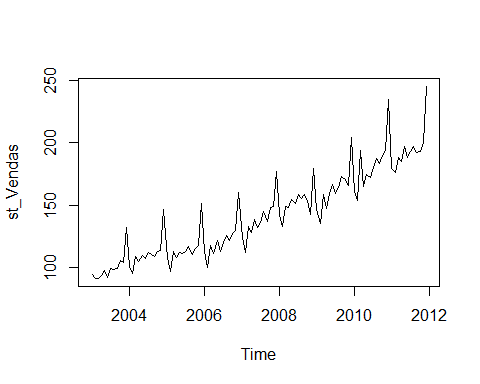
## Análise de serie temporal do IBGE sobre o volume de vendas do comercio varejista ampliado.

Serie em:[IBGE-series estatisticas]<http://seriesestatisticas.ibge.gov.br/series.aspx?no=2&op=0&vcodigo=MC67&t=volume-vendas-comercio-varejista-ampliado-tipos> extraída em 02/02/2018.

carregando arquivo e preparando os dados

plotagem da serie

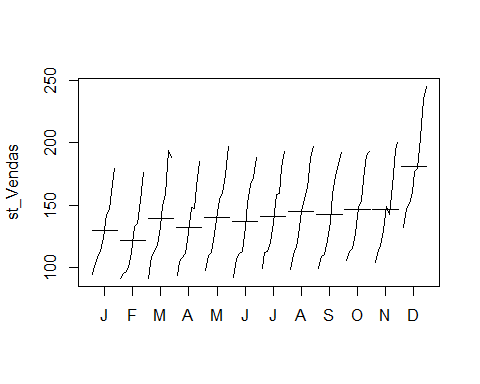
ts.plot(st\_Vendas)



plotagem da serie, nota-se variação sazonal muito semelhante até o ano de 2008.

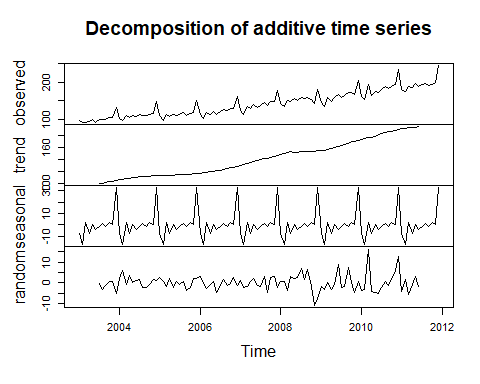
Gráfico para análise de médias e tendencia

monthplot(st\_Vendas)



decomposição da série nas principais componentes

plot(decompose(st\_Vendas))



observase que o valor mínimo de vendas está proximo de 100 e o máximo de 190.

A componente de tendencia afeta fortemente a série (em torno de 70%).

A componente sazonal presente na série afeta em torno de 25% e o RB é de pequeno valor e não aparenta estar contaminado pela componente sazonal

Precisa-se modelar as componentes de tendencia e sazonal, o que amerita o uso do modelo SARIMA

função de autocorrelação BETS.corrgram(ts, lag.max = 12, type = “correlation”, mode = “simple”, ci = 0.95, style = “plotly”, knit = F)

ainda nao eh estacionaria, pelo grafico, já que não tem decrescimento exponencial ou senoidal

BETS.corrgram(st\_Vendas,   
 lag.max = 36) # 36 = 3 anos

testes de RU para determinação o número de diferenciações necessárias para transformar a ST em estacionária usarei ADF (Augmented Dickey Fuller)

adf.drift1 <- ur.df(y = st\_Vendas,   
 type = c("drift"),  
 lags = 24,   
 selectlags = "AIC")  
  
summary(adf.drift1)

