**MBA Executivo em Economia e Gestão: Business Analytics e Big Data**

***“Análise de Séries Temporais”***

Prof. Dr. João Luiz Chela

**Trabalho: Séries Temporais usando R**

17/06/2018

Equipe:

Everton Pacheco, João Roberto Perin

Natália Rosa Silva, Suely Fischer Omura

**SÉRIES TEMPORAIS USANDO R**

1. **Utilizando o arquivo “Serie\_Dados.csv” realize as seguintes etapas:**

**1.a. Crie a série temporal dos retornos LN, ou seja, r = Ln(P\_t+1/P\_t).**

**1.b. Para cada ação construa o histograma dos retornos. Comente o resultado dos histogramas, verifique também o desvio padrão e a média de cada série.**

*> par(mfrow=c(2,2))*

*> hist(VALE5, main="Histograma VALE5", col="sky blue")*

*> hist(GOLL4, main="Histograma GOLL4", col="steelblue")*

*> hist(AMBV4, main="Histograma AMBV4", col="dodgerblue")*

*> hist(ITUB4, main="Histograma ITUB4", col="navy")*

*>*

*> par(mfrow=c(2,2))*

*> hist(BBDC4, main="Histograma BBDC4", col="indianred")*

*> hist(BVMF3, main="Histograma BVMF3", col="rosybrown")*

*> hist(RAPT4, main="Histograma RAPTA4", col="salmon")*

*> hist(MYPK3, main="Histograma MYPK3", col="lavenderblush3")*

*>*

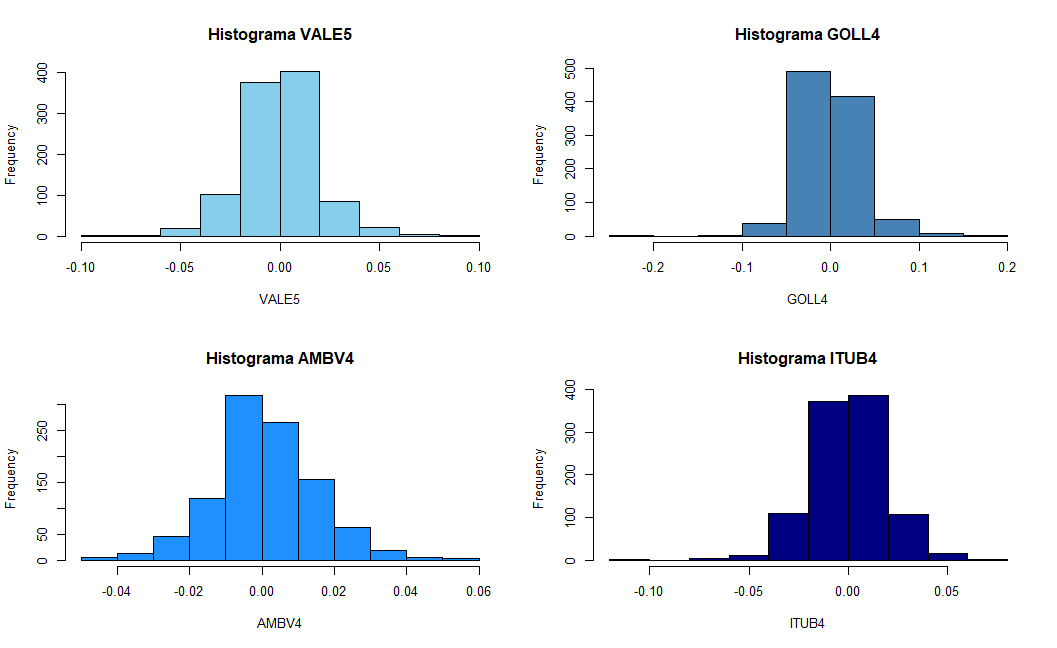
*> par(mfrow=c(2,2))*

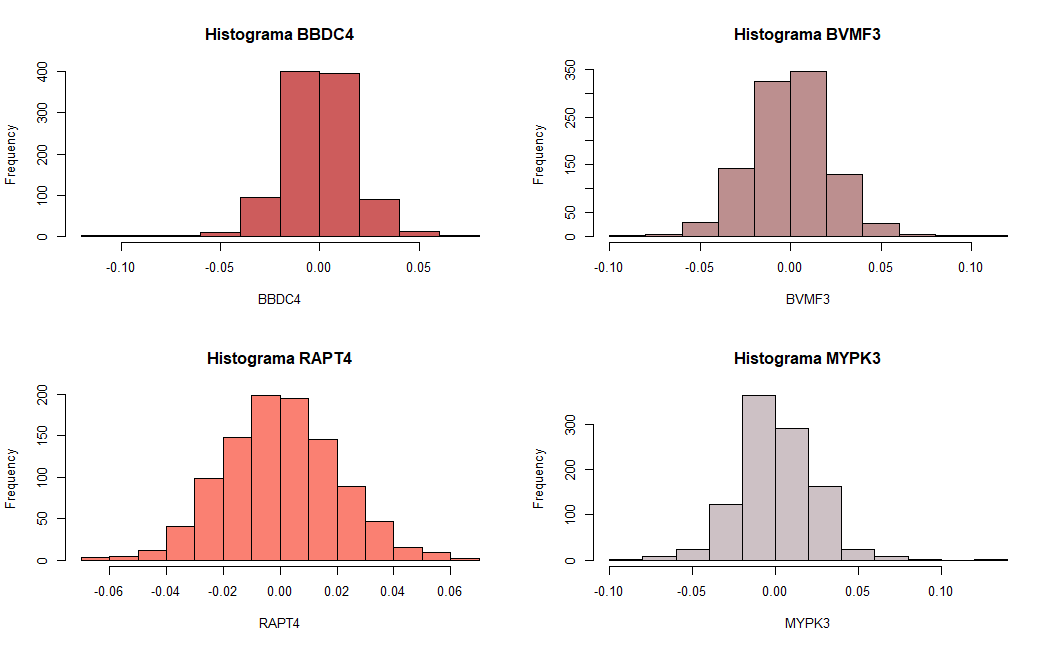
*> hist(GOAU4, main="Histograma GOAU4", col="powderblue")*

