

Projeto de Séries Temporais

Caio Gabriel Barreto Balieiro

Modelo de Séries Temporais para vendas de vinhos

Os dados descrevem as vendas totais de vinho australiano por vinicultores em garrafas ≤ 1 litro.

Esta série temporal registra as vendas de vinho por vinicultores australianos entre janeiro de 1980 e agosto de 1994. Este conjunto de dados é encontrado no pacote de forecast do R (wineind).

O objetivo deste projeto é:

Estudar os componentes de tendência e sazonalidade (se houver) da série de vendas de vinho.

```
# Pacotes utilizados
```

```
# É importante destacar que para este projeto funcionar, há a necessidade de instalar os pacotes abaixo
```

```
library(forecast) # Fazer previsão
```

Fazer a previsão de vendas de vinhos para o restante do ano de 1994 e 1995.

```
## Registered S3 method overwritten by 'quantmod':
```

```
##   method             from
```

```
##   as.zoo.data.frame zoo
```

```
library(MLmetrics) # Calcular a métrica MAPE
```

```
##
```

```
## Attaching package: 'MLmetrics'
```

```
## The following object is masked from 'package:base':
```

```
##
```

```
##   Recall
```

```
library(knitr) # Formatar tabela em formato Markdown

# Carregado o conjunto de dados extraídos do pacote forecast

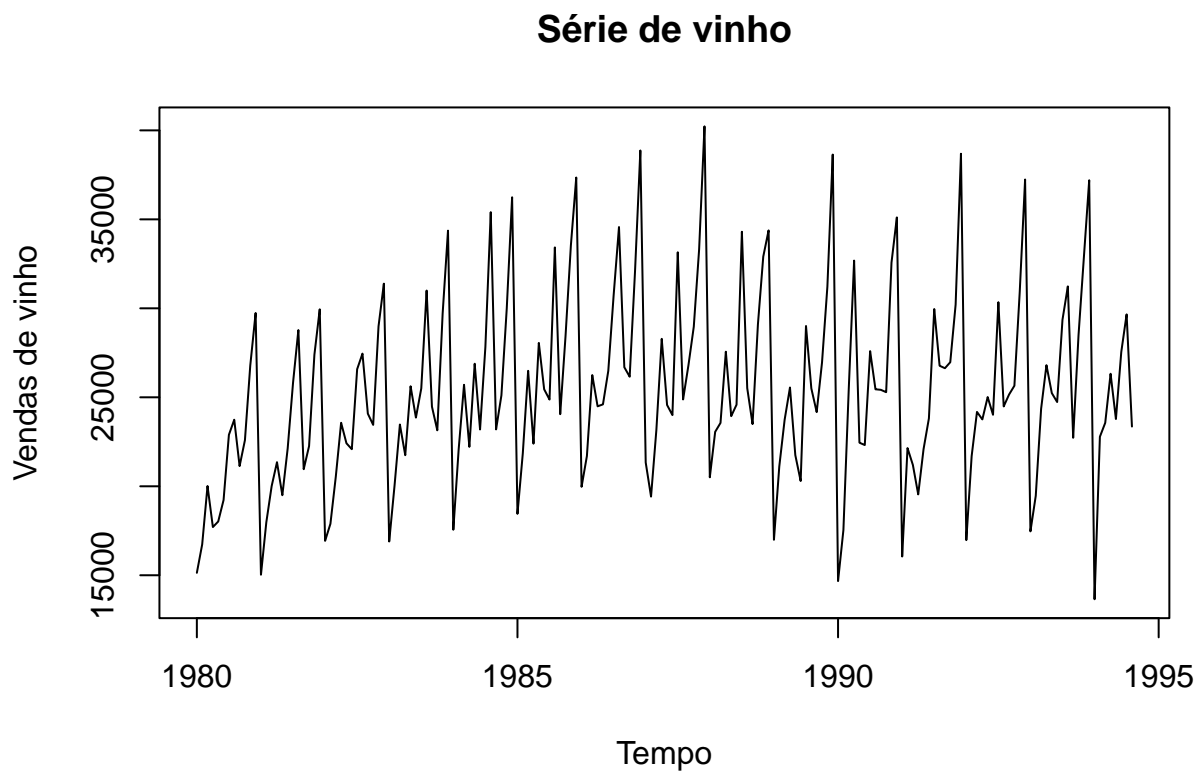
data('wineind')

wineind
```

| ## | | Jan | Feb | Mar | Apr | May | Jun | Jul | Aug | Sep | Oct | Nov | Dec |
|----|------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| ## | 1980 | 15136 | 16733 | 20016 | 17708 | 18019 | 19227 | 22893 | 23739 | 21133 | 22591 | 26786 | 29740 |
| ## | 1981 | 15028 | 17977 | 20008 | 21354 | 19498 | 22125 | 25817 | 28779 | 20960 | 22254 | 27392 | 29945 |
| ## | 1982 | 16933 | 17892 | 20533 | 23569 | 22417 | 22084 | 26580 | 27454 | 24081 | 23451 | 28991 | 31386 |
| ## | 1983 | 16896 | 20045 | 23471 | 21747 | 25621 | 23859 | 25500 | 30998 | 24475 | 23145 | 29701 | 34365 |
| ## | 1984 | 17556 | 22077 | 25702 | 22214 | 26886 | 23191 | 27831 | 35406 | 23195 | 25110 | 30009 | 36242 |
| ## | 1985 | 18450 | 21845 | 26488 | 22394 | 28057 | 25451 | 24872 | 33424 | 24052 | 28449 | 33533 | 37351 |
| ## | 1986 | 19969 | 21701 | 26249 | 24493 | 24603 | 26485 | 30723 | 34569 | 26689 | 26157 | 32064 | 38870 |
| ## | 1987 | 21337 | 19419 | 23166 | 28286 | 24570 | 24001 | 33151 | 24878 | 26804 | 28967 | 33311 | 40226 |
| ## | 1988 | 20504 | 23060 | 23562 | 27562 | 23940 | 24584 | 34303 | 25517 | 23494 | 29095 | 32903 | 34379 |
| ## | 1989 | 16991 | 21109 | 23740 | 25552 | 21752 | 20294 | 29009 | 25500 | 24166 | 26960 | 31222 | 38641 |
| ## | 1990 | 14672 | 17543 | 25453 | 32683 | 22449 | 22316 | 27595 | 25451 | 25421 | 25288 | 32568 | 35110 |
| ## | 1991 | 16052 | 22146 | 21198 | 19543 | 22084 | 23816 | 29961 | 26773 | 26635 | 26972 | 30207 | 38687 |
| ## | 1992 | 16974 | 21697 | 24179 | 23757 | 25013 | 24019 | 30345 | 24488 | 25156 | 25650 | 30923 | 37240 |
| ## | 1993 | 17466 | 19463 | 24352 | 26805 | 25236 | 24735 | 29356 | 31234 | 22724 | 28496 | 32857 | 37198 |
| ## | 1994 | 13652 | 22784 | 23565 | 26323 | 23779 | 27549 | 29660 | 23356 | | | | |