##   
## ###############################################   
## # Augmented Dickey-Fuller Test Unit Root Test #   
## ###############################################   
##   
## Test regression drift   
##   
##   
## Call:  
## lm(formula = z.diff ~ z.lag.1 + 1 + z.diff.lag)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -17.8129 -3.4059 0.4511 2.8084 14.4397   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -2.436116 4.527063 -0.538 0.592587   
## z.lag.1 0.056546 0.041364 1.367 0.176978   
## z.diff.lag1 -0.559269 0.147104 -3.802 0.000352 \*\*\*  
## z.diff.lag2 -0.561102 0.162527 -3.452 0.001054 \*\*   
## z.diff.lag3 -0.337501 0.178597 -1.890 0.063882 .   
## z.diff.lag4 -0.390888 0.174865 -2.235 0.029327 \*   
## z.diff.lag5 -0.296366 0.183232 -1.617 0.111306   
## z.diff.lag6 -0.255181 0.188643 -1.353 0.181487   
## z.diff.lag7 -0.323772 0.188507 -1.718 0.091306 .   
## z.diff.lag8 -0.298081 0.188984 -1.577 0.120266   
## z.diff.lag9 -0.019517 0.191986 -0.102 0.919387   
## z.diff.lag10 -0.511226 0.201167 -2.541 0.013789 \*   
## z.diff.lag11 -0.371567 0.211701 -1.755 0.084607 .   
## z.diff.lag12 0.135538 0.223276 0.607 0.546233   
## z.diff.lag13 0.025126 0.218002 0.115 0.908649   
## z.diff.lag14 0.028408 0.211750 0.134 0.893751   
## z.diff.lag15 -0.143893 0.203096 -0.708 0.481524   
## z.diff.lag16 -0.128458 0.202771 -0.634 0.528933   
## z.diff.lag17 -0.170658 0.196930 -0.867 0.389799   
## z.diff.lag18 -0.246345 0.188189 -1.309 0.195777   
## z.diff.lag19 -0.145189 0.188323 -0.771 0.443917   
## z.diff.lag20 -0.189747 0.187125 -1.014 0.314861   
## z.diff.lag21 -0.464068 0.187647 -2.473 0.016400 \*   
## z.diff.lag22 -0.007434 0.206159 -0.036 0.971361   
## z.diff.lag23 -0.144717 0.186064 -0.778 0.439915   
## z.diff.lag24 0.367329 0.173505 2.117 0.038627 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 6.532 on 57 degrees of freedom  
## Multiple R-squared: 0.9086, Adjusted R-squared: 0.8686   
## F-statistic: 22.68 on 25 and 57 DF, p-value: < 2.2e-16  
##   
##   
## Value of test-statistic is: 1.367 2.2747   
##   
## Critical values for test statistics:   
## 1pct 5pct 10pct  
## tau2 -3.46 -2.88 -2.57  
## phi1 6.52 4.63 3.81

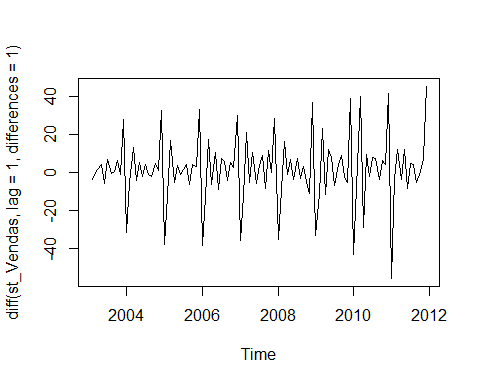
Ao analisar a estatística de teste (tau2 = 1,367) notamos que seu valor é superior ao valor crítico associado ao nível de confiança de 95% (-2,88).

Dessa forma, conclui-se que a ST não é estacionária (não rejeição da hipotese nula)

Dado que a nossa ST é não estacionária, vamos tentar torná-la estacionária fazendo uma diferenciação e vamos observar o gráfico e a FAC novamente.