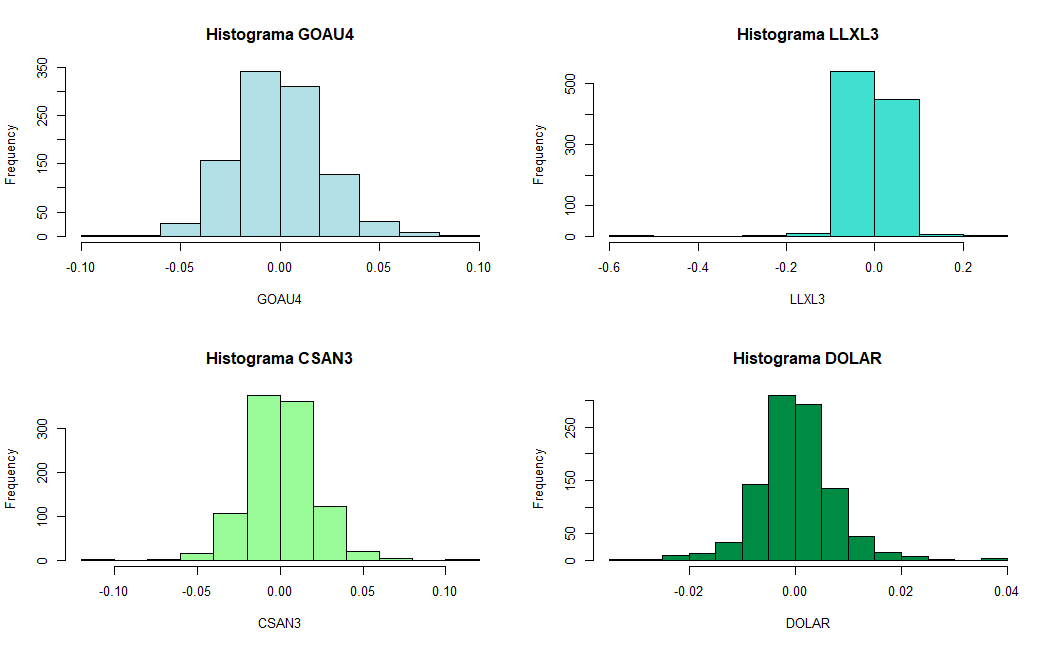
*> hist(LLXL3, main="Histograma LLXL3", col="turquoise")*

*> hist(CSAN3, main="Histograma CSAN3", col="palegreen")*

*> hist(DOLAR, main="Histograma DOLAR", col="springgreen4")*







* *Os histogramas de todas as séries indicam que a variação dos retornos segue uma distribuição normal. No entanto, a série que apresenta a distribuição mais normal é a RAPT4.*
* *As séries que apresentam outliers são: GOLL4, ITUB4, MYPK3, LLXL3, CSAN3 e DOLAR.*

*> mean(VALE5)*

*[1] 0.0001292354*

*> mean(GOLL4)*

*[1] -0.0004912977*

*> mean(AMBV4)*

*[1] 0.001271991*

*> mean(ITUB4)*

*[1] -3.20353e-05*

*> mean(BBDC4)*

*[1] 0.0002094082*

*> mean(BVMF3)*

*[1] 9.446902e-05*

*> mean(RAPT4)*

*[1] 0.0003500063*

*> mean(MYPK3)*

*[1] 0.001023929*

*> mean(GOAU4)*

*[1] -0.0001970191*

*> mean(LLXL3)*

*[1] -0.001223711*

*> mean(CSAN3)*

*[1] 0.0008826716*

*> mean(DOLAR)*

*[1] 0.0002535005*

*>*

*> sd(VALE5)*

*[1] 0.01839278*

*> sd(GOLL4)*

*[1] 0.03247557*

*> sd(AMBV4)*

*[1] 0.01426501*

*> sd(ITUB4)*

*[1] 0.01832851*

*> sd(BBDC4)*

*[1] 0.0170888*

*> sd(BVMF3)*

*[1] 0.02196125*

*> sd(RAPT4)*

*[1] 0.02013209*

*> sd(MYPK3)*

*[1] 0.02255484*

*> sd(GOAU4)*

*[1] 0.02253028*

*> sd(LLXL3)*

*[1] 0.04117323*

*> sd(CSAN3)*

*[1] 0.01927904*

*> sd(DOLAR)*

*[1] 0.007316637*

* *A série com a menor média é a ITUB4 e a série com o menor desvio padrão é a AMBV4.*

