 1 - \delta_{j,1}B - \ \dots - \delta_{j,r}B^r$ são polinómios em B, b_j é um operador de recuo no tempo para o início do efeito da j-ésima intervenção e, por fim, N_t é a série temporal sem qualquer tipo de intervenção e representa-se na forma de um modelo SARIMA, isto é,

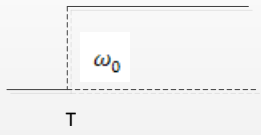
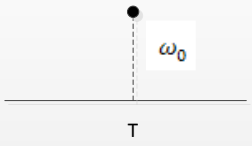
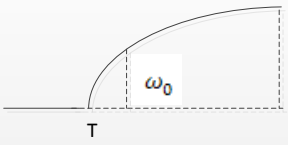
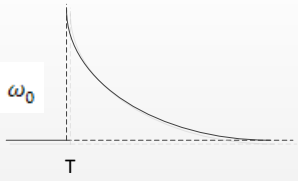
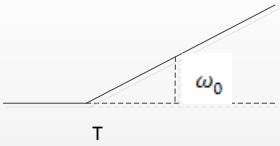
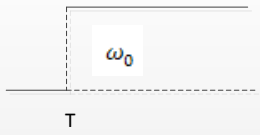
$$N_t = \frac{\theta_q(B)\Theta_Q(B)}{\phi_p(B)\Phi_P(B^S)(1-B)^d(1-B^S)^D}\varepsilon_t$$

Portanto temos que ter algum cuidado se estamos mesmo perante um caso de intervenção porque, por vezes, a série pode ter só de facto tendência e assim podemos tirar falsas conclusões. Também devemos ter em conta com outros dois tipos de factores: a sazonalidade e o erro aleatório.

Efeitos de intervenção

De seguida apresentaremos num quadro resumo os tipos mais comuns de manifestações, de durações e de funções de transferência que uma intervenção pode ter.



v(B)	Permanente	Temporária
ω_0		
$\frac{\omega_0}{1-\delta B}$ $\lvert\delta\rvert < 1$		
$\frac{\omega_0}{1-B}$		

Por simplicidade, podemos considerar o caso de uma única função de transferência, que é dada por:

$$Y_t = C + v(B)X_t + N_t$$

onde C é uma constante, $v(B) = \frac{w(B)}{\delta(B)}$ em que $w(B) = w_0 + w_1B + w_2B^2 + \ \ldots + w_rB^r$ e $\delta(B) = 1 - \delta_1B - \delta_2B^2 - \ \ldots - \delta_sB^s$.

Sendo que uma das expressões mais utilizadas é dada por:

$$Y_t = C + \frac{w_0}{1-\delta_1B}X_t + N_t$$

Que se pode escrever de forma equivalente por:

$$Y_t = C + (w_0X_t + w_0\delta_1X_{t-1} + w_0\delta_1^2X_{t-2} + w_0\delta_1^3X_{t-3} + \ \ldots\ )N_t$$

Para terminar, podemos ainda referir que é possível ter um modelo de intervenção com uma combinação de funções de impulso e de degrau, como poderemos ver no exemplo apresentado a seguir:

$$Y_t = C + \frac{w_0}{1 - \delta B} P_t + w_1 S_t + N_t$$

Métodos de Previsão

Neste capítulo irei enunciar alguns métodos de previsão mais conhecidos actualmente. Algumas destas aplicações só devem ser usadas dependendo da natureza dos dados em questão como veremos de seguida.

Séries de nível constante
Como o próprio nome indica, caso o conjunto de dados apresente o nível da série constante, deveremos utilizar um dos três processos que se seguem:

Naive 1
Este é um dos métodos preditivos mais simples que existe. Utiliza o valor do período corrente como previsão para o período seguinte, isto é:

$$\hat{X}_t = X_{t-1}$$

onde $t = 1, \dots, N$, em que N é a dimensão total da amostra.

Cálculo da previsão

$$\hat{X}_t(h) = \hat{X}_t \,, \forall h > 0$$

Naive 2
Este método é uma versão mais avançada do método anterior. Considera que a próxima previsão é baseada na diferença do valor do dobro do período $t-1$ com o valor do período $t-2$, ou seja:

$$\hat{X}_t = X_{t-1} + X_{t-1} - X_{t-2}$$

onde $t = 1, \dots, N$.

Cálculo da previsão

$$\hat{X}_t(h) = \hat{X}_t \,, \forall h > 0$$

Principais Vantagens Naive

- Método simples e de rápida execução;
- Não necessita de um software para a usa aplicação;
- Eficaz quando se tem o número de observações pequeno;
- Implementação de custo reduzido.

Principais Desvantagens Naive

- Assume-se que não há alterações na série ao longo do tempo.

Médias Móveis Simples (MMS)
Este método consiste em calcular a média das últimas r observações da série:

$$\hat{X}_t = \frac{X_t + X_{t-1} + \ \dots \ + X_{t-r+1}}{r}$$

onde $t = 1, \dots, N$ e $r \in \mathbb{Z}^+$.

Cálculo da previsão

$$\hat{X}_t(h) = \hat{X}_t \, , \forall h > 0$$

Principais Vantagens

- Método flexível e de simples aplicação;
- Pode-se aplicar quando se tem o número de observações pequeno.

Principais Desvantagens

- Só deve ser usado para séries estacionárias;
- São dados pesos iguais a todas as observações incluídas no cálculo da média;
- Dificuldade em determinar o valor de r.

Alisamento Exponencial Simples (AES)
Este método não é mais do que uma média móvel ponderada de todas as observações passadas da série:

$$\hat{X}_t = \alpha X_t + \alpha(1-\alpha)X_{t-1} + \alpha(1-\alpha)^2 X_{t-2} + \ \dots \ \Leftrightarrow \ \hat{X}_t = \alpha X_t + (1-\alpha)\hat{X}_{t-1}$$

onde $t = 1, \dots, N$ e têm-se que $\hat{X}_0 = X_1$ e $\alpha \in [0,1]$.

Convém realçar o facto de que, quanto maior for a constante de suavização, designada por α, maior é o peso dado às observações mais recentes.

Cálculo da previsão

$$\hat{X}_t(h) = \hat{X}_t \, , \forall h > 0$$

Principais Vantagens

- Modelo flexível e de fácil compreensão;
- Aplicação não dispendiosa;
- Quando $\alpha = \frac{2}{r-1}$ fornece previsões semelhantes ao método anterior (MMS) com parâmetro igual a r.

Principais Desvantagens

- Dificuldade em determinar o valor da constante de suavização.

Séries que apresentam tendência

Caso os dados considerados apresentem uma tendência crescente ou decrescente ao nível da série, deveremos utilizar o método que é apresentado a seguir:

Holt-Winters (HW)

Este método é uma extensão do método anterior, só que entra em linha de conta com a tendência da série.

$$\hat{X}_t = \alpha X_t + (1-\alpha)\left(\hat{X}_{t-1} + \hat{T}_{t-1}\right)$$
$$\hat{T}_t = \beta\left(\hat{X}_t - \hat{X}_{t-1}\right) + (1-\beta)\hat{T}_{t-1}$$

onde $t = 2, \dots, N$, $\hat{T}_2 = X_2 - X_1$, $\hat{X}_2 = X_2$, $\alpha \in [0,1]$ e $\beta \in [0,1]$.

Cálculo da previsão

$$\hat{X}_t(h) = \hat{X}_t + h\hat{T}_t \,, \forall h > 0$$

Principais Vantagens

- São semelhantes ao método referido anteriormente.

Principais Desvantagens

- Dificuldade em determinar os valores das constantes de suavização, α e β.

Séries que apresentam tendência e sazonalidade

Caso a nossa série apresente sazonalidade, o método proposto a seguir é o mais indicado para realizar previsões ao nível da série.

Holt-Winters Sazonal (HWS)

Este método utiliza-se quando se pretende modelar o nível, a tendência e a sazonalidade de uma série.

Como os padrões sazonais podem ser considerados aditivos ou multiplicativos, então utilizaremos as duas variantes deste método.

Sazonalidade Aditiva

Neste método o factor sazonal S_t e a tendência T_t são aditivos, isto é:

$$X_t = \mu_t + T_t + S_t + \varepsilon_t \ , \qquad t = 1, \dots, N$$

E apresentaremos a seguir, as equações respectivas.

$$\hat{X}_t = \alpha\left(X_t - \hat{S}_{t-s}\right) + (1-\alpha)\left(\hat{X}_{t-1} + \hat{T}_{t-1}\right)$$
$$\hat{T}_t = \beta\left(\hat{X}_t - \hat{X}_{t-1}\right) + (1-\beta)\hat{T}_{t-1}$$
$$\hat{S}_t = \gamma\left(X_t - \hat{X}_t\right) + (1-\gamma)\hat{S}_{t-s}$$

onde $t = s + 1, \dots, N$, s é o período da série sazonal , $\alpha \in [0,1]$,
$\beta \in [0,1]$ e $\gamma \in [0,1]$.

Cálculo da previsão

$$\hat{X}_t(h) = \hat{X}_t + h\hat{T}_t + \hat{S}_{t+h-s} \ , \ \ h = 1,2, \dots, s$$
$$\hat{X}_t(h) = \hat{X}_t + h\hat{T}_t + \hat{S}_{t+h-2s} \ , \ \ h = s+1, \dots, 2s$$
$$(\dots)$$

Sazonalidade Multiplicativa

Alteremos o método anterior e consideremos agora que o factor sazonal S_t é multiplicativo, ou seja,

$$X_t = \mu_t S_t + T_t + \varepsilon_t \ , \qquad t = 1, \dots, N$$

E apresentaremos a seguir, as equações respectivas.

$$\hat{X}_t = \alpha\left(\frac{X_t}{\hat{S}_{t-s}}\right) + (1-\alpha)\left(\hat{X}_{t-1} + \hat{T}_{t-1}\right)$$
$$\hat{T}_t = \beta\left(\hat{X}_t - \hat{X}_{t-1}\right) + (1-\beta)\hat{T}_{t-1}$$
$$\hat{S}_t = \gamma\left(\frac{X_t}{\hat{X}_t}\right) + (1-\gamma)\hat{S}_{t-s}$$

onde $t = s + 1, \dots, N$, s é o período da série sazonal , $\alpha \in [0,1]$,
$\beta \in [0,1]$, $\gamma \in [0,1]$, $\hat{S}_j = \frac{X_j}{\left(\frac{1}{s}\right)\sum_{k=1}^{s} X_k}, j = 1,2, \dots, s$, $\hat{X}_s = \frac{1}{s}\sum_{k=1}^{s} X_k$ e $\hat{T}_s = 0$.

Cálculo da previsão

$$\hat{X}_t(h) = (\hat{X}_t + h\hat{T}_t)\hat{S}_{t+h-s} \ , \ \ h = 1,2, \dots, s$$
$$\hat{X}_t(h) = (\hat{X}_t + h\hat{T}_t)\hat{S}_{t+h-2s} \ , \ \ h = s+1, \dots, 2s$$
$$(\dots)$$

Principais Vantagens

- São semelhantes ao método de Holt-Winters.

Principais Desvantagens

- Dificuldade em determinar os valores das constantes de suavização α ,β e γ;
- Dificuldade em calcular a média, variância ou intervalos de confiança de previsão.

Metodologia de Box-Jenkins – Modelo ARIMA

Como alternativa aos métodos anteriores, surgiu em 1970 uma nova metodologia que foi implementada por Box e Jenkins e daí deriva o seu nome.
Basicamente ao aplicar este método, a série temporal inicial tem que ser estacionária (requisito obrigatório), caso não seja deve-se fazer uma ou duas diferenciações no máximo para torná-la. Ao fazer este procedimento, ficamos com modelos autorregressivos integrados de médias móveis designados normalmente por *ARIMA(p,d,q)*. Caso a série apresente sazonalidade, este método pode ser estendido para incluir os termos autoregressivos sazonais e de médias móveis sazonais, e deste modo, ficamos com modelos designados por autorregressivos integrados de médias móveis sazonais, *SARIMA(p,d,q)x(P,D,Q).*

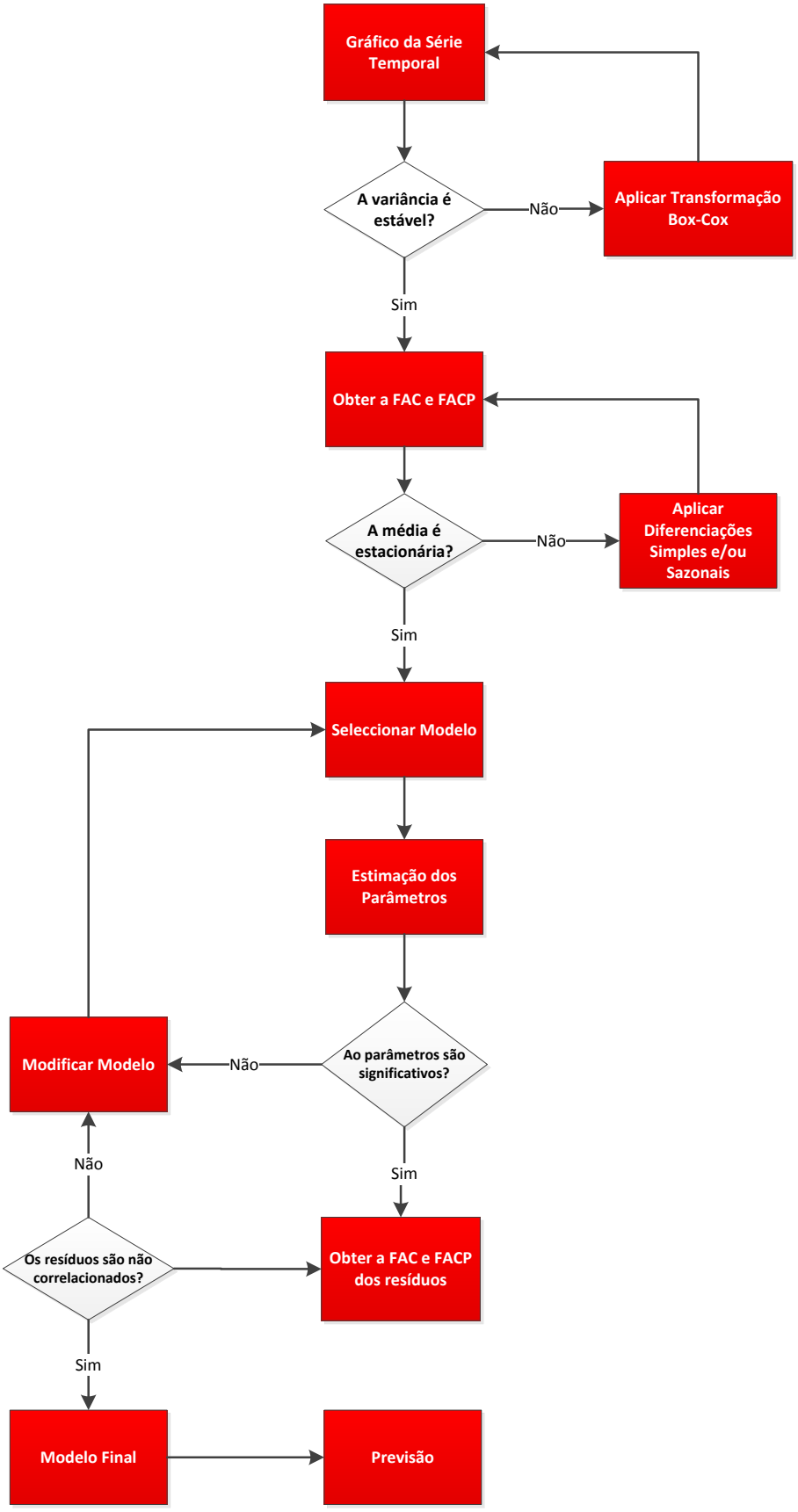
Assim, a construção do modelo vai ser baseada num ciclo iterativo em que a escolha para a estrutura do modelo é assente nos próprios dados. As etapas principais do ciclo são:

1. Fazer um estudo inicial dos dados através de uma análise gráfica (por exemplo);
2. Verificar se a variância da série está estabilizada, caso não esteja, utilizar a aplicação Box-Cox para estabilizá-la;
3. Obter as funções de autocorrelação e autocorrelação parcial dos dados;
4. Verificar se a série é estacionária, caso não seja, deve-se aplicar diferenciações simples e/ou sazonais para torná-la;
5. Identificação de um modelo tendo por base a análise das funções de autocorrelação e autocorrelação parcial;
6. Subtrair a média da série diferenciada aos dados (depende dos softwares);
7. Estimação dos parâmetros do modelo selecionado;
8. Verificação ou diagnóstico do modelo ajustado através de uma análise de resíduos;
9. Caso o modelo obtido seja adequado, deve-se avançar para o objectivo final: realizar as previsões.

Convém realçar o facto de que, caso o modelo final não seja adequado, o ciclo é repetido voltando-se à fase de identificação. Por vezes é identificado mais do que um modelo para o mesmo conjunto de dados e portanto tendo em conta todos os modelos ajustados, devemos escolher o modelo que fornece o menor erro quadrático médio de previsão essencialmente. Existem outros critérios que os modelos em causa devem obedecer como o princípio da parcimónia, ou seja, só devemos ter no modelo os parâmetros necessários e que sejam significativos, e devemos penalizar os modelos que possuem uma maior variância residual.
Como último ponto, convém referir que neste método a fase de identificação é muito importante e faz toda a diferença nos resultados obtidos, pois várias pessoas podem identificar modelos diferentes para a mesma série temporal.

Veremos a seguir um esquema simplificado do referido método em causa para uma melhor compreensão para o leitor.

Metodologia de Box-Jenkins



Principais Vantagens

- Óptimos resultados para previsões a curto prazo.

Principais Desvantagens

- Requer alguma experiência e algum conhecimento na sua aplicação;
- Exige o manuseamento de programas para a sua utilização;
- A série temporal deve ter no mínimo 100 observações.

Ao aplicar este método descrito anteriormente, por vezes acontece que os erros podem aparentar estarem autocorrelacionados ou então terem uma distribuição não estacionária no tempo. Podemos confirmar estas suposições ao observar os gráficos das funções de autocorrelação dos resíduos ao quadrado ou então, pelo simples gráfico dos resíduos standardizados.
Quando os erros se comportam assim, existem várias alternativas a utilizar. Caso aconteça, irei recorrer à família dos chamados modelos não lineares ARCH/GARCH.

Regressão com Erros ARMA

Este é um outro método de previsão que poderemos aplicar e consiste em estimar a tendência da série através de um polinómio e de uma função periódica, sendo que esta última pode ser uma combinação linear de funções seno e cosseno ortogonais. Posteriormente modelamos os erros sendo que estes pertencem à classe ARIMA. Não esquecer que caso se aplique diferenciações à série, neste caso em específico aos erros, sejam elas simples e/ou sazonais ficamos com erros pertencentes à classe ARMA (daí deriva o nome do processo).
Deste modo, fazemos a análise da mesma maneira que o método anterior e assim que obtermos o modelo final, realizamos as previsões e somamos estas últimas com os valores que estimamos pela tendência por este novo processo considerado.

Assim, tomemos a expressão inicial que será dada por:

$$X_t = T_t + \varepsilon_t\,, \quad t = 1, \dots, N$$

e representaremos por $\hat{T}_t$ a estimativa de T_t que nos é dada pela seguinte expressão:

$$\begin{aligned}\hat{T}_t = a_1 + a_2 t + a_3 t^2 + \ \dots + a_{l+1} t^l + b_1 \cos\left(\frac{2\pi t f_1}{N}\right) \\ + c_1 \operatorname{sen}\left(\frac{2\pi t f_1}{N}\right) + \ b_2 \cos\left(\frac{2\pi t f_2}{N}\right) \\ + c_2 \operatorname{sen}\left(\frac{2\pi t f_2}{N}\right) + \ \dots + b_r \cos\left(\frac{2\pi t f_r}{N}\right) + c_r \operatorname{sen}\left(\frac{2\pi t f_r}{N}\right)\end{aligned}$$

onde l representa o grau do polinómio, $\{a_j\}_{j\geq 1} \in \mathbb{R}$ são os coeficientes do polinómio anterior, $r \in \mathbb{Z}^+$ e $r < 5$ é o número de harmónicas consideradas, aos valores de

$2\pi f_r/N$ são as chamadas frequências de Fourier em que f_r representa os índices das frequências de Fourier que são dadas por $N/365$ (se tivermos a considerar os dados como diários) e pelos seus múltiplos que correspondem ao número de harmónicas consideradas inicialmente e por fim, $\{b_j\}_{j\geq 1} \in \mathbb{R}$ e $\{c_j\}_{j\geq 1} \in \mathbb{R}$ são designados por coeficientes de Fourier.

As estimativas dos parâmetros que correspondem aos coeficientes que se encontram na expressão anterior, são obtidas pelo método dos mínimos quadrados ou pelo método dos mínimos quadrados generalizados, sendo que considerei para o meu trabalho este último processo referido.

Como nota convém lembrar que qualquer função periódica se pode escrever como combinação linear de funções seno e cosseno ortogonais e para que estas funções trigonométricas sejam ortogonais, temos que considerar obrigatoriamente as frequências de Fourier e não outras quaisquer.

Caso Prático

Veremos agora uma aplicação prática das metodologias estudadas anteriormente.

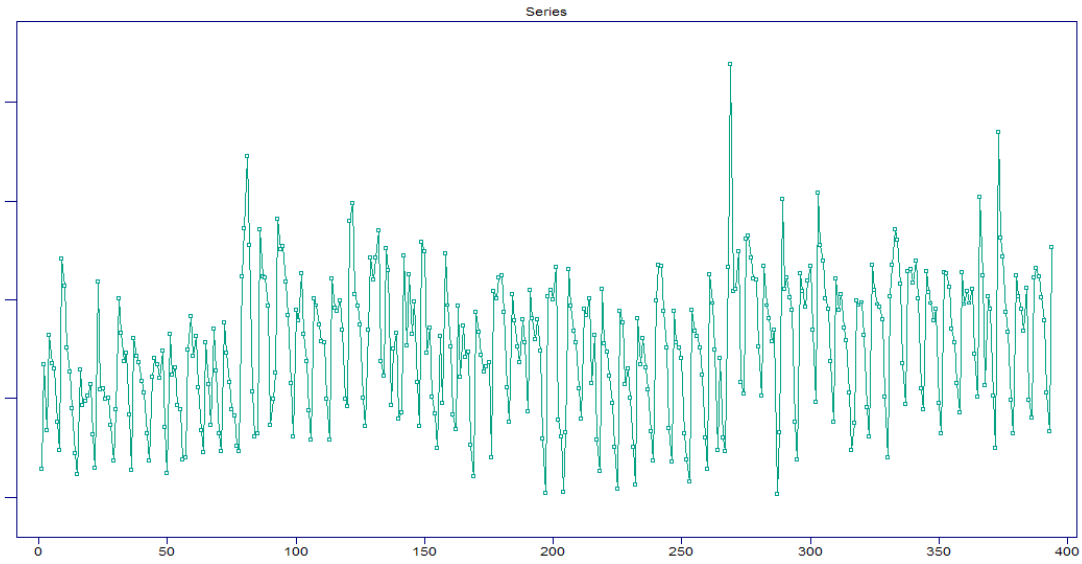
Uma vez que estamos a tratar de dados reais e de natureza confidencial, irei considerar para os meus conjuntos de dados nomes fictícios e também irei omitir determinados valores. Dito isto, chamarei ao meu primeiro conjunto de dados a Linha A, ao meu segundo conjunto de dados a Linha B e assim sucessivamente.

Estas linhas referem-se ao número de chamadas atendidas diariamente. A única diferença entre elas reside no facto de que, cada uma está relacionada com um determinado sector de serviços disponíveis pela empresa de telecomunicações em questão.

Assim, irei fazer agora uma análise detalhada relativamente ao meu primeiro conjunto de dados.

Linha A

Esta linha encontra-se disponível 24h por dia, 365 dias por ano. Considerei para este conjunto um total de 394 observações que correspondem a um intervalo temporal que vai desde 3 de Outubro de 2010 (Domingo) a 31 de Outubro de 2011 (Segunda-Feira).



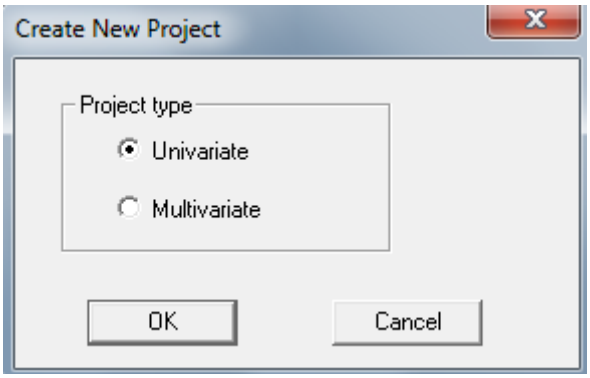
Dado o gráfico anterior, é importante fazemos uma análise inicial à série. Assim, passo a citar algumas observações que podemos retirar inicialmente da nossa série em estudo:

- A série apresenta uma média e uma variância não constantes no tempo e portanto devemos estar perante um processo não estacionário certamente;
- É provável a série ter uma componente sazonal, uma vez que no gráfico podemos observar oscilações para cima e para baixo consecutivamente, o que indica que estamos diante de um movimento cíclico (talvez semanal);
- Pode-se considerar o nível da série constante em certos intervalos de tempo;
- A série não apresenta uma tendência significativa;
- Há presença de algumas perturbações na série;

Para visualizarmos o nosso conjunto de dados inicial basta realizar o seguinte comando no package estatístico ITSM 2000:

1º) $File \rightarrow Project \rightarrow Open$

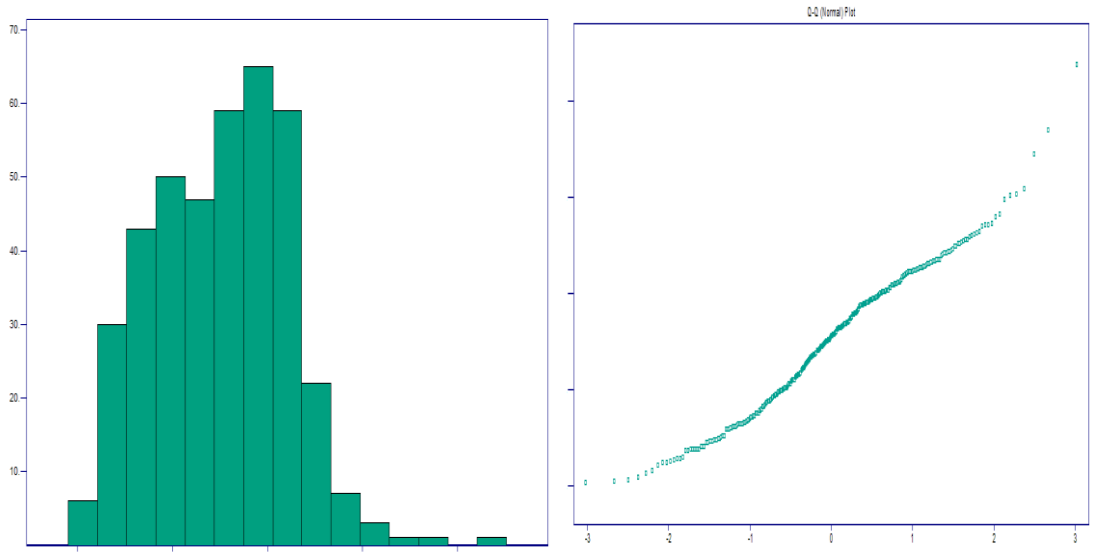
2º)



3º) $Seleccionar\ o\ ficheiro\ "Linha\ A.txt"\ e\ clicar\ em\ Abrir$

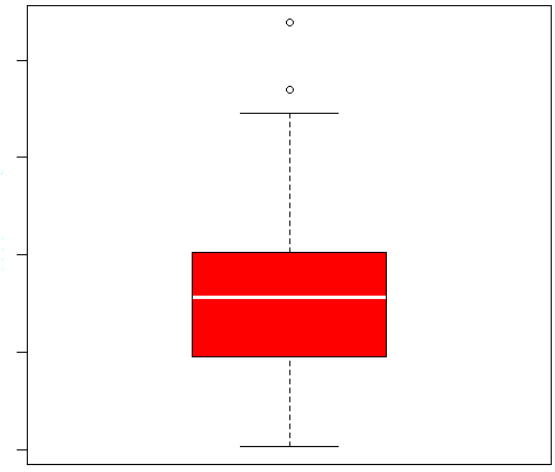
Para complementar as informações anteriores, apresento ainda os gráficos do histograma e do QQ-Plot (Normal) do nosso conjunto de dados e que podem ser visualizados no programa ITSM 2000 se efectuarmos os seguintes comandos:

1º) $Statistics \rightarrow Histogram \rightarrow Default$
2º) $Statistics \rightarrow QQ - Plot\ (normal)$



Através do gráfico do histograma podemos reparar que os dados apresentam uma distribuição assimétrica positiva e que se aproximam de uma distribuição normal. Podemos chegar à mesma conclusão observando o gráfico QQ-Plot (Normal), embora seja notório a presença de alguns valores atípicos.

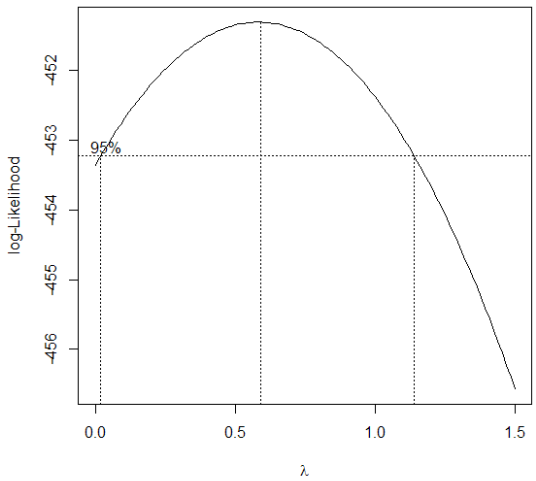
Para detectar a presença de outliers na série, utilizou-se o gráfico Box-Plot que se encontra no programa R.



Pela figura anterior, notamos a presença de dois valores que são discrepantes quando comparados com o resto dos valores da amostra. O valor mais elevado foi registado no dia 28/06/2011 (Terça-Feira) enquanto que o outro valor foi registado no dia 10/10/2011 (Segunda-Feira). Pelo o que tive acesso dos relatórios diários disponibilizados, nestes dias houve problemas nesta linha e portanto houve um aumento no número de chamadas em relação aos dias normais.

Uma vez que já fizemos uma breve descrição da série inicial, estamos nas condições para aplicar a Metodologia de Box-Jenkins ao nosso conjunto de dados.
Dito isto, a primeira condição a se verificar é se a série apresenta uma variância constante no tempo. Como verificamos pelo primeiro gráfico apresentado neste capítulo, a série apresenta uma variância não constante e com o objectivo de estabilizá-la, utiliza-se normalmente a transformação Box-Cox.

Com a aplicação Box-Cox que se encontra no programa R, iremos saber de uma forma mais simples e precisa se é necessário fazer ou não a transformação no programa ITSM 2000.



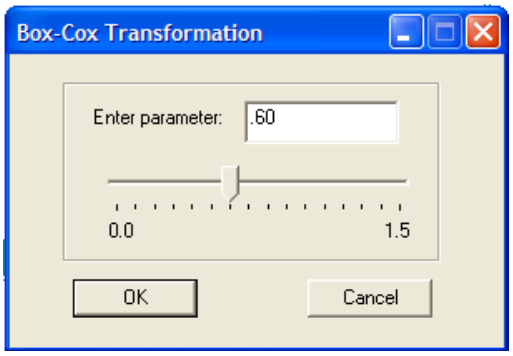
Ao observar o gráfico anterior, no eixo dos x encontram-se os valores de lambda para os quais se deve fazer a transformação enquanto que no eixo dos y os valores são respeitos à função log-verosimilhança. Dito assim, reparamos que para os valores de lambda situados entre 0 e 1.2 encontra-se a solução óptima para um nível de significância de 5%, sendo que devemos considerar para lambda um valor perto de 0.5 visto que atingimos a função log-verosimilhança quando ela é máxima.

Perante esta situação, uma vez que no programa ITSM 2000 a série original tem o valor inicial de lambda igual a 1, efectuei deste modo a transformação e considerei para o valor de lambda 0.6.