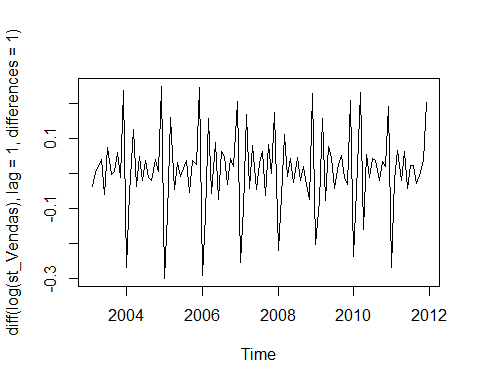
tornando a serie estacionaria…tirando a tendencia

ts.plot(diff(st\_Vendas,   
 lag = 1,   
 differences = 1))



A ST aparenta ser estacionária, a variancia tem um crescimento pequeno nos últimos anos, condicente com a teoria económica de crescimento vamos usar a funcao log para tirar o crescimento da variancia

ts.plot(diff(log(st\_Vendas),   
 lag = 1,   
 differences = 1))



O gráfico mostra agora a ST estacionária na média e na variancia, acorde a teoria de Box & Jenkins

Avaliando a estacionariedade da parte sazonal

BETS.corrgram(diff(log (st\_Vendas),  
 lag = 1,   
 differences = 1), lag.max = 60)

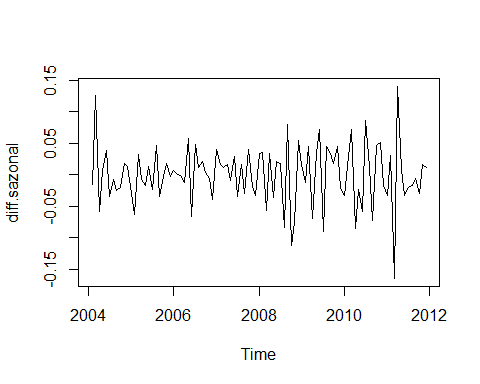
nota-se cortes nos lags 12, 24 e 36…

diferença simples

diff.simples <- diff(log(st\_Vendas),   
 lag = 1,   
 differences = 1)

diferença sazonal (12)

diff.sazonal <- diff(diff.simples,   
 lag =12,   
 differences = 1)  
ts.plot(diff.sazonal)



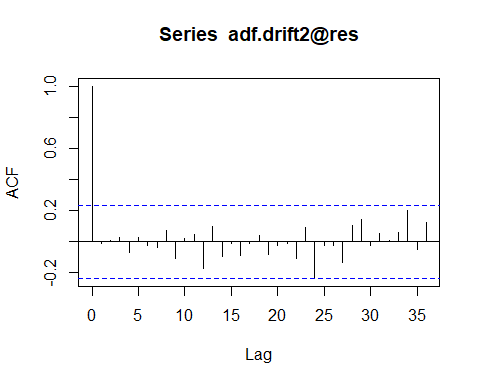
BETS.corrgram(diff.sazonal,   
 lag.max = 36)  
  
BETS.corrgram(diff.sazonal,   
 lag.max = 36,   
 type = "partial")

Refazendo o teste de RU depois das tranformações

adf.drift2 <- ur.df(y = diff(diff(log(st\_Vendas),   
 lag = 1),   
 lag = 12),  
 type = "drift",   
 lags = 24,   
 selectlags = "AIC")  
  
summary(adf.drift2)