Para realizar esse projeto, foi utilizado as principais bibliotecas de séries temporais da linguagem R sendo elas: forecast, MLmetrics. O pacote knitr é para visualização melhor dos dados em tabelas no formato Markdown. Os dados carregados são do pacote forecast, são dados de vendas totais de vinho por mês.

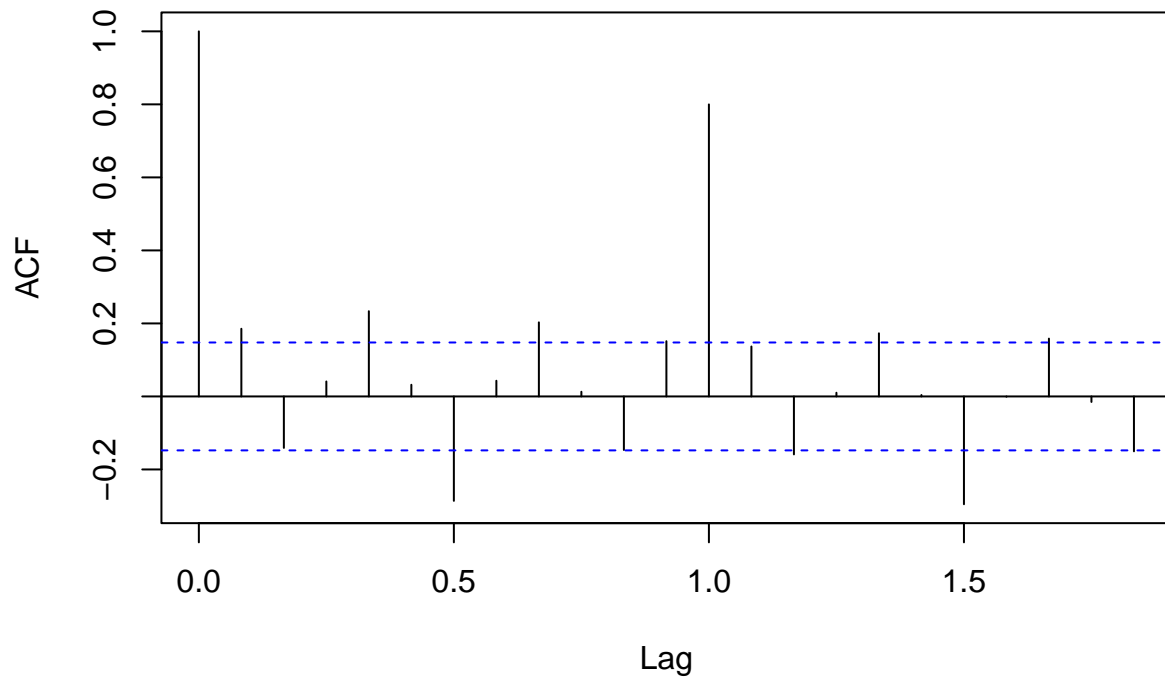
```
# Gráfico da série temporal
ts.plot(wineind, xlab = 'Tempo', ylab = 'Vendas de vinho', main='Série de vinho' )
```



Pode-se notar pelo gráfico da série de vendas de vinho, que as vendas de vinhos totais tiveram um aumento no início do acompanhamento de 1980 até próximo a 1990. Logo após isto tem uma leve queda de vendas. Entretanto no fim se estabiliza. Agora, vamos avaliar o gráfico de ACF que apresenta os principais componentes da série temporal.

```
acf(wineind)
```