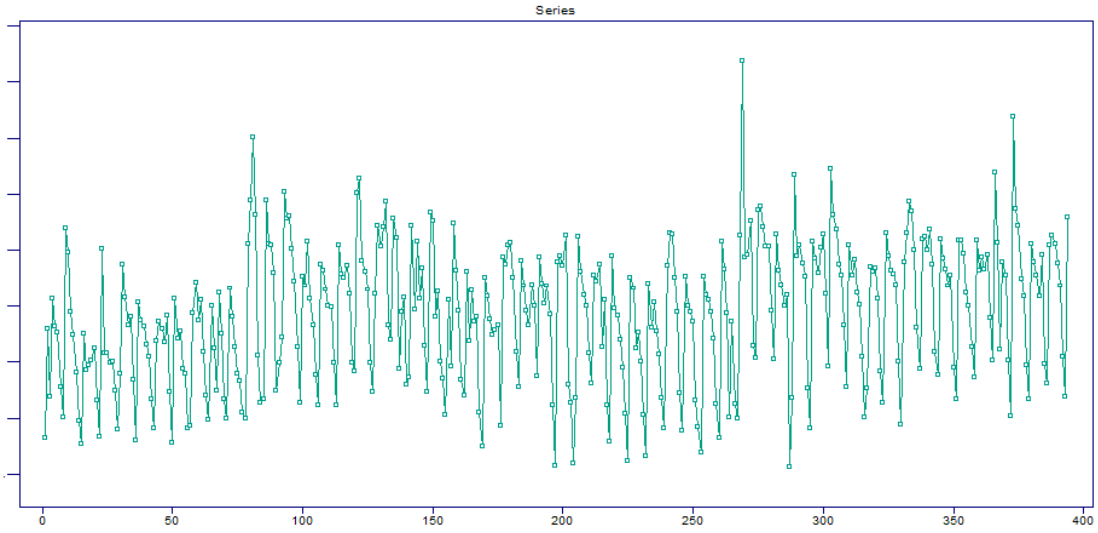
A seguir apresento os passos que efectuei no programa:

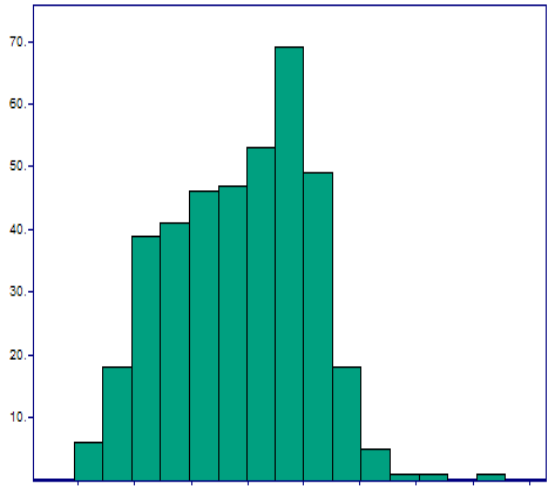
$\mathbf{1º}) \; Transform \rightarrow Box - Cox$

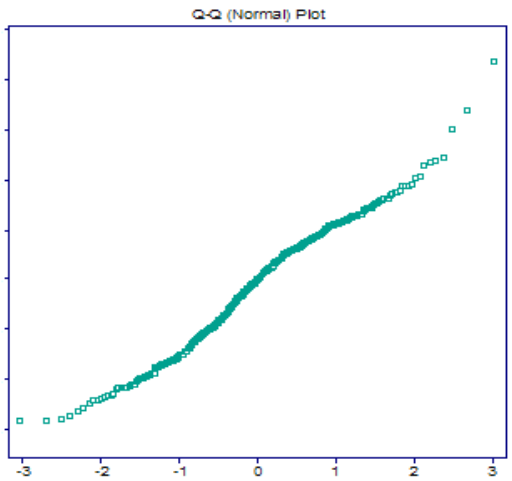
$\mathbf{2º})$

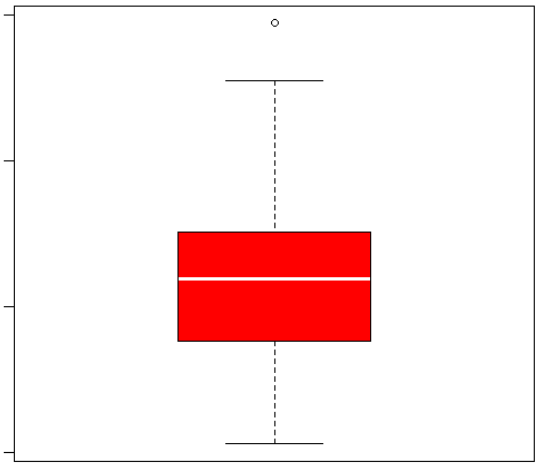


Veremos a seguir os gráficos apresentados atrás mas agora tendo em conta a transformação anterior. Convém não esquecer que a partir deste momento vamos considerar os dados quando aplicada a transformação Box-Cox.



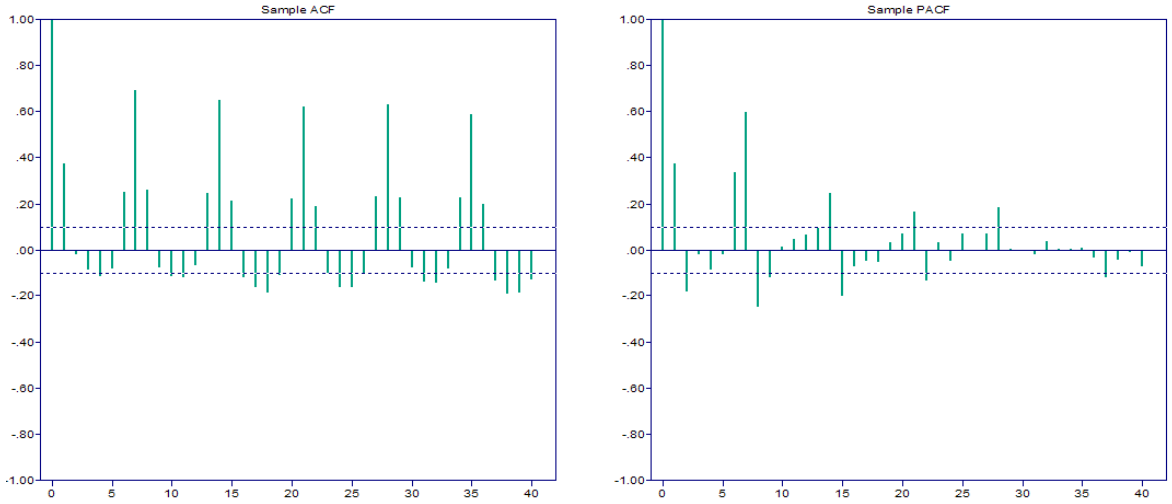






O passo a realizar a seguir é verificar se a média da série é constante. Graficamente conferimos que a série não é constante relativamente ao seu nível. Para confirmar tal facto, veremos a seguir os gráficos das funções de autocorrelação e autocorrelação parcial da nossa série que é dado pelo seguinte comando:

1º) $Statistics \rightarrow ACF/PACF \rightarrow Sample$



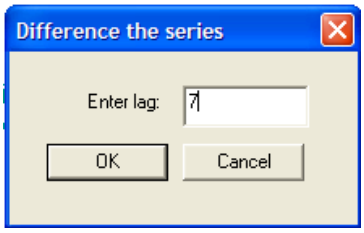
Pelo gráfico da função de autocorrelação verificamos que a série não é estacionária uma vez que apresenta fortes correlações, sendo elas muito significativas devido aos altos valores que tomam. Em ambas as funções o lag de 0 toma o valor de 1 (por definição) e podemos reparar, principalmente pelo gráfico da FAC, que no lag 7 e nos seus múltiplos há uma grande correlação. Estas indicações apontam para a presença de uma componente sazonal semanal tendo por período 7.

Assim, de modo a tornar a série em estacionária, temos que aplicar uma diferenciação de ordem 7 de modo a podermos eliminar os vestígios de sazonalidade da série.

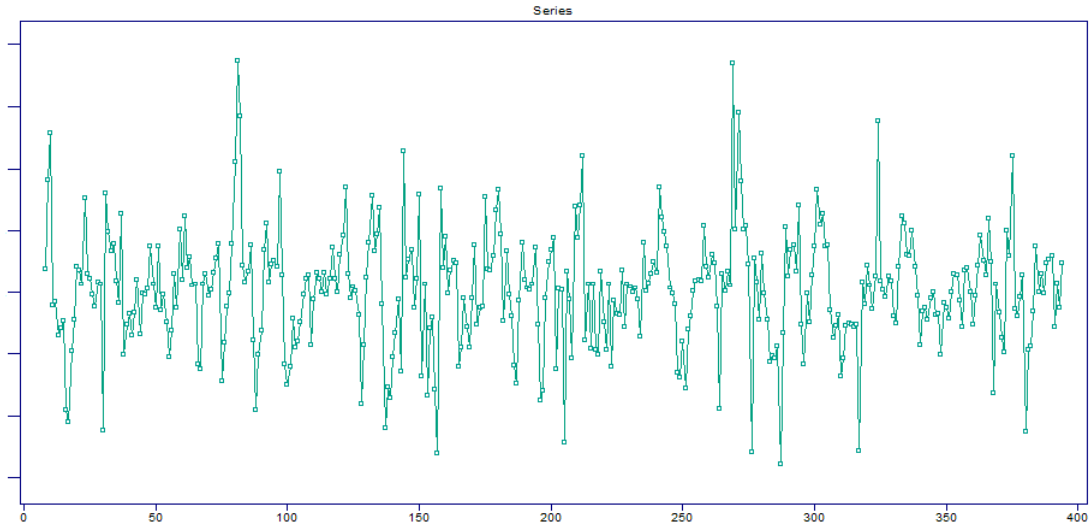
Para efectuar esta diferenciação utilizamos o seguinte comando:

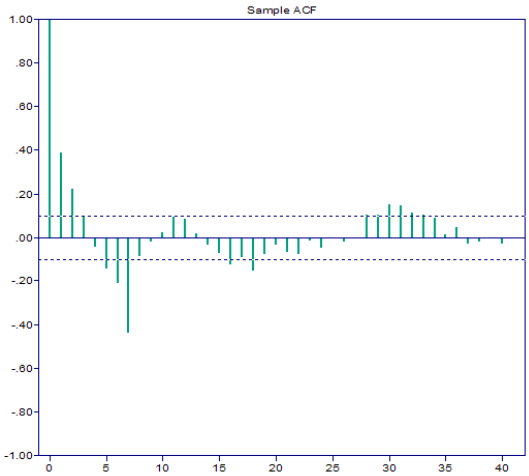
1º) $Transform \rightarrow Difference$

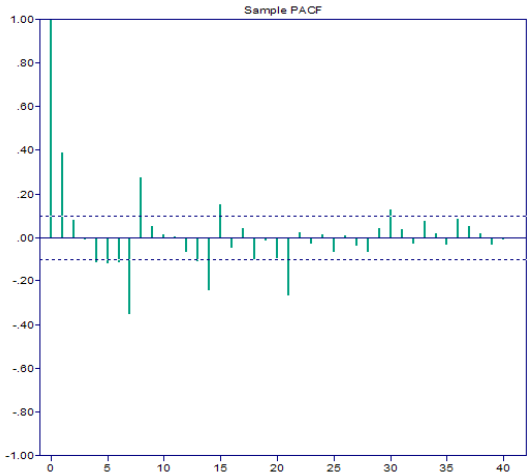
2º)



Apresentamos a seguir, os novos gráficos da série diferenciada e das funções de autocorrelação e autocorrelação parcial da mesma.







Nesta altura estamos nas condições exigidas pela Metodologia de Box-Jenkins pois a nossa série já se encontra estacionária visto que a FAC decresce lentamente para zero. Dito isto, chegamos a uma etapa em que temos de identificar o modelo que melhor descreve o nosso conjunto de dados.

Após observar os gráficos da FAC e FACP, os valores para p e q não são claros e por isso vamos considerar várias hipóteses para o par (p,q), tais como $\{(1,0);(1,1);(0,2)\}$. Já para os valores de (P,Q) podemos observar pelo gráfico da FAC que temos uma queda brusca depois do lag de 7 e que no gráfico da FACP temos um rápido decaimento exponencial para o lag de 7 e para os seus restantes múltiplos e por isso sugerimos para os valores de (P,Q) o conjunto (0,1).

Por tudo o que foi dito anteriormente, sabemos que estamos perante um modelo SARIMA que pertence à classe dos modelos não estacionários na média que incluem a componente sazonal.

Antes de avançarmos para a parte da estimação dos parâmetros do modelo, convém dizer que o programa estatístico ITSM 2000 pergunta se podemos subtrair a média do processo antes da referida estimação, com o propósito de tornar a média da série constante (próxima de zero) e normalmente realizamos essa operação.

Uma vez que a selecção do modelo depende muito da capacidade de análise do estatístico em questão, acabei, na minha opinião, por selecionar 3 possíveis modelos para o nosso conjunto de dados.

Possíveis Modelos	AICC	BIC
$SARIMA\ (1,0,0)\times(0,1,1)_7$	3106.97	3088.09
$SARIMA\ (1,0,1)\times(0,1,1)_7$	3101.80	3092.69
$SARIMA\ (0,0,2)\times(0,1,1)_7$	3117.17	3112.56

De entre estes 3 modelos propostos, o modelo que seleccionei para representar o meu conjunto de dados foi o seguinte:

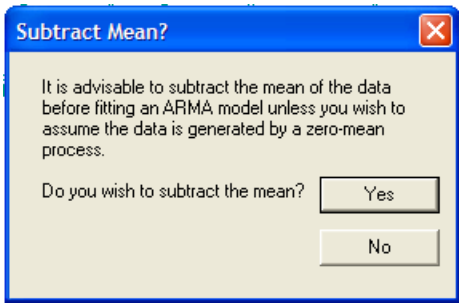
$$SARIMA\ (1,0,1)\times(0,1,1)_7$$

visto que de entre todos os modelos é o que apresenta o menor valor de AICC (quanto menor, melhor será o modelo).

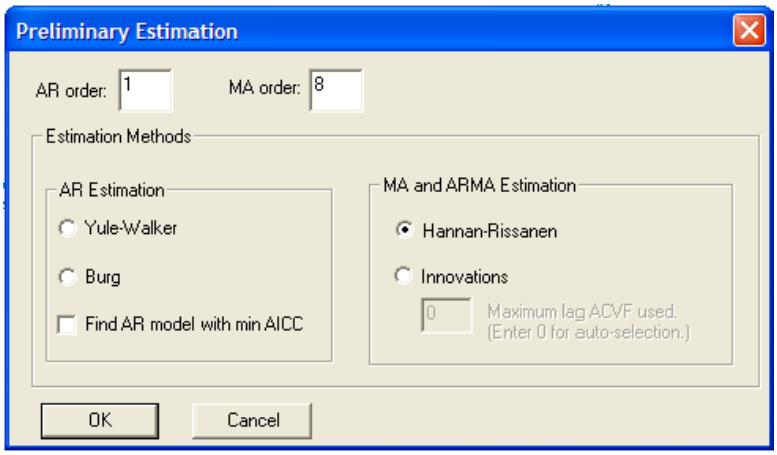
A seguir apresento os passos que se realizaram no programa ITSM 2000 para a obtenção do nosso modelo escolhido e o resultante output.

1º) $Model \rightarrow Estimation \rightarrow Preliminary$

2º)

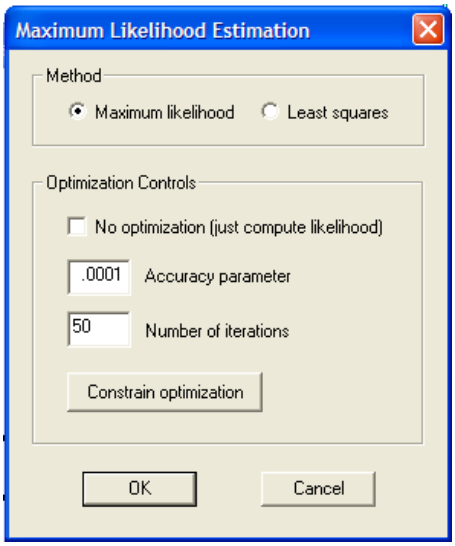


3º)

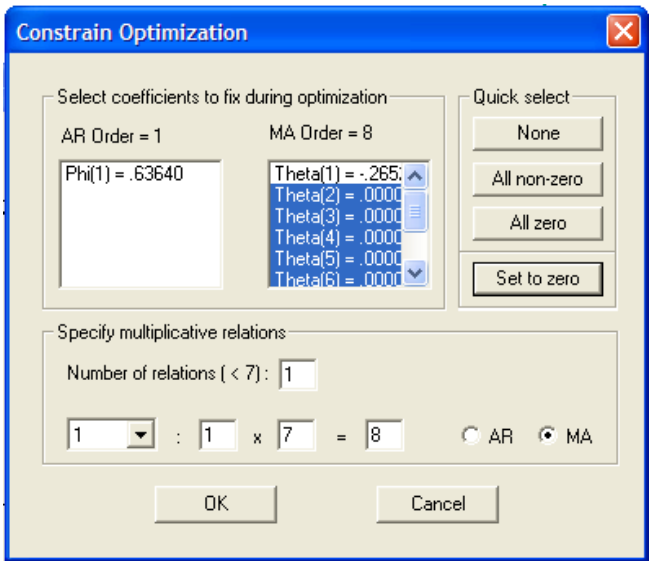


4º) $Model \rightarrow Estimation \rightarrow Max\ likelihood$

5º)



6º)



7º)

==========
ITSM::(Maximum likelihood estimates)
==========

Method: Maximum Likelihood

ARMA Model:
X(t) = .7220 X(t-1)
+ Z(t) - .3136 Z(t-1) + .0000 Z(t-2) + .0000 Z(t-3)
+ .0000 Z(t-4) + .0000 Z(t-5) + .0000 Z(t-6) - .9609 Z(t-7)
+ .3014 Z(t-8)

WN Variance = .165203E+03

AR Coefficients
.722005

Standard Error of AR Coefficients
.036469

MA Coefficients
-.313643 .000000 .000000 .000000
.000000 .000000 -.960885 .301375

Standard Error of MA Coefficients
.014527 .000000 .000000 .000000
.000000 .000000 .014034 .000000

(Residual SS)/N = .165203E+03

AICC = .310180E+04
BIC = .309269E+04

-2Log(Likelihood) = .309370E+04

Accuracy parameter = .000130000

Number of iterations = 11

Number of function evaluations = 59

Optimization stopped within accuracy level.

A expressão geral e a expressão ajustada do nosso modelo é dada respectivamente por:

$$X_t = \phi_1 X_{t-1} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_7 \varepsilon_{t-7} + \theta_8 \varepsilon_{t-8}$$

$$X_t = 0.7220 X_{t-1} + \varepsilon_t - 0.3136 \varepsilon_{t-1} - 0.9609 \varepsilon_{t-7} + 0.3014 \varepsilon_{t-8}$$

em que X_t representa a série original diferenciada por 7 em que se subtraiu a média dessa série e ε_t representa os erros associados quando ajustamos o modelo aos dados.

Relativamente aos passos atrás mencionados, vou fazer uns breves esclarecimentos para não deixar margem de dúvidas na sua interpretação:

Quanto ao passo 3, caso o leitor recorde-se, inicialmente tínhamos um processo não estacionário mas como efectuamos uma diferenciação sazonal e verificamos pelas FAC e FACP que o processo ficou estacionário, então classificamos este modelo como pertencente a esta classe e diz-se, neste caso, um ARMA (1,8) e portanto vem daí a colocação destes valores neste 3º passo. A estimação preliminar dos parâmetros faz-se pelo método Hannan-Rissanen e a dita estimação dos parâmetros é feita pelo método de máximo verossimilhança como podemos observar no passo 5.
No 6º passo, como estamos a considerar um caso em particular em que temos um ARMA (1,8) mas não queremos utilizar todos os seus parâmetros então teremos de igualar alguns coeficientes a zero para que o programa perceba essa nossa intenção e neste caso serão igualados a zero alguns coeficientes do polinómio de médias móveis que vão desde θ_2 até θ_6. Por fim, ainda neste passo temos que definir uma relação multiplicativa presente no nosso modelo uma vez que temos $\theta_1 \times \theta_7 = \theta_8$ e o próprio software não reconhece essa igualdade existente.

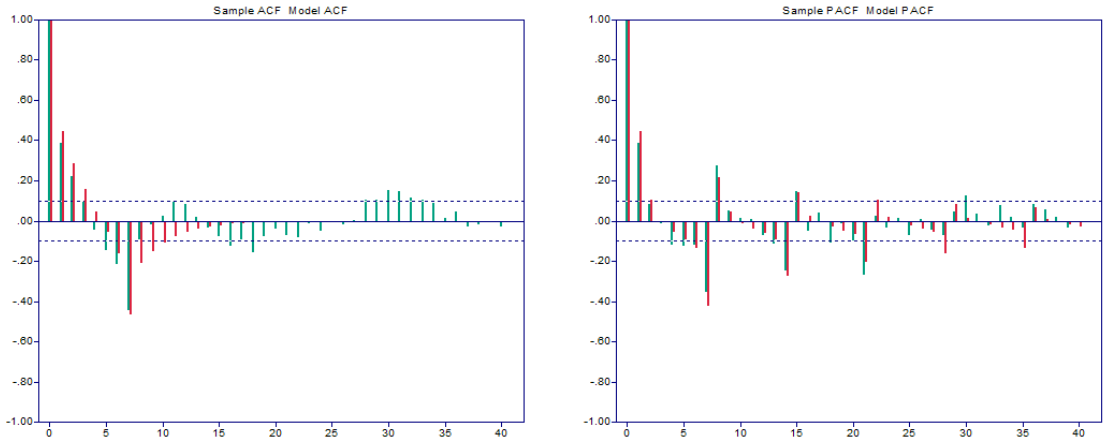
Relativamente ao output disponibilizado podemos reparar que os 4 parâmetros do modelo são significativos visto que temos sempre verificada a seguinte inequação:

$$|Coeficiente\ Estimado| \geq 1.96 \times |Erro\ Padrão\ Coeficiente\ Estimado|$$

e quanto ao programa ITSM 2000 este apresenta, diria eu, uma grande falha pois não apresenta nenhum gráfico em que se vê o modelo ajustado final relativamente aos nossos dados iniciais.

Como alternativa ilustra os gráficos das FAC e FACP relativas ao modelo teórico e ao modelo estimado proposto, que se encontraram representados nos gráficos pela cor verde e pela cor vermelha respectivamente. Para obter estes gráficos basta realizar o seguinte comando no package estatístico ITSM 2000:

$$Statistics \rightarrow ACF/PACF \rightarrow Sample/Model$$

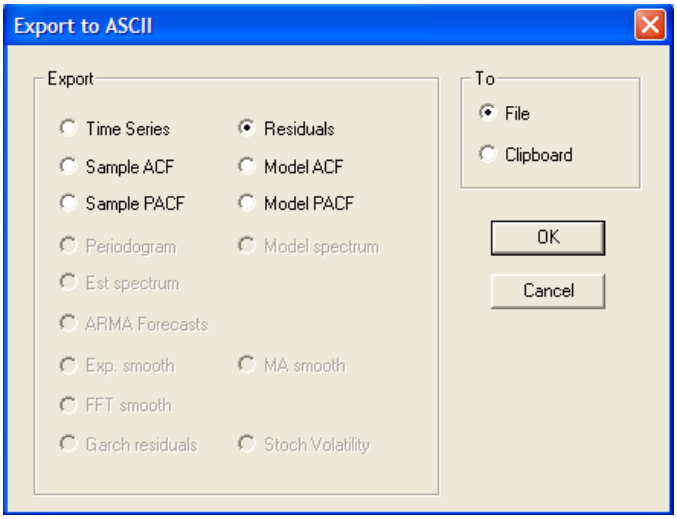


Pela observação dos gráficos anteriores poderemos dizer que o nosso modelo se ajusta de uma forma satisfatória ao nosso conjunto de dados e para comprovar tal afirmação, iremos agora verificar se os pressupostos relativamente aos modelos lineares são cumpridos ou não no nosso caso. Para isso iremos fazer uma análise detalhada aos resíduos do nosso modelo.

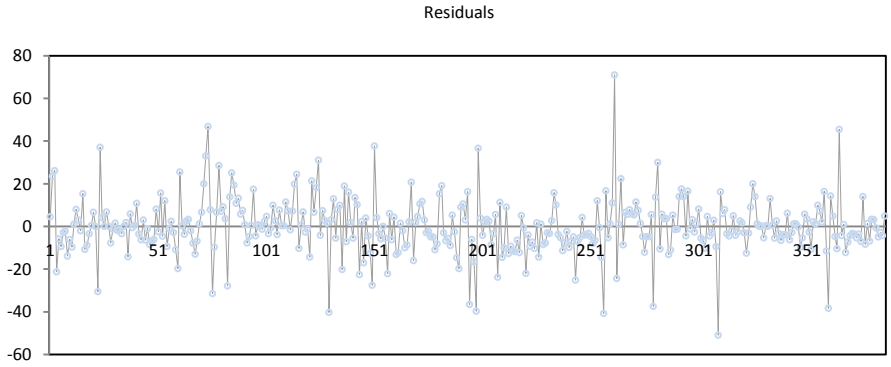
Ao manusear o programa ITSM 2000 deparo que este não apresenta graficamente os resíduos mas existe outra forma de obter os seus valores se realizarmos o seguinte comando no programa:

1º) $File \rightarrow Export$

2º)



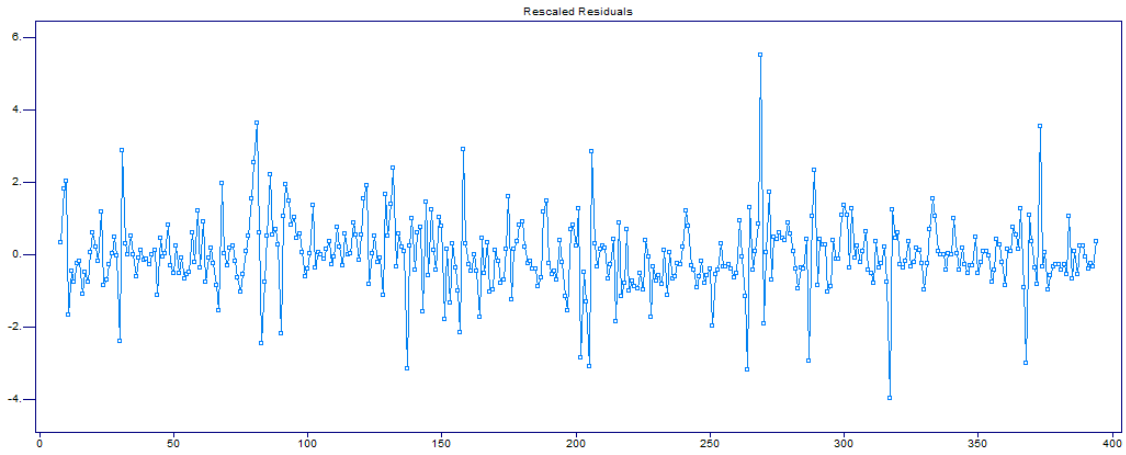
Assim, já estamos nas condições de representá-los graficamente recorrendo ao uso do programa Excel.



Alternativamente, o package estatístico ITSM 2000 apresenta os resíduos mas é na forma de standardizados, isto é, a cada resíduo subtrai-se a média total deles e a esse valor dividimos pelo desvio-padrão total dos mesmos. Dessa maneira os resíduos tendem a seguir uma distribuição normal padrão, cuja média é 0 e variância é 1.

A representação gráfica dos resíduos standardizados é dada pelo seguinte comando:

$Statistics \rightarrow Residual\ Analysis \rightarrow Plot$



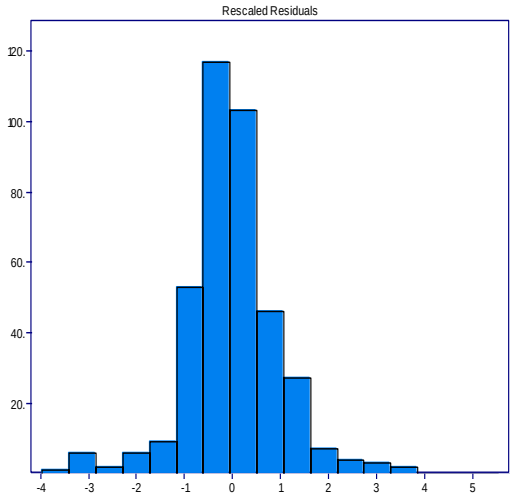
Ao analisar a estrutura dos resíduos podemos dizer, de uma forma geral, que estes se distribuem aleatoriamente em torno de zero, o que nos dá uma indicação de que o nosso modelo ajustado conseguiu "capturar" de uma forma satisfatória toda a informação contida nos dados. Perante esta situação, os resíduos são classificados como independentes.

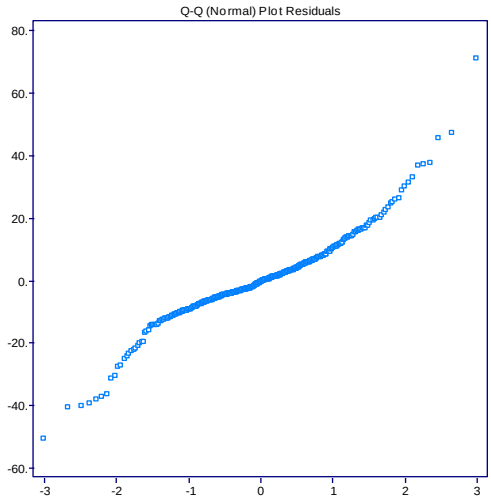
Relativamente ao gráfico anterior, podemos notar a existência de alguns valores mais elevados, sendo que estes podem dever-se a alguns outliers que não retirei do meu conjunto de dados inicialmente. Deste modo, ao ajustarmos o nosso modelo este pode não ter conseguido atingir esses valores atípicos e daí os altos valores registrados no gráfico dos resíduos standardizados.

Outra condição que os resíduos têm que satisfazer é que estes devem possuir uma distribuição aproximadamente Normal. Para isso irei mostrar o gráfico do Histograma dos resíduos standardizados e o gráfico do QQ-Plot (Normal) dos resíduos de forma a comprovar essa tal veracidade.

Para obter os gráficos mencionados anteriormente, basta-nos realizar as seguintes operações no programa ITSM 2000:

1º) $Statistics \rightarrow Residual\ Analysis \rightarrow Histogram \rightarrow Default$
2º) $Statistics \rightarrow Residual\ Analysis \rightarrow QQ - Plot\ (normal)$



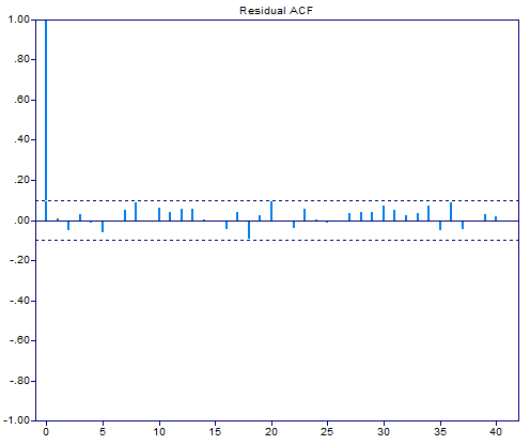


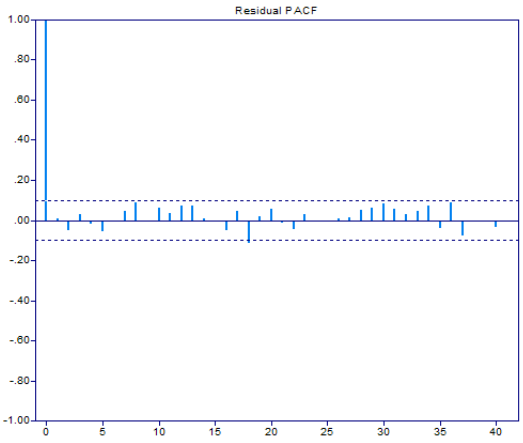
Através do gráfico do Histograma e do QQ-Plot (Normal) podemos dizer que os resíduos standardizados aparentam seguir uma distribuição Normal. Ainda pelo gráfico do QQ-Plot (Normal) podemos observar alguns valores mais dispersos da recta traçada, sendo que estes podem ser considerados como possíveis candidatos a outliers.

Quanto aos restantes pressupostos que os resíduos têm de verificar, falta-nos ainda averiguar se eles são de facto não correlacionados. Sendo assim, iremos ver a seguir os gráficos das FAC e FACP dos resíduos para verificar se tal requisito acontece.

Para obter estes gráficos basta realizar o seguinte comando no package estatístico ITSM 2000:

$$Statistics \rightarrow Residual\ Analysis \rightarrow\ ACF/PACF$$





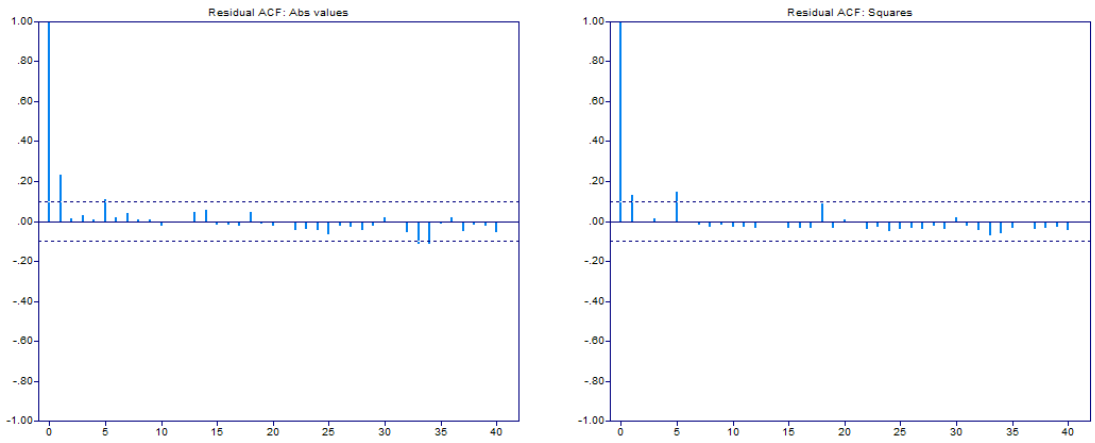
Pela análise dos gráficos anteriores podemos afirmar que os resíduos têm um comportamento análogo ao do ruído branco, isto é, podemos dizer que os resíduos

são não correlacionados, pois podemos reparar que os valores das correlações apresentam-se como sendo muito baixas.

Por fim, vamos apresentar a última condição que os resíduos têm que satisfazer: se eles são de facto independentes ou não. Para isso, iremos precisar de observar os gráficos da FAC dos valores absolutos e dos valores ao quadrado dos resíduos.

Para obter os respectivos gráficos, basta realizar a seguinte operação no programa ITSM 2000:

$$Statistics \rightarrow Residual\ Analysis \rightarrow\ ACF\ Abs\ values/Squares$$



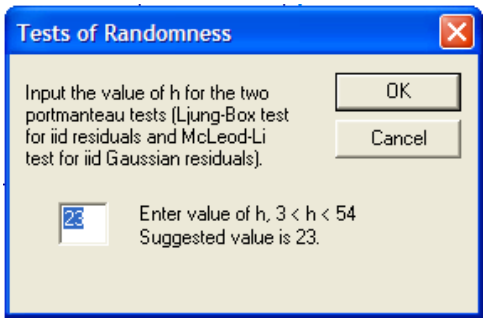
Ao analisar os gráficos anteriores, podemos dizer que os resíduos apresentam-se como sendo independentes, pois não se verifica valores elevados de correlação em ambos os gráficos, à excepção do lag 0 apresentado no gráfico da FAC dos valores absolutos dos resíduos.

Uma outra forma de confirmar as nossas conclusões a que chegamos acerca das características dos resíduos é através da realização de testes.