##   
## ###############################################   
## # Augmented Dickey-Fuller Test Unit Root Test #   
## ###############################################   
##   
## Test regression drift   
##   
##   
## Call:  
## lm(formula = z.diff ~ z.lag.1 + 1 + z.diff.lag)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.113293 -0.022799 0.007386 0.026332 0.085920   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.001290 0.005027 0.257 0.79843   
## z.lag.1 -2.392338 0.804970 -2.972 0.00433 \*\*  
## z.diff.lag1 0.967005 0.786414 1.230 0.22388   
## z.diff.lag2 0.618829 0.756582 0.818 0.41680   
## z.diff.lag3 0.555728 0.720832 0.771 0.44392   
## z.diff.lag4 0.465794 0.669161 0.696 0.48920   
## z.diff.lag5 0.456159 0.614510 0.742 0.46095   
## z.diff.lag6 0.512886 0.563424 0.910 0.36650   
## z.diff.lag7 0.436938 0.503930 0.867 0.38954   
## z.diff.lag8 0.397217 0.432846 0.918 0.36265   
## z.diff.lag9 0.586323 0.359713 1.630 0.10862   
## z.diff.lag10 0.414260 0.255440 1.622 0.11038   
## z.diff.lag11 0.425706 0.147228 2.891 0.00542 \*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.04164 on 57 degrees of freedom  
## Multiple R-squared: 0.8068, Adjusted R-squared: 0.7662   
## F-statistic: 19.84 on 12 and 57 DF, p-value: 4.083e-16  
##   
##   
## Value of test-statistic is: -2.972 4.4194   
##   
## Critical values for test statistics:   
## 1pct 5pct 10pct  
## tau2 -3.51 -2.89 -2.58  
## phi1 6.70 4.71 3.86

acf(adf.drift2@res,   
 lag.max = 36,   
 drop.lag.0 = T)



teste de RU tau2= -2.972 é inferior ao valor crítico de -2.89, com 95% de probabilidade pelo que podemos concluir que a ST é estacionária

Modelado da ST com SARIMA

automático para ver a sugestão de modelo

auto.arima(st\_Vendas)

## Series: st\_Vendas   
## ARIMA(0,1,1)(0,1,2)[12]   
##   
## Coefficients:  
## ma1 sma1 sma2  
## -0.6243 -0.5146 -0.3001  
## s.e. 0.0843 0.1719 0.1797  
##   
## sigma^2 estimated as 30.72: log likelihood=-301.62  
## AIC=611.23 AICc=611.68 BIC=621.45

# Sugestão de modelo: ARIMA(0,1,1)(0,1,2)[12], AIC=611.23

fit.vendas1 <- Arima(st\_Vendas,  
 order = c(0,1,1),   
 seasonal = c(0,1,2),  
 lambda = 0)  
summary(fit.vendas1)

## Series: st\_Vendas   
## ARIMA(0,1,1)(0,1,2)[12]   
## Box Cox transformation: lambda= 0   
##   
## Coefficients:  
## ma1 sma1 sma2  
## -0.5852 -0.6397 -0.3600  
## s.e. 0.0846 0.2445 0.1453  
##   
## sigma^2 estimated as 0.001101: log likelihood=179.58  
## AIC=-351.15 AICc=-350.71 BIC=-340.94  
##   
## Training set error measures:  
## ME RMSE MAE MPE MAPE MASE  
## Training set -0.03930447 4.743846 3.220691 -0.02232722 2.16296 0.2669357  
## ACF1  
## Training set 0.004746829

Resultados: AIC=-351.15 AICc=-350.71 BIC=-340.94

outro modelo

fit.vendas2 <- Arima(st\_Vendas,   
 order = c(0,1,1),   
 seasonal = c(0,1,1),  
 lambda = 0)  
summary(fit.vendas2)

## Series: st\_Vendas   
## ARIMA(0,1,1)(0,1,1)[12]   
## Box Cox transformation: lambda= 0   
##   
## Coefficients:  
## ma1 sma1  
## -0.5990 -0.7235  
## s.e. 0.0868 0.1409  
##   
## sigma^2 estimated as 0.001318: log likelihood=176.67  
## AIC=-347.34 AICc=-347.08 BIC=-339.68  
##   
## Training set error measures:  
## ME RMSE MAE MPE MAPE MASE  
## Training set 0.02211907 5.20587 3.61367 0.00718641 2.428774 0.2995064  
## ACF1  
## Training set 0.02710828

# Resultados: AIC=-347.34 AICc=-347.08 BIC=-339.68

BETS.t\_test(fit.vendas1)

## Coeffs Std.Errors t Crit.Values Rej.H0  
## ma1 -0.5852160 0.08455046 6.921500 1.983264 TRUE  
## sma1 -0.6396968 0.24450927 2.616248 1.983264 TRUE  
## sma2 -0.3600249 0.14525972 2.478491 1.983264 TRUE