Series wineind

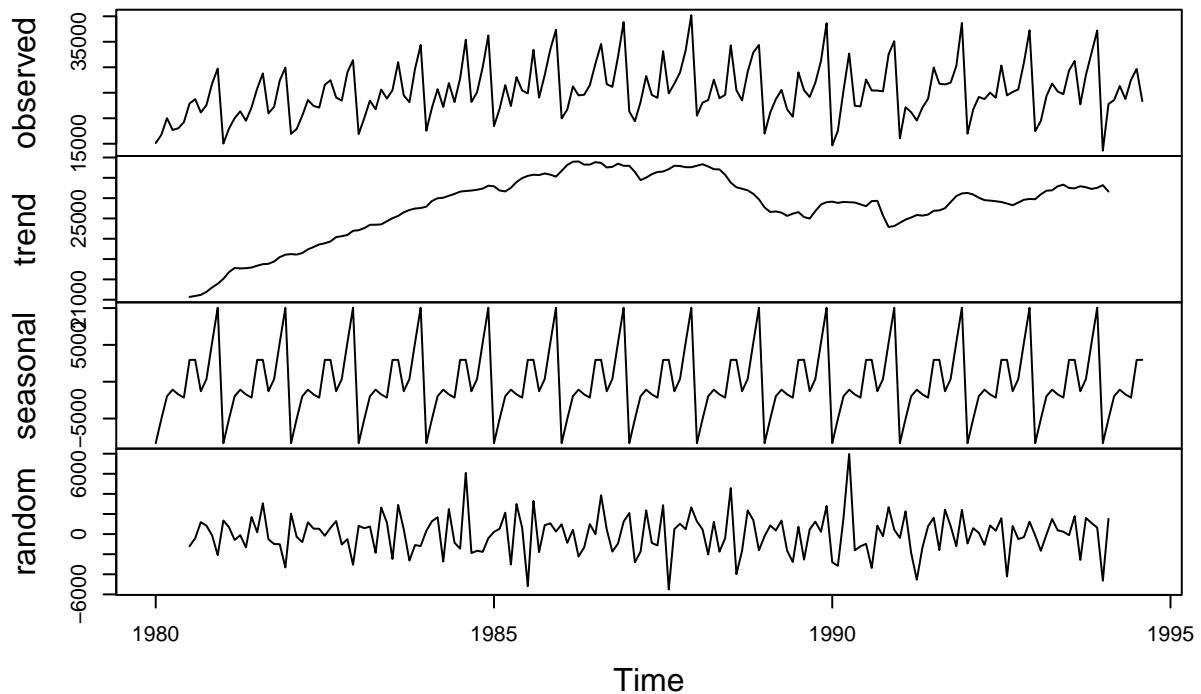


O acf da série de vendas indica que repetições de picos de autocorrelação altos nos lags multiplos de 6, isto da um indicativo de presença de componente sazonal na série, isto é, fatores climaticos e econômicos assim como de datas (tais como feriados e eventos importantes) estão influenciando para que ocorram um movimento de queda e aumento de vendas nos anos acompanhados na série, ou seja, um padrão de repetição onde as vendas totais caem e aumentam ao longo dos meses.

```
# Decomposição da série temporal de vinho em tendencia, sazonalidade e ruido
dec <- decompose(wineind)

plot(dec)
```

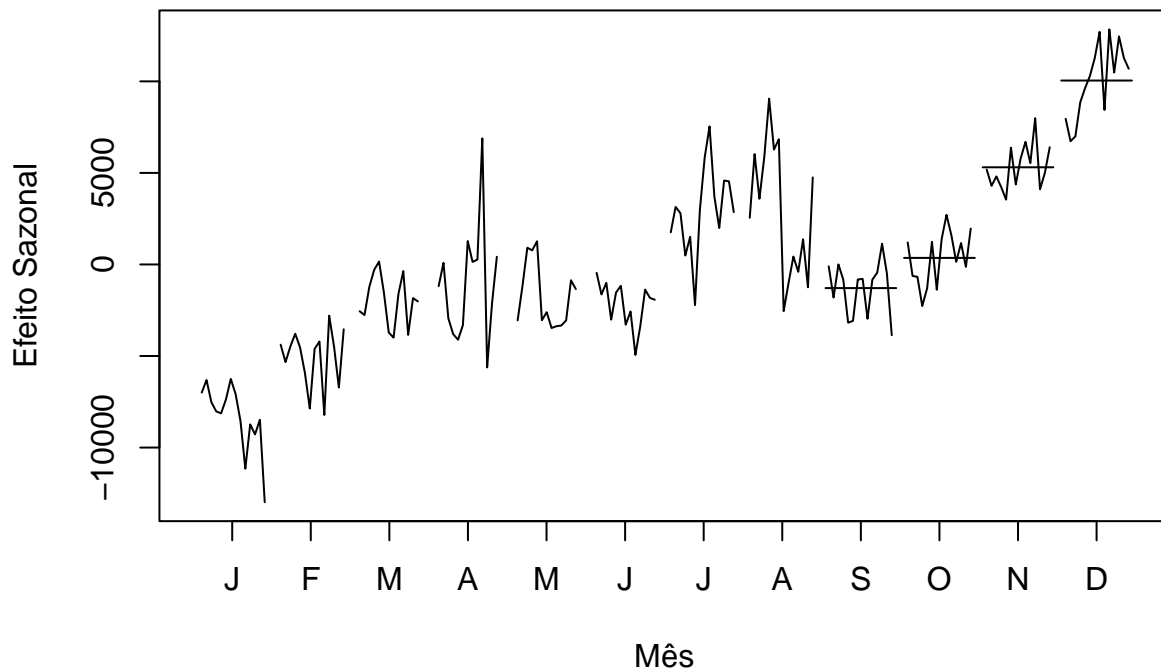
Decomposition of additive time series



O gráfico acima separa de maneira prática os componentes que compoem esta série temporal. Como observado anteriormente a parte de tendência da série apresenta um aumento nos anos iniciais e ao fim tem uma leve redução e se estabiliza. Você pode notar no componente de sazonalidade os movimentos de onda, característico de séries com componente sazonal (em breve será estudado com detalhes esse comportamento sazonal da série).

```
vinho_saz <- wineind - dec$trend  
  
monthplot(vinho_saz, ylab= "Efeito Sazonal", xlab="Mês",  
          main ="Vendas de vinho")
```

Vendas de vinho



Acima temos uma análise descritiva muito interessante estudando apenas o componente sazonal da série. Pode-se notar que as menores vendas (valores negativos do gráfico) dentre os anos de 1980 e 1994 são dos meses de Janeiro a Março, e, os maiores valores de vendas totais são registrados nos meses de novembro e dezembro.

```
# splitando os dados em treino e teste
```

```
train <- wineind[1:150]  
test  <- wineind[151:176]
```