Alguns destes testes podem ser obtidos no package estatístico ITSM 200 através do seguinte comando:

$$\mathbf{1}º)\ Statistics \rightarrow Residual\ Analysis \rightarrow\ Tests\ of\ Randomness$$

2º)



De seguida, apresento o output disponibilizado pelo programa:

==========
ITSM::(Tests of randomness on residuals)
==========

Ljung - Box statistic = 25.004 Chi-Square (20), p-value = .20127

McLeod - Li statistic = 25.008 Chi-Square (23), p-value = .34988

Turning points = .24600E+03~AN(.25667E+03,sd = 8.2751), p-value = .19740

Diff sign points = .19600E+03~AN(.19300E+03,sd = 5.6862), p-value = .59778

Rank test statistic = .35974E+05~AN(.37346E+05,sd = .12713E+04), p-value = .28067

Order of Min AICC YW Model for Residuals = 0

Como podemos verificar caso seja aplicado o teste de Ljung-Box aos nossos resíduos, este apresenta um p-value de 0.20127, ou seja, não rejeitamos a hipótese nula e portanto podemos afirmar que não há evidência para afirmar que os resíduos são correlacionados, o que vem comprovar a nossa conclusão dada inicialmente.

Relativamente ao teste de McLeod-Li, este apresenta um p-value de 0.34988, ou seja, não rejeitamos a hipótese nula e portanto podemos afirmar que não há evidência para afirmar que o nosso modelo é não-linear do tipo ARCH. Dito isto, o modelo linear (SARIMA) é o mais indicado para descrever o nosso conjunto de dados.
Assim, por tudo o que foi dito e mostrado anteriormente podemos dizer que os pressupostos dos resíduos são verificados relativamente aos modelos lineares.

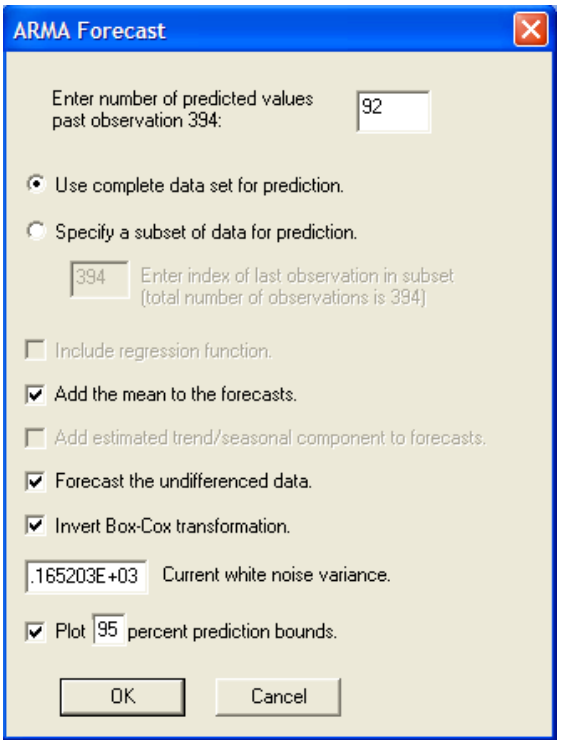
Neste momento, encontramo-nos na fase final da Metodologia de Box-Jenkins uma vez que já temos um modelo que se ajusta satisfatoriamente ao nosso conjunto de dados e que cumpre com os requisitos propostos. Como tal estamos nas condições para avançar para a nossa etapa final: realizar as previsões.

Deste modo, pretende-se prever os valores desta linha para os meses de Novembro e Dezembro de 2011 e para o mês de Janeiro de 2012, o que perfaz um total de 92 observações diárias.

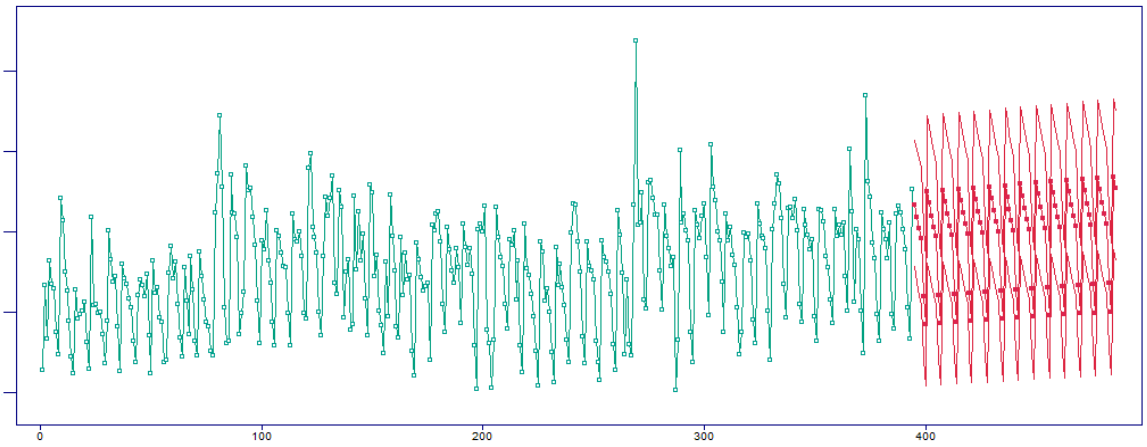
Para obtermos as representações gráficas das previsões, basta realizarmos a seguinte operação no programa ITSM 2000:

1º) $Forecasting \rightarrow ARMA$

2º)

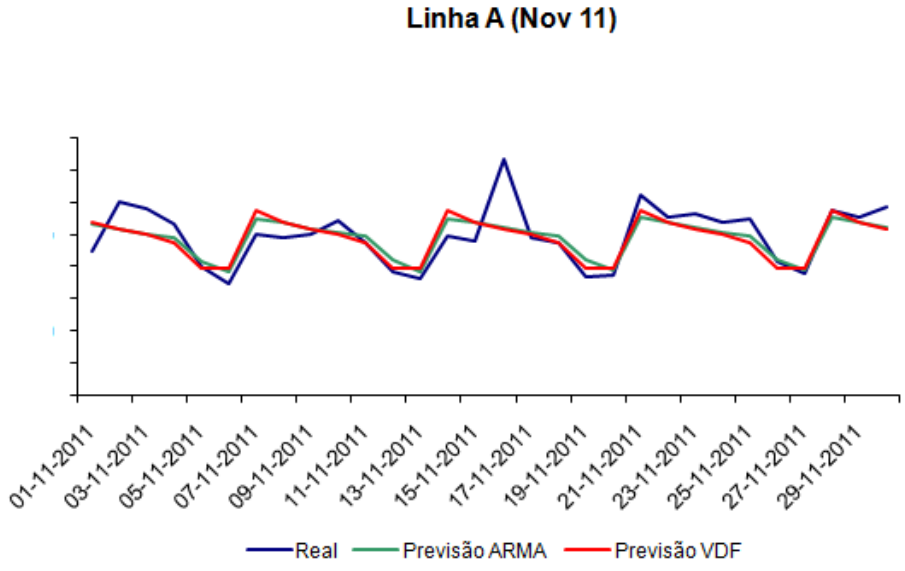


Como nota convém referir que também pedi ao programa que apresentasse os intervalos de confiança de 95% para as minhas previsões e por isso seleccionei a última caixa como podemos ver na figura apresentada em cima.

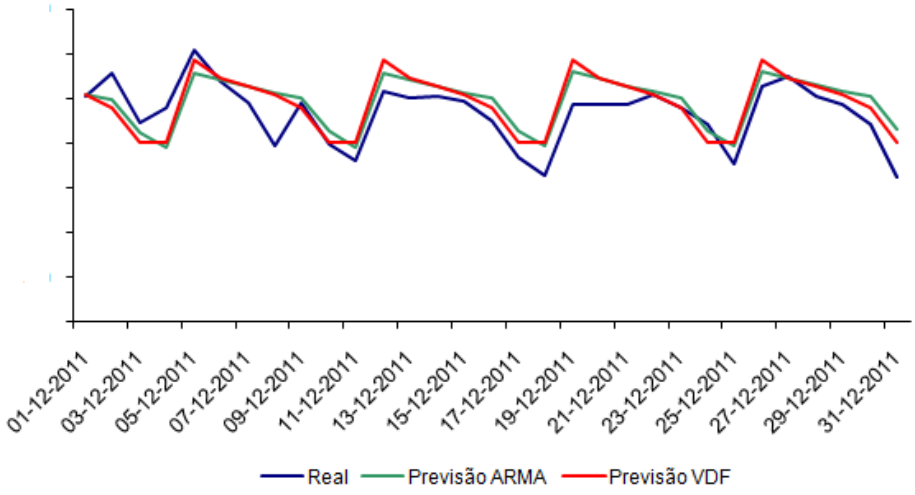


Para obtermos os valores das nossas previsões e os intervalos de confiança de 95% respectivos, temos que clicar com o botão do lado direito do rato sobre o gráfico anterior no programa ITSM 2000 e escolher a opção Info para obter os referidos valores.

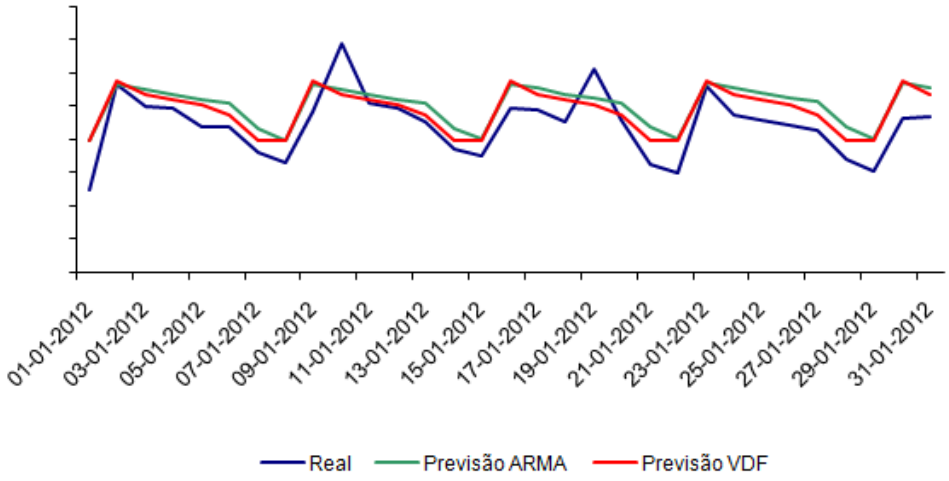
Neste momento, estamos nas condições de mostrar os gráficos com os valores das previsões obtidas pelo modelo SARIMA, as previsões obtidas pela Vodafone pelo método Holt-Winters Sazonal e os valores reais desta linha para cada um dos meses mencionados e para o total dos 3 meses em questão.

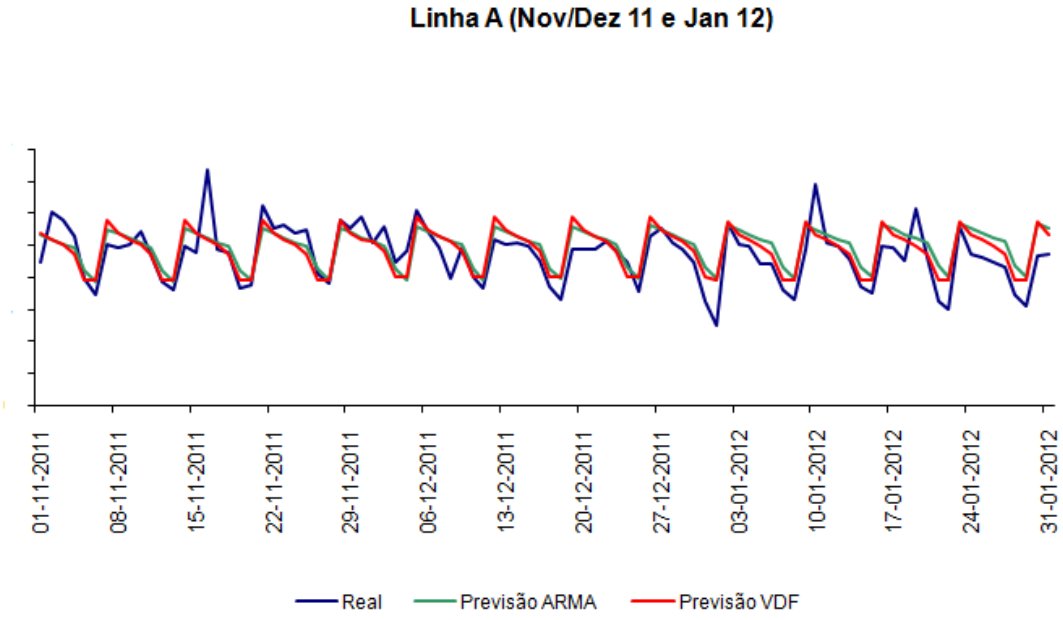


Linha A (Dez 11)

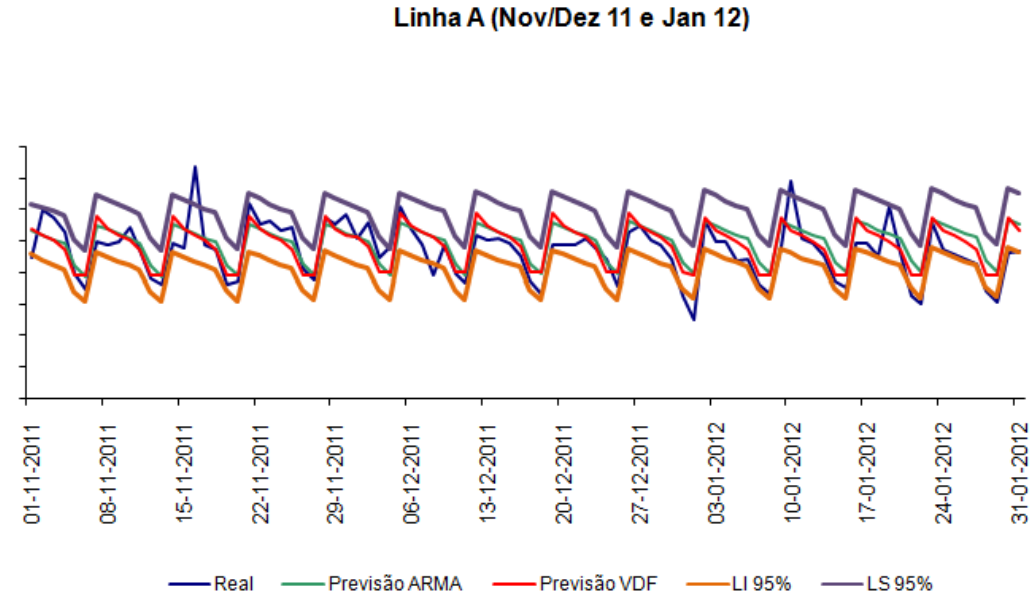


Linha A (Jan 12)





Tendo por base o gráfico anterior, vamos agora introduzir nesse mesmo gráfico os intervalos de confiança de 95% dados pela aplicação do meu método de previsão.



Como podemos confirmar pelo gráfico anterior, praticamente os intervalos de confiança de 95% abrangem os valores reais desta linha apresentada (tirando 2 situações mais evidentes), provando deste modo a importância que eles transmitem nos variadíssimos métodos de previsão. Assim, ao considerar estes valores podemos garantir com 95% de certeza que os valores reais desta linha encontram-se nesse intervalo.

Neste momento, estamos com todas as condições reunidas para fazermos as comparações necessárias entre os dois métodos propostos para os vários meses referidos. As tabelas apresentadas a seguir contêm os resumos para cada um dos meses e no seu total relativamente a cada método de previsão.

Linha A (Nov 11)											
Desvio Mensal		Desvio Diário		Desvio Diário Absoluto		EQM		EAM		REQM	
ARMA	VDF	ARMA	VDF	ARMA	VDF	ARMA	VDF	ARMA	VDF	ARMA	VDF
0,76	1,84	-0,80	0,42	8,73	8,62	343748,41	378891,67	448,30	449,37	586,30	615,54

Linha A (Dez 11)											
Desvio Mensal		Desvio Diário		Desvio Diário Absoluto		EQM		EAM		REQM	
ARMA	VDF	ARMA	VDF	ARMA	VDF	ARMA	VDF	ARMA	VDF	ARMA	VDF
-5,31	-4,69	-6,30	-5,48	9,31	9,06	247336,09	248733,66	405,44	398,55	497,33	498,73

Linha A (Jan 12)											
Desvio Mensal		Desvio Diário		Desvio Diário Absoluto		EQM		EAM		REQM	
ARMA	VDF	ARMA	VDF	ARMA	VDF	ARMA	VDF	ARMA	VDF	ARMA	VDF
-12,71	-8,59	-15,41	-11,00	17,71	13,65	617565,81	457811,03	716,97	555,97	785,85	676,62

Linha A (Nov/Dez 11 e Jan 12)											
Desvio Mensal		Desvio Diário		Desvio Diário Absoluto		EQM		EAM		REQM	
ARMA	VDF	ARMA	VDF	ARMA	VDF	ARMA	VDF	ARMA	VDF	ARMA	VDF
-5,59	-3,69	-7,57	-5,42	11,96	10,46	403526,21	361626,47	524,39	468,17	635,24	601,35

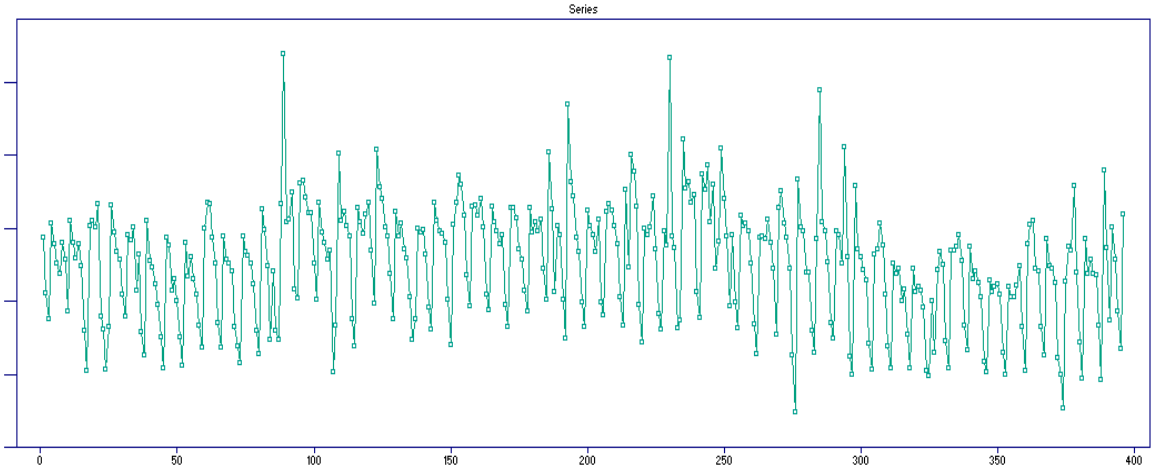
Ao olharmos para os resultados obtidos podemos observar que no método ARMA em relação aos desvios dia/mensal, este apresenta maiores desvios em comparação com o método da Vodafone (excepto num valor - Desvio Mensal de Nov 11).

Quanto aos valores apresentados pelas medidas de desempenho há um certo equilíbrio entre os 2 métodos propostos. Relativamente ao mês de Novembro de 2011 e ao mês de Dezembro de 2011 o método ARMA apresenta melhores resultados pois exibe valores mais baixos para essas medidas relativamente ao método da Vodafone (excepto na medida EAM de Dez 11). Em relação às restantes datas (Jan 12 e no total dos 3 meses) o método da Vodafone indica melhores resultados em relação ao nosso método apresentado.

Contudo, por tudo o que foi dito anteriormente fica a dúvida de qual o melhor método a escolher. Esta escolha não é tão evidente como pode parecer, pois depende de variadíssimos factores e de condições com que se pretende inicialmente. Assim, de entre os 2 métodos utilizados, poderíamos incidir tanto sobre o método da Vodafone como para o nosso método proposto pela Metodologia de Box-Jenkins.

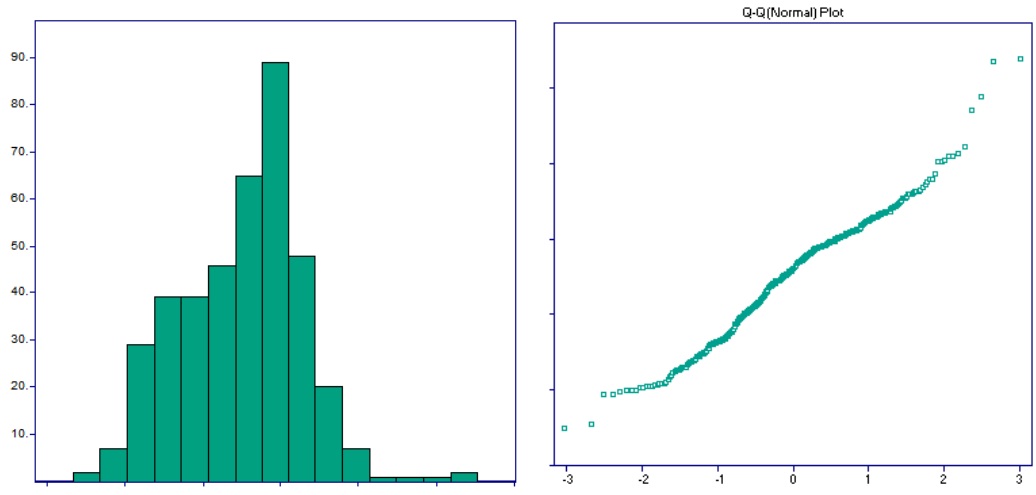
Ainda referente a esta linha, um dos outros objectivos para cumprir era realizar as previsões referentes aos meses de Verão, nomeadamente aos meses de Junho, Julho e Agosto de 2012.

Como tal, consideremos agora outro conjunto de dados que contém 396 observações a que correspondem a um intervalo temporal que vai desde 1 de Abril de 2011 (Sexta-feira) a 30 de Abril de 2012 (Segunda-feira).



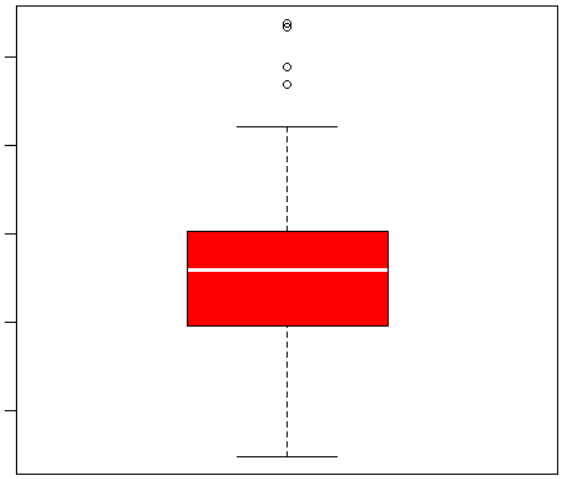
Dado o gráfico anterior, é importante fazemos uma análise inicial à série como fizemos inicialmente para o nosso primeiro conjunto de dados mas como podemos observar, esta série comporta-se de modo semelhante à descrita anteriormente e portanto vamos ter conclusões análogas relativamente ao caso anterior estudado.

Para complementar as informações anteriores, apresento ainda os gráficos do histograma e do QQ-Plot (Normal) deste nosso conjunto de dados considerado.



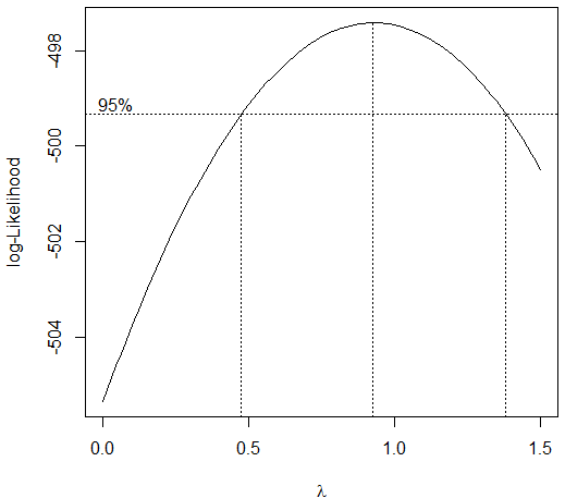
Através dos gráficos do histograma e do QQ-Plot (Normal), podemos dizer que voltamos, novamente, a ter conclusões semelhantes ao caso anterior, uma vez que os dados apresentam ter uma distribuição assimétrica positiva e que portanto aproximam-se de uma distribuição normal e que há evidência no nosso conjunto de dados a presença de outliers.

Para detectar a presença de valores atípicos na série, utilizou-se o gráfico Box-Plot.



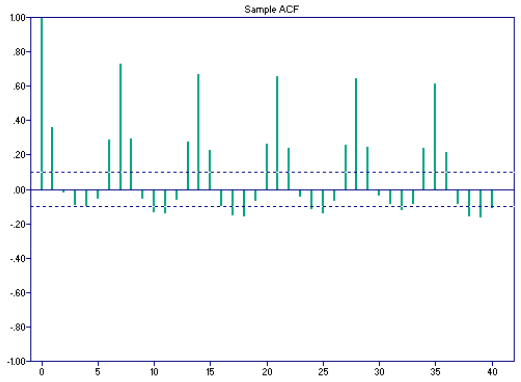
Pela figura anterior, notamos a presença de vários valores que são discrepantes quando comparados com o resto dos valores da amostra. Quanto aos valores mais elevados, estes foram registrados nos dias 28/06/2011 (Terça-Feira), 10/10/2011 (Segunda-Feira), 16/11/2011 (Quarta-Feira) e 10/01/2012 (Terça-feira). Pelo o que tive acesso dos relatórios diários disponibilizados, nestes dias houve problemas nesta linha e portanto houve um aumento no número de chamadas em relação aos dias ditos normais.

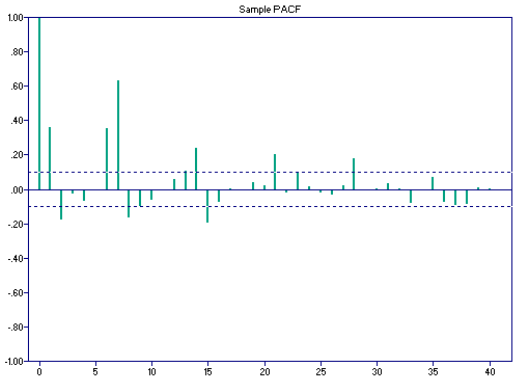
De seguida, iremos aplicar a Transformação Box-Cox para sabermos se é necessário ou não estabilizar a variância da série.



Ao observar o gráfico anterior reparamos que para os valores de lambda situados entre 0.5 e 1.4 encontra-se a solução óptima para um nível de significância de 5%, sendo que devemos considerar para lambda um valor perto de 0.9 visto que atingimos a função log-verosimilhança quando ela é máxima.

Perante esta situação, uma vez que no programa ITSM 2000 a série original tem o valor inicial de lambda igual a 1, não efectuei deste modo a transformação.
Para confirmar se a média da série é constante, veremos a seguir os gráficos das funções de autocorrelação e autocorrelação parcial da mesma.

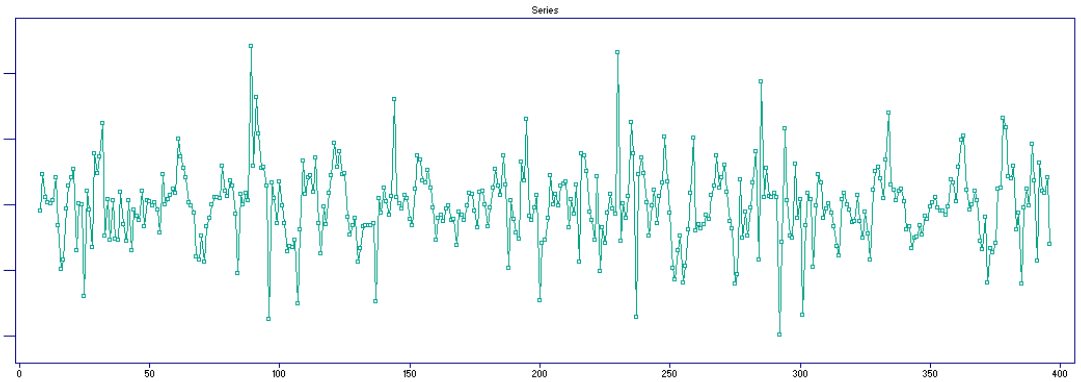


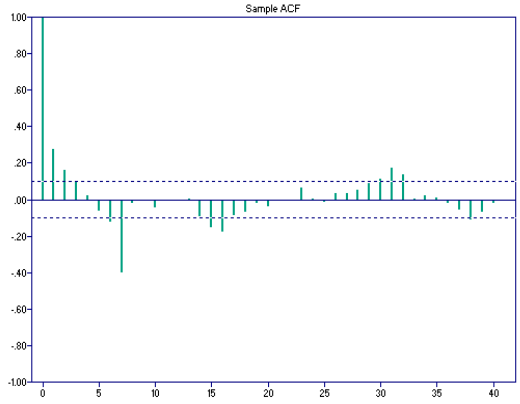


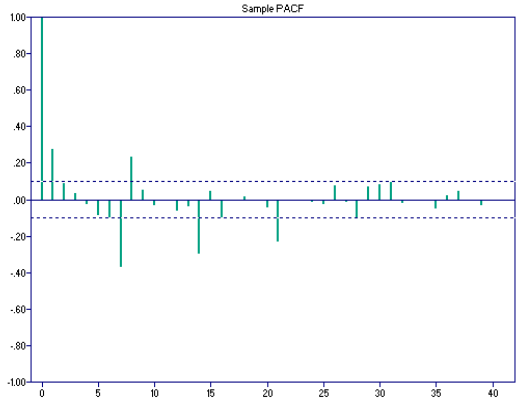
Pelo gráfico da função de autocorrelação verificamos que a série não é estacionária uma vez que apresenta fortes correlações, sendo elas muito significativas devido aos altos valores que tomam. Podemos reparar, principalmente pelo gráfico da FAC, que no lag 7 e nos seus múltiplos há uma grande correlação e portanto, todas estas indicações apontam para a presença de uma componente sazonal semanal tendo por período 7.

Assim, de modo a tornar a série em estacionária, teremos que aplicar uma diferenciação de ordem 7 de modo a podermos eliminar os vestígios de sazonalidade da série.

Apresentamos a seguir, os novos gráficos da série diferenciada e das funções de autocorrelação e autocorrelação parcial da mesma.







Nesta altura estamos nas condições exigidas pela Metodologia de Box-Jenkins pois a nossa série já se encontra estacionária visto que a FAC decresce lentamente para zero. Dito isto, chegamos a uma etapa em que temos de identificar o modelo que melhor descreve o nosso conjunto de dados.

Ao olhar directamente para os gráficos das FAC e FACP anteriores, a identificação do modelo em causa será feita da mesma maneira que no primeiro caso prático demonstrado e portanto diria que o modelo que melhor descreve esta série será dado por um:

$$SARIMA\ (1,0,1) \times (0,1,1)_7$$

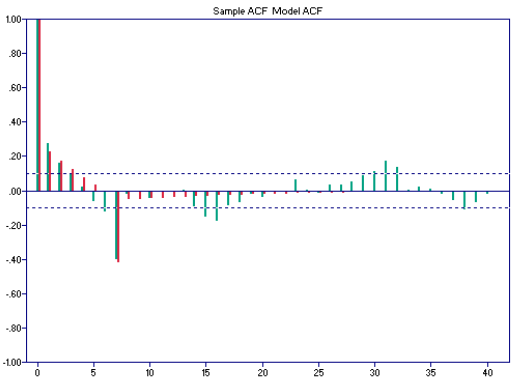
e os respectivos valores de AICC e de BIC são dados por 5837.71 e 5827.21.

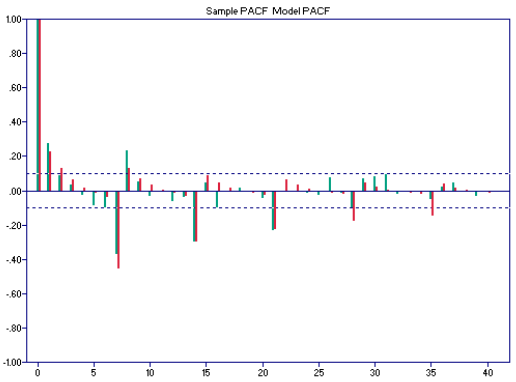
A expressão ajustada do nosso modelo é dada pela seguinte expressão:

$$X_t = 0.9259X_{t-1} + \varepsilon_t - 0.6798\varepsilon_{t-1} - 0.9679\varepsilon_{t-7} + 0.6580\varepsilon_{t-8}$$

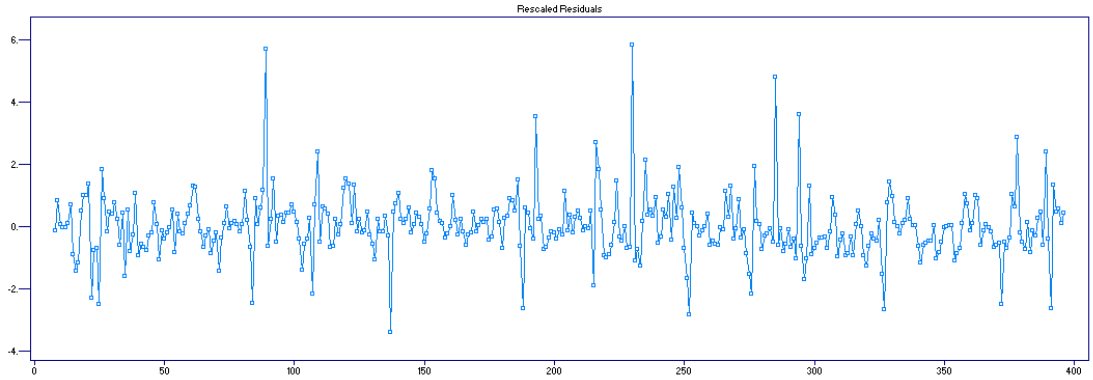
que podemos ver pelo output que se encontra em anexo e outra informação que também podemos retirar de lá é que os 4 parâmetros do modelo são dados como significativos.

A seguir são apresentados os gráficos das FAC e FACP relativas ao modelo teórico e ao modelo estimado que propus.