**1.c. Calcule o ACF e PACF de cada série de retornos. Comente os resultados.**

*> par(mfrow=c(2,2))*

*> acf(VALE5, main="Série VALE5", col="skyblue")*

*> pacf(VALE5, main="Série VALE5", col="skyblue")*

*> acf(GOLL4, main="Série GOLL4", col="steelblue")*

*> pacf(GOLL4, main="Série GOLL4", col="steelblue")*

*>*

*> par(mfrow=c(2,2))*

*> acf(AMBV4, main="Série AMBV4", col="dodgerblue")*

*> pacf(AMBV4, main="Série AMBV4", col="dodgerblue")*

*> acf(ITUB4, main="Série ITUB4", col="navy")*

*> pacf(ITUB4, main="Série ITUB4", col="navy")*

*>*

*> par(mfrow=c(2,2))*

*> acf(BBDC4, main="Série BBDC4", col="indianred")*

*> pacf(BBDC4, main="Série BBDC4", col="indianred")*

*> acf(BVMF3, main="Série BVMF3", col="rosybrown")*

*> pacf(BVMF3, main="Série BVMF3", col="rosybrown")*

*>*

*> par(mfrow=c(2,2))*

*> acf(RAPT4, main="Série RAPT4", col="salmon")*

*> pacf(RAPT4, main="Série RAPT4", col="salmon")*

*> acf(MYPK3, main="Série MYPK3", col="lavenderblush3")*

*> pacf(MYPK3, main="Série MYPK3", col="lavenderblush3")*

*>*

*> par(mfrow=c(2,2))*

*> acf(GOAU4, main="Série GOAU4", col="powderblue")*

*> pacf(GOAU4, main="Série GOAU4", col="powderblue")*

*> acf(LLXL3, main="Série LLXL3", col="turquoise")*

*> pacf(LLXL3, main="Série LLXL3", col="turquoise")*

*>*

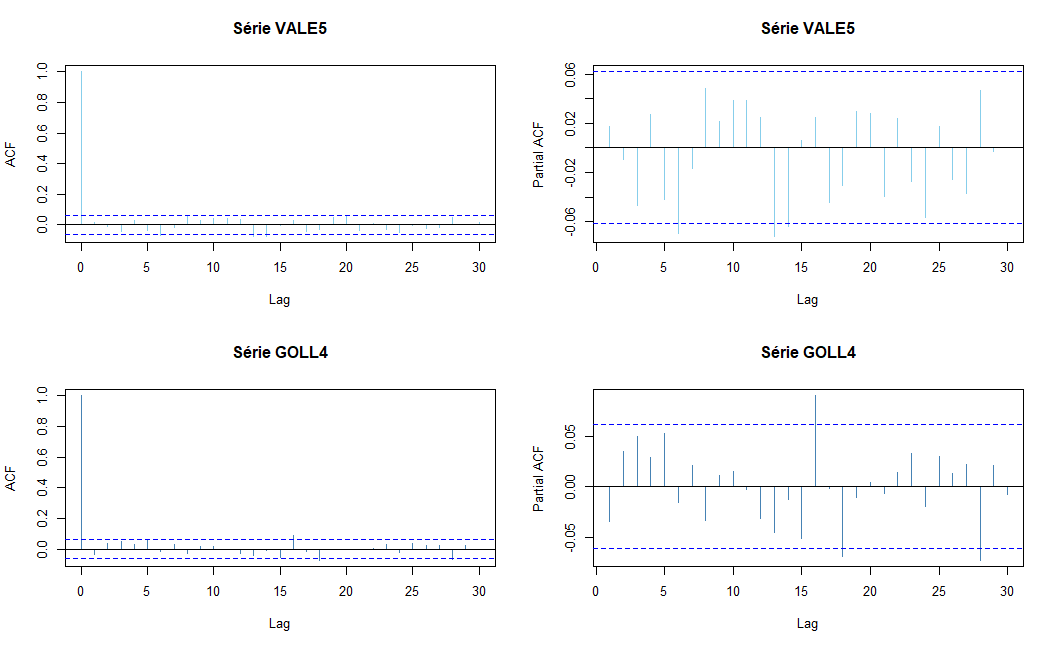
*> par(mfrow=c(2,2))*

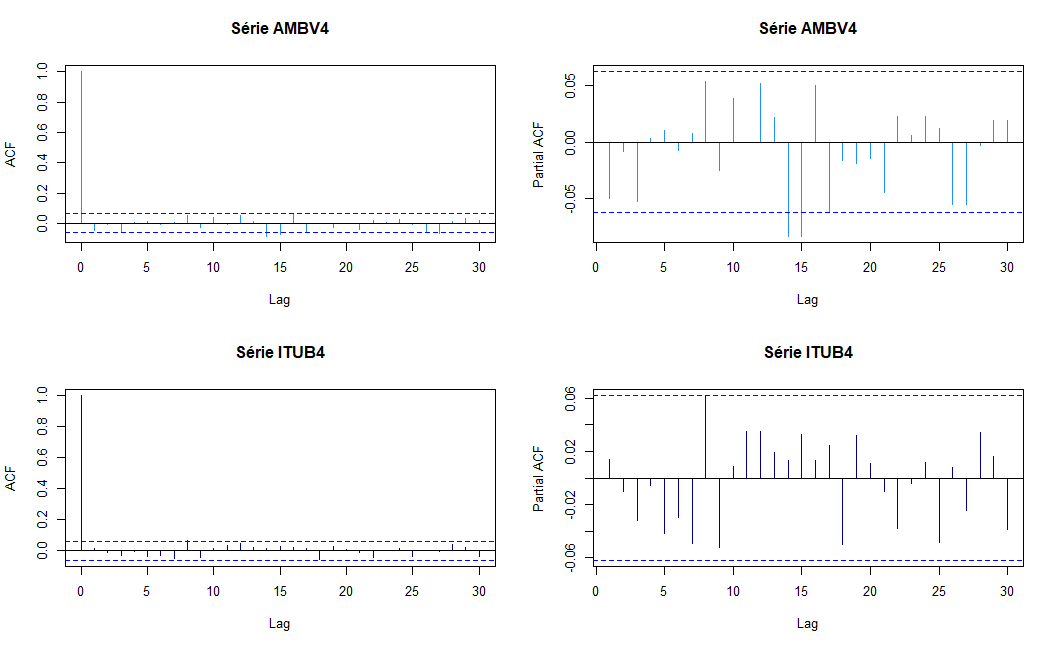
*> acf(CSAN3, main="Série CSAN3", col="palegreen")*