BETS.t\_test(fit.vendas2)

## Coeffs Std.Errors t Crit.Values Rej.H0  
## ma1 -0.5989926 0.08681092 6.899969 1.983038 TRUE  
## sma1 -0.7235438 0.14094308 5.133589 1.983038 TRUE

accuracy(fit.vendas1)

## ME RMSE MAE MPE MAPE MASE  
## Training set -0.03930447 4.743846 3.220691 -0.02232722 2.16296 0.2669357  
## ACF1  
## Training set 0.004746829

accuracy(fit.vendas2)# ver MAPE mean absolute percentage error 2,43% erro!!

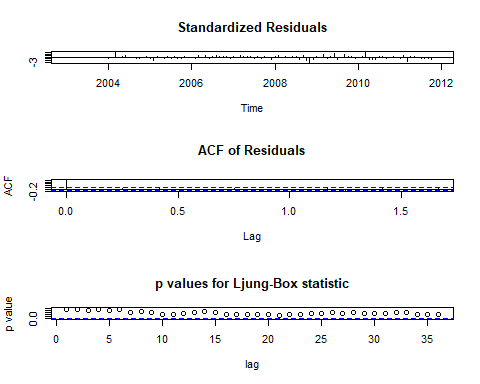
## ME RMSE MAE MPE MAPE MASE  
## Training set 0.02211907 5.20587 3.61367 0.00718641 2.428774 0.2995064  
## ACF1  
## Training set 0.02710828

Diagnóstico -

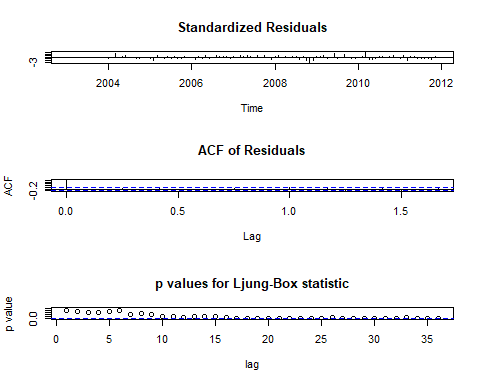
Verificar:

* Ausência de autocorrelação linear
* Ausência de heterocedasticidade condicional;
* Normalidade.

diag1 <- tsdiag(fit.vendas1, gof.lag = 36)



diag2 <- tsdiag(fit.vendas2, gof.lag = 36)



analise para diag2:

1- os dados aparentam estar distribuídos simetricamente em torno da média zero, indicação de distribuição normal. Não temos nenhuma informação discrepante fora do intervalo [-3,3], apenas um dado em dez 2009.

2- FAC dos resíduos sem presença de lags significantes

3- O valor de p do teste não rejeita a H0 e diminui significativamente após a desfazagem 14

1- Verificar a autocorrelação linear

Box.test(x = fit.vendas1$residuals, lag = 24,type = "Ljung-Box", fitdf = 2)

##   
## Box-Ljung test  
##   
## data: fit.vendas1$residuals  
## X-squared = 23.257, df = 22, p-value = 0.3873

Box.test(x = fit.vendas2$residuals, lag = 24,type = "Ljung-Box", fitdf = 2)

##   
## Box-Ljung test  
##   
## data: fit.vendas2$residuals  
## X-squared = 34.854, df = 22, p-value = 0.04013

analise para fit.vendas2

p-value = 0.04013 a 95% de confiança não rejeitamos a hipótese nula de não existência de autocorrelação serial até o lag 24.

2 - Ausência de heterocedasticidade condicional

BETS.arch\_test(fit.vendas1$residuals,lags = 12)

## statistic p.value htk  
## 1 6.884544 0.865148 TRUE

BETS.arch\_test(fit.vendas2$residuals,lags = 12)

## statistic p.value htk  
## 1 8.562208 0.7397986 TRUE

analise para fit.vendas2

Conforme mostrado pelo teste, a hipótese nula é que não há presença de efeito ARCH. Dessa forma, dado o valor do p-valor = 0.7398, não rejeitamos a hipótese nula a 95% de confiança, ou seja, a variância é estacionária.