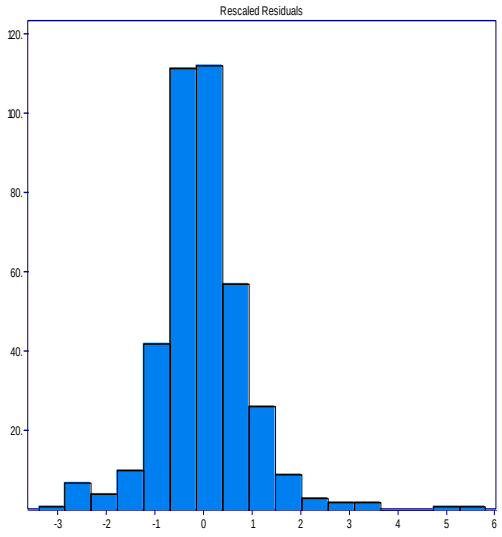


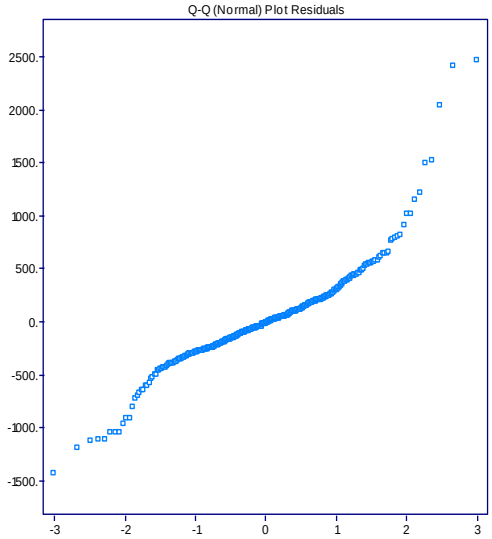
Ao observar os gráficos anteriores poderemos dizer que o nosso modelo se ajusta de uma forma satisfatória ao nosso conjunto de dados. Dito isto, agora só nos falta verificar se os resíduos cumprem os pressupostos relativamente aos modelos lineares e para tal iremos fazer uma análise detalhada aos resíduos do nosso modelo.



Ao examinar a estrutura dos resíduos standardizados poderemos dizer, de uma maneira geral, que estes se distribuem aleatoriamente em torno de zero e portanto podemos considerar os resíduos como independentes.

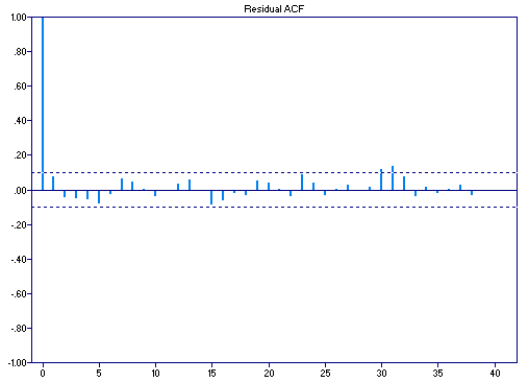
A seguir iremos verificar se os resíduos se aproximam de uma distribuição Normal e para tal é necessário observar o gráfico do Histograma dos resíduos standardizados e o gráfico do QQ-Plot (Normal) dos resíduos.

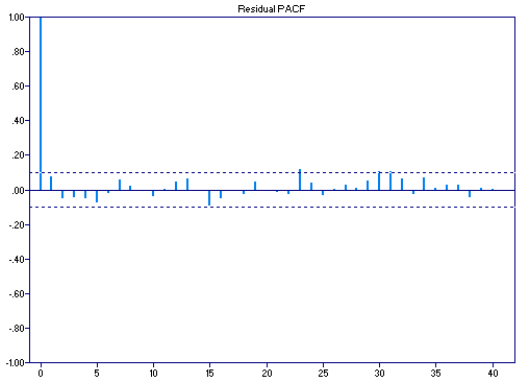




Pelo gráfico do Histograma podemos dizer que os resíduos standardizados aparentam seguir uma distribuição aproximadamente Normal e também podemos retirar a mesma conclusão referente ao segundo gráfico, pois ao observá-lo notamos que a grande maioria dos pontos encontram-se sobre a recta traçada. Como nota ressalvo ainda o facto de podermos observar possíveis candidatos a outliers no gráfico QQ-Plot (Normal).

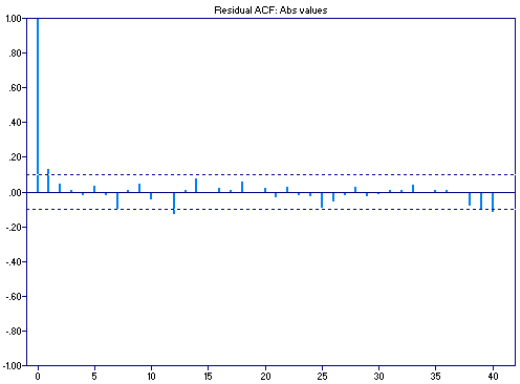
Como ainda nos falta verificar se os resíduos são de facto não correlacionados ou não, apresentamos já a seguir os gráficos das FAC e FACP dos resíduos para verificar se tal condição acontece.

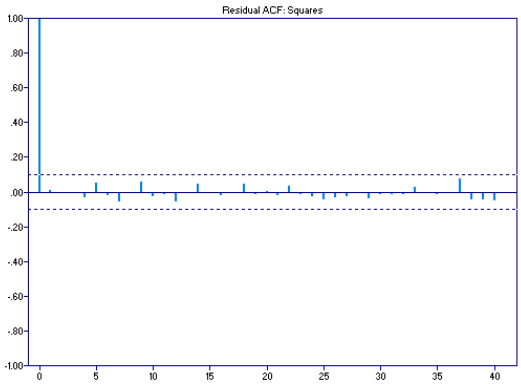




Pela análise dos gráficos anteriores podemos afirmar que os resíduos são não correlacionados, pois podemos reparar que os valores das correlações apresentam-se como sendo muito baixas.

Por último, vamos agora apresentar a última condição que os resíduos têm que satisfazer: se eles são de facto independentes ou não. Dito isto, iremos observar os gráficos da FAC dos valores absolutos e dos valores ao quadrado dos resíduos.





Ao analisar os gráficos anteriores, podemos dizer que os resíduos apresentam-se como sendo independentes, pois não se verifica valores elevados de correlação em ambos os gráficos.

Para provar que, de facto, os resíduos são não correlacionados e que o modelo linear é ou não o mais adequado para o nosso conjunto de dados, pedi ao programa ITSM 2000 para efectuar alguns testes para confirmar tal afirmações. O output desses testes encontra-se em anexo.

Quando foi aplicado o teste de Ljung-Box aos nossos resíduos, este apresentou um valor de p-value de 0.14378, ou seja, não rejeitamos a hipótese nula e portanto

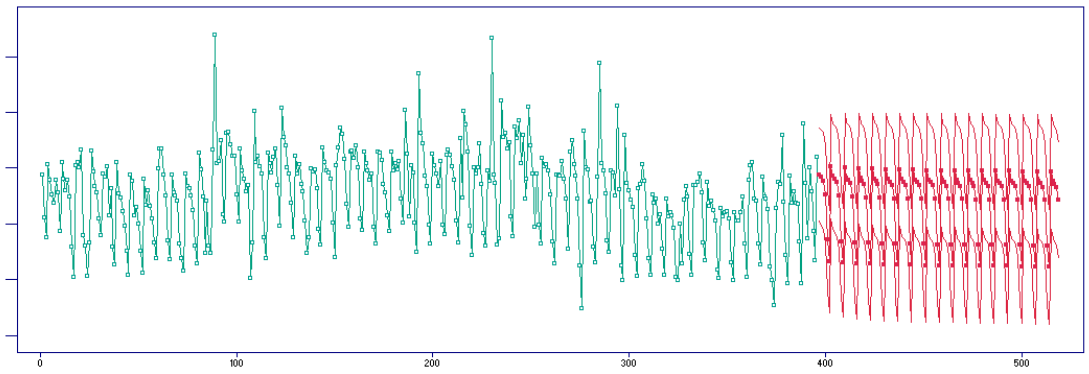
podemos concluir que há evidência para afirmar que os resíduos são não correlacionados, o que vem confirmar a nossa conclusão dada inicialmente.

Quanto ao teste de McLeod-Li, este apresenta um p-value de 0.98590, ou seja, não rejeitamos a hipótese nula e portanto podemos concluir que há evidência para afirmar que o nosso modelo é linear, o que vem novamente confirmar as nossas suposições.

Por tudo o que foi dito anteriormente, sabemos que o modelo linear (SARIMA) é o que melhor descreve o nosso conjunto de dados e também podemos dizer que os pressupostos dos resíduos relativamente aos modelos lineares foram cumpridos. Por conseguinte, estamos nas condições para avançar para a etapa final referente ao cálculo das previsões tendo por base o nosso modelo ajustado.

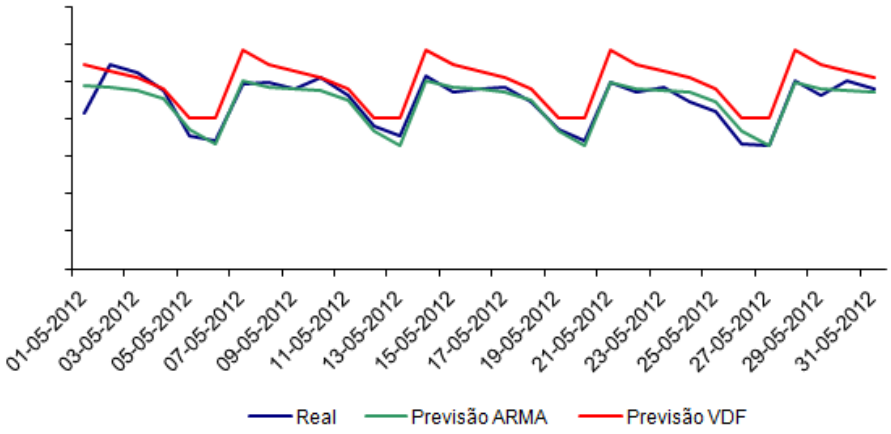
Convém relembrar o leitor que pretendemos prever o número das chamadas atendidas diariamente para esta linha relativos aos meses de Junho, Julho e Agosto de 2012.

Mas uma vez que o nosso conjunto de dados só vai até ao final de Abril de 2012, teremos certamente que realizar as previsões referentes ao mês de Maio de 2012 e portanto vamos prever os valores desta linha para os meses de Maio, Junho, Julho e Agosto de 2012, o que perfaz um total de 123 observações diárias.

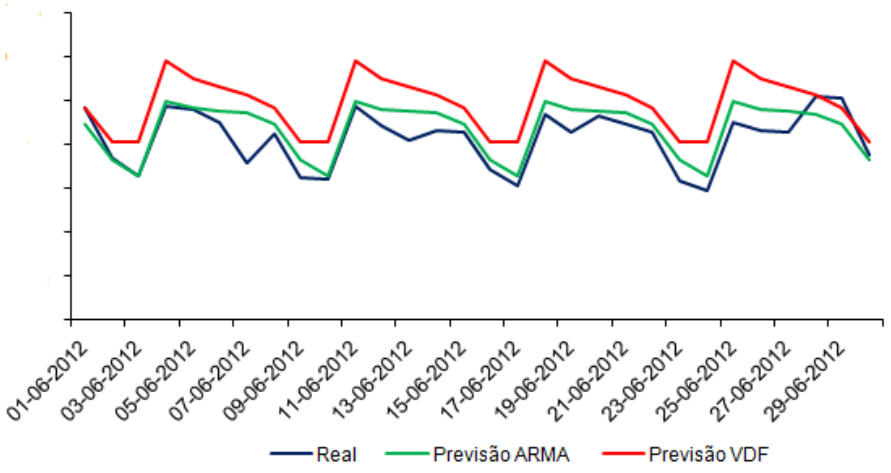


A seguir irei mostrar os gráficos com os valores das previsões obtidas pelo modelo SARIMA, as previsões obtidas pela Vodafone pelo método Holt-Winters Sazonal e os valores reais desta linha para cada um dos meses mencionados e para o total dos 3 meses em questão.

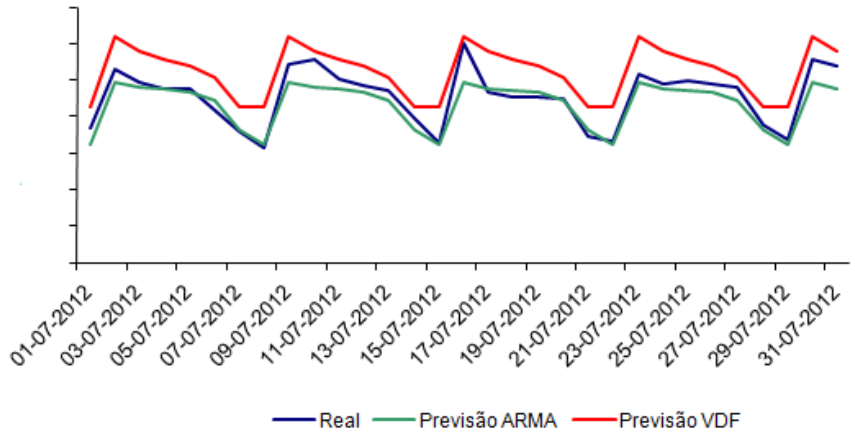
Linha A (Maio 12)



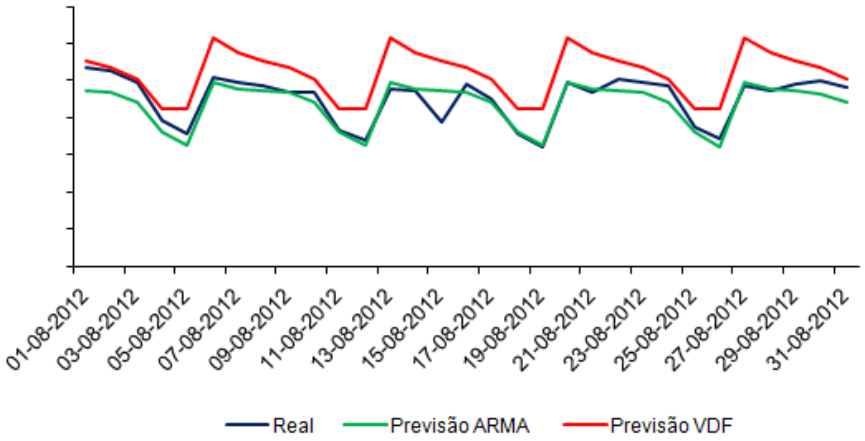
Linha A (Jun 12)



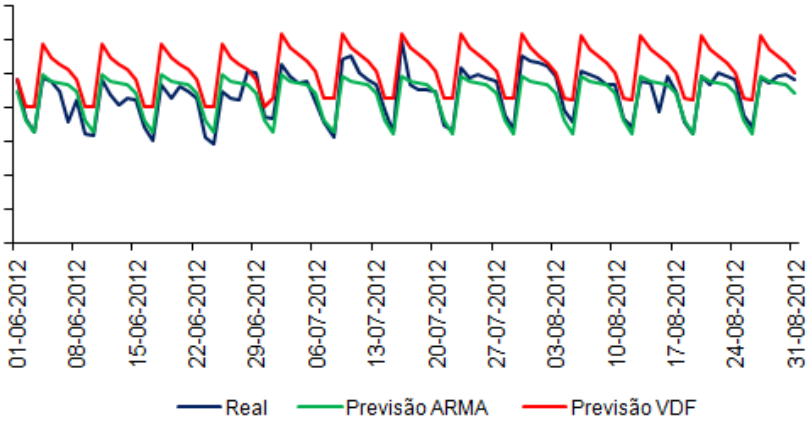
Linha A (Jul 12)



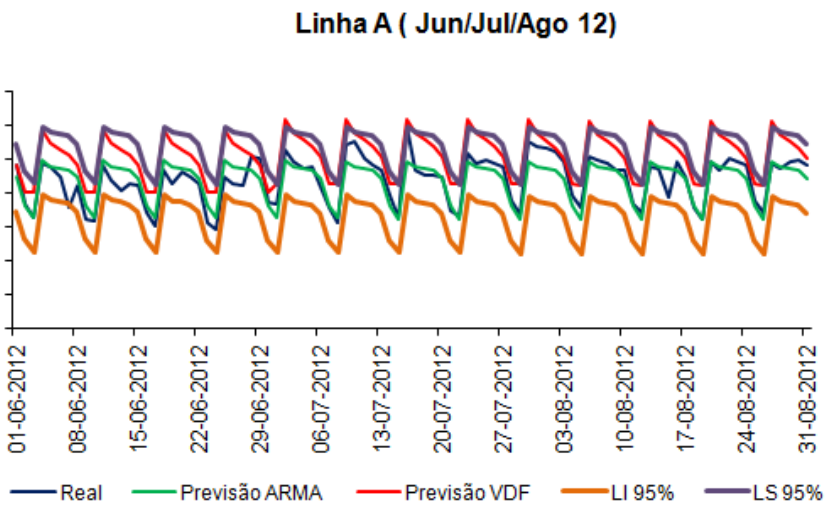
Linha A (Ago 12)



Linha A (Jun/Jul/Ago 12)



Tendo por base o gráfico anterior, vamos agora introduzir nesse mesmo gráfico os intervalos de confiança de 95% dados pela aplicação do meu método de previsão.



Dado o gráfico anterior podemos garantir com 95% de certeza que os valores reais desta linha encontram-se nesse intervalo, sendo que se trata de uma conclusão muito importante a tomar em conta.

Neste instante, estão reunidas todas as condições para realizarmos as comparações necessárias entre os dois métodos apresentados para os vários meses mencionados anteriormente. As tabelas apresentadas a seguir são uma síntese para cada um dos meses e no seu total relativamente a cada método de previsão.

Linha A (Maio 12)											
Desvio Mensal		Desvio Diário		Desvio Diário Absoluto		EQM		EAM		REQM	
ARMA	VDF	ARMA	VDF	ARMA	VDF	ARMA	VDF	ARMA	VDF	ARMA	VDF
0,64	-10,76	0,39	-11,29	3,98	11,71	59806,96	334383,81	176,05	503,44	244,55	578,26

Linha A (Jun 12)											
Desvio Mensal		Desvio Diário		Desvio Diário Absoluto		EQM		EAM		REQM	
ARMA	VDF	ARMA	VDF	ARMA	VDF	ARMA	VDF	ARMA	VDF	ARMA	VDF
-5,43	-19,45	-5,98	-20,36	7,99	20,64	157213,76	803306,78	321,00	814,54	396,50	896,27

Linha A (Jul 12)											
Desvio Mensal		Desvio Diário		Desvio Diário Absoluto		EQM		EAM		REQM	
ARMA	VDF	ARMA	VDF	ARMA	VDF	ARMA	VDF	ARMA	VDF	ARMA	VDF
4,35	-15,31	3,82	-16,23	5,47	16,23	121749,76	550323,33	263,82	695,69	348,93	741,84

Linha A (Ago 12)											
Desvio Mensal		Desvio Diário		Desvio Diário Absoluto		EQM		EAM		REQM	
ARMA	VDF	ARMA	VDF	ARMA	VDF	ARMA	VDF	ARMA	VDF	ARMA	VDF
3,31	-15,24	3,08	-15,87	5,16	15,87	93498,03	621498,03	234,54	687,50	305,77	788,35

Linha A (Jun/Jul/Ago 12)											
Desvio Mensal		Desvio Diário		Desvio Diário Absoluto		EQM		EAM		REQM	
ARMA	VDF	ARMA	VDF	ARMA	VDF	ARMA	VDF	ARMA	VDF	ARMA	VDF
1,00	-16,55	0,38	-17,46	6,19	17,55	123794,51	656800,71	272,60	731,69	351,84	810,43

Ao olharmos para os resultados obtidos podemos dizer que o método ARMA apresenta melhores resultados em comparação com o método da Vodafone, pois o nosso modelo de previsão apresenta menores desvios diários e mensais em relação aos valores reais da linha.

Podemos retirar uma conclusão análoga quanto aos valores apresentados pelas medidas de desempenho porque novamente o método ARMA volta a ser considerado melhor que o método da Vodafone para todos os meses considerados.

Deste modo, torna-se evidente que o melhor método a considerar para esta linha seja o nosso em alternativa ao método de Holt-Winters Sazonal considerado pela Vodafone.

Outro Método Alternativo – Modelo de Regressão

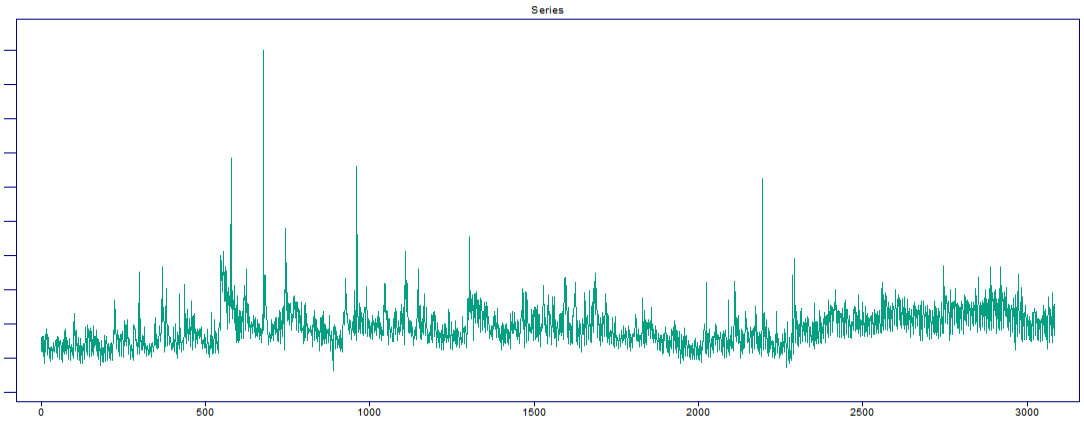
Uma vez que no Verão há normalmente um aumento no número de chamadas realizadas pelos clientes destas empresas de telecomunicações devido a diversos factores, tais como: são os meses em que geralmente as famílias costumam escolher para tirar as suas férias; aonde há fortes campanhas e promoções de Verão lançadas por este tipo de empresas e reciprocamente há uma grande adesão por parte dos seus clientes; existem muitas festas e eventos e o tempo é convidativo para se estar com os amigos e familiares; aonde há uma maior chegada de emigrantes e de turistas ao nosso país, entre outros variadíssimos motivos.
Devido a este fenómeno, colocaram a hipótese de que o modelo anterior não seria capaz de obter valores satisfatórios para os meses de Julho e Agosto para todas as linhas existentes na empresa.

Assim, tornou-se necessário a implementação de um outro processo de previsão de modo a tentar melhorar a questão colocada anteriormente.

Como chamada de atenção refiro que o modelo linear SARIMA (que ajustamos anteriormente) até se ajusta razoavelmente bem a esta linha para os meses de Verão, mas para título de exemplo irei aplicar à mesma o modelo de regressão para tiramos conclusões acerca do conjunto dos três métodos aplicados para esta linha.

Como alternativa, foi-me sugerido aplicar um método de regressão em que os erros são da classe ARMA.

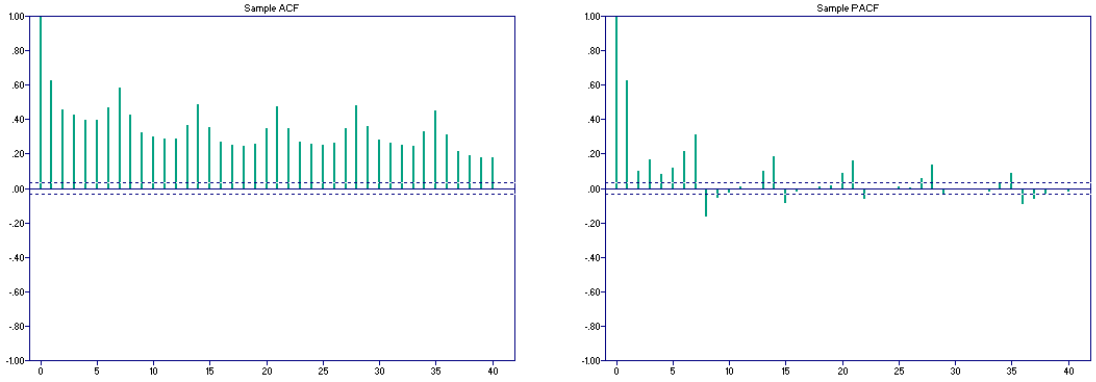
Para tal, considerei para conjunto inicial dos meus dados um total de 3083 observações que correspondem a um intervalo temporal que vai desde 22 de Novembro de 2003 (Domingo) a 30 de Abril de 2012 (Terça-Feira).



Tendo em conta o gráfico anterior podemos retirar várias conclusões da série cronológica em causa, tais como:

- A série apresenta uma média e uma variância não constantes no tempo e portanto devemos estar perante um processo não estacionário certamente;
- A série possui uma componente sazonal, uma vez que no gráfico podemos observar oscilações para cima e para baixo consecutivamente, o que indica que estamos diante de um movimento cíclico (talvez semanal);
- Pode-se considerar o nível da série constante em certos intervalos de tempo;
- A série apresenta uma tendência significativa em alguns períodos (nomeadamente no último ano considerado);
- Existência de valores elevados na série, o que pode indicar a presença de outliers na mesma.

De modo a comprovar que o nível da série não possui uma média constante ao longo do tempo, iremos ver a seguir os gráficos das FAC e FACP do nosso conjunto de dados inicial.

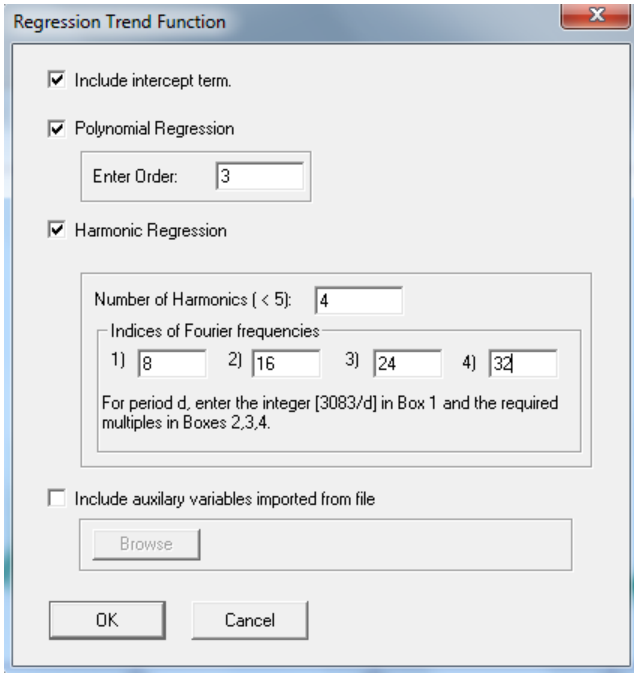


Pelo gráfico da FAC verificamos que a série não é estacionária uma vez que apresenta fortes correlações, sendo elas muito significativas no lag de 7 e em todos os seus múltiplos, o que nos dá a indicação de uma componente sazonal semanal tendo por período o valor de 7.

O passo a realizar a seguir é especificar como vai ser a nossa recta de tendência que iremos ajustar aos nossos dados e portanto teremos que defini-la através do seguinte comando dado pelo programa ITSM 2000:

1º) $Regression \rightarrow Specify$

2º)



Relativamente à figura anterior, podemos fazer algumas considerações de modo a que o leitor não fique com dúvidas sobre quais os campos a preencher. O termo intercept encontra-se seleccionado por defeito e o algarismo 3 foi introduzido uma vez que corresponde ao grau do polinómio que corresponde à recta de tendência, o que indica que esta última será uma função cúbica. Relativamente ao número de harmónicas consideramos 4 e os respondentes índices das frequências de Fourier são dados por 8, 16, 24 e 32. Como temos a informação de que o número total de observações que temos é de 3083, então teremos 3083/365 ≈ 8.47 e portanto consideremos por 4 vezes os múltiplos de 8 (daí a existência de tais valores).

Para estimar a tendência da série, basta realizar o seguinte comando no programa ITSM 2000:

1º) $Regression \rightarrow Estimation \rightarrow Generalized\ LS$

E podemos visualizar a seguir o resultante output da instrucção dada precedentemente:

```
=======================================
ITSM::(Regression estimates)
=======================================

Method: Generalized Least Squares

 Y(t) = M(t) + X(t)

Trend Function:
M(t) = .23906772E+04 t^0 + 4.0407312 t^1 - .0030309314 t^2 + .00000065542765 t^3
    - .14990069E+03 Cos(2*pi*t8/N) + 20.708735 Sin(2*pi*t8/N)
    + .15644268E+03 Cos(2*pi*t16/N) - 13.061779 Sin(2*pi*t16/N)
    - 17.824051 Cos(2*pi*t24/N) + 76.798394 Sin(2*pi*t24/N)
    + 90.871577 Cos(2*pi*t32/N) - .11856703E+03 Sin(2*pi*t32/N)

ARMA Model:
X(t) = Z(t)

WN Variance = 1.000000

        Coeff     Value         Std Error
          0  .23906772E+04     .07315970
          1    4.04073117      .00020561
          2    -.00303093      .00000015
          3     .00000066  .33035154E-10
          4 -.14990069E+03     .02547286
          5   20.70873488      .02587464
          6  .15644268E+03     .02547017
          7  -13.06177863      .02557429
          8  -17.82405105      .02547002
          9   76.79839360      .02551659
         10   90.87157739      .02547000
         11 -.11856703E+03     .02549623
```

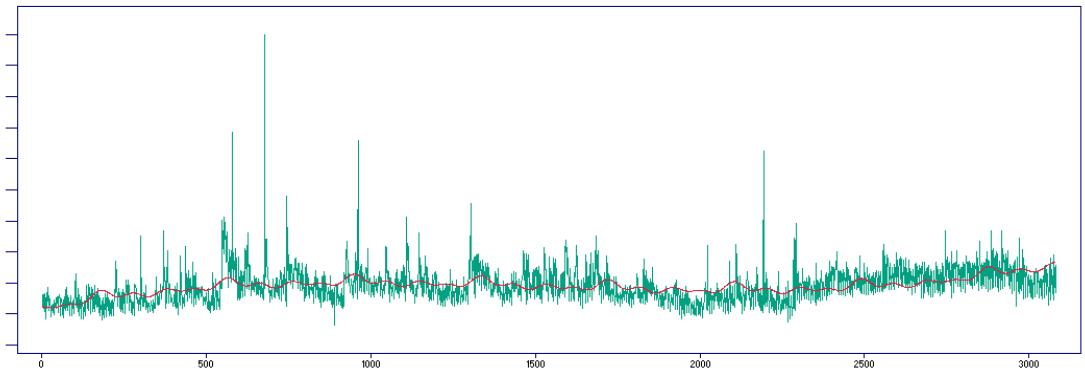
Pelo output acima, vemos que podemos modelar a tendência da nossa série pela seguinte expressão:

$$\hat{T}_t = 2390.6772 + 4.0407312\,t - 0.0030309314\,t^2 + 0.00000065542765\,t^3 - 0.14990069\,cos\left(\frac{2\pi t8}{N}\right) + 20.708735\,sen\left(\frac{2\pi t8}{N}\right) + 156.44268\,cos\left(\frac{2\pi t16}{N}\right) - 13.061779\,sen\left(\frac{2\pi t16}{N}\right) - 17.824051\,cos\left(\frac{2\pi t24}{N}\right) + 76.798394\,sen\left(\frac{2\pi t24}{N}\right) + 90.871577\,cos\left(\frac{2\pi t32}{N}\right) + 118.56703\,sen\left(\frac{2\pi t32}{N}\right)$$

em que o N representa o número total de observações e t=1,…N.

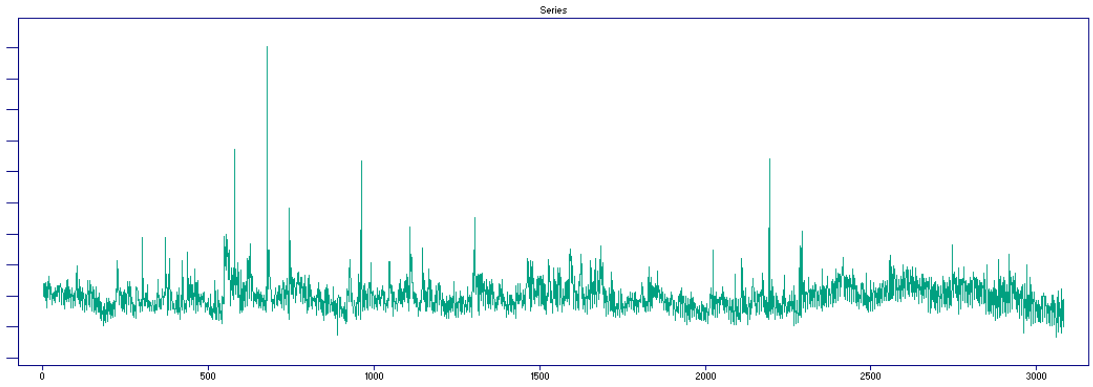
Caso o leitor queira ver graficamente a tendência que foi traçada ao nosso conjunto de dados tem que efectuar o seguinte comando no package estatístico ITSM 2000:

1º) $Regression \rightarrow Show\ fit$

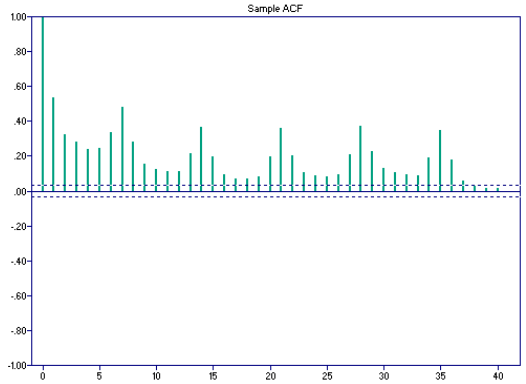


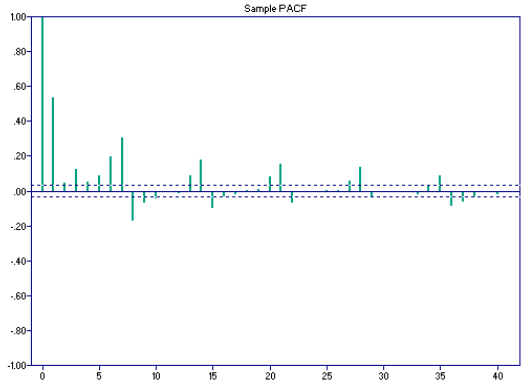
Dito isto, falta-nos agora modelar os erros provenientes do ajustamento da recta de tendência aos nossos dados.

Assim, iremos exportar os resíduos resultantes da aplicação anterior para um ficheiro de texto e abrir esse mesmo ficheiro novamente no programa ITSM 2000.



Mostramos a seguir os gráficos das funções de autocorrelação e autocorrelação parcial referente à série ilustrada em cima.