Para a realização das análises dos modelos de séries temporais foi-se dividido a base em treino e teste. Para teste foi escolhido utilizar 150 meses e para teste 26 meses.

```
# Transformando os dados de train em objeto do tipo ts
```

```
tswineind<-ts(train, start=c(1980,1), frequency=12)
```

```
# Utilizando a função auto.arima para identificar o 'melhor' modelo para a série temporal de vinhos
```

```
tswineindbest <- auto.arima(tswineind,approximation = FALSE)
```

```
# Modelo escolhido ARIMA(0,1,1)(0,1,1)[12]
```

```
tswineindbest
```

```
## Series: tswineind
```

```
## ARIMA(0,1,1)(0,1,1)[12]
##
## Coefficients:
##          ma1      sma1
##       -0.8990  -0.5484
## s.e.    0.0337   0.1068
##
## sigma^2 = 5688118: log likelihood = -1261.97
## AIC=2529.93   AICc=2530.11   BIC=2538.69
```

```
# Utilizando o modelo de Espaço de Estados
```

```
fit_BSM <- StructTS(tswineind, "BSM")
fit_BSM$coef
```

```
##          level      slope      seas      epsilon
## 256977.713    1047.205 2330407.485 1435733.381
```

Os modelos apresentados neste projeto e que iram compor a escolha dos modelos para a previsão da série temporal de vendas são: Arima e Modelo de Espaço de Estados. Para o modelo Arima foi utilizado a função `auto.arima()` para a escolha do ‘melhor’ modelo (via AIC). Para o modelo de espaço de estados foi adotado o modelo que permite a inclusão de uma parte de tendência e sazonalidade.

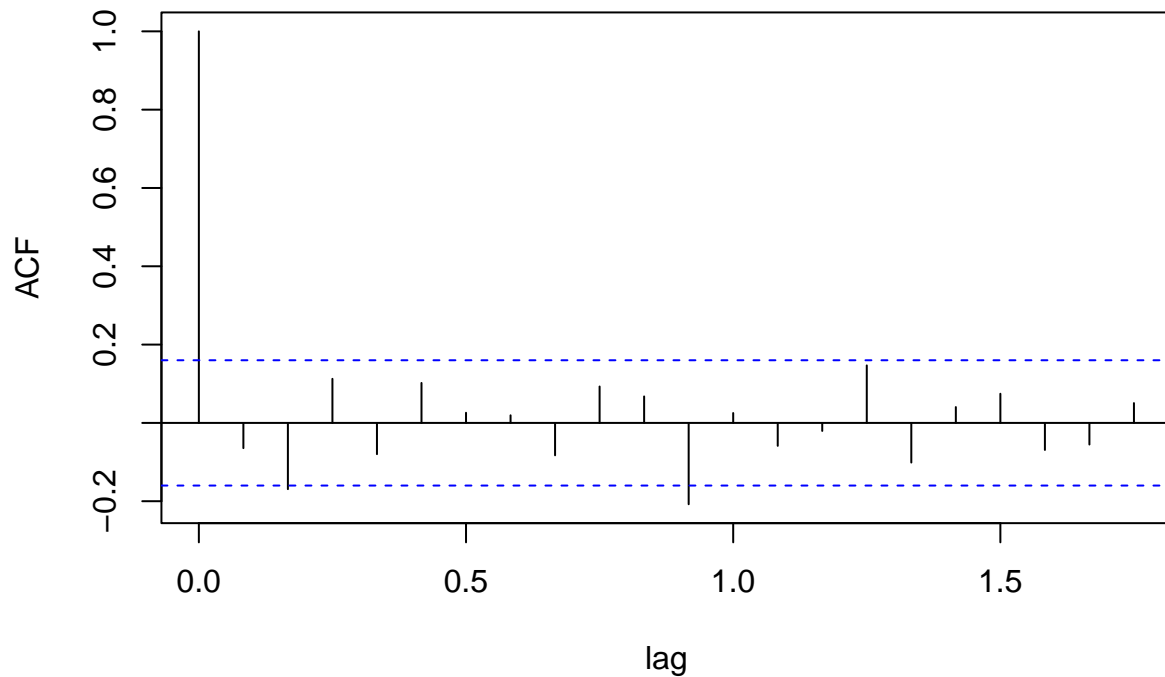
```
# Análise de Resíduos do Modelo Arima
```

```
res <- tswineindbest$residuals
```

```
# Gráfico de ACF
```

```
acf(tswineindbest$residuals, main = 'Autocorrelação', xlab = 'lag')
```

Autocorrelação

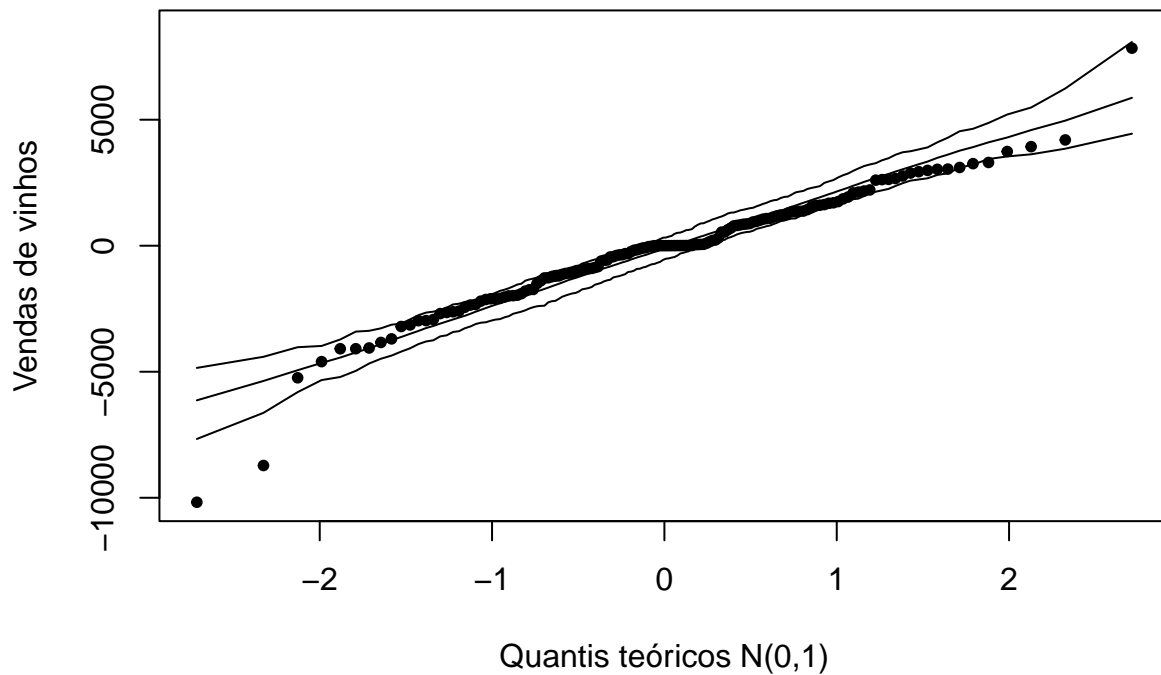


```
# Calculando o gráfico de envelope para distribuição normal

n <- length(res)
xb <- mean(res) # mu
sx <- sd(res) # sigma

## 2. Gráfico QQ com envelope
nsim <- 100 # Número de simulações
conf <- 0.95 # Coef. de confiança
# Dados simulados ~ normal
dadossim <- matrix(rnorm(n * nsim, mean = xb, sd = sx), nrow = n)
dadossim <- apply(dadossim, 2, sort)
# Limites da banda e média
infsup <- apply(dadossim, 1, quantile, probs = c((1 - conf) / 2,
                                              (1 + conf) / 2))

xbsim <- rowMeans(dadossim)
# Gráfico
faixay <- range(res, dadossim)
qq0 <- qqnorm(res, main = "", xlab = "Quantis teóricos N(0,1)", pch = 20,
              ylab = "Vendas de vinhos", ylim = faixa)
eixox <- sort(qq0$x)
lines(eixox, xbsim)
lines(eixox, infsup[1,])
lines(eixox, infsup[2,])
```