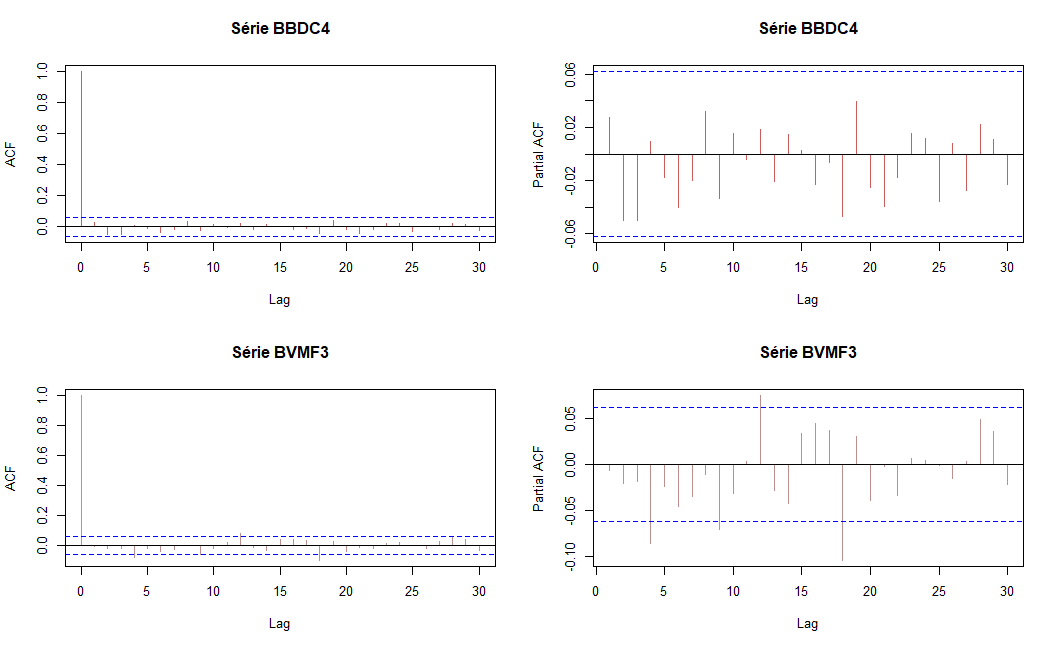
*> pacf(CSAN3, main="Série CSAN3", col="palegreen")*

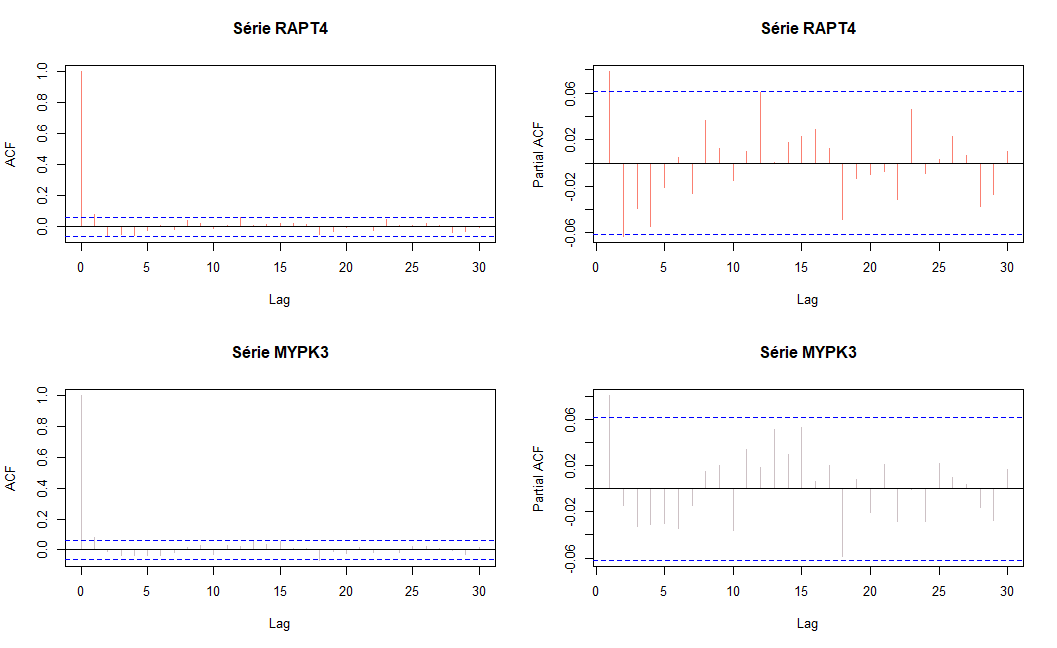
*> acf(DOLAR, main="Série DOLAR", col="springgreen4")*

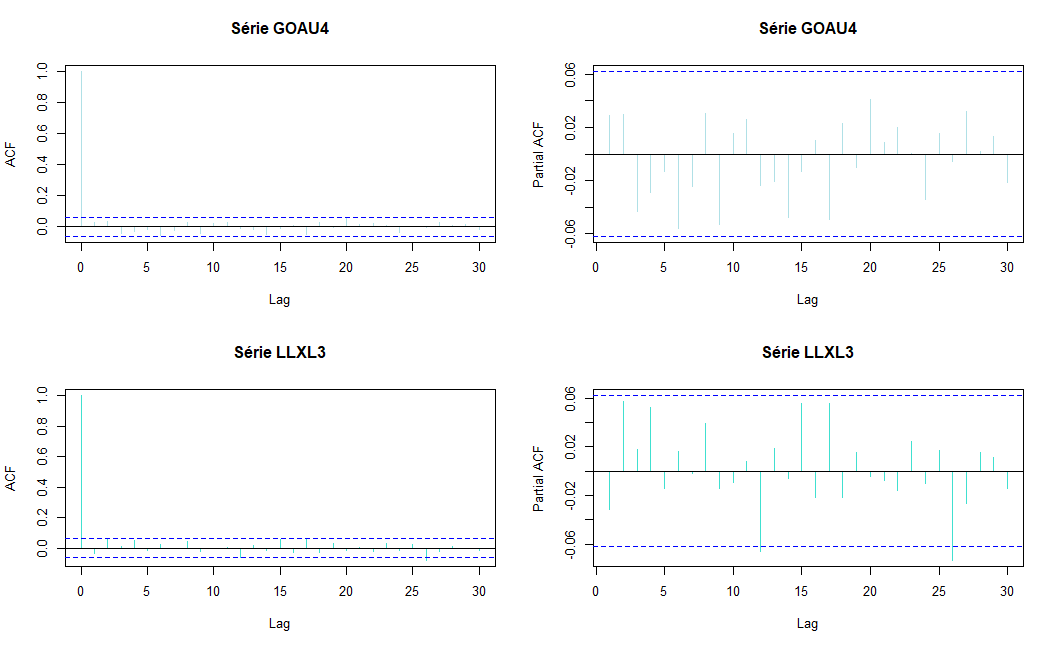
*> pacf(DOLAR, main="Série DOLAR", col="springgreen4")*