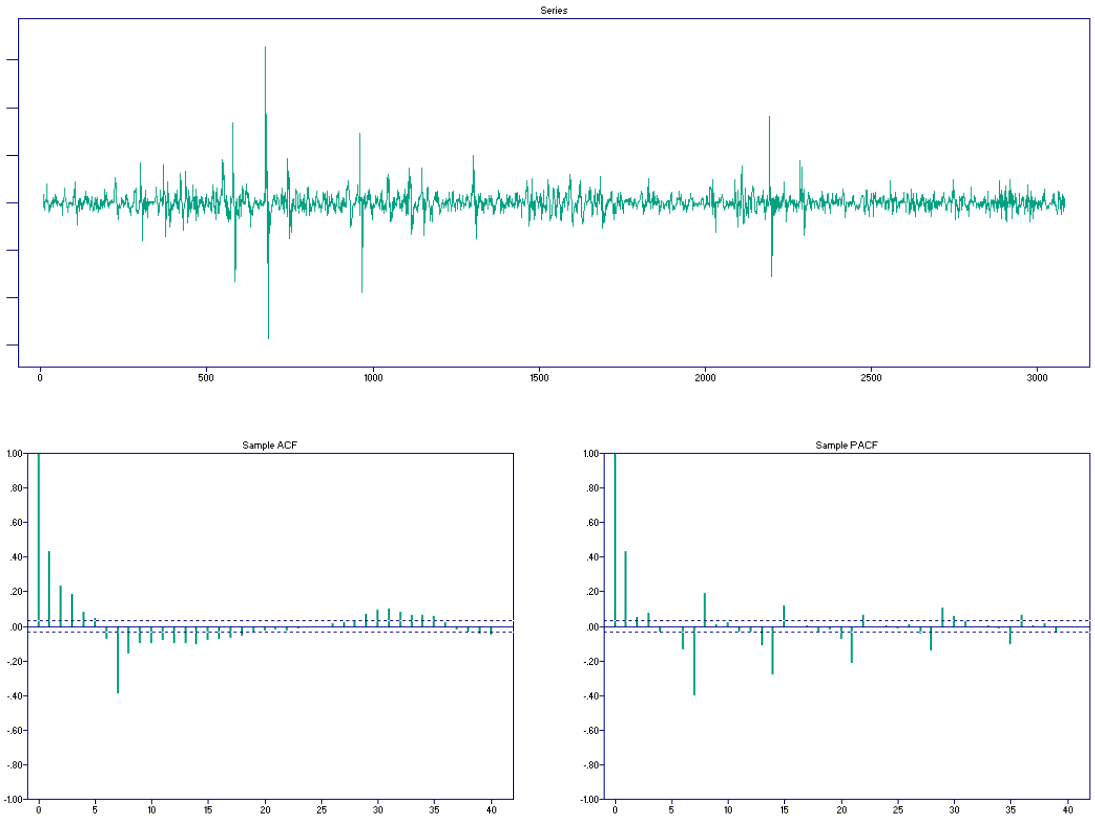




Pelo gráfico da FAC verificamos que a série não é estacionária, pois podemos ver pelo gráfico que os valores das correlações decrescem lentamente para zero. Também podemos observar em ambos os gráficos que no lag 7 e nos seus múltiplos há um elevado valor de correlação e portanto estamos perante a existência de uma componente sazonal semanal de período 7.

Agora iremos aplicar uma diferenciação de ordem 7 de modo a tornar a nossa série em estacionária, eliminando desta maneira quaisquer indícios de sazonalidade da mesma.

Apresentamos a seguir, os novos gráficos da série diferenciada e das funções de autocorrelação e autocorrelação parcial.



Nesta etapa já estamos nas condições para identificar o modelo que melhor descreve o nosso conjunto de dados, pois a série já se encontra estacionarizada.

Ao olhar directamente para os gráficos das FAC e FACP anteriores e ao fim de alguma análise cuidada sobre o nossa série diria que o modelo que melhor descreve esta série será dado por um:

$$SARIMA\ (1,0,2) \times (0,1,1)_7$$

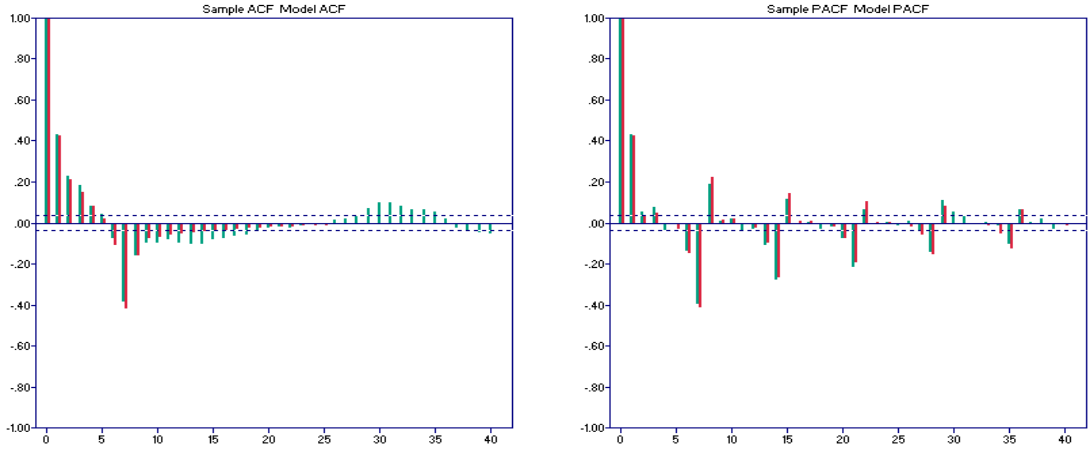
e os valores de AICC e de BIC são dados respectivamente por 48861.4 e 48350.7.

A expressão ajustada do nosso modelo é dada pela seguinte expressão:

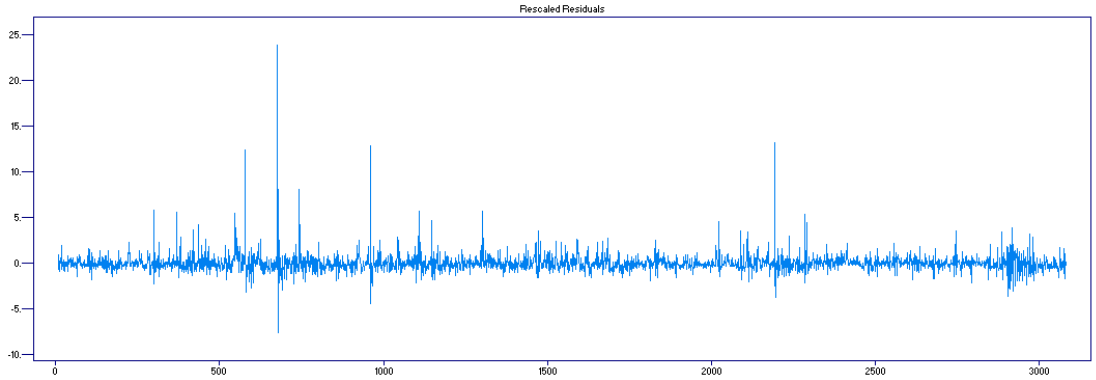
$$X_t = 0.8914X_{t-1} + \varepsilon_t - 0.4340\varepsilon_{t-1} - 0.1531\varepsilon_{t-2} - 1.087\varepsilon_{t-7} + 0.4716\varepsilon_{t-8} + 0.1664\varepsilon_{t-9}$$

que podemos confirmar no output anexado e outra informação a reter é que os 6 parâmetros contidos no modelo são dados como significativos.

A seguir podemos ver os gráficos das FAC e FACP relativas ao modelo teórico e ao modelo estimado alcançado.

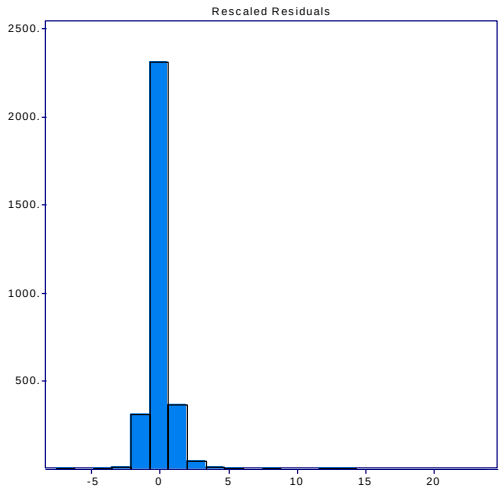


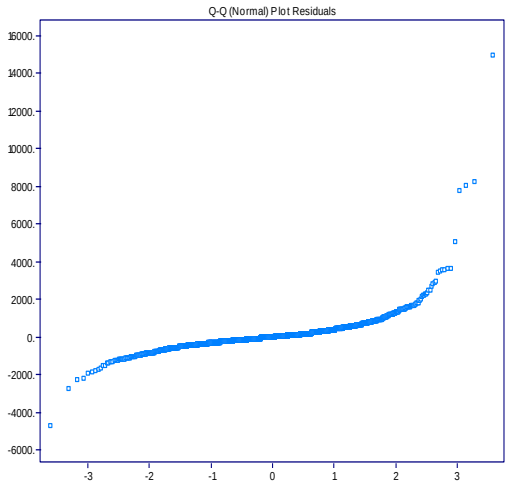
Ao observar os gráficos anteriores podemos dizer que o nosso modelo se ajusta de uma forma aceitável ao nosso conjunto de dados. Assim, agora só nos falta verificar se os resíduos cumprem os pressupostos relativamente aos modelos lineares, mas para isso acontecer é necessário realizar-se uma análise detalhada aos resíduos do nosso modelo.



Ao vermos a estrutura dos resíduos standardizados poderemos dizer que estes se distribuem aleatoriamente à volta do valor de zero e portanto podemos assegurar que os resíduos podem vir a serem independentes.

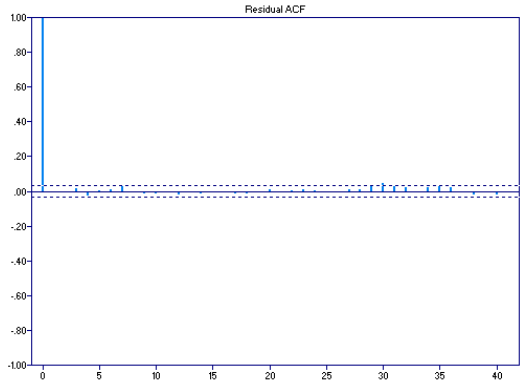
Para se verificar se os resíduos se aproximam de uma distribuição Normal, é preciso observar o gráfico do Histograma dos resíduos standardizados e o gráfico do QQ-Plot (Normal) dos resíduos.

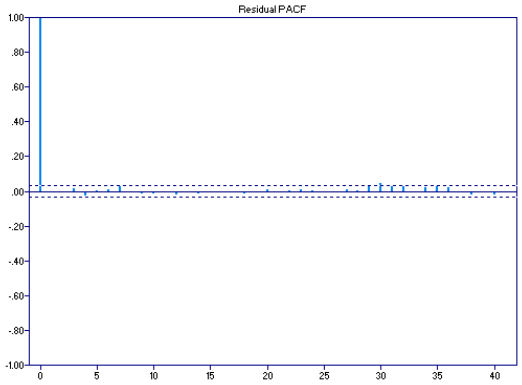




Pelo gráfico do Histograma podemos dizer que deve vir rejeitada a hipótese de que os resíduos standardizados seguem uma distribuição aproximadamente Normal, uma vez que obtemos uma distribuição muito afuzilada e esta apresenta também uma das caudas muito pesada. Pelo gráfico do QQ-Plot (Normal) podemos chegar à mesma conclusão, pois a recta que contém os pontos não aparenta ter um declive de valor 1. Ainda a respeito deste último gráfico podemos observar vários valores que podem ser aspirantes a outliers.

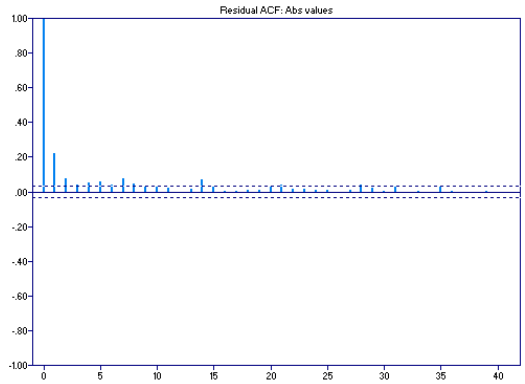
Já a seguir apresento os gráficos das FAC e FACP dos resíduos para averiguar se os resíduos são de facto não correlacionados ou não.

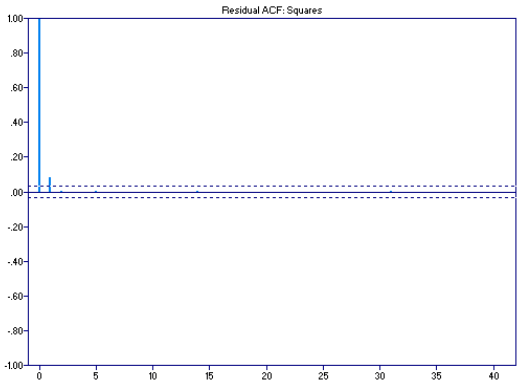




Pela análise dos gráficos anteriores podemos dizer que os resíduos são realmente não correlacionados.

Por fim, falta-nos verificar a última condição que os resíduos têm que satisfazer: se eles são de facto independentes ou não. Para tal, iremos observar já a seguir os gráficos da FAC dos valores absolutos e dos valores ao quadrado dos resíduos.





Ao visualizar estes dois gráficos podemos reparar que não existem praticamente correlações com valores muito elevados e por isso podemos afirmar que os resíduos comportam-se como sendo independentes, apesar de haver um ligeiro valor de correlação mais elevado no lag 1 no primeiro gráfico apresentado mas não me parece ser suficiente para assumir a sua dependência.

Uma vez que assumi que os resíduos eram não correlacionados e que o modelo linear deverá ser o modelo mais indicado para o nosso conjunto de dados, solicitei ao programa ITSM 2000 para realizar alguns testes para confirmar as minhas certezas. Estes testes encontram-se no output em anexo.

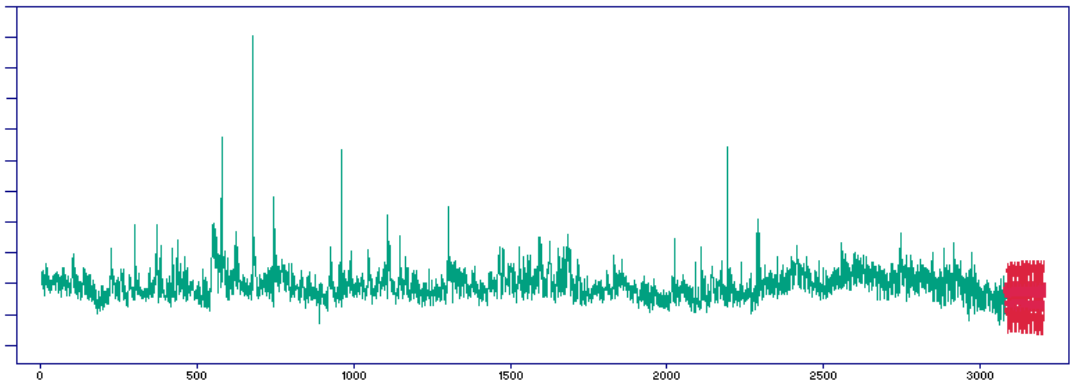
Quando o teste de Ljung-Box é aplicado aos nossos resíduos, este revela um valor para o p-value de 0.64855, ou seja, não rejeitamos a hipótese nula e logo não há evidência para afirmar que os resíduos são correlacionados, o que vem confirmar a nossa suposição dada inicialmente.

Quanto ao teste de McLeod-Li, este apresenta um valor de p-value de 0.38903 o que nos indica que não rejeitamos a hipótese nula e portanto não há evidência para afirmar que o nosso modelo é não-linear.

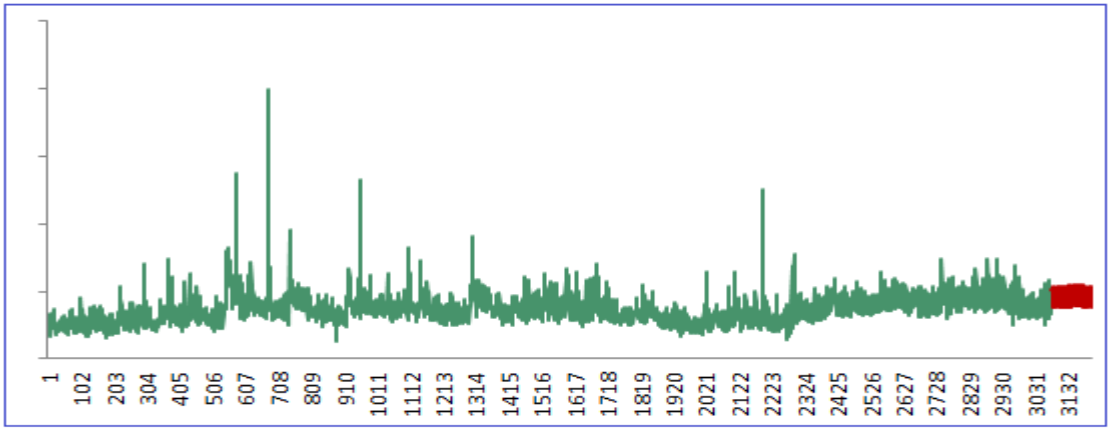
Neste momento, sabemos que os nossos resíduos satisfazem os pressupostos relativamente aos modelos lineares e o modelo SARIMA é considerado o melhor modelo que se ajusta ao nosso conjunto de dados.

Por tudo o que foi dito anteriormente, já podemos avançar para o cálculo das previsões dos erros do nosso modelo estimado. Convém lembrar o leitor que os valores totais das previsões referentes ao nosso modelo de regressão será a soma das previsões dos erros (que veremos a seguir) mais os valores das componentes da tendência que estimamos no início deste processo.

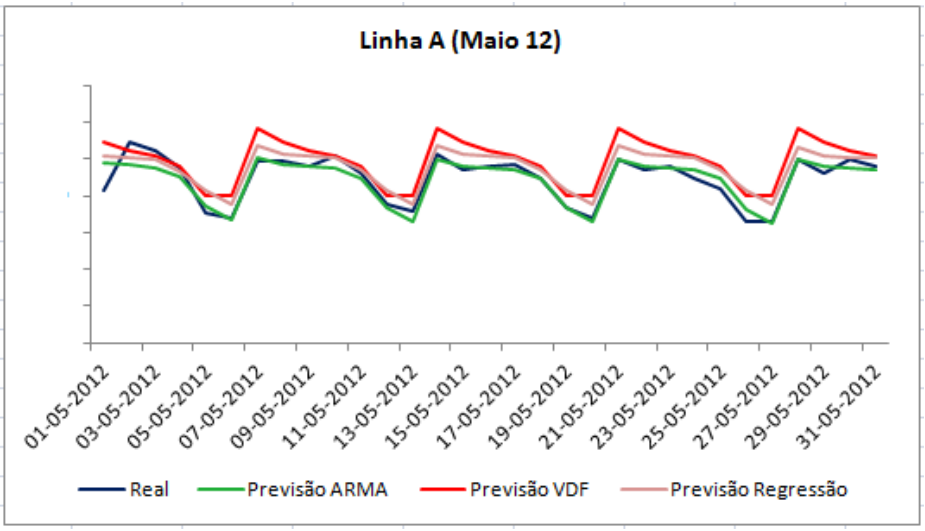
Convém recordar que pretendemos prever o número das chamadas atendidas diariamente para esta linha relativos aos meses de Maio, Junho, Julho e Agosto de 2012 mas agora considerando um modelo de regressão.

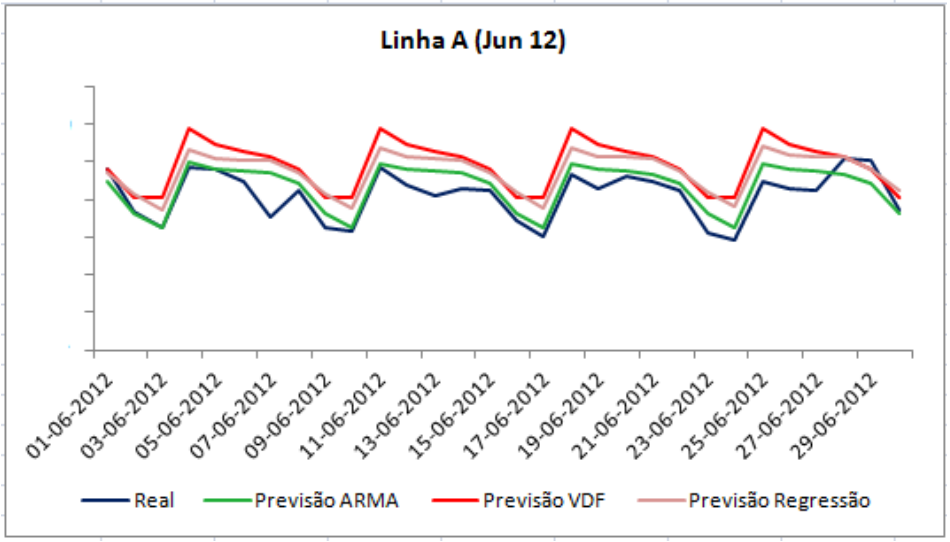


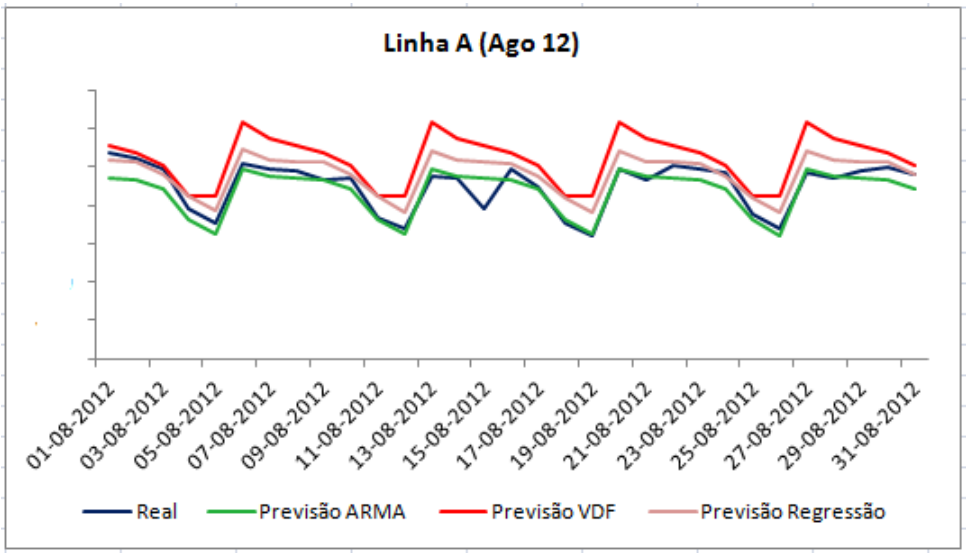
Dado o gráfico anterior vamos agora somar a estas previsões os valores estimados de cada componente da tendência da série inicial, sendo o gráfico com as previsões totais apresentado a seguir.

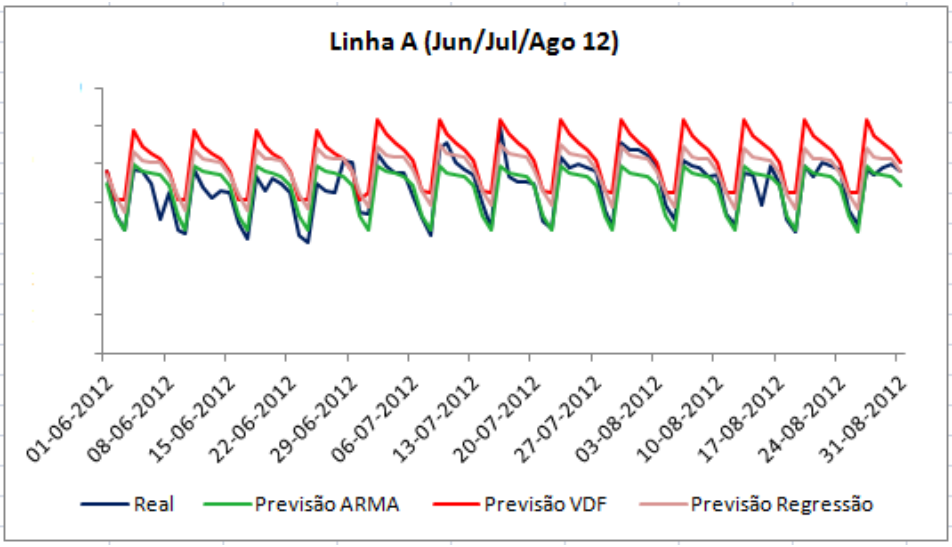


A seguir irei comparar as previsões anteriores que obtive com as previsões obtidas pelo meu método anterior a este, ou seja, pelo método ARMA e também irei comparar com estas duas, os valores das previsões obtidas pela Vodafone pelo método Holt-Winters Sazonal. Por fim, também irei introduzir os valores reais desta linha para cada um dos meses referidos e para o total dos 3 meses em que estamos a estudar.


Linha A (Maio 12)
01-05-2012
03-05-2012
05-05-2012
07-05-2012
09-05-2012
11-05-2012
13-05-2012
15-05-2012
17-05-2012
19-05-2012
21-05-2012
23-05-2012
25-05-2012
27-05-2012
29-05-2012
31-05-2012
Real
Previsão ARMA
Previsão VDF
Previsão Regressão


Linha A (Jun 12)
01-06-2012
03-06-2012
05-06-2012
07-06-2012
09-06-2012
11-06-2012
13-06-2012
15-06-2012
17-06-2012
19-06-2012
21-06-2012
23-06-2012
25-06-2012
27-06-2012
29-06-2012
Real
Previsão ARMA
Previsão VDF
Previsão Regressão


Linha A (Ago 12)
01-08-2012
03-08-2012
05-08-2012
07-08-2012
09-08-2012
11-08-2012
13-08-2012
15-08-2012
17-08-2012
19-08-2012
21-08-2012
23-08-2012
25-08-2012
27-08-2012
29-08-2012
31-08-2012
Real
Previsão ARMA
Previsão VDF
Previsão Regressão



Neste momento já nos encontramos em condições para avaliar o desempenho dos três métodos apresentados para os vários meses citados trás. As tabelas apresentadas a seguir dão-nos um resumo actual dos métodos aplicados.

Linha A (Mai 12)

	Desvio Mensal	Desvio Diário	Desvio Diário Absoluto	EQM	EAM	REQM
ARMA	0,64	0,39	3,98	59806,96	176,05	244,55
VDF	-10,76	-11,29	11,71	334383,81	503,44	578,26
REG.	-6,79	-7,47	8,36	162803,07	350,42	403,49

Linha A (Jun 12)

	Desvio Mensal	Desvio Diário	Desvio Diário Absoluto	EQM	EAM	REQM
ARMA	-5,43	-5,98	7,99	157213,76	321,00	396,50
VDF	-19,45	-20,36	20,64	803306,78	814,54	896,27
REG.	-15,41	-16,52	16,98	519285,24	657,70	720,61

Linha A (Jul 12)

	Desvio Mensal	Desvio Diário	Desvio Diário Absoluto	EQM	EAM	REQM
ARMA	4,35	3,82	5,47	121749,76	263,82	348,93
VDF	-15,31	-16,23	16,23	550323,33	695,69	741,84
REG.	-7,44	-8,61	9,73	209853,29	402,21	458,10

Linha A (Ago 12)

	Desvio Mensal	Desvio Diário	Desvio Diário Absoluto	EQM	EAM	REQM
ARMA	3,31	3,08	5,16	93498,03	234,54	305,77
VDF	-15,24	-15,87	15,87	621498,03	687,50	788,35
REG.	-7,10	-7,79	8,41	182531,49	352,08	427,24

Linha A (Jun/Jul/Ago 12)						
	Desvio Mensal	Desvio Diário	Desvio Diário Absoluto	EQM	EAM	REQM
ARMA	1,00	0,38	6,19	123794,51	272,60	351,84
VDF	-16,55	-17,46	17,55	656800,71	731,69	810,43
REG.	-9,76	-10,91	11,65	301548,75	468,63	549,13

Após uma breve observação aos quadros anteriores verifico que o método que se destaca mais pelos seus valores de desvio diário e mensal é o método ARMA, pois este apresenta menores desvios em relação aos valores reais da linha em questão. Podemos dizer de igual modo que relativamente às medidas de desempenho, o modelo que mostra melhores resultados foi o modelo linear SARIMA. Assim, quanto aos três métodos utilizados, o segundo que apresenta melhores resultados é o modelo de Regressão e por último é o método de Holt-Winters Sazonal aplicado pela empresa Vodafone. Desta maneira, podemos constatar que o modelo de Regressão não veio trazer grandes melhorias relativamente ao modelo linear SARIMA que ajustamos inicialmente para os meses de Verão.

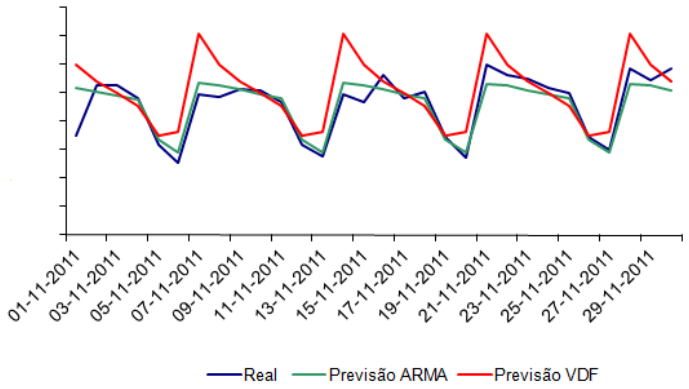
Outras linhas consideradas

Apresentarei a seguir, um breve resumo para as restantes linhas que faltam considerar, pois a aplicação do método faz-se sempre do mesmo modo.
Na tabela abaixo, podemos ver para cada linha qual o modelo que considerei e o respectivo valor de AICC.

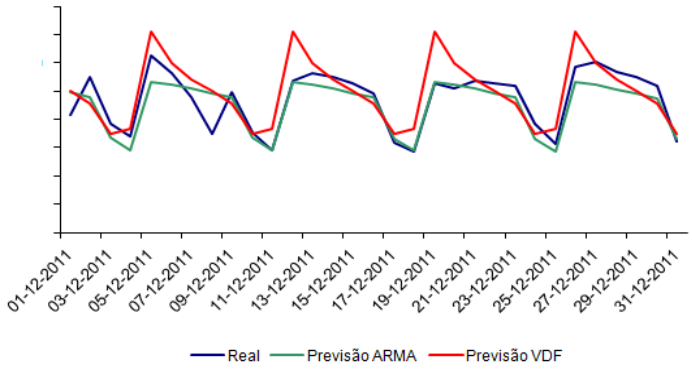
Linhas	Efectou-se Transformação Box-Cox?	Modelo Ajustado	AICC
B	Não	SARIMA(1,0,1)x(0,1,1)$_7$	6373.50
C	Não	SARIMA(1,0,3)x(0,1,1)$_7$	-297.912
D	Não	SARIMA(1,0,1)x(0,1,2)$_7$	5403.94
E	Não	SARIMA(1,0,2)x(0,1,1)$_7$	3443.04
F	Não	SARIMA(1,0,1)x(0,1,2)$_7$	5391.12
G	Não	SARIMA(1,0,2)x(0,1,1)$_7$	4584.37
H	Não	SARIMA(1,0,1)x(0,1,1)$_7$	1793.67
I	Não	SARIMA(1,0,1)x(0,1,2)$_7$	-496.58
J	Não	SARIMA(1,0,1)x(0,1,2)$_7$	5350.39
K	Não	SARIMA(0,0,1)x(0,1,1)$_7$	5852.30

Veremos a seguir a representação gráfica das previsões feitas pelo método ARMA e pelo método Holt-Winters Sazonal (VDF) para Novembro e Dezembro de 2011 e Janeiro de 2012, relativas às linhas apresentadas anteriormente.