Com base na análise dos resíduos do modelo Arima, pode-se notar que os resíduos apresentam característica de ruído branco, isto é, os pontos de autocorrelação estão quase todos dentro da banda de confiança do gráfico de ACF e o gráfico de envelope da normal contempla grande parte dos pontos da série temporal, apenas indicando três pontos que são possíveis pontos de anomalia da série (valores discrepantes). Em outras palavras o modelo está adequado para realizar previsão.

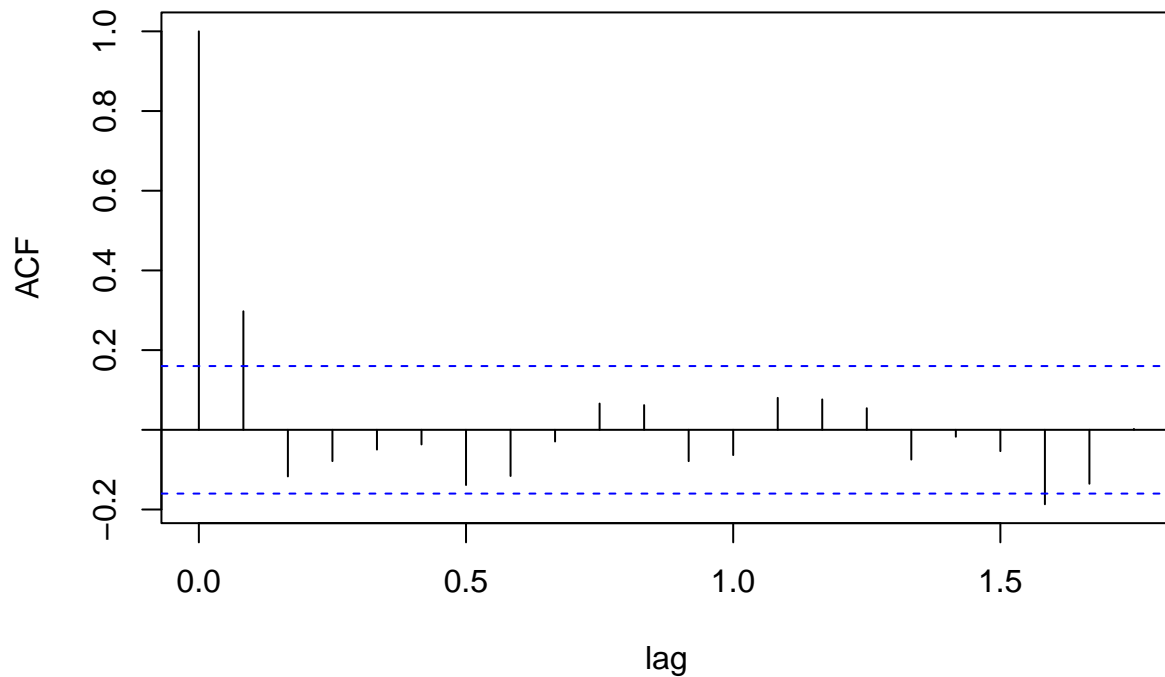
```
# Análise de Resíduos do modelo de espaço de Estados

res <- fit_BSM$residuals

# Gráfico de ACF

acf(fit_BSM$residuals, main = 'Autocorrelação', xlab = 'lag')
```