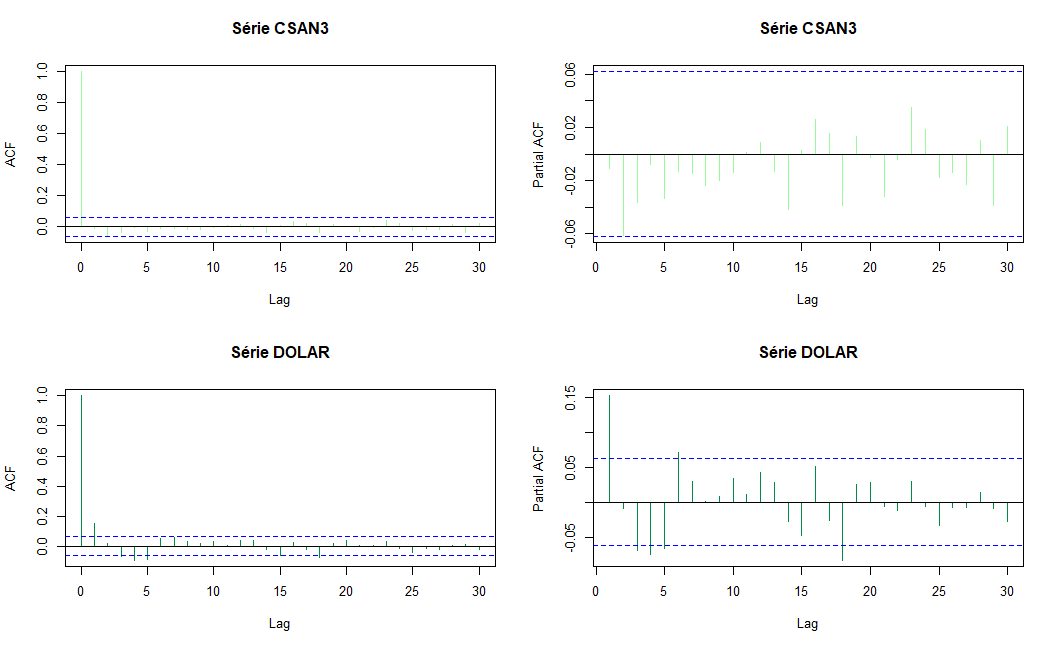












* *Os gráficos ACF de todas as séries apresentaram o mesmo resultado: a variável somente tem correlação com ela mesma, ou seja, trata-se de ruído branco e não faz sentido usá-la (passeio aleatório). A explicação para tal ocorrência está no fato de que foi utilizado o retorno dos valores de cada ação, cuja distribuição é normal.*
* *Os gráficos PACF de todas as séries confirmaram a informação, pois não sugerem nehuma lag para ser usada como defasagem ideal.*

1. **Para cada um dos processos abaixo gere 200 observações. Faça um gráfico da série, ACF e PACF. Comente os resultados.**