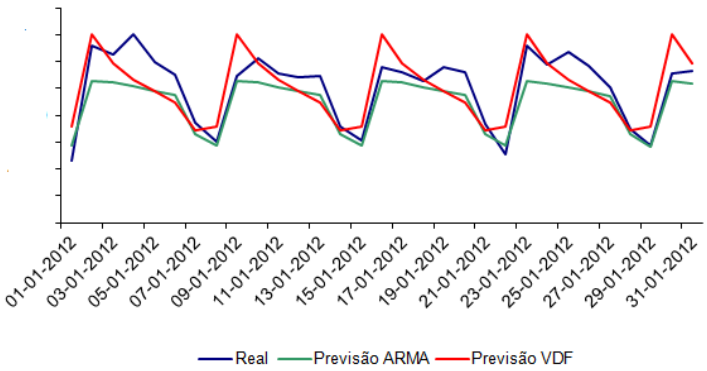
Linha B (Nov 11)



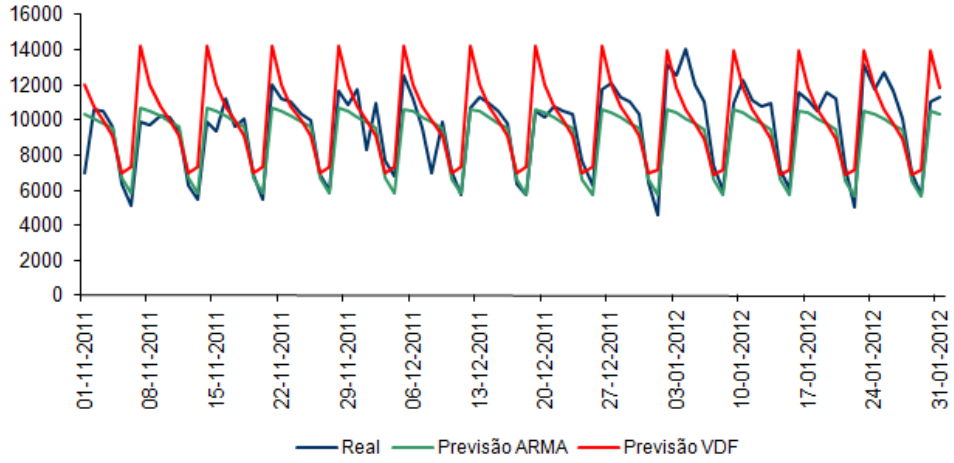
Linha B (Dez 11)



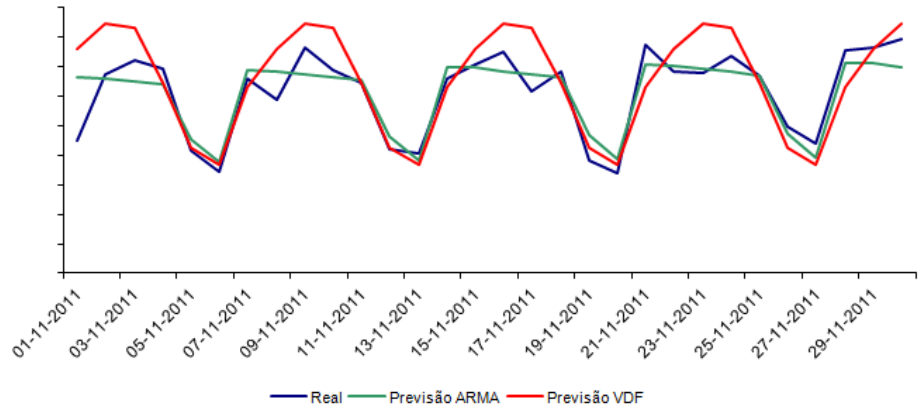
Linha B (Jan 12)



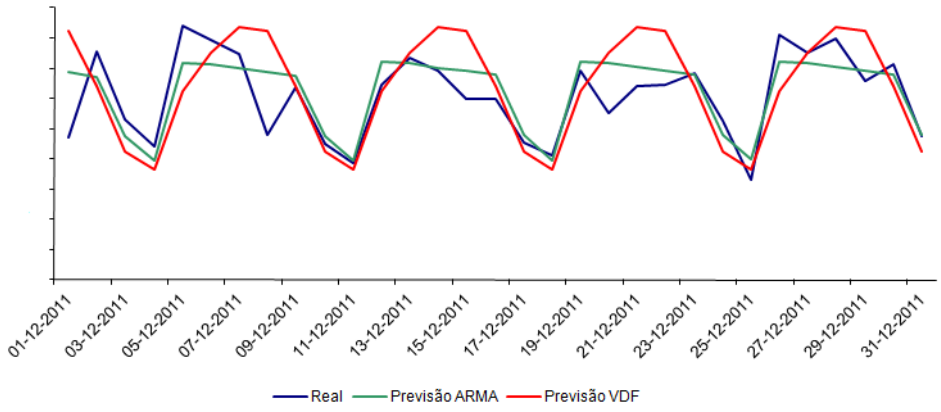
Linha B (Nov / Dez 11 e Jan 12)



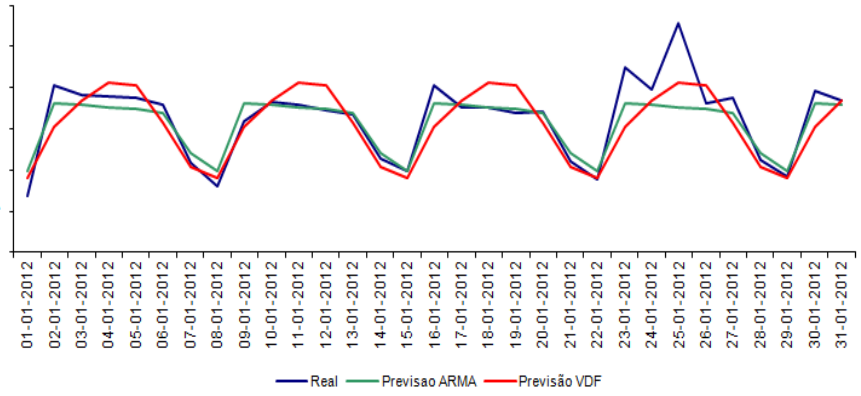
Linha C (Nov 11)



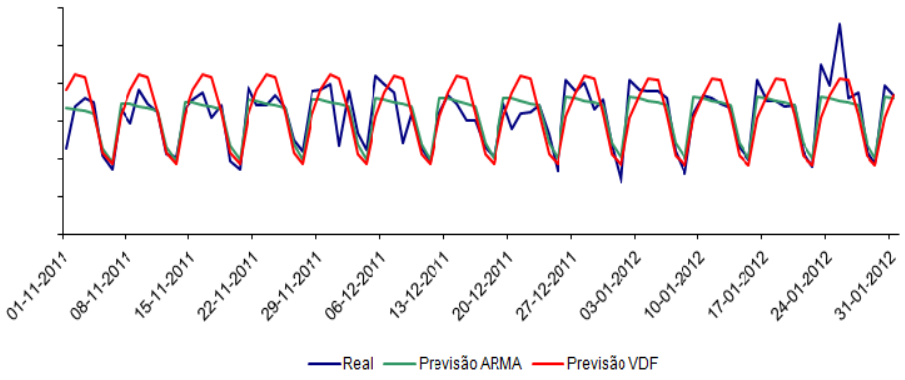
Linha C (Dez 11)



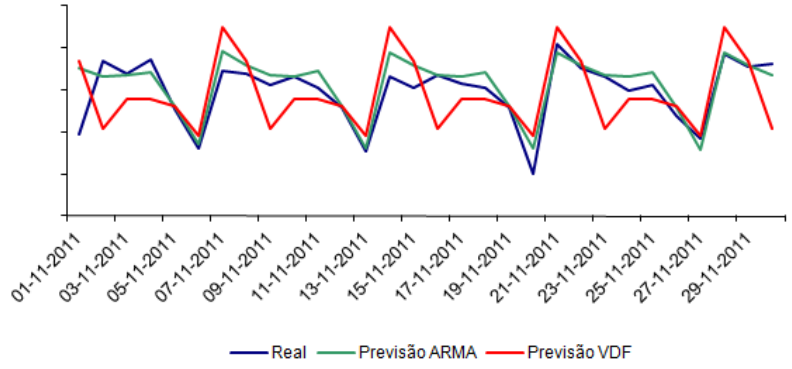
Linha C (Jan 12)



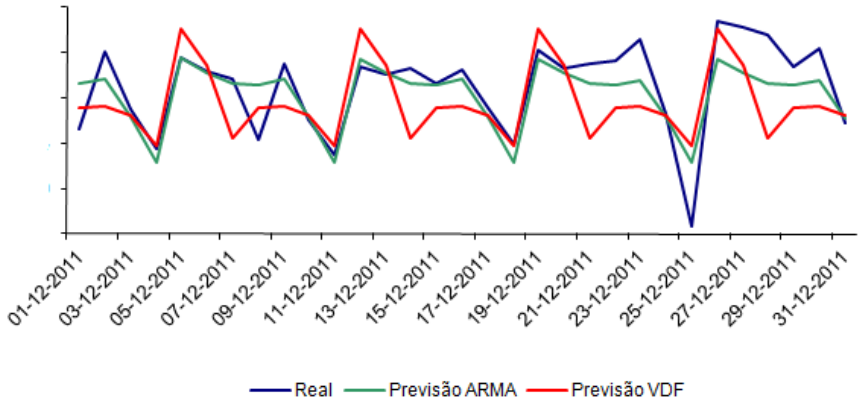
Linha C (Nov/Dez 11 e Jan 12)



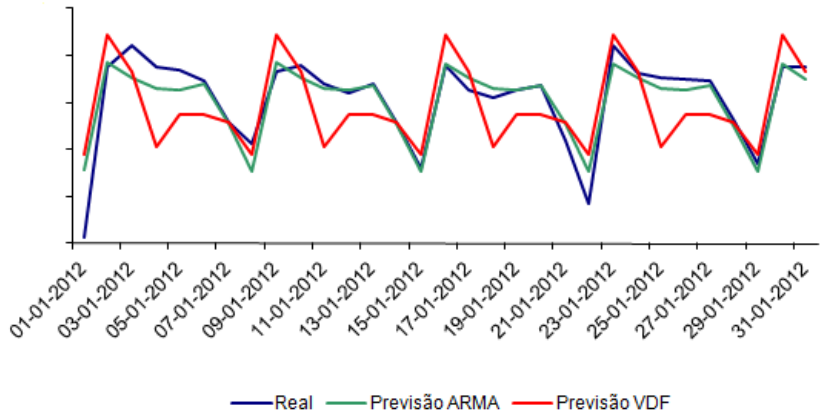
Linha D (Nov 11)



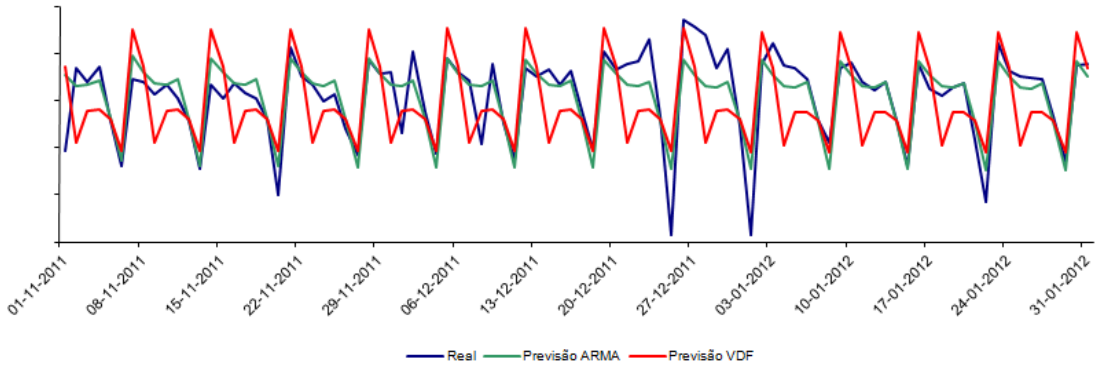
Linha D (Dez 11)



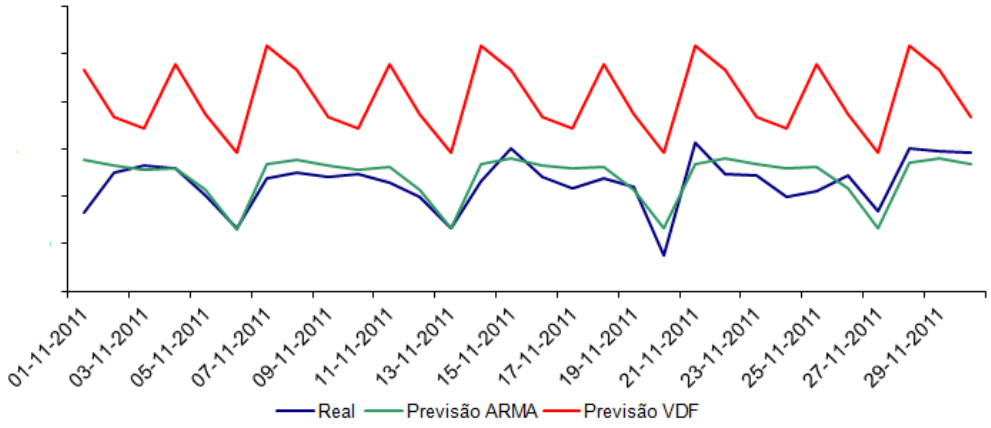
Linha D (Jan 12)



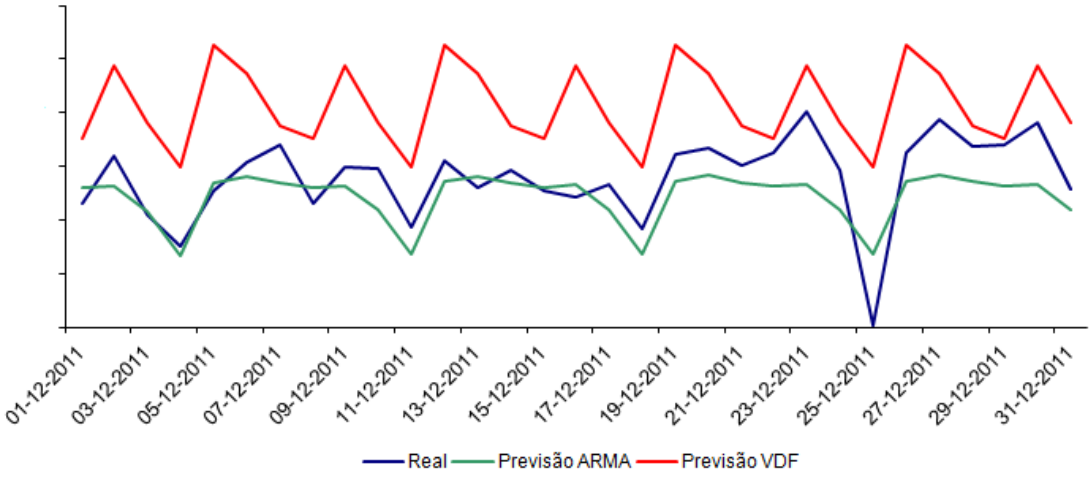
Linha D (Nov/Dez 11 e Jan 12)



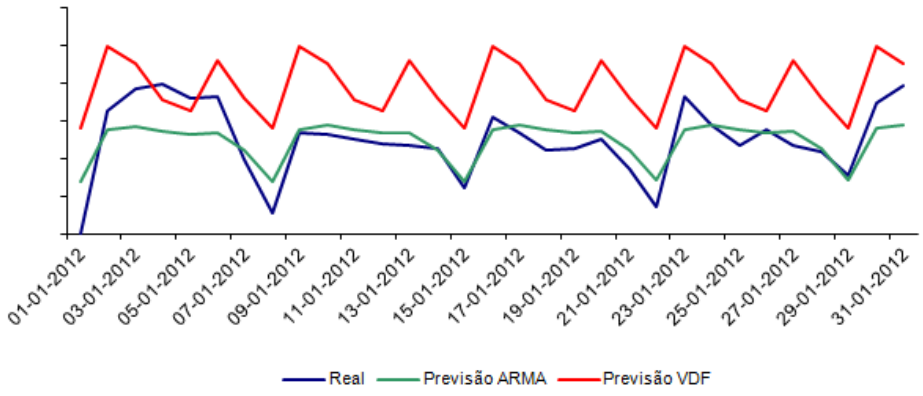
Linha E (Nov 11)



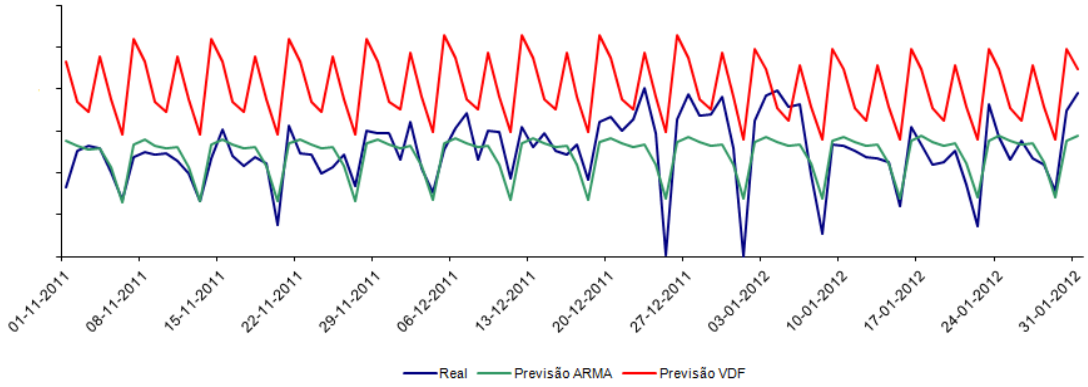
Linha E (Dez 11)



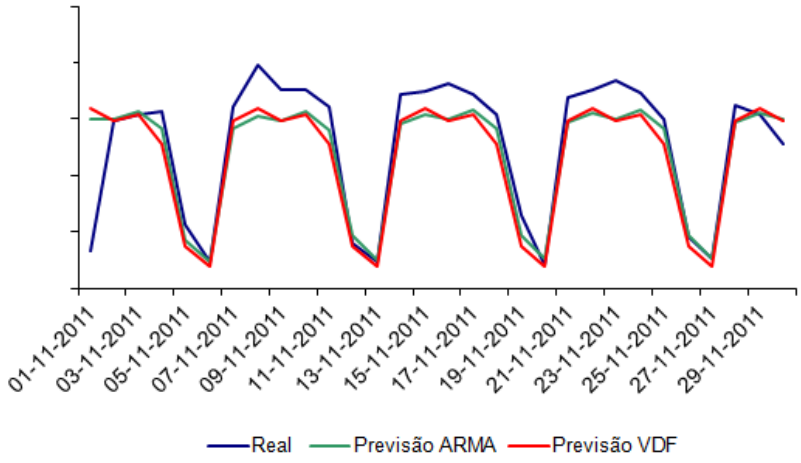
Linha E (Jan 12)



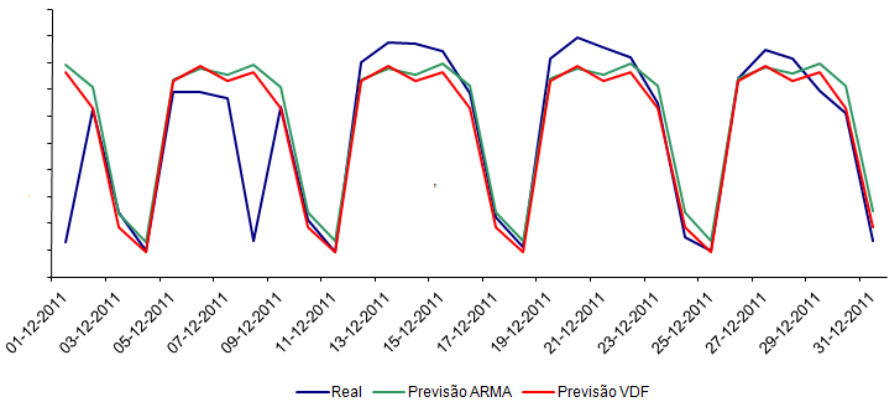
Linha E (Nov/Dez 11 e Jan 12)



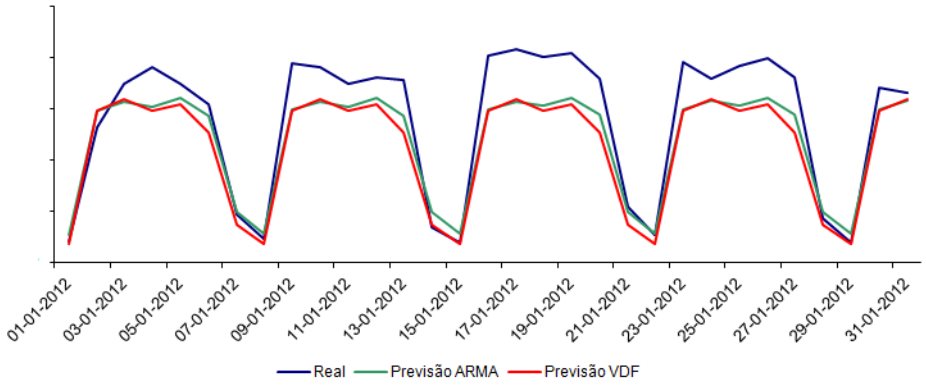
Linha F (Nov 11)



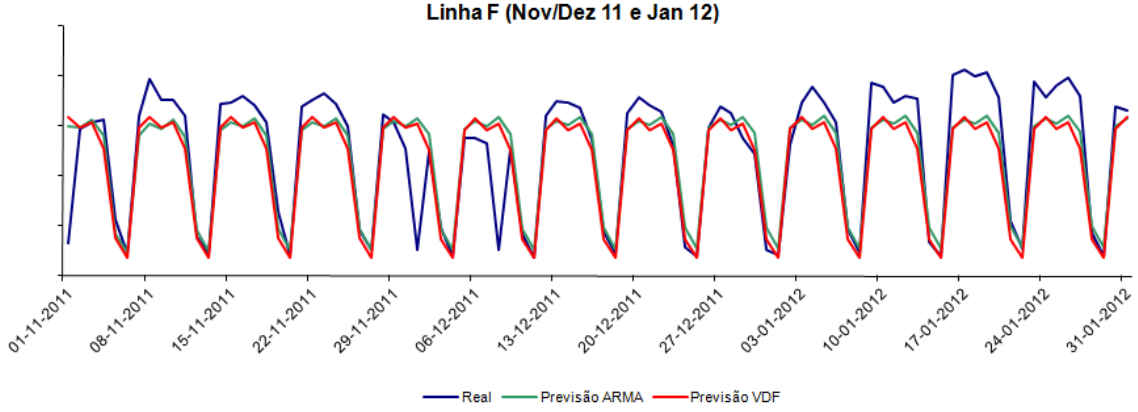
Linha F (Dez 11)



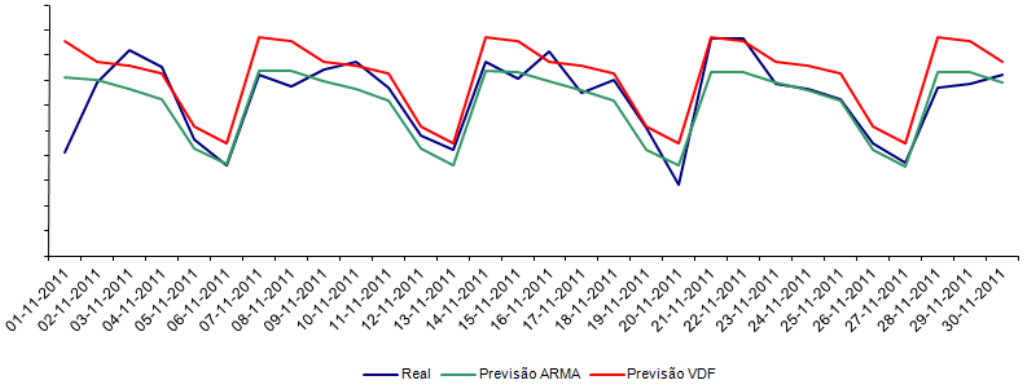
Linha F (Jan 12)



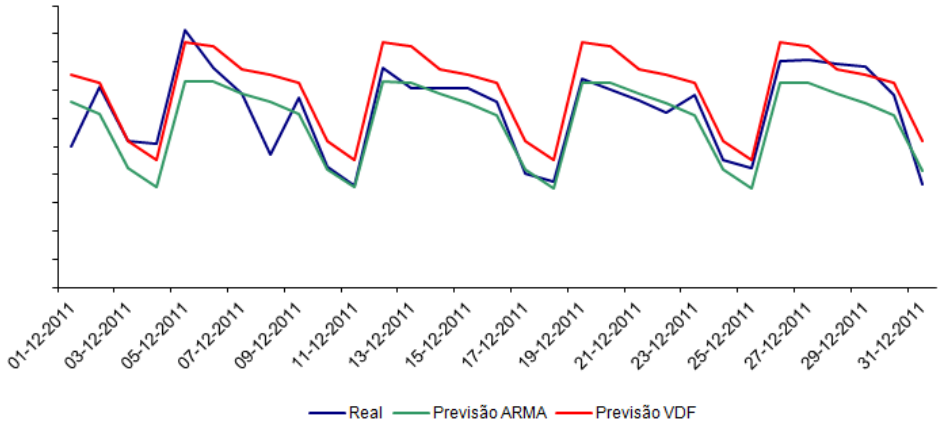
Linha F (Nov/Dez 11 e Jan 12)



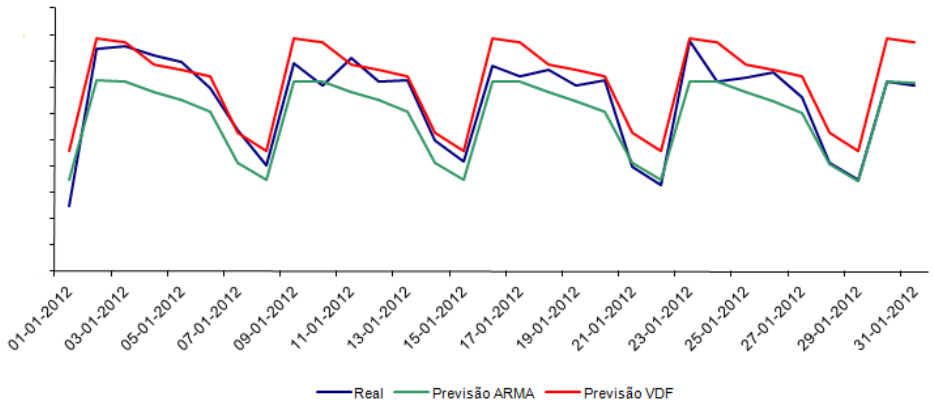
Linha G (Nov 11)



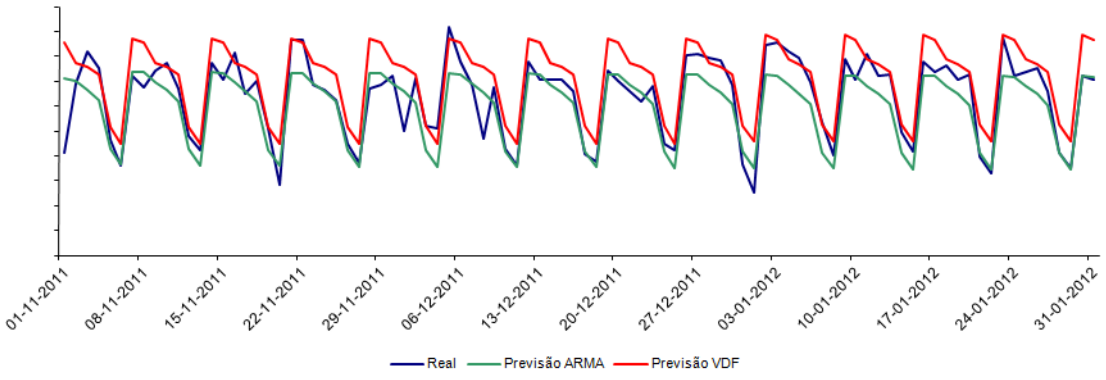
Linha G (Dez 11)

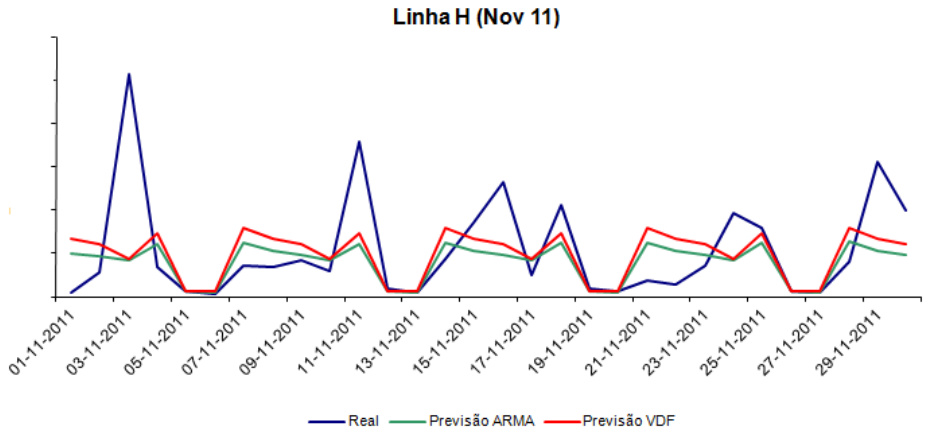


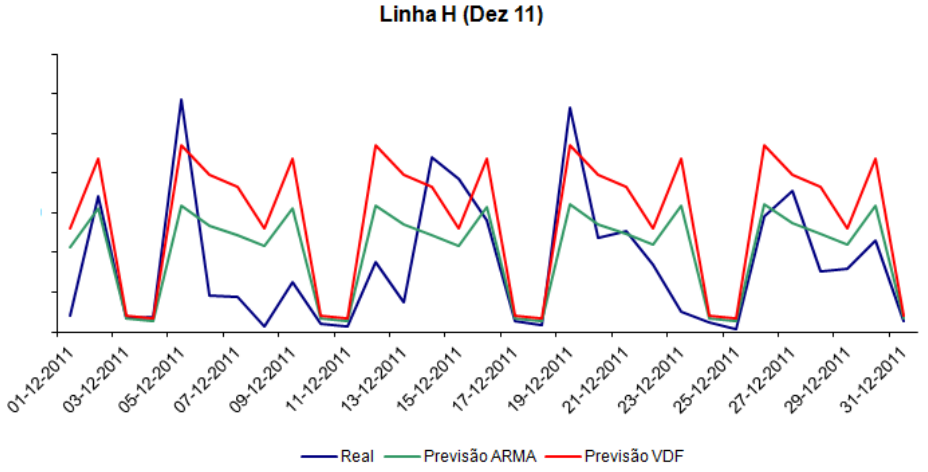
Linha G (Jan 12)

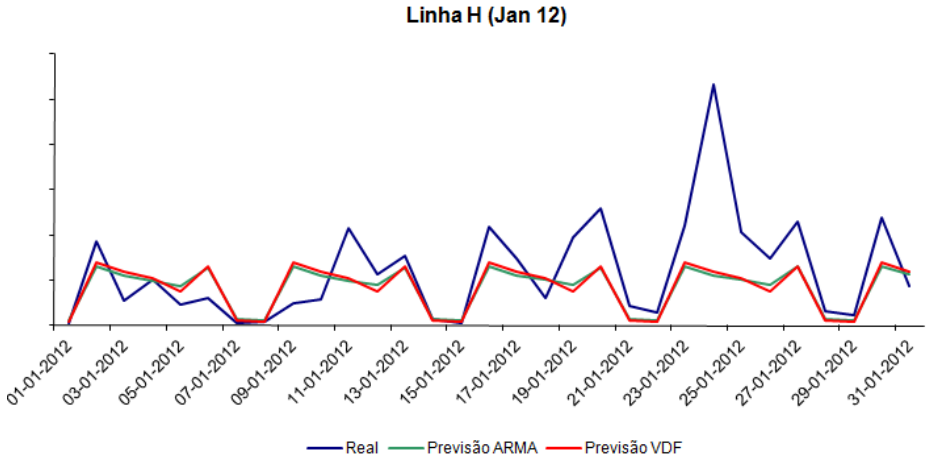


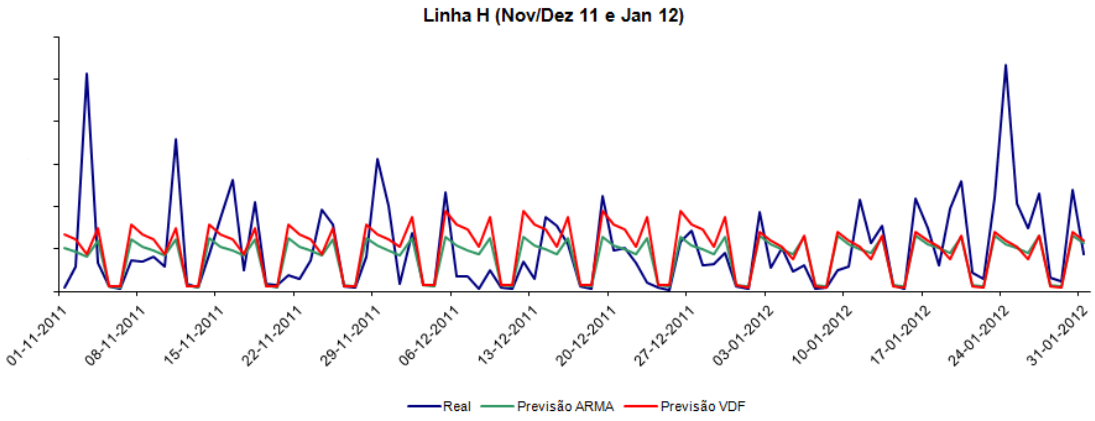
Linha G (Nov/Dez 11 e Jan 12)

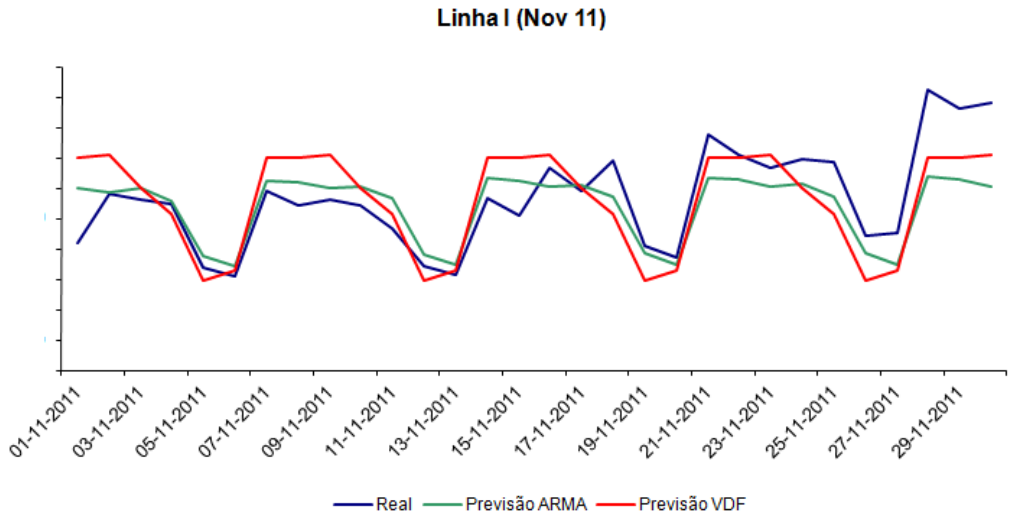


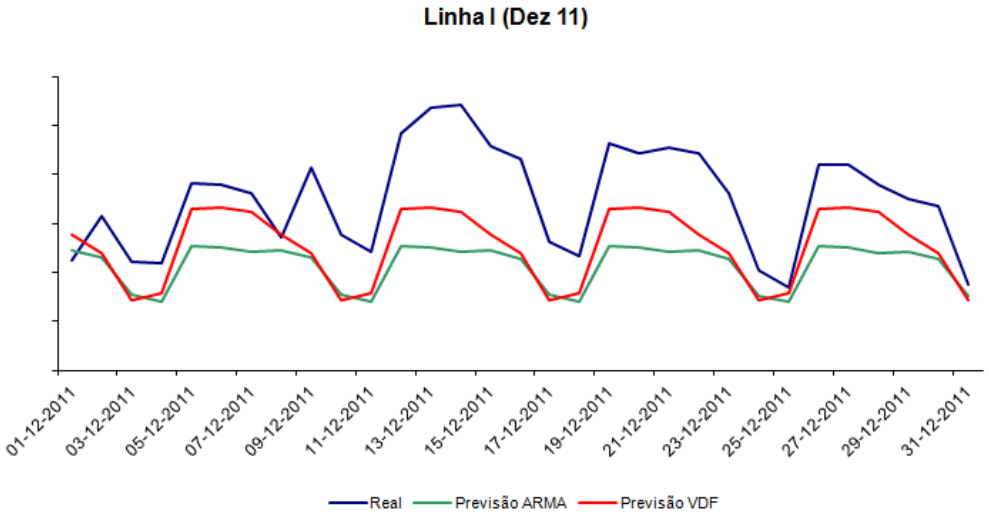

Linha H (Nov 11)
01-11-2011
03-11-2011
05-11-2011
07-11-2011
09-11-2011
11-11-2011
13-11-2011
15-11-2011
17-11-2011
19-11-2011
21-11-2011
23-11-2011
25-11-2011
27-11-2011
29-11-2011
Real
Previsão ARMA
Previsão VDF