Autocorrelação

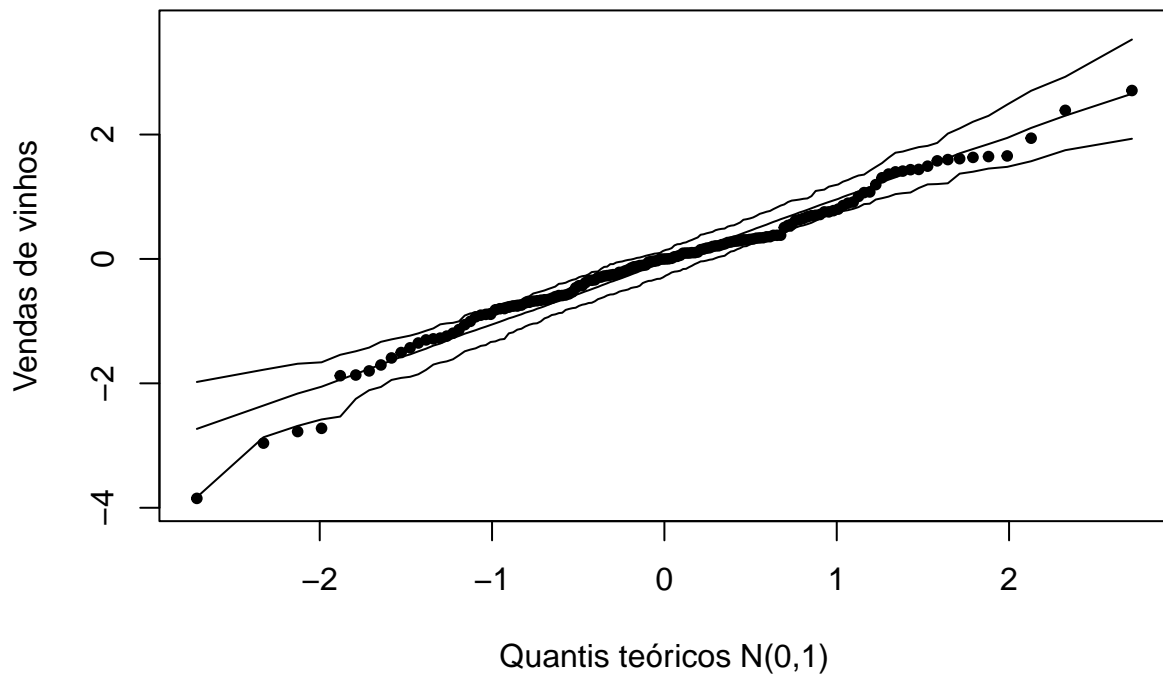


```
# Calculando o gráfico de envelope para distribuição normal

n <- length(res)
xb <- mean(res) # mu
sx <- sd(res) # sigma

## 2. Gráfico QQ com envelope
nsim <- 100 # Número de simulações
conf <- 0.95 # Coef. de confiança
# Dados simulados ~ normal
dadossim <- matrix(rnorm(n * nsim, mean = xb, sd = sx), nrow = n)
dadossim <- apply(dadossim, 2, sort)
# Limites da banda e média
infsup <- apply(dadossim, 1, quantile, probs = c((1 - conf) / 2,
                                              (1 + conf) / 2))

xbsim <- rowMeans(dadossim)
# Gráfico
faixay <- range(res, dadossim)
qq0 <- qqnorm(res, main = "", xlab = "Quantis teóricos N(0,1)", pch = 20,
              ylab = "Vendas de vinhos", ylim = faixa)
eixox <- sort(qq0$x)
lines(eixox, xbsim)
lines(eixox, infsup[1,])
lines(eixox, infsup[2,])
```



De maneira similar ao modelo Arima, o modelo de Espaço de Estados, mostrou que há indícios de que os resíduos da série são um ruído branco, isto é, o modelo também é adequado para realizar previsão para o conjunto de dados de vendas de vinho.

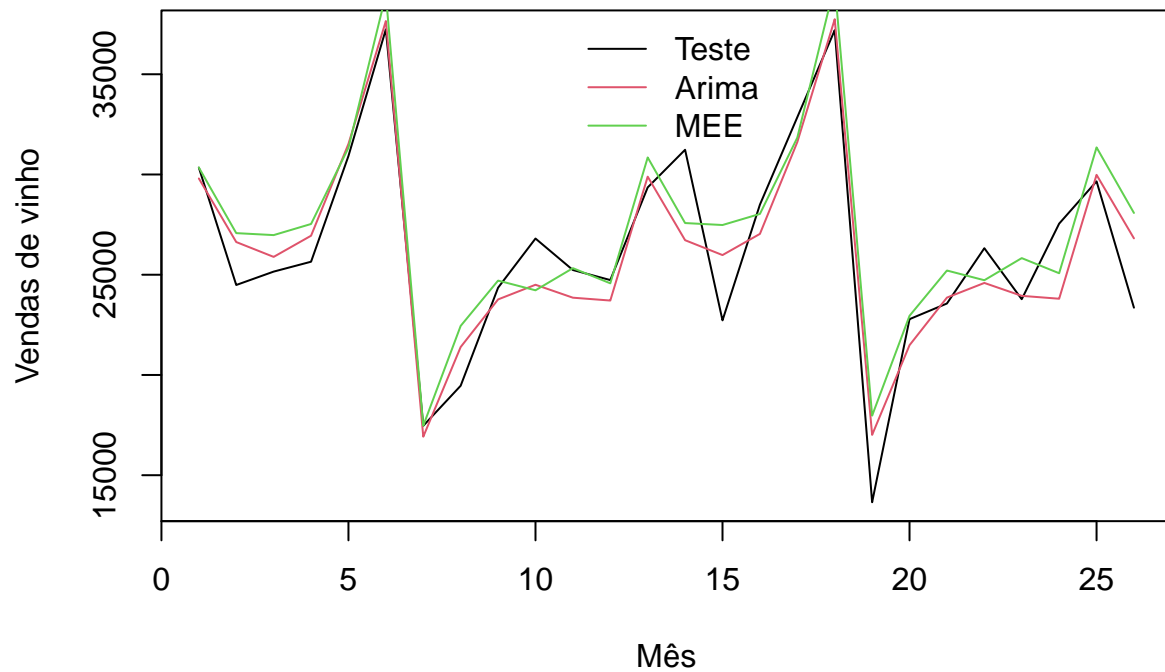
```
# Previsão dos 26 meses do conjunto de dados teste

prev <-forecast(tswineindbest, 26)

prev_BSM <-forecast(fit_BSM, 26)

plot(1:26, test, type = 'l', xlab="Mês", ylab = "Vendas de vinho", main = "Previsão dos modelos" )
lines(1:26, prev$mean, col =2)
lines(1:26,prev_BSM$mean, col=3)
legend("top",c("Teste","Arima", 'MEE'), col=1:3, lty=1, bty="n")
```