**2.a. Série aleatória, observações iid da distribuição: N(0,1)**

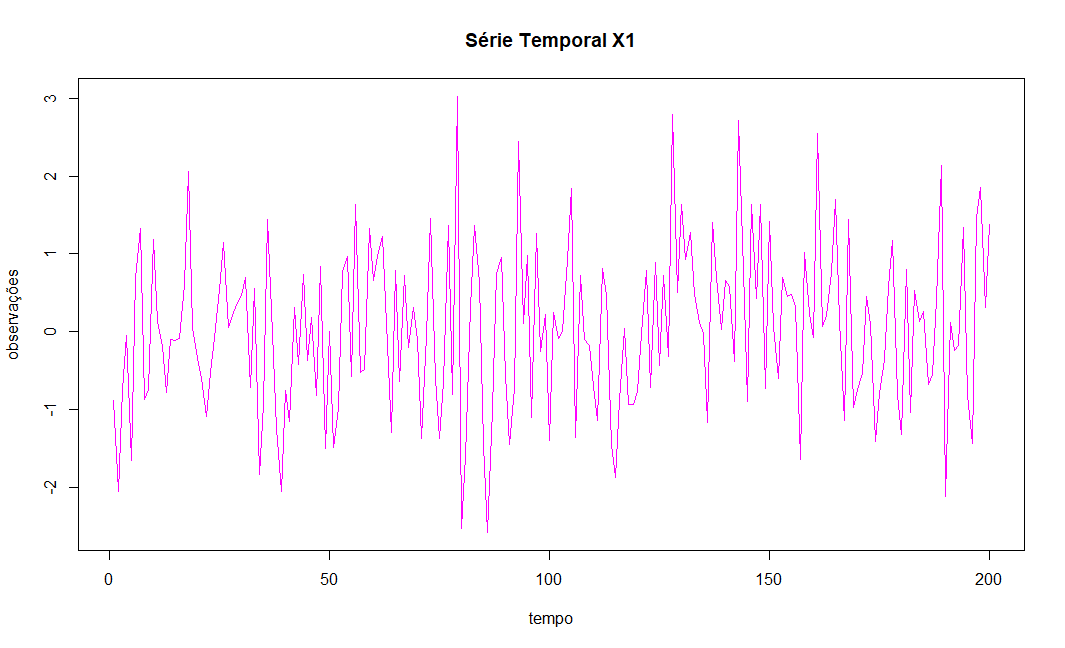
*> x1<-rnorm(200,0,1)*

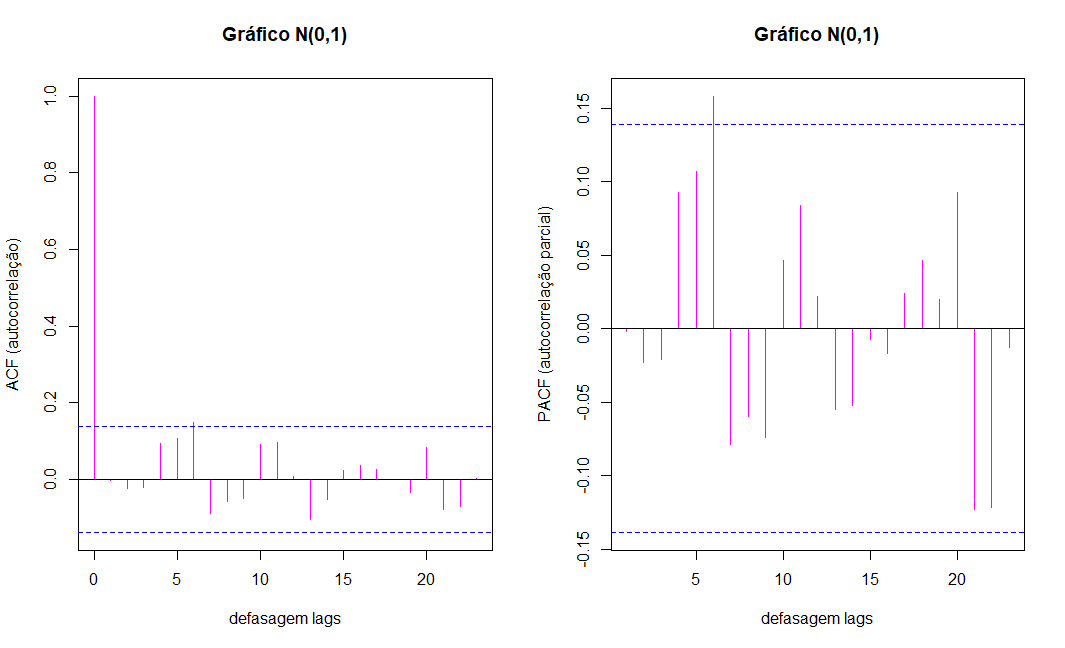
*> plot.ts(x1, main="Série Temporal X1", xlab='tempo',ylab='observações', col=6)*

*> par(mfrow=c(1,2))*

*> acf(x1, main="Gráfico N(0,1)", xlab="defasagem lags", ylab="ACF (autocorrelação)", col=6)*

*> pacf(x1, main="Gráfico N(0,1)", xlab="defasagem lags", ylab="PACF (autocorrelação parcial)", col=6)*





* *Gráfico ACF indica que a variável somente tem correlação com ela mesma, então é ruído branco. Nesse caso, não se trata de uma série estacionária e não faz sentido usá-la (passeio aleatório).*
* *Gráfico PACF confirma a informação, uma vez que não sugere nenhuma lag para ser usada como defasagem ideal.*