Linha H (Dez 11)
01-12-2011
03-12-2011
05-12-2011
07-12-2011
09-12-2011
11-12-2011
13-12-2011
15-12-2011
17-12-2011
19-12-2011
21-12-2011
23-12-2011
25-12-2011
27-12-2011
29-12-2011
31-12-2011
Real
Previsão ARMA
Previsão VDF

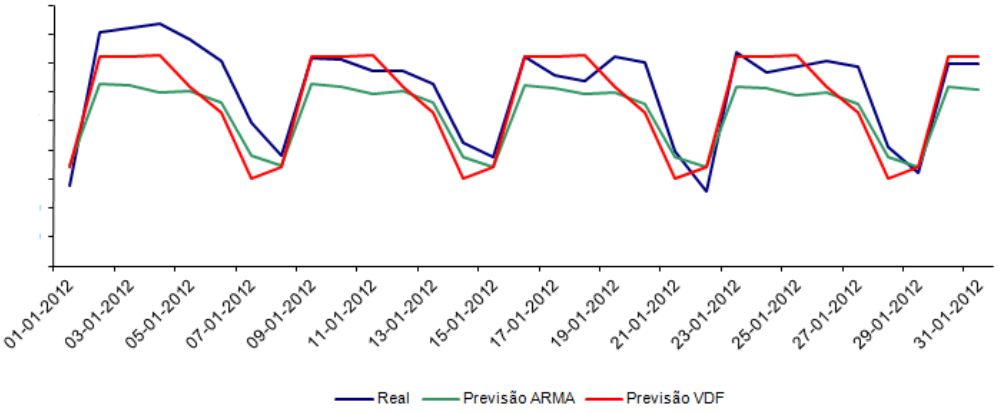

Linha H (Jan 12)
01-01-2012
03-01-2012
05-01-2012
07-01-2012
09-01-2012
11-01-2012
13-01-2012
15-01-2012
17-01-2012
19-01-2012
21-01-2012
23-01-2012
25-01-2012
27-01-2012
29-01-2012
31-01-2012
Real
Previsão ARMA
Previsão VDF


Linha H (Nov/Dez 11 e Jan 12)
01-11-2011
08-11-2011
15-11-2011
22-11-2011
29-11-2011
06-12-2011
13-12-2011
20-12-2011
27-12-2011
03-01-2012
10-01-2012
17-01-2012
24-01-2012
31-01-2012
Real
Previsão ARMA
Previsão VDF

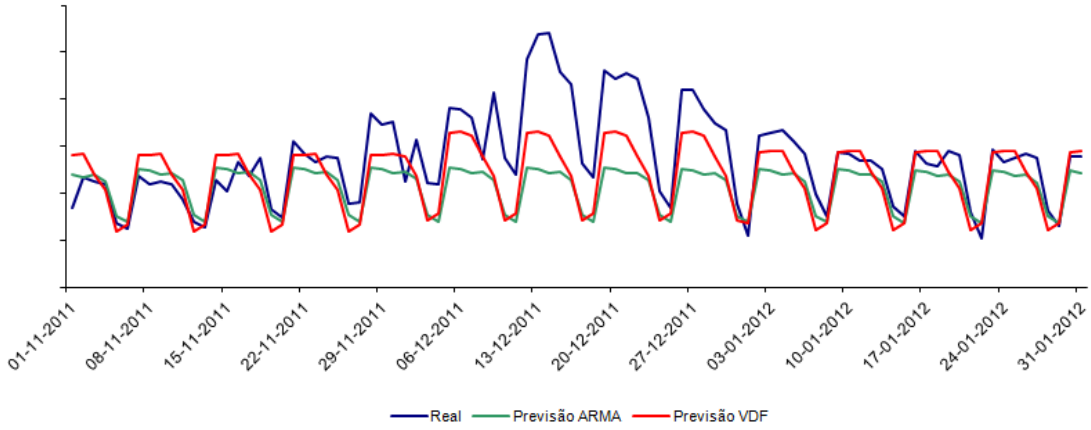

Linha I (Nov 11)
01-11-2011
03-11-2011
05-11-2011
07-11-2011
09-11-2011
11-11-2011
13-11-2011
15-11-2011
17-11-2011
19-11-2011
21-11-2011
23-11-2011
25-11-2011
27-11-2011
29-11-2011
Real
Previsão ARMA
Previsão VDF


Linha I (Dez 11)
01-12-2011
03-12-2011
05-12-2011
07-12-2011
09-12-2011
11-12-2011
13-12-2011
15-12-2011
17-12-2011
19-12-2011
21-12-2011
23-12-2011
25-12-2011
27-12-2011
29-12-2011
31-12-2011
Real
Previsão ARMA
Previsão VDF

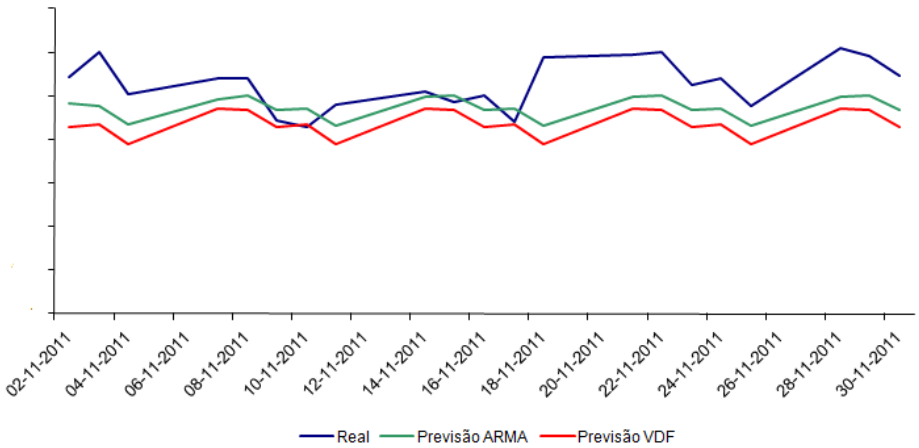
Linha I (Jan 12)



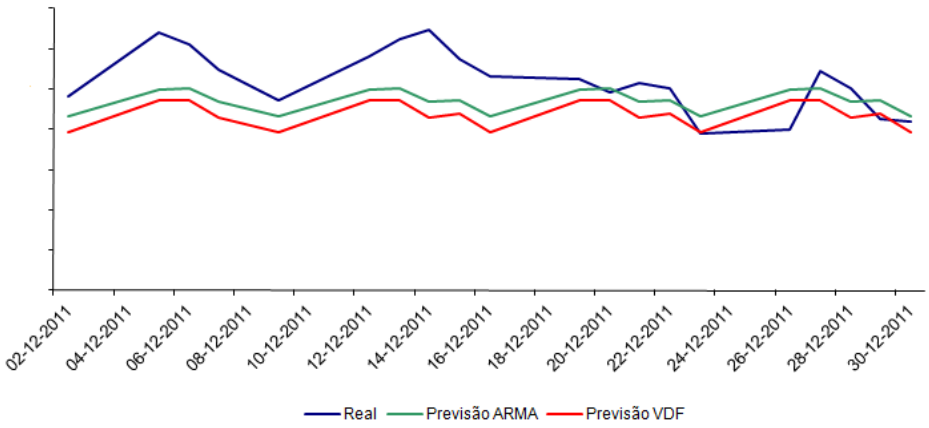
Linha I (Nov/Dez 11 e Jan 12)



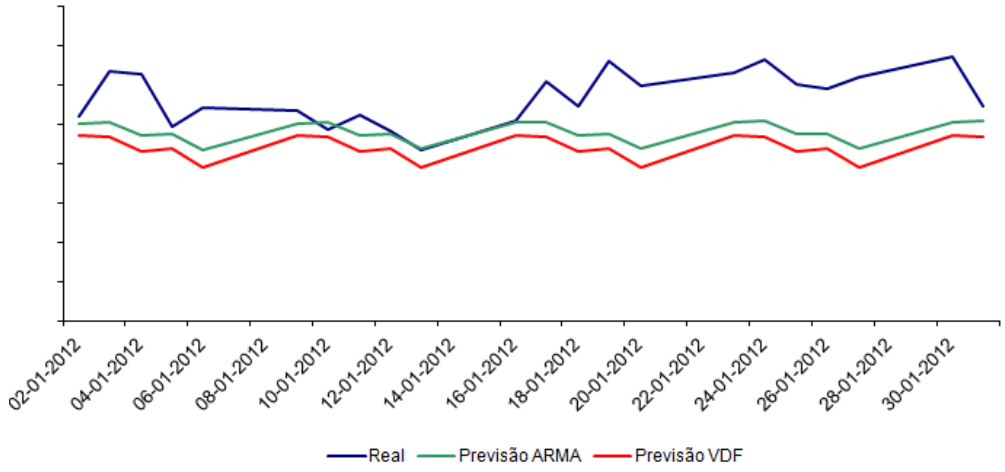
Linha J (Nov 11)



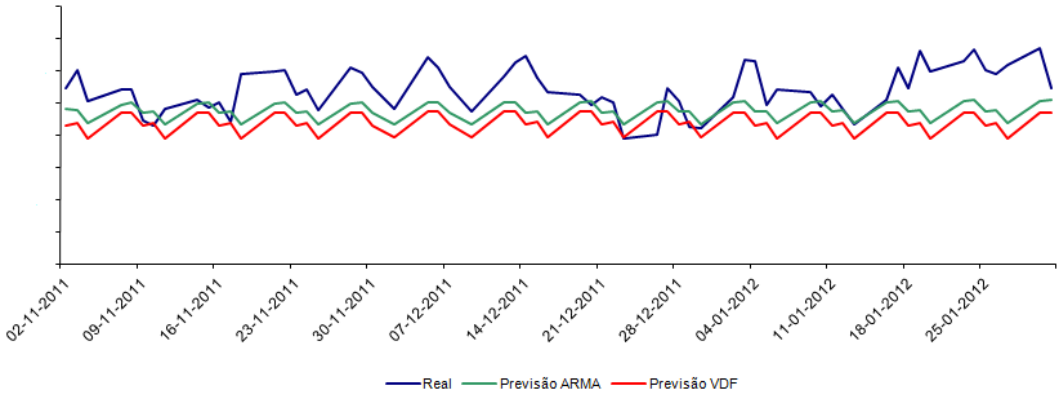
Linha J (Dez 11)



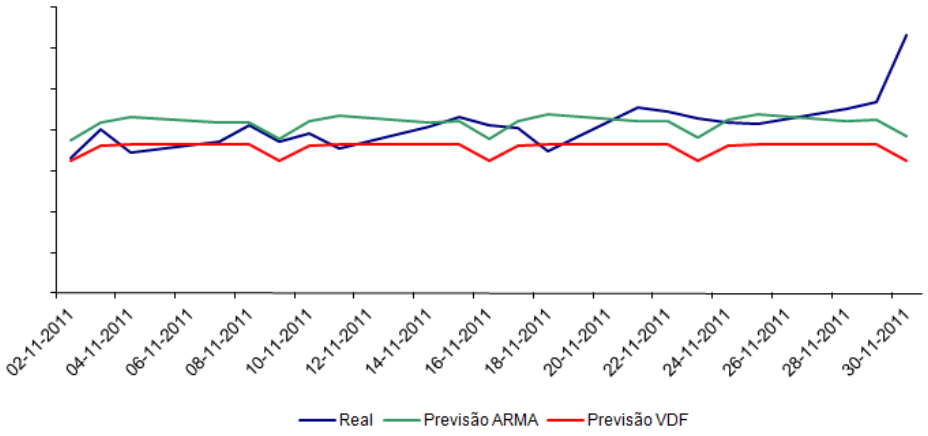
Linha J (Jan 12)



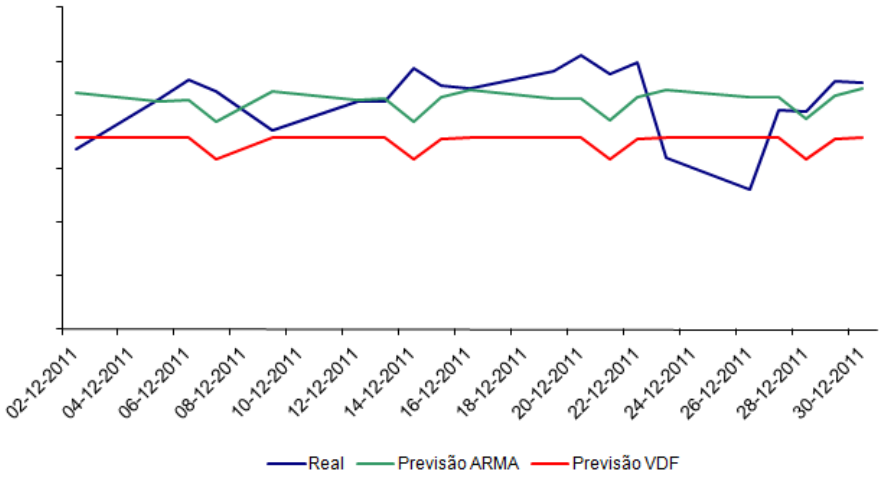
Linha J (Nov/Dez 11 e Jan 12)



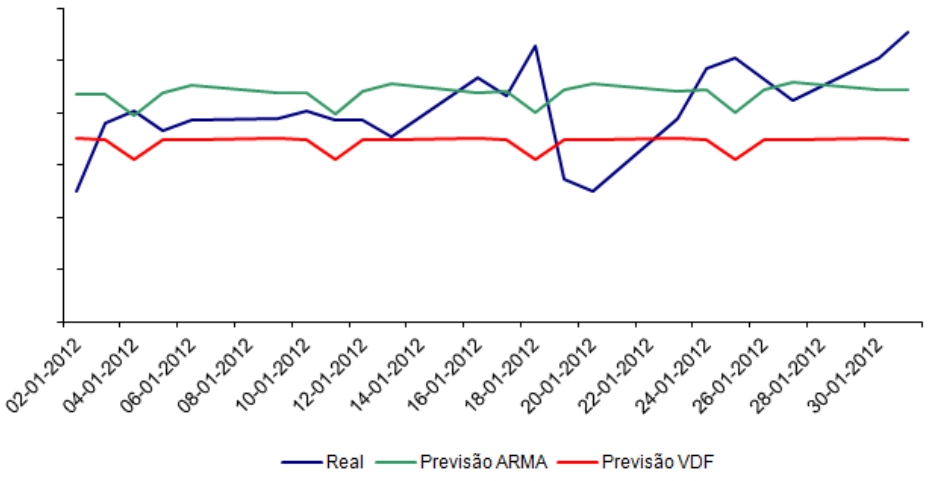
Linha K (Nov 11)

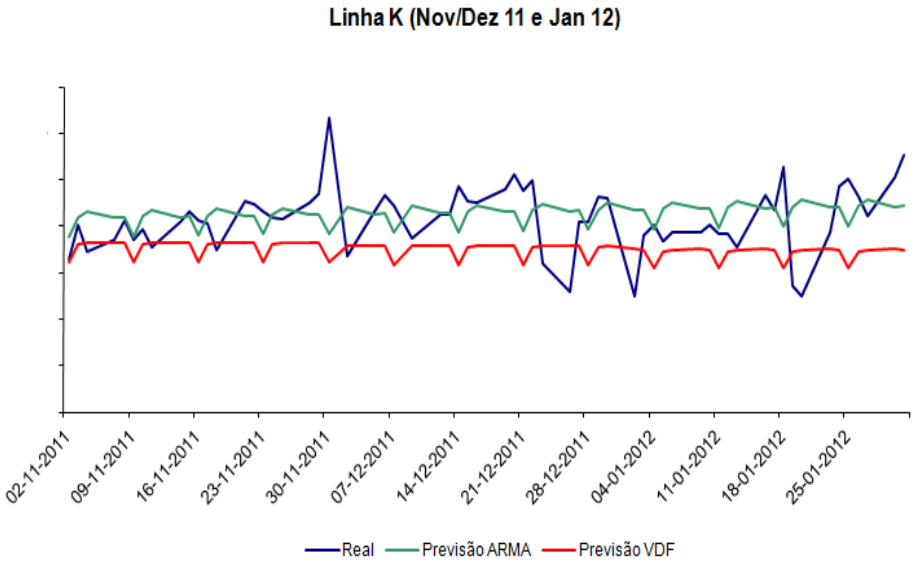


Linha K (Dez 11)



Linha K (Jan 11)





A seguir são apresentados os quadros com os resultados dos métodos aplicados.

Nov-11	Desvio Mensal		Desvio Diário		Desvio Diário Absoluto		EQM		EAM		REQM	
	ARMA	VDF	ARMA	VDF	ARMA	VDF	ARMA	VDF	ARMA	VDF	ARMA	VDF
Linha A	0,76	1,84	-0,80	0,42	8,73	8,62	343748,41	378891,67	448,30	449,37	586,30	615,54
Linha B	-0,14	10,94	1,21	12,91	7,45	16,26	834997,41	3548815,92	659,09	1351,50	913,78	1883,83
Linha C	-0,27	6,58	1,41	6,95	8,85	13,70	1021,39	2941,79	24,77	41,19	31,96	54,24
Linha D	5,65	-2,10	7,95	2,47	11,00	22,38	40646,91	156588,64	131,96	306,84	201,61	395,71
Linha E	39,76	78,07	45,13	86,59	45,13	86,59	9656,34	35476,37	90,77	178,23	98,27	188,35
Linha F	-6,28	-8,51	5,72	-0,10	23,44	27,26	80530,47	94347,64	187,00	210,30	283,78	307,16
Linha G	-3,61	12,65	-1,84	15,56	10,75	16,87	7803,06	15342,37	63,72	90,78	88,33	123,86
Linha H	-25,58	-9,49	58,20	94,46	96,23	126,51	318961,65	318393,68	344,44	370,14	564,77	564,26
Linha I	-4,59	-1,30	-1,31	0,16	14,09	18,29	50058,23	61301,91	169,71	204,03	223,74	247,59
Linha J	-10,19	-17,07	-9,36	-16,34	11,77	16,52	22077,82	43421,64	128,96	181,84	148,59	208,38
Linha K	0,06	-14,51	2,00	-12,87	10,46	14,07	117537,96	202538,47	224,10	321,47	342,84	450,04

Dez-11	Desvio Mensal		Desvio Diário		Desvio Diário Absoluto		EQM		EAM		REQM	
	ARMA	VDF	ARMA	VDF	ARMA	VDF	ARMA	VDF	ARMA	VDF	ARMA	VDF
Linha A	-5,31	-4,69	-6,30	-5,48	9,31	9,06	247336,09	248733,66	405,44	398,55	497,33	498,73
Linha B	-4,46	5,59	-3,53	6,66	8,70	13,23	1036939,51	2266281,30	809,48	1185,68	1018,30	1505,42
Linha C	2,22	4,30	3,80	5,41	10,91	18,07	1705,83	4931,56	31,51	52,69	41,30	70,23
Linha D	-5,07	-10,63	24,32	27,05	40,56	54,62	84224,67	221562,34	212,90	360,11	290,21	470,70
Linha E	-12,56	48,04	94,64	283,74	122,94	283,74	3842,43	22837,84	52,18	134,04	61,99	151,12
Linha F	9,64	3,12	42,59	27,91	48,46	41,08	131979,29	117980,56	202,76	179,43	363,29	343,48
Linha G	-5,14	12,34	-4,02	14,31	10,74	15,80	6955,55	11556,71	65,32	87,27	83,40	107,50
Linha H	18,05	64,26	120,68	189,86	136,47	195,75	67379,66	124740,93	192,37	273,79	259,58	353,19
Linha I	-38,63	-27,72	-35,81	-26,29	36,40	27,93	600171,35	333718,85	683,43	504,60	774,71	577,68
Linha J	-9,23	-15,86	-7,57	-14,40	12,95	16,62	27584,69	45497,58	140,04	183,18	166,09	213,30
Linha K	-0,59	-18,52	2,34	-16,15	14,05	21,71	123871,96	271908,50	266,51	475,25	351,95	521,45

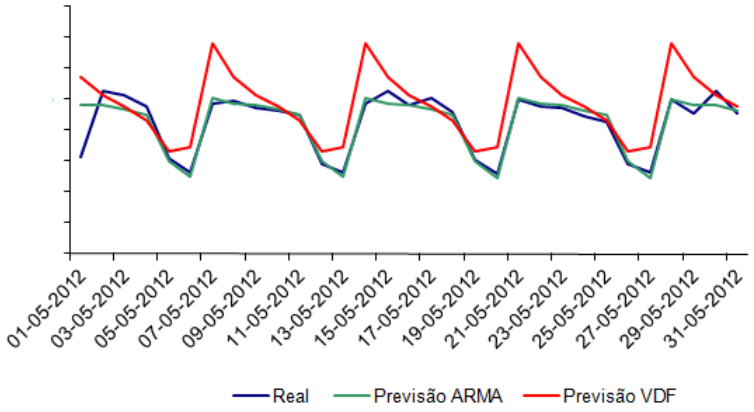
Jan-12	Desvio Mensal		Desvio Diário		Desvio Diário Absoluto		EQM		EAM		REQM	
	ARMA	VDF	ARMA	VDF	ARMA	VDF	ARMA	VDF	ARMA	VDF	ARMA	VDF
Linha A	-12,71	-8,59	-15,41	-11,00	17,71	13,65	617565,81	457811,03	716,97	555,97	785,85	676,62
Linha B	-11,47	-0,92	-9,51	1,95	11,91	14,75	2389958,76	2752582,68	1272,45	1363,89	1545,95	1659,09
Linha C	-3,31	-4,20	0,14	-2,34	9,10	12,13	2191,05	3196,45	28,59	41,91	46,81	56,54
Linha D	-0,95	-5,06	34,54	40,59	43,17	63,13	35472,74	151148,93	123,16	309,83	188,34	388,78
Linha E	-0,82	57,06	447,80	969,59	461,99	970,92	3719,68	25246,22	48,37	146,58	60,99	158,89
Linha F	-14,16	-18,94	-4,96	-17,52	18,61	18,90	83850,06	110300,97	240,85	275,32	289,57	332,12
Linha G	-9,42	10,24	-7,83	13,35	11,14	14,11	7407,21	8465,73	71,48	72,74	86,07	92,01
Linha H	-30,95	-30,26	5,81	2,79	59,14	58,80	241544,06	238903,33	311,09	313,55	491,47	488,78
Linha I	-13,56	-7,46	-11,01	-6,74	15,08	14,06	46580,81	36733,57	188,34	158,37	215,83	191,66
Linha J	-15,48	-22,37	-14,41	-21,40	14,78	21,40	48226,61	82776,60	179,56	254,24	219,61	287,71
Linha K	6,74	-16,26	12,05	-12,10	20,93	21,87	211296,93	295250,38	364,75	455,98	459,67	543,37

3 Meses	Desvio Mensal		Desvio Diário		Desvio Diário Absoluto		EQM		EAM		REQM	
	ARMA	VDF	ARMA	VDF	ARMA	VDF	ARMA	VDF	ARMA	VDF	ARMA	VDF
Linha A	-5,59	-3,69	-7,57	-5,42	11,96	10,46	403526,21	361626,47	524,39	468,17	635,24	601,35
Linha B	-5,60	4,94	-4,00	7,11	9,37	14,73	1426997,49	2848361,53	916,44	1299,80	1194,57	1687,71
Linha C	-0,51	2,06	1,79	3,30	9,63	14,64	1646,14	3698,06	2606,20	4168,19	40,57	60,81
Linha D	-0,34	-6,10	22,43	23,60	31,80	46,97	53587,25	176649,01	156,27	46,39	231,49	420,30
Linha E	-3,22	59,90	185,58	450,55	201,80	451,00	2957,56	27770,62	43,05	152,67	54,38	166,65
Linha F	-4,59	-9,03	14,54	3,47	30,24	29,10	98985,05	107686,48	210,46	221,81	314,62	328,16
Linha G	-6,12	11,71	-4,59	14,39	10,88	15,58	7384,10	11749,64	66,87	83,52	85,93	108,40
Linha H	-17,63	-0,86	61,60	95,72	97,29	127,03	208103,10	226356,33	281,96	318,60	456,18	475,77
Linha I	-21,92	-14,51	-16,20	-11,08	21,94	20,11	234250,70	144816,11	349,09	289,92	483,99	380,55
Linha J	-11,84	-18,65	-10,55	-17,49	13,20	18,26	32957,35	57823,67	150,15	207,55	181,54	240,47
Linha K	2,13	-16,42	5,62	-13,64	15,25	19,22	152289,98	256936,29	286,68	417,26	390,24	506,89

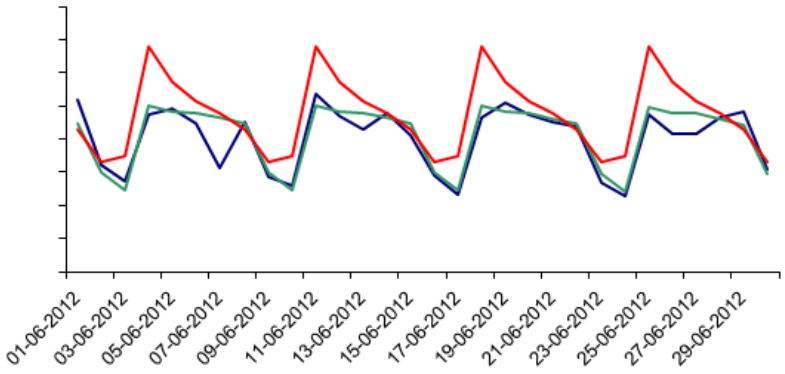
Analogamente apresento os resultados obtidos para os meses de Maio, Junho, Julho e Agosto de 2012 para as linhas B, C, D e E.

Linhas	Efectou-se Transformação Box-Cox?	Modelo Ajustado	AICC
B	Não	SARIMA(1,0,2)x(0,1,1)$_7$	6409.39
C	Não	SARIMA(1,0,1)x(0,1,1)$_7$	4044.43
D	Não	SARIMA(1,0,1)x(0,1,2)$_7$	5333.47
E	Não	SARIMA(1,0,2)x(0,1,1)$_7$	4114.20

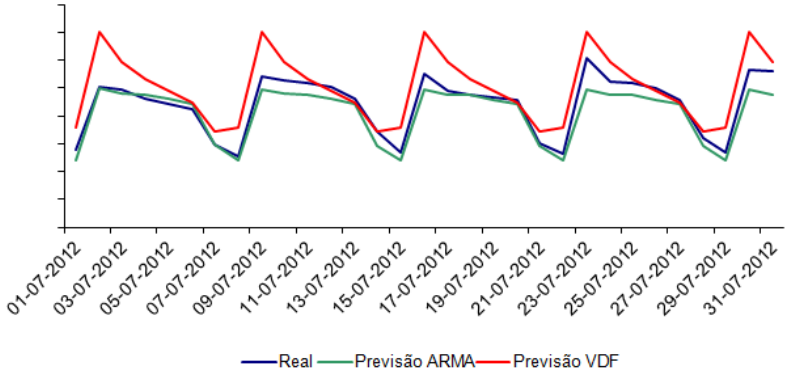
Linha B (Maio 12)



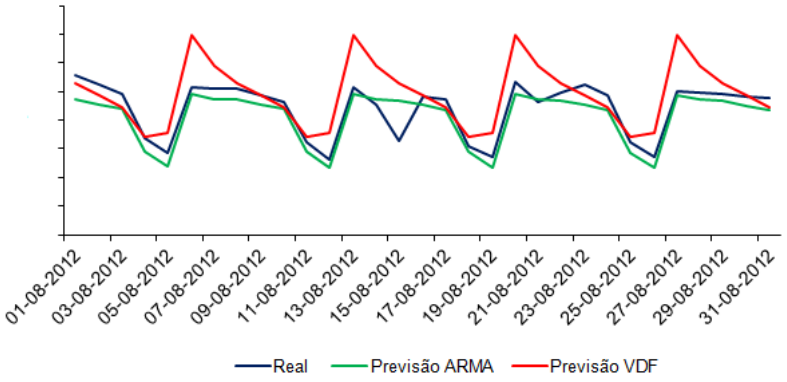
Linha B (Jun 12)



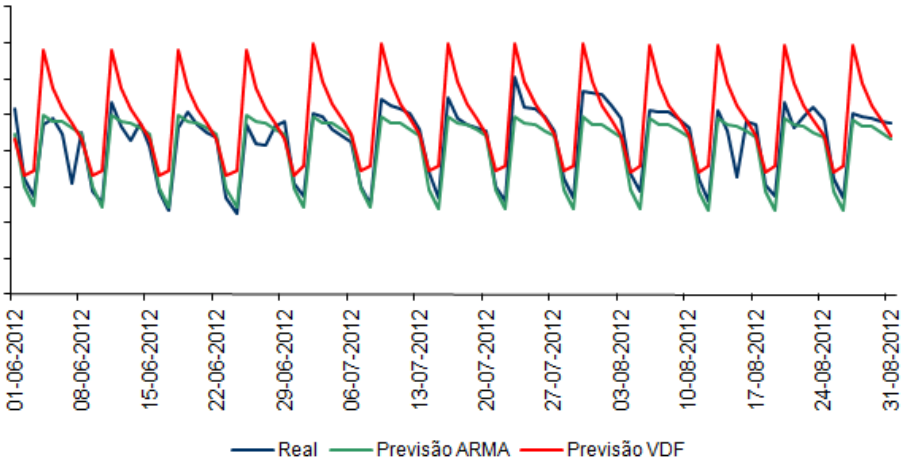
Linha B (Jul 12)



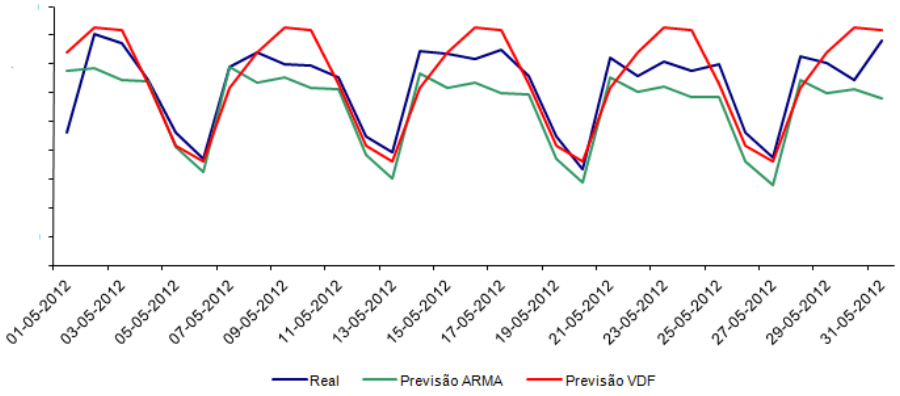
Linha B (Ago 12)



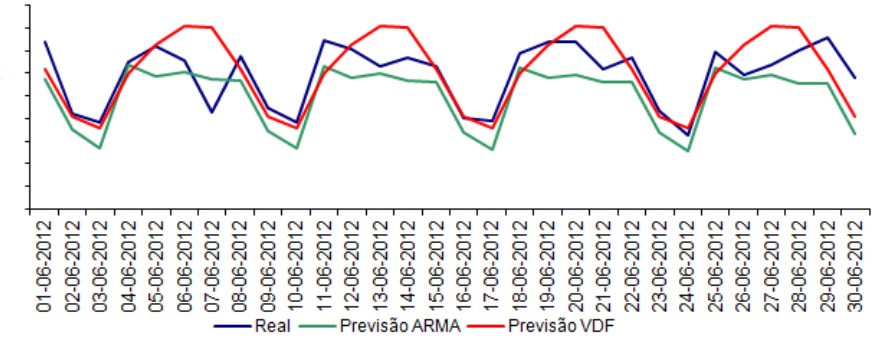
Linha B (Jun /Jul/ Ago 12)



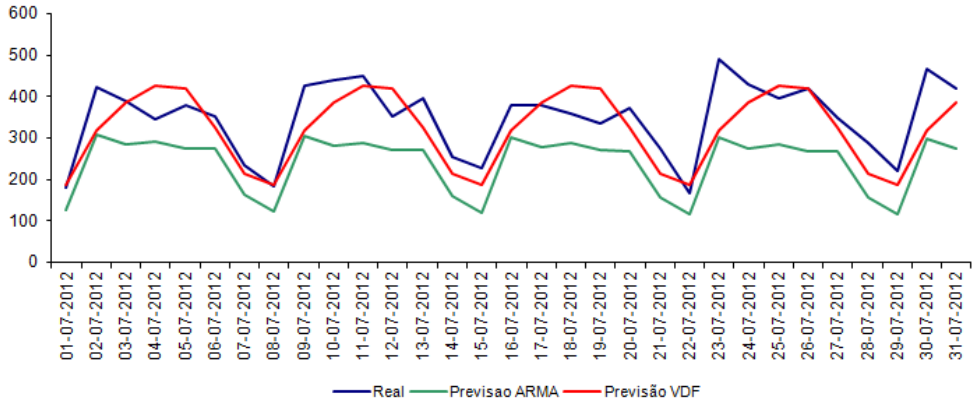
Linha C (Maio 12)



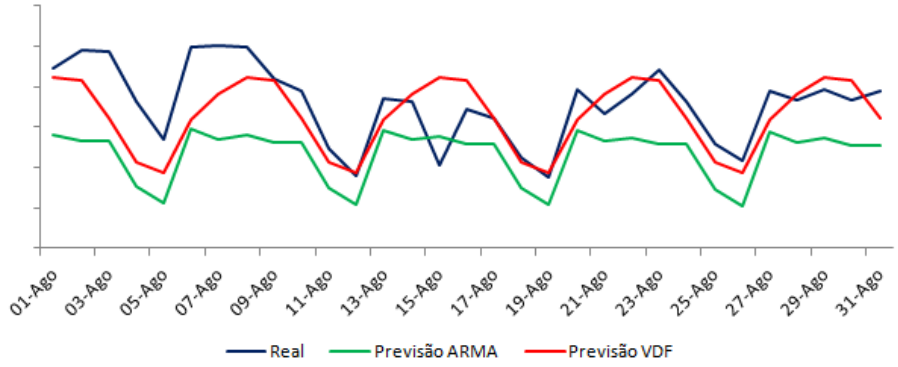
Linha C (Jun 12)