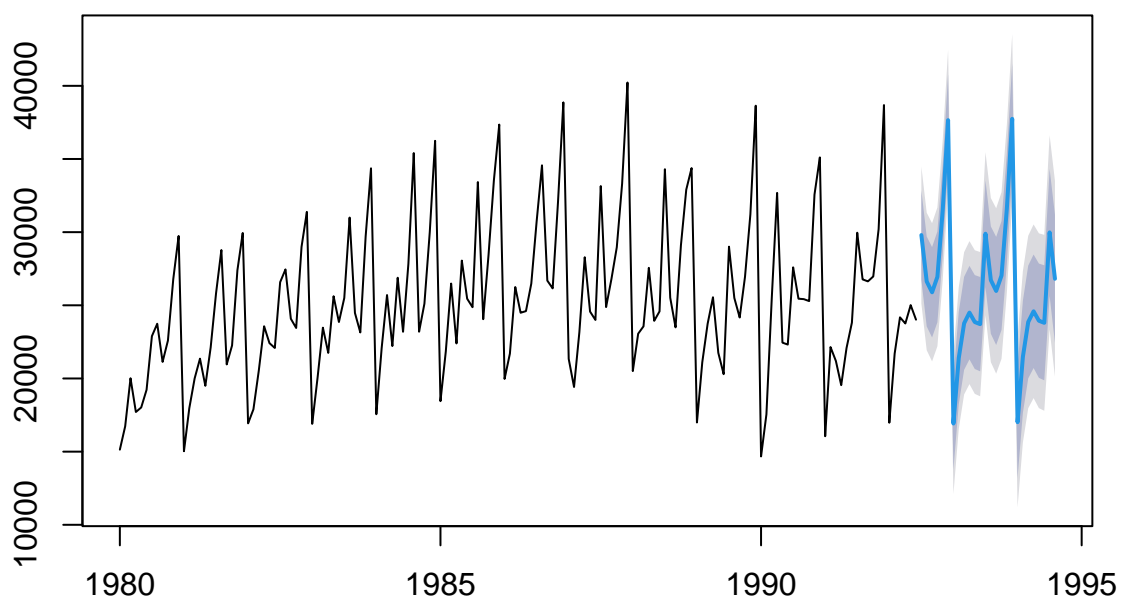
Previsão dos modelos



O modelo Arima e modelo de Espaço de Estados apresentam graficamente resultados similares em relação a aplicação da previsão para os dados de treino. Nota-se por exemplo que os modelos apresentam uma boa previsão para os dados teste, mostrando que os modelos são razoáveis para realizar previsão.

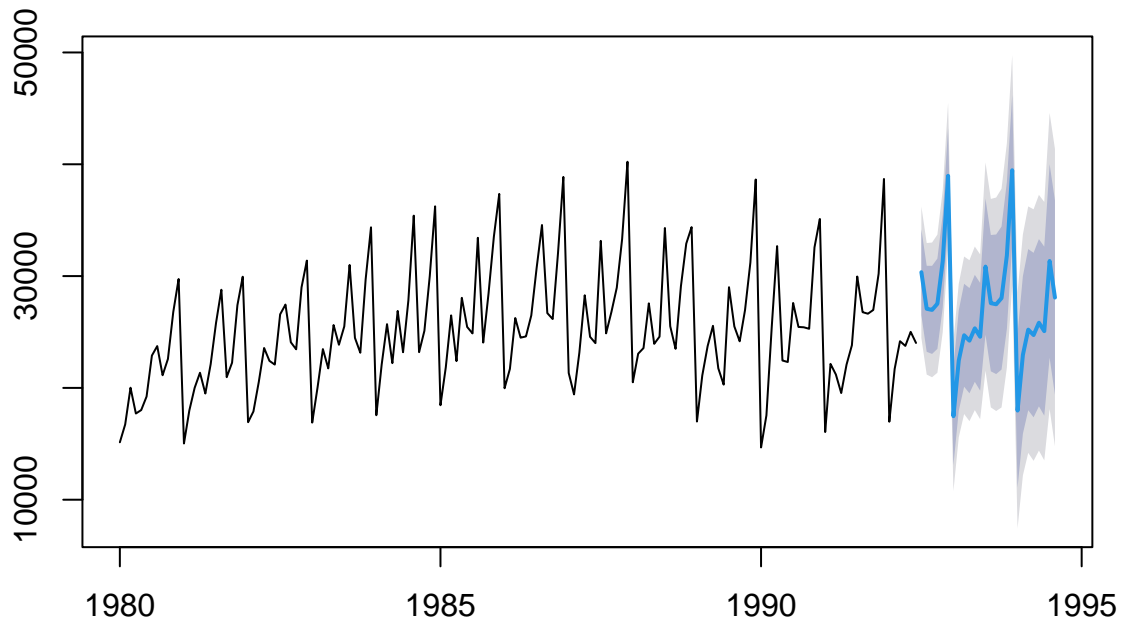
```
# Gráfico de previsão considerando os dados de treino e a previsão do modelo Arima  
plot(prev)
```

Forecasts from ARIMA(0,1,1)(0,1,1)[12]



```
# Gráfico de previsão considerando os dados de treino e a previsão do modelo de espaço de estados  
plot(prev_BSM)
```

Forecasts from Basic structural model



Acima temos os gráficos de previsão considerando os dois modelos, e pelo menos a forma dos gráficos apontam certa similaridade entre os modelos.

```
# Tabela contendo os valores previstos pelo modelo Arima e o valor real (dados teste)
previsao <- data.frame(prev, test, prev_BSM$mean )
previsao <- previsao[,c(6, 1,7)]
names(previsao) <- c('Teste', 'Arima', 'MEE')
kable(previsao, format = "markdown")
```