**2.b. Série com tendência estocástica: xt = xt-1 + N(1, 52)**

*> e<-rnorm(200,1,25)*

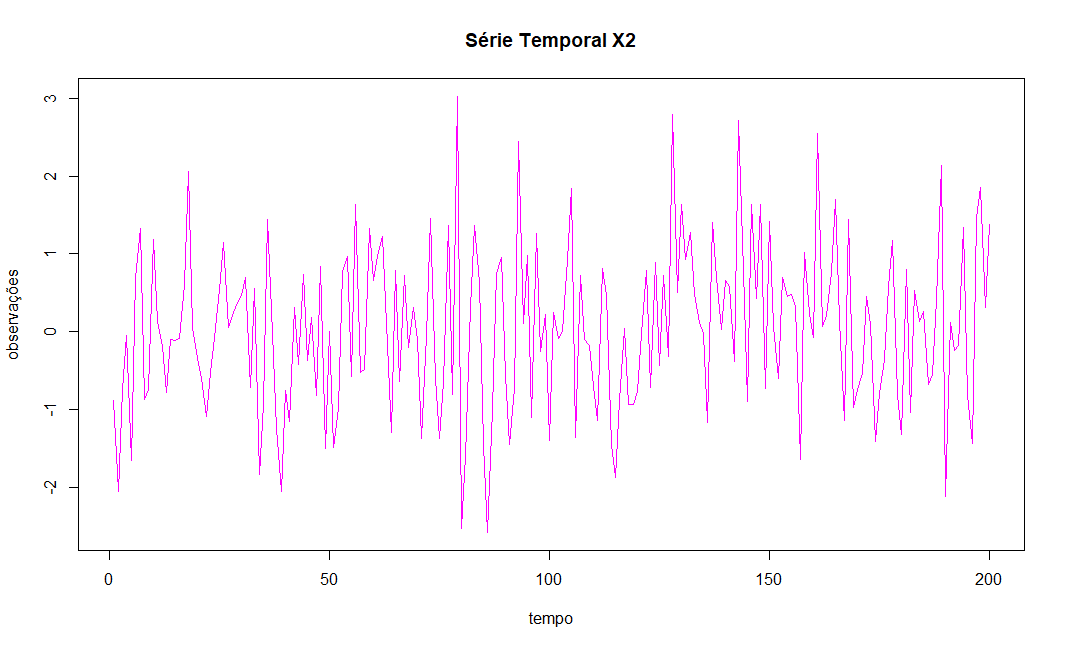
*> x2<-cumsum(e)*

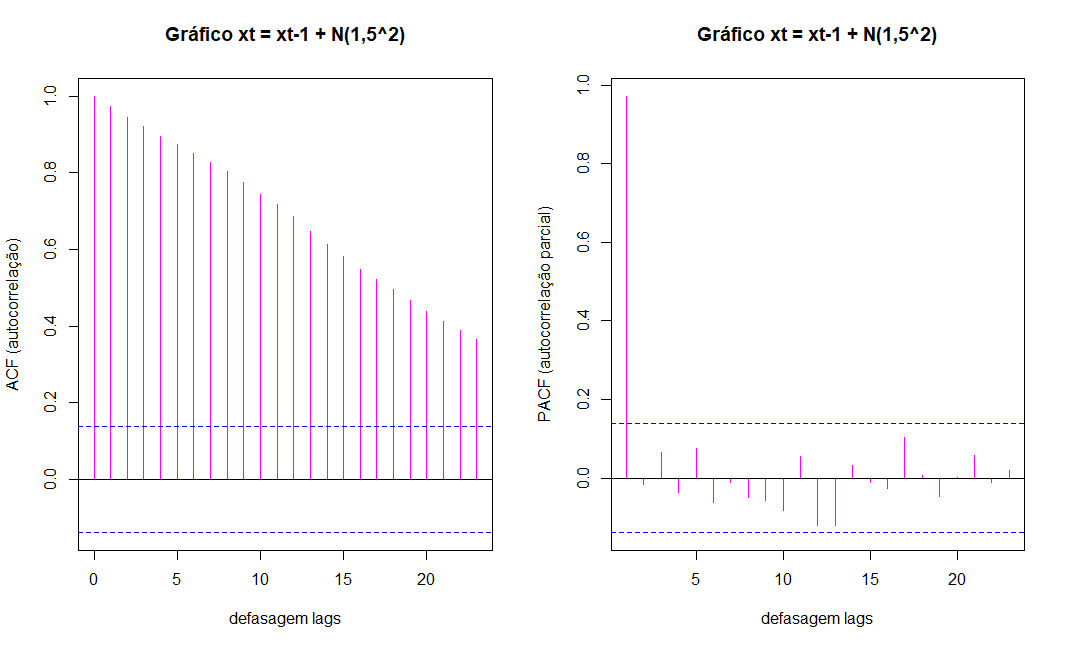
*> plot.ts(x1, main="Série Temporal X2", xlab='tempo',ylab='observações', col=6)*

*> par(mfrow=c(1,2))*

*> acf(x2,main='Gráfico xt = xt-1 + N(1,5^2)',xlab='defasagem lags',ylab='ACF (autocorrelação)', col=6)*

*> pacf(x2,main='Gráfico xt = xt-1 + N(1,5^2)',xlab='defasagem lags',ylab='PACF (autocorrelação parcial)', col=6)*





* *Gráfico ACF indica que não se trata de uma série estacionária, pois ela não converge. Dessa forma, é necessário trabalhar com a diferença dos valores.*
* *Gráfico PACF confirma a informação, uma vez que não sugere nenhuma lag para ser usada como defasagem ideal para o modelo.*

**2.c. Série com correlação de curto-prazo: xt = 0,7xt-1 + N(0,1)**

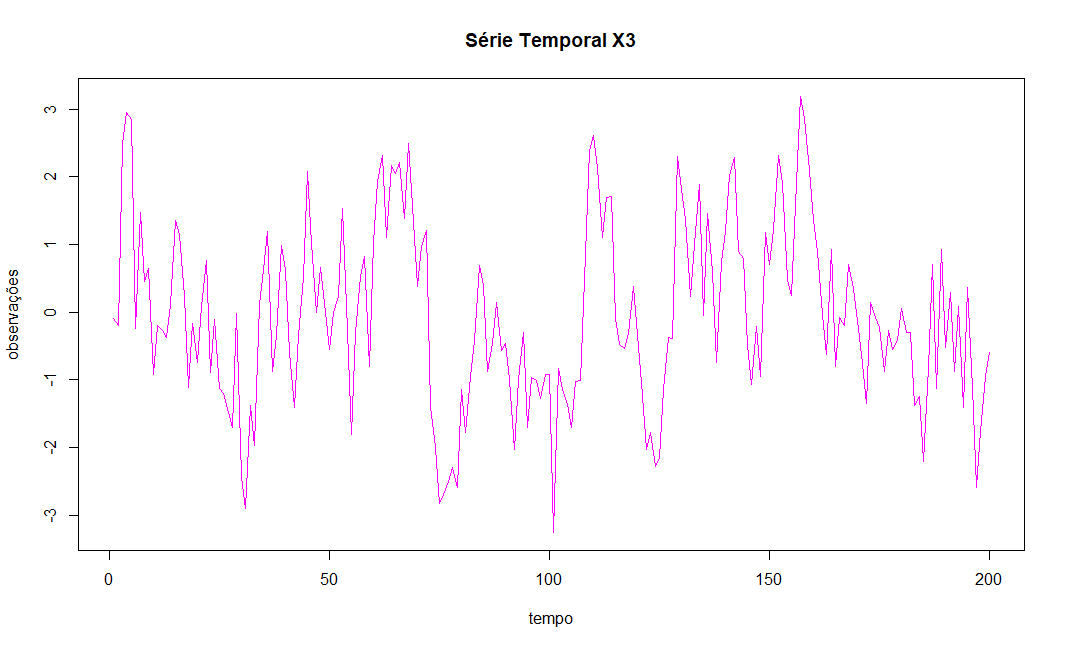
*> x3<-arima.sim(n = 200, list(ar = 0.7))*