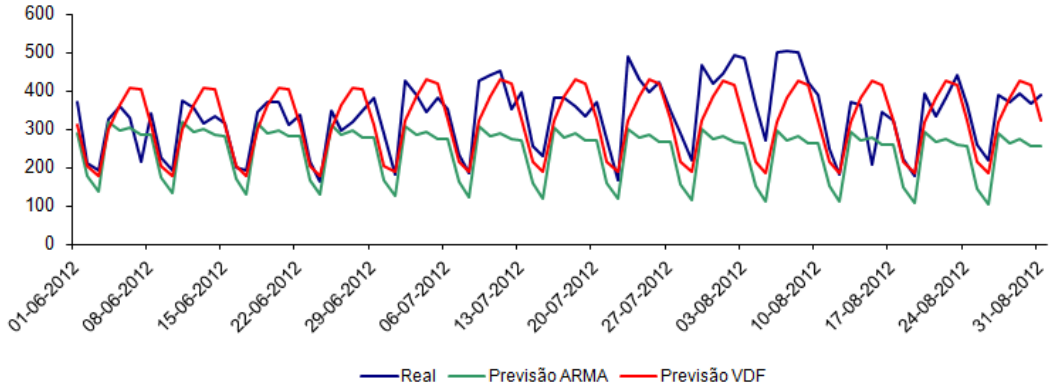
Linha C (Jul 12)



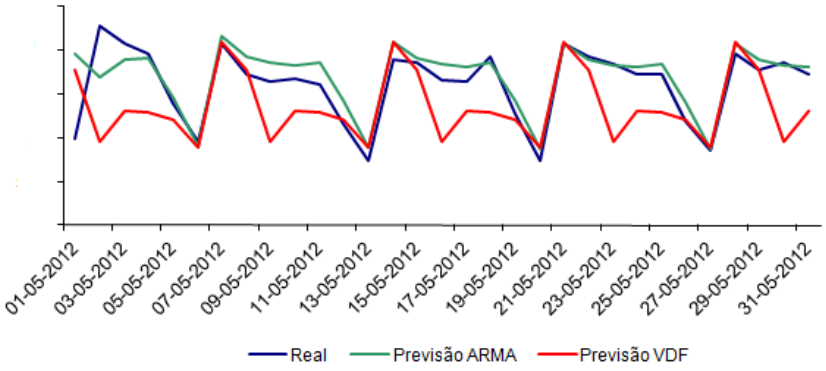
Linha C (Ago 12)



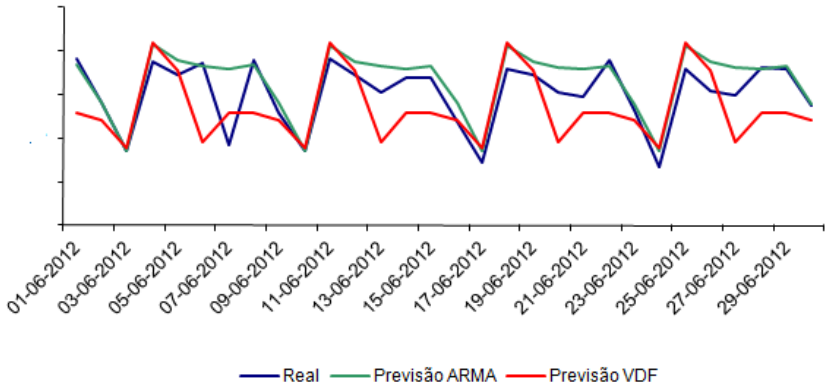
Linha C (Jun / Jul / Ago 12)



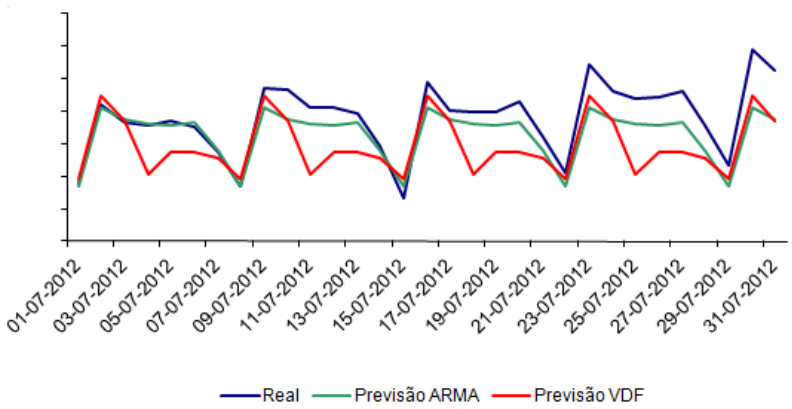
Linha D (Maio 12)



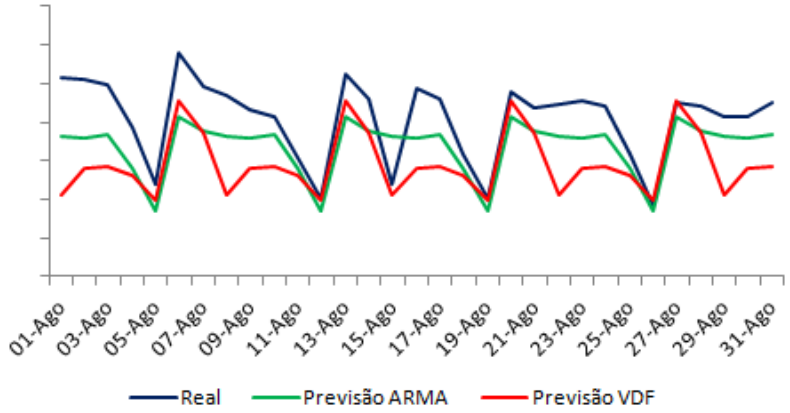
Linha D (Jun 12)



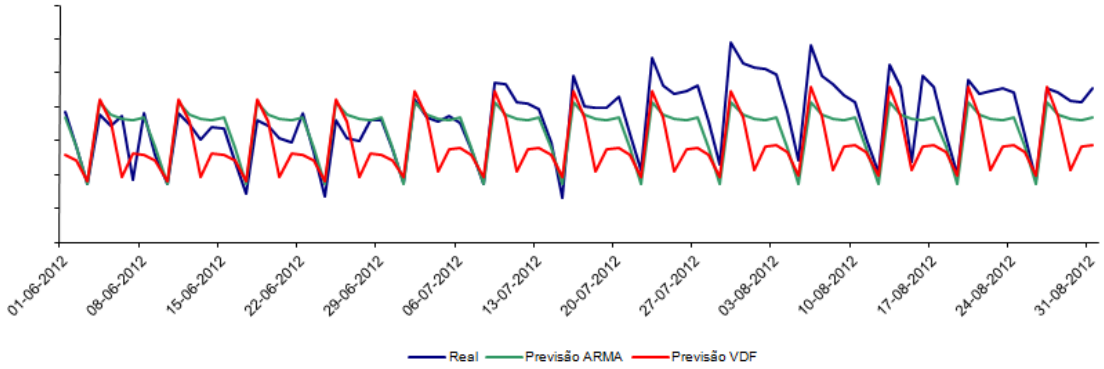
Linha D (Jul 12)



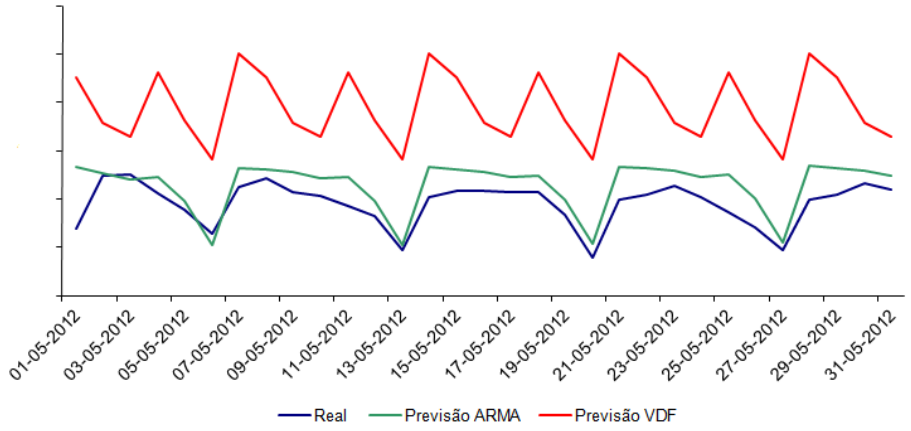
Linha D (Ago 12)



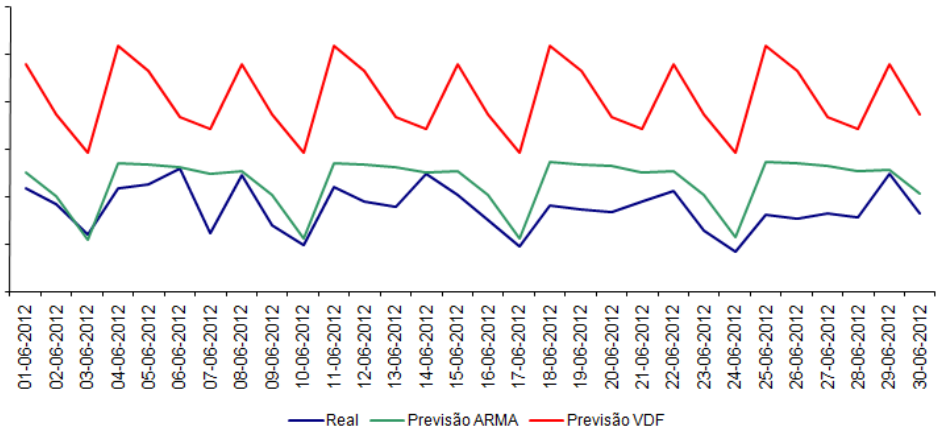
Linha D (Jun / Jul / Ago 12)



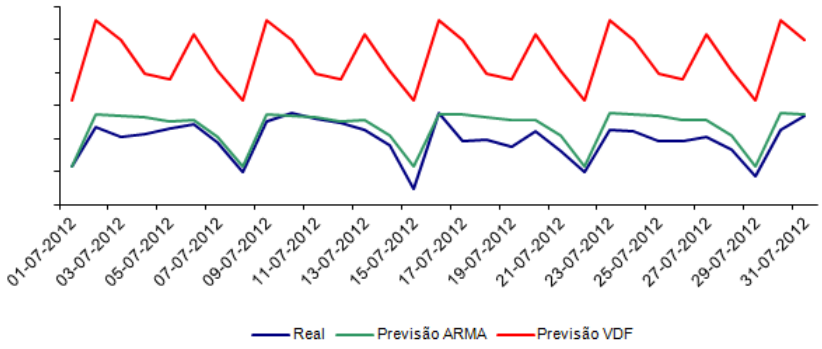
Linha E (Maio 12)

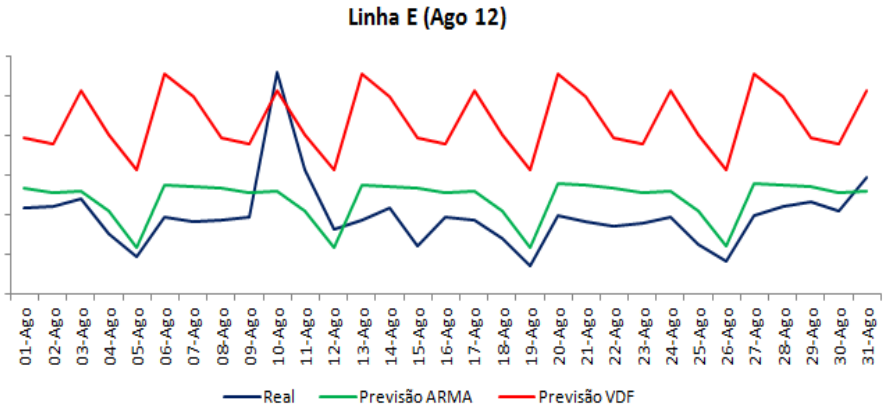


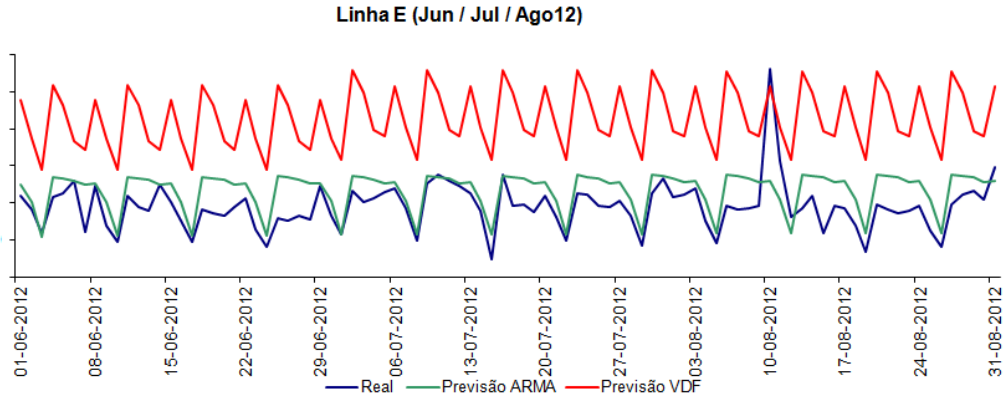
Linha E (Jun 12)



Linha E (Jul 12)







Mai-12	Desvio Mensal		Desvio Diário		Desvio Diário Absoluto		EQM		EAM		REQM	
	ARMA	VDF	ARMA	VDF	ARMA	VDF	ARMA	VDF	ARMA	VDF	ARMA	VDF
Linha A	0,64	-10,76	0,39	-11,29	3,98	11,71	59806,96	334383,81	176,05	503,44	244,55	578,26
Linha B	-0,03	-13,72	-0,51	-15,39	5,50	17,84	558856,35	3665536,53	433,87	1404,43	747,57	1914,56
Linha C	11,35	-3,44	11,16	-3,42	14,14	11,44	2327,55	2165,53	42,33	35,20	48,24	46,54
Linha D	5,96	15,57	8,43	11,65	11,60	21,64	59692,96	241030,55	160,99	357,80	244,32	490,95
Linha E	20,18	-105,60	21,45	-116,49	22,90	116,49	2222,88	44979,08	40,51	200,82	47,15	212,08

Jun-12	Desvio Mensal		Desvio Diário		Desvio Diário Absoluto		EQM		EAM		REQM	
	ARMA	VDF	ARMA	VDF	ARMA	VDF	ARMA	VDF	ARMA	VDF	ARMA	VDF
Linha A	-5,43	-19,45	-5,98	-20,36	7,99	20,64	157213,76	803306,78	321,00	814,54	396,50	896,27
Linha B	-2,62	-17,46	-3,23	-19,23	7,68	21,52	677506,24	4163012,35	585,60	1631,03	823,11	2040,35
Linha C	15,55	-2,73	15,74	-3,50	17,99	15,18	3292,22	3581,39	51,15	44,60	57,38	59,84
Linha D	-10,20	10,31	-12,00	7,38	12,90	20,00	57313,78	146736,02	171,01	308,00	239,40	383,06
Linha E	-31,05	-130,66	-34,04	-141,19	34,56	141,19	4407,10	56472,47	55,29	230,05	66,39	237,64

Jul-12	Desvio Mensal		Desvio Diário		Desvio Diário Absoluto		EQM		EAM		REQM	
	ARMA	VDF	ARMA	VDF	ARMA	VDF	ARMA	VDF	ARMA	VDF	ARMA	VDF
Linha A	4,35	-15,31	3,82	-16,23	5,47	16,23	121749,76	550323,33	263,82	695,69	348,93	741,84
Linha B	6,59	-13,82	6,47	-14,98	7,23	15,62	668863,51	2632681,38	647,19	1276,11	817,84	1622,55
Linha C	30,74	7,20	31,13	6,25	31,13	14,14	12818,99	4212,42	107,08	51,08	113,22	64,90
Linha D	12,86	22,81	10,79	19,59	13,30	24,10	119148,89	325846,87	268,40	472,02	345,18	570,83
Linha E	-18,34	-123,56	-22,84	-143,92	23,08	143,92	1937,15	62285,47	36,74	243,38	44,01	249,57

Ago-12	Desvio Mensal		Desvio Diário		Desvio Diário Absoluto		EQM		EAM		REQM	
	ARMA	VDF	ARMA	VDF	ARMA	VDF	ARMA	VDF	ARMA	VDF	ARMA	VDF
Linha A	3,31	-15,24	3,08	-15,87	5,16	15,87	93498,03	621498,03	234,54	687,50	305,77	788,35
Linha B	6,46	-11,66	6,41	-13,02	9,53	16,55	861681,12	3357814,07	781,90	1377,57	928,27	1832,43
Linha C	34,67	8,02	33,71	5,35	35,94	17,57	19389,00	6622,53	128,82	61,24	139,24	81,38
Linha D	17,71	28,42	15,98	25,99	19,42	17,57	203009,02	512945,33	402,63	586,47	450,57	716,20
Linha E	-21,59	-121,52	-31,81	-146,40	39,96	146,94	7918,54	65015,13	73,89	242,55	88,99	254,98

(Jun/Jul/ Ago 12)	Desvio Mensal		Desvio Diário		Desvio Diário Absoluto		EQM		EAM		REQM	
	ARMA	VDF	ARMA	VDF	ARMA	VDF	ARMA	VDF	ARMA	VDF	ARMA	VDF
Linha A	1,00	-16,55	0,38	-17,46	6,19	17,55	123794,51	656800,71	272,60	731,69	351,84	810,43
Linha B	3,72	-14,19	3,28	-15,70	8,15	17,86	736652,94	3376040,54	672,50	1426,03	858,28	1837,40
Linha C	27,75	4,62	26,98	2,77	28,47	15,64	11926,24	4818,75	96,17	52,39	109,21	69,42
Linha D	8,46	21,55	5,11	17,77	15,23	23,58	127242,48	330485,20	281,87	457,10	356,71	574,88
Linha E	-23,31	-124,99	-29,52	-143,86	32,51	144,05	4758,04	61309,71	55,31	238,76	68,98	247,61

Conclusão

Com base nos resultados obtidos verificamos que conseguimos melhorar alguns modelos relativamente aos modelos actuais considerados pela Vodafone para as várias linhas que foram pedidas e portanto posso dizer que um dos principais objectivos da realização deste estágio foi cumprido pois obtive melhores performances nos meus modelos que considerei. Outra das metas a cumprir era a aplicação de um novo método de previsão. Relativo a este ponto, sugeri uma Metodologia que toma o nome de Box-Jenkins em detrimento ao Método de Holt-Winters Sazonal proposto pela Vodafone. Por fim, um dos últimos pontos dos objectivos era a introdução deste novo método baseado nos modelos lineares SARIMA na empresa mas tal não fui possível concretizar em tempo útil.

Discussão

Por vezes podemos estar divididos em escolher qual o melhor método a seleccionar para o nosso conjunto de dados (o que acontece muitas vezes) mas relativamente a esta questão há uma solução com que se pode resolver o problema: unir os métodos de previsão atribuindo um determinado peso a cada um deles. Logicamente, que esta solução iria ser muito trabalhosa mas seria engraçado analisar os resultados obtidos.

Projectos Futuros

Relativamente a esta parte existe muita coisa que se pode fazer futuramente, pois trata-se de uma área muito abrangente e que atinge vários domínios. Eu neste trabalho só trabalhei com séries univariadas mas seria interessante ver a que resultados chegaríamos caso tivesse trabalhado com séries multivariadas, pois estas deveriam explicar melhor porque contém mais informações acerca dos dados. Uma outra abordagem que é muito popular nos dias de hoje para o cálculo das previsões são as chamadas redes neurais. Relativamente a este tema, ainda há muito que se diga pois existem alguns autores que defendem a sua aplicação enquanto outros corroboram. A respeito deste tema, ainda realizei algumas pesquisas e cheguei a obter códigos em R mas acabei por não colocar neste relatório final porque obtive várias dúvidas na sua programação. Uma outra abordagem que se podia ter feito e encontra-se referida na teórica é a analise de intervenção a séries temporais.

Bibliografia

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

BROCKWELL, P.J. & DAVIS, R.A. Introduction to Time Series and Forecasting Second Edition, USA, Springer, 2002.

MURTEIRA, B. J. F & MULLER, D.A. & TURKMAN, K.F. Análise de Sucessões Cronológicas. Portugal, McGraw-Hill, 1993.

TSAY, R.S. Analysis of Financial Time Series Third Edition, USA, Wiley, 2010.

Apontamentos da cadeira de Processos de Previsão e Decisão do Professor António José Rodrigues referentes ao Ano Lectivo 2010/2011.

PESTANA, D.D. & VELOSA, S.F. Introdução à Probabilidade e à Estatística Volume I, Lisboa, Fundação Calouste Gulbenkian, 2006.

CHOONG, JOE. Powerful Forecasting with MS Excel, Kindle Edition, 2012.

CABRAL, M.S. & GONÇALVES, M.H. Análise de Dados Longitudinais, Lisboa, SPE, 2011.

Anexos

Output - Linha A (Meses Verão)

```
=======================================
ITSM::(Maximum likelihood estimates)
=======================================

Method: Maximum Likelihood


ARMA Model:
X(t) = .9259 X(t-1)
    + Z(t) - .6798 Z(t-1) + .0000 Z(t-2) + .0000 Z(t-3)
    + .0000 Z(t-4) + .0000 Z(t-5) + .0000 Z(t-6) - .9679 Z(t-7)
    + .6580 Z(t-8)

WN Variance = .179594E+06

AR Coefficients
      .925945

Standard Error of AR Coefficients
      .020962

MA Coefficients
     -.679778      .000000      .000000      .000000
      .000000      .000000     -.967934      .657980


Standard Error of MA Coefficients
      .013169      .000000      .000000      .000000
      .000000      .000000      .012455      .000000


(Residual SS)/N = .179594E+06

AICC = .583771E+04
BIC  = .582721E+04

-2Log(Likelihood) = .582961E+04

Accuracy parameter = .0000700000

Number of iterations = 26

Number of function evaluations = 150
```

Uncertain minimum.

Output dos Testes – Linha A (Meses de Verão)

==========
ITSM::(Tests of randomness on residuals)
==========

Ljung - Box statistic = 26.705 Chi-Square (20), p-value = .14378

McLeod - Li statistic = 10.716 Chi-Square (23), p-value = .98590

Turning points = .22600E+03~AN(.25800E+03,sd = 8.2966), p-value = .00011

Diff sign points = .20300E+03~AN(.19400E+03,sd = 5.7009), p-value = .11440

Rank test statistic = .34589E+05~AN(.37733E+05,sd = .12812E+04), p-value = .01413

Order of Min AICC YW Model for Residuals = 1

Output Regressão - Linha A (Meses de Verão)

```
========================================
ITSM::(Maximum likelihood estimates)
========================================

Method: Maximum Likelihood

ARMA Model:
X(t) = .8914 X(t-1)
    + Z(t) - .4340 Z(t-1) - .1531 Z(t-2) + .0000 Z(t-3)
    + .0000 Z(t-4) + .0000 Z(t-5) + .0000 Z(t-6) - 1.087 Z(t-7)
    + .4716 Z(t-8) + .1664 Z(t-9)

WN Variance = .388609E+06

AR Coefficients
     .891431

Standard Error of AR Coefficients
     .015276

MA Coefficients
    -.434013     -.153125      .000000      .000000
     .000000      .000000    -1.086565      .471584
     .166380

Standard Error of MA Coefficients
     .023993      .021432      .000000      .000000
     .000000      .000000      .009534      .000000
     .000000

(Residual SS)/N = .388609E+06

AICC = .488614E+05
BIC  = .483507E+05

-2Log(Likelihood) = .488514E+05

Accuracy parameter = .0000200000

Number of iterations = 12

Number of function evaluations = 105

Uncertain minimum.
```

Output dos Testes Regressão – Linha A (Meses de Verão)

ITSM::(Tests of randomness on residuals)
==========

Ljung - Box statistic = 17.068 Chi-Square (20), p-value = .64855

McLeod - Li statistic = 25.311 Chi-Square (24), p-value = .38903

Turning points = .19550E+04~AN(.20493E+04,sd = 23.378), p-value = .00005

Diff sign points = .15610E+04~AN(.15375E+04,sd = 16.013), p-value = .14222

Rank test statistic = .24243E+07~AN(.23647E+07,sd = .28440E+05), p-value = .03589

Order of Min AICC YW Model for Residuals = 0

Código R

Uma vez que existe uma infinidade de programas que realizam estes estudos relacionados com as previsões, achei por bem considerar outro programa alternativo ao ITSM 2000. Assim, irei apresentar a seguir um código em linguagem R que faz equivalentemente o mesmo que o programa anterior que utilizei. Este código baseia-se essencialmente no método de previsão da Metodologia de Box-Jenkins referente à linha A.

```
##### Metodologia de Box-Jenkins #####

library(forecast)

dados<-read.table("16918.txt") # lê o ficheiro
dados<-dados[,1]  # armazena os dados num vector

dados<-ts(dados) # transforma os dados em uma série temporal
dados

# Sumário dos dados
summary(dados)
#fivenum(dados) - outra versão

# Número total de observações da série
n<-length(dados)
n

# Gráfico da série temporal
plot(ts(dados),xlab="Observações",ylab="Chamadas Atendidas",
main="Série Temporal Linha A",col="blue")
#plot.ts(dados) - comando alternativo

#### Curiosidades ###
# Teste Kwiatkowski-Phillips-Schmidt-Shin (KPSS)
library(tseries)
# H0: A série apresenta tendência
kpss.test(dados,null="Trend")
# Rej. Ho para alfas de 5% e 10% -> Não há evidência para afirmar que a série tem tendência
significativa

# H0: A série é estacionária
kpss.test(dados,null="Level")
# Rej. Ho para todos os niveis usuais de alfa-> Não há evidência para afirmar que a série é
estacionária
######################

# Histograma e QQ-Plot (Normal)
hist(dados, col="blue", main="Histograma")
```

```
qqnorm(dados, col="blue")

# Verificar a existência de Outliers
boxplot(dados,outline=T,col="blue",medcol="white",ylab="nº de chamadas
atendidas",horizontal=F)
# Pela análise do Boxplot há presença de 2 possiveis outliers no conjunto de dados

# Serve para verificamos se devemos fazer a Transformação Box-Cox ou não
library(MASS)
boxcox(dados~1,lambda = seq(0, 1.5, 1/10))
# A informação dada pelo gráfico, indica-nos que devemos aplicar uma Transfornação Box-
Cox para
# unm lambda de 0.6

# Transformação de Box-Cox para lambda=0.6
BC<-((dados^0.6)-1)/0.6

############## NOTA ###############
# Caso tivessemos considerado lambda=0, aplicariamos o logaritmo à série inicial
# ldados<-log(dados)
###################################

# Gráfico da série temporal quando aplicada a Transformação de Box-Cox para lambda=0.6
plot(BC,xlab="Observações",ylab="Chamadas Atentidas",col="blue")

# Histograma e QQ-Plot (Normal)
hist(BC, col="blue", main="Histograma")
qqnorm(BC, col="blue")

# Verificar a existência de Outliers novamente
boxplot(BC,outline=T,col="blue",medcol="white",ylab="nº de chamadas
atendidas",horizontal=F)
# Passamos a ter um só valor outlier

par(mfrow=c(2,1))

# Gráfico da Função de Autocorrelação (FAC)
acf(BC,main="Função de Autocorrelação",lag.max=40)

# Gráfico da Função de Autocorrelação Parcial (FACP)
pacf(BC,main="Função de Autocorrelação Parcial",lag.max=40)

### Observações ###
# Pelo gráfico da FAC repara-se que o processo não é estacionário (uma vez que
# que a função FAC decresce lentamente para zero) e nota-se uma componente
# sazonal semanal de 7 dias
####################

# Diferenciar a série por 7 (de modo a retirar a componente sazonal)
dados1<-diff(BC,lag = 7)
```

```
# Gráfico da série diferenciada
plot(dados1,xlab="Observações",ylab="Chamadas Oferecidas",main=" Série
Diferenciada por 7",col="blue")

par(mfrow=c(2,1))
# Gráfico da FAC da 7ª Diferença
acf(dados1,main="Função de Autocorrelação da 7ª Diferença",lag.max=40)

# Gráfico da FACP da 7ª Diferença
pacf(dados1,main="Função de Autocorrelação Parcial da 7ª Diferença",lag.max=40)

# Modelos Possíveis
#SARIMA(1,0,0)x(0,1,1)[7] #ARMA(1,7) - quando diferenciado
#SARIMA(1,0,1)x(0,1,1)[7] #ARMA(1,8) - quando diferenciado
#SARIMA(0,0,2)x(0,1,1)[7] #MA(9) - quando diferenciado

# Ajustar um SARIMA(1,0,0)x(0,1,1)7
mod1<-arima(BC,order=c(1,0,0), seasonal=list(order=c(0,1,1),period=7),method="ML")
mod1
summary(mod1)
# AICC = 3116.76

# Ajustar um SARIMA(1,0,1)x(0,1,1)7
mod2<-arima(BC,order=c(1,0,1), seasonal=list(order=c(0,1,1),period=7),method="ML")
mod2
summary(mod2)
# AICC = 3108.8

# Ajustar um SARIMA(0,0,2)x(0,1,1)7
mod3<-arima(BC,order=c(0,0,2), seasonal=list(order=c(0,1,1),period=7),method="ML")
mod3
summary(mod3)
# AIC = 3132.26

# O que apresenta menor valor de AICC é o 2º modelo (mod2) - vamos considerar este
modelo

# Dá os valores da série ajustada
valores_ajustados<-fitted(mod2)

# Gráfico da série diferenciada por 7 com série ajustada
plot(ts(dados1),xlab="Observações",ylab="Chamadas Atendidas",
main="Série Temporal Linha A",col="blue")
lines(valores_ajustados,col="orange")

# Legenda
legend("topleft", legend=c("Valores Série Dif. 7","Valores Ajustados
SARIMA(1,0,1)x(0,1,1)7"),lty=c(1,1),
col=c("blue","orange"), bty=c("n","n"), lwd=c(1,1))
```

```
## Análise aos Resíduos

residuos<-residuals(mod2)
residuos

# Pressupostos do Modelo linear
# residuos são v.a. de média zero e de variância constante(hipotese de homocedasticidade)
# residuos não correlacionados  e independentes
# residuos segue uma dist normal ~ N(0,sigma^2)

# Gráfico dos resíduos
plot(residuos,main="Resíduos do Modelo",col="purple")
legend("topleft", legend=c("Resíduos Modelo SARIMA(1,0,1)(0,1,1)7"),lty=1,
col=c("purple"), bty=c("n"), lwd=1)
abline(h=0)

# Resíduos Standardizados
rs<-(residuos-mean(residuos))/sd(residuos)
rs

# Gráfico dos resíduos standardizados
plot(rs,main="Resíduos do Modelo",col="purple")
legend("topleft", legend=c("Resíduos Standardizados Modelo SARIMA(1,0,1)(0,1,1)7"),lty=1,
col=c("purple"), bty=c("n"), lwd=1)
abline(h=0)

par(mfrow=c(2,1))

# Histograma dos resíduos standardizados
hist(rs)

# QQ-plot(Normal) dos resíduos
qqnorm(residuos)

library(tseries)

# Testes para testar a normalidade
# H0: têm dist. normal  vs  H1: não tem dist. normal

# Teste de Shapiro-Wilk
shapiro.test(residuos)

# Teste Jarque-Bera
jarque.bera.test(residuos)

# Teste de Kolmogorov-Smirnov
ks.test(residuos,"pnorm")

# Os testes rejeitam a normalidade dos resíduos

# Gráficos da FAC e da FACP dos resíduos
```

```
par(mfrow=c(2,1))
acf(residuos,main="ACF dos Resíduos")
pacf(residuos,main="PACF dos Resíduos")

# Podemos concluir que os residuos têm um comportamento análogo ao do WN, ou seja, os
resíduos são
# não correlacionados

residuos_valor_absoluto<-abs(residuos)
quadrados_residuos<-residuos^2

# Gráfico da FAC dos resíduos em valor absoluto e dos resíduos ao quadrados
par(mfrow=c(2,1))
acf(residuos_valor_absoluto,main="Função de Autocorrelação dos Residuos em valor
absoluto",lag.max=40)
acf(quadrados_residuos,main="Função de Autocorrelação dos Residuos ao
quadrado",lag.max=40)

# Testar se os resíduos são não correlacionados
# Teste Ljung-Box
AutocorTest(residuos,lag=ceiling(log(length(residuos))),type=c("Ljung-Box"), df=9)
Box.test(residuos,type = "Ljung-Box",lag=9)
# Não Rej. Ho -> os residuos são não correlacionados

# Testes que permitem verificar se existe heteroscedasticidade condicional entre os residuos
library(FinTS)
#H0: Sem efeitos ARCH
ArchTest(residuos,lag=14)

library(TSA)
McLeod.Li.test(mod2)
# Não Rej. Ho em ambos os testes

### Previsão ###

# Pedi uma previsão 92 passos à frente – Meses de Nov/Dez 11 e Jan 12
valores_preditos<-predict(mod2,92)$pred
valores_preditos

valores_reais<-read.table("real_linhaA.txt")

se_valores_preditos<-predict(mod2,92)$se
se_valores_preditos

# Intervalos de Confiança para 95%

LI<-valores_preditos-1.96*se_valores_preditos
LS<-valores_preditos+1.96*se_valores_preditos

plot(dados,xlab="Observações",ylab="Chamadas Atendidas",col="blue")
previsao<-ts(valores_preditos,start=395,end=487)
```

```
lines(previsao,col="red")

lines(LI,col="purple",lty=5)
lines(LS,col="purple",lty=5)
lines(valores_reais,col="green")

legend("topleft", legend=c("Valores Preditos","IC para Valores Preditos","Valores
Reais"),lty=c(1,4,1),
col=c("red","purple","green"), bty=c("n","p","n"), lwd=c(1,1,1))

# Dá os valores das medidas de desempenho relativas ao modelo que consideramos
summary(mod2)
```