* Normalidade dos resíduos

o teste de Jarque-Bera para normalidade, verifica se a amostra de dados tem uma curtosis e simetria semelhante a uma distribuição normal. Se os dados vem de uma dist. normal, o teste JB tem, asintóticamente, uma distribuição chi-quadrado com dois graus de liberdade, e assim a estatística pode ser usada para verificar a hipótese H0, de que os dados provem de uma dist. normal.

A H0 do teste diz que a assimetria e a curtosis tem de ser zero (ou 3 para curtosis, que é o mesmo). Qualquer diferença com essa hipótese, incrementa o valor da estatística JB.

Para amostras pequenas de dados, a aproximação de chi-quadrado é muito sensível a variações, causando a rejeição ainda que H0 seja verdadeira e levando assim a erros do tipo I.

Para um tamanho de amostra de aprox. 100 elementos o p-valor calculado equivalente ao verdadeiro nível de alfa é: 0.062 (valores aprox por simulação Monte Carlo).

Teste JB

r jb.norm.test(fit.vendas2$residuals, nrepl = 2000)

## ## Jarque-Bera test for normality ## ## data: fit.vendas2$residuals ## JB = 21.31, p-value = 0.0015

p-value = 0.002 como se vê o teste rejeita H0 de que a amostra provem de dist. normal

vamos verificar qual das duas condições do teste é violada:

Curtosis

kurtosis.norm.test(fit.vendas2$residuals, nrepl=2000)

##   
## Kurtosis test for normality  
##   
## data: fit.vendas2$residuals  
## T = 5.1661, p-value = 0.0015

p-value = 0.0015, a curtosis é rejeitada

simetria

skewness.norm.test(fit.vendas2$residuals, nrepl=2000)

##   
## Skewness test for normality  
##   
## data: fit.vendas2$residuals  
## T = -0.10403, p-value = 0.631

p-value = 0.6495 a simetria é aceita ou seja os resíduos tem simetria mas não tem a curtose de 0 (3) necessária

Pelo número baixo de resíduos, aceita-se a hipótese de normalidade dos mesmos, por possívelmente estar em presença de um erro tipo I (rejeição quando H0 é verdadeira, ou falso positivo)

Para amostras de tamanho pequeno (maior a 4), o pacote nortest tem a função lillie.test - *Lilliefors (Kolmogorov-Smirnov) test for normality*.

lillie.test(fit.vendas2$residuals)

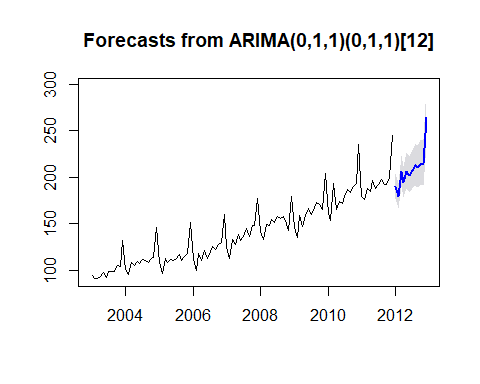
##   
## Lilliefors (Kolmogorov-Smirnov) normality test  
##   
## data: fit.vendas2$residuals  
## D = 0.064556, p-value = 0.3277

o valor p-value = 0.3277, estabelece que não podemos rejeitar a hipótese de que a amostra de resíduos é normal

Assim sendo passamos a etapa da previsão de vendas para os próximos

12 meses, com nível de confiança de 95%

previsao <- forecast(fit.vendas2,   
 h =12, # 12 meses  
 level=0.95)  
plot(previsao)



write.csv(data.frame(previsao), "previsao\_vendas.csv")  
write.xlsx(data.frame(previsao), "previsao\_vendas.xlsx")

# Fim do script