| | Teste | Arima | MEE |
|----------|-------|----------|----------|
| Jul 1992 | 30345 | 29801.73 | 30347.29 |
| Aug 1992 | 24488 | 26631.78 | 27075.58 |
| Sep 1992 | 25156 | 25890.33 | 26977.26 |
| Oct 1992 | 25650 | 26946.32 | 27531.07 |
| Nov 1992 | 30923 | 31525.12 | 31335.09 |
| Dec 1992 | 37240 | 37654.75 | 38966.42 |
| Jan 1993 | 17466 | 16922.03 | 17471.22 |
| Feb 1993 | 19463 | 21393.84 | 22448.22 |
| Mar 1993 | 24352 | 23768.50 | 24708.10 |
| Apr 1993 | 26805 | 24500.51 | 24223.54 |
| May 1993 | 25236 | 23855.44 | 25327.08 |
| Jun 1993 | 24735 | 23712.69 | 24572.44 |
| Jul 1993 | 29356 | 29892.57 | 30849.78 |
| Aug 1993 | 31234 | 26722.61 | 27578.08 |
| Sep 1993 | 22724 | 25981.16 | 27479.76 |

| | Teste | Arima | MEE |
|----------|-------|----------|----------|
| Oct 1993 | 28496 | 27037.16 | 28033.57 |
| Nov 1993 | 32857 | 31615.95 | 31837.59 |
| Dec 1993 | 37198 | 37745.59 | 39468.92 |
| Jan 1994 | 13652 | 17012.87 | 17973.72 |
| Feb 1994 | 22784 | 21484.68 | 22950.72 |
| Mar 1994 | 23565 | 23859.33 | 25210.60 |
| Apr 1994 | 26323 | 24591.35 | 24726.04 |
| May 1994 | 23779 | 23946.27 | 25829.58 |
| Jun 1994 | 27549 | 23803.52 | 25074.94 |
| Jul 1994 | 29660 | 29983.40 | 31352.28 |
| Aug 1994 | 23356 | 26813.45 | 28080.58 |

Para melhorar a visualização dos resultados do modelo Arima e MEE foi criado uma tabela para mostra os valores de previsão e os dados testes. Nota-se que os valores estão bem próximos dos valores de teste, mostrando novamente que os modelos são úteis para esse projeto.

```
# Calculando a métrica MAPE e MSE para os modelos
```

```
100*MAPE(prev$mean, test)
```

```
## [1] 6.327908
```

```
100*MAPE(prev_BSM$mean, test)
```

```
## [1] 7.543211
```

```
MSE(prev$mean, test)
```

```
## [1] 3756667
```

```
MSE(prev_BSM$mean, test)
```

```
## [1] 5251444
```

```
# Vejo que a porcentagem do erro médio do modelo Arima considerado aqui é de 6% e do modelo de espaço d
```

Para verificar qual modelo apresenta a melhor previsão foi utilizado duas métricas famosas, sendo elas: MAPE e MSE, deixo registrado aqui que qualquer métrica poderia ser utilizada, mas, para este problema foram essas duas. Nota-se tanto pelo MAPE quanto pelo MSE que o modelo Arima apresenta vantagem em relação ao modelo MEE. Entretanto, é importante destacar que o modelo MEE também apresenta uma razoável previsão. Portanto, para analisar os meses a previsão do modelo Arima para os meses restantes de 1994 até 1995.

```
# Utilizando o modelo Arima (0,1,1)(0,1,1)12 para a produção.
```

```
modelo <- arima(wineind, order = c(0,1,1), seasonal = c(0,1,1))
```

```
# Realizando previsão para os próximos 16 meses
```

```
previsão <- forecast(modelo, 16)
```

```
kable(previsão, format = "markdown")
```

| | Point Forecast | Lo 80 | Hi 80 | Lo 95 | Hi 95 |
|----------|----------------|----------|----------|----------|----------|
| Sep 1994 | 24461.26 | 21474.54 | 27447.98 | 19893.46 | 29029.05 |
| Oct 1994 | 27229.90 | 24228.55 | 30231.26 | 22639.73 | 31820.08 |
| Nov 1994 | 31923.22 | 28907.30 | 34939.14 | 27310.77 | 36535.67 |
| Dec 1994 | 37332.16 | 34301.75 | 40362.57 | 32697.55 | 41966.77 |
| Jan 1995 | 15873.28 | 12828.45 | 18918.12 | 11216.62 | 20529.95 |
| Feb 1995 | 21424.36 | 18365.17 | 24483.54 | 16745.73 | 26102.98 |
| Mar 1995 | 23842.84 | 20769.37 | 26916.32 | 19142.37 | 28543.32 |
| Apr 1995 | 25829.36 | 22741.66 | 28917.06 | 21107.13 | 30551.58 |
| May 1995 | 24122.97 | 21021.12 | 27224.82 | 19379.09 | 28866.85 |
| Jun 1995 | 25370.80 | 22254.85 | 28486.74 | 20605.37 | 30136.23 |
| Jul 1995 | 29757.06 | 26627.09 | 32887.04 | 24970.18 | 34543.95 |
| Aug 1995 | 26244.80 | 23100.86 | 29388.74 | 21436.56 | 31053.05 |
| Sep 1995 | 24491.29 | 21042.18 | 27940.41 | 19216.32 | 29766.26 |
| Oct 1995 | 27259.94 | 23786.86 | 30733.01 | 21948.32 | 32571.55 |
| Nov 1995 | 31953.25 | 28456.38 | 35450.13 | 26605.25 | 37301.26 |
| Dec 1995 | 37362.19 | 33841.68 | 40882.70 | 31978.04 | 42746.35 |

O modelo Arima apresentou métricas que mostravam que seus erros de previsão em relação a venda eram menores que o modelo MEE. Dessa forma, a previsão dos 16 meses (meses restantes de 1994 e o ano completo de 1995) foram realizadas com o modelo Arima. Com base nos resultados, pode-se notar que os meses finais de 1994 apresentam mais vendas que os demais meses de 1994, situação similar ocorre em 1995, em que os meses próximos do final do ano apresentam a maior quantidade de vendas de vinhos australianos.

No que se refere ao melhor período de venda de vinho e lucro em relação a venda, com base no que foi apresentado neste projeto, os meses do final do ano são os melhores períodos de venda de vinhos, isto porque a venda tem período sazonal, ou seja, sempre aumentam no final do ano, em época de comemoração e datas especiais.