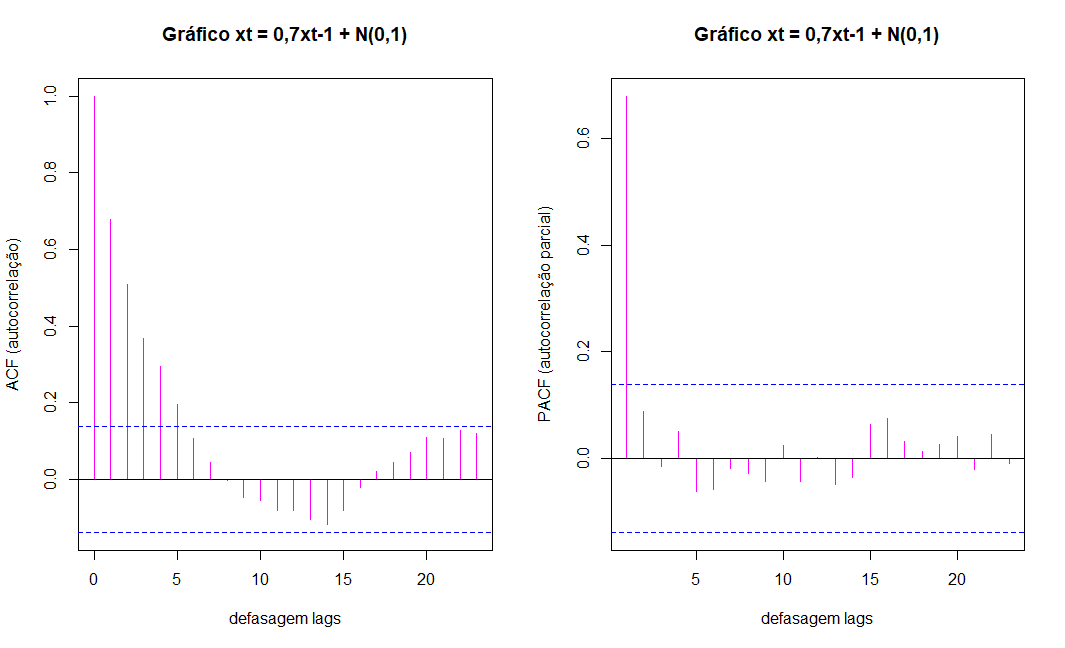
*> plot.ts(x3, main="Série Temporal X3", xlab='tempo',ylab='observações', col=6)*

*> par(mfrow=c(1,2))*

*> acf(x3, main="Gráfico xt = 0,7xt-1 + N(0,1)", xlab="defasagem lags", ylab="ACF (autocorrelação)", col=6)*

*> pacf(x3, main="Gráfico xt = 0,7xt-1 + N(0,1)", xlab="defasagem lags", ylab="PACF (autocorrelação parcial)", col=6)*





* *Gráfico ACF indica que se trata de uma série estacionária, pois ela converge.*
* *Gráfico PACF sugere a lag 1, como a defasagem ideal para o modelo.*

**2.d. Série com correlações negativas: xt = -0,8xt-1 + N(0,1**

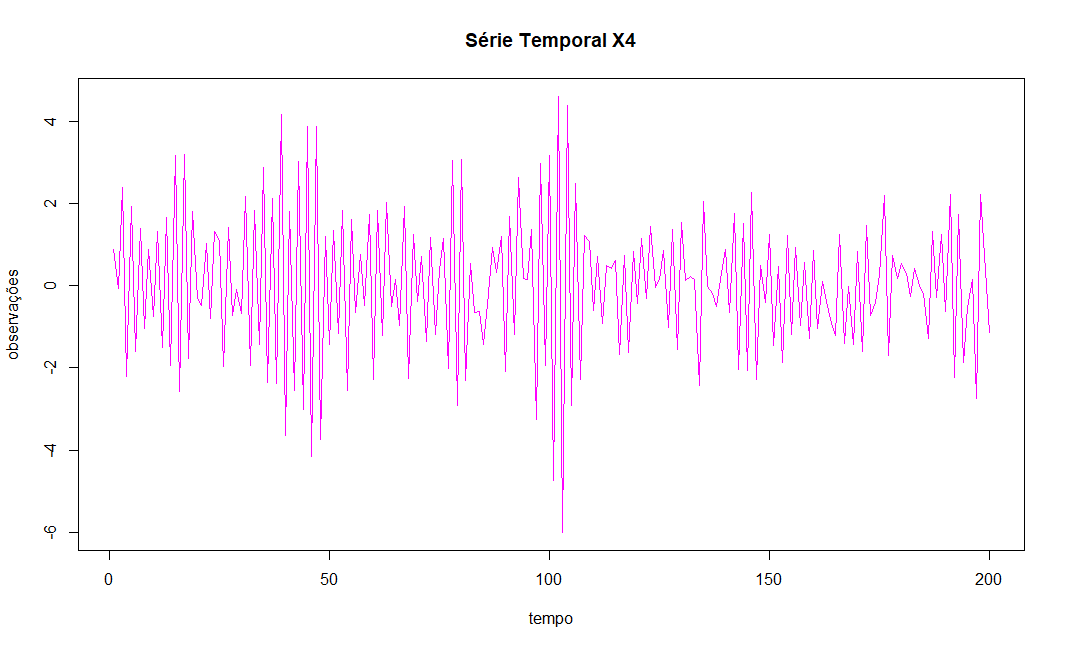
*> x4<-arima.sim(n = 200, list(ar = -0.8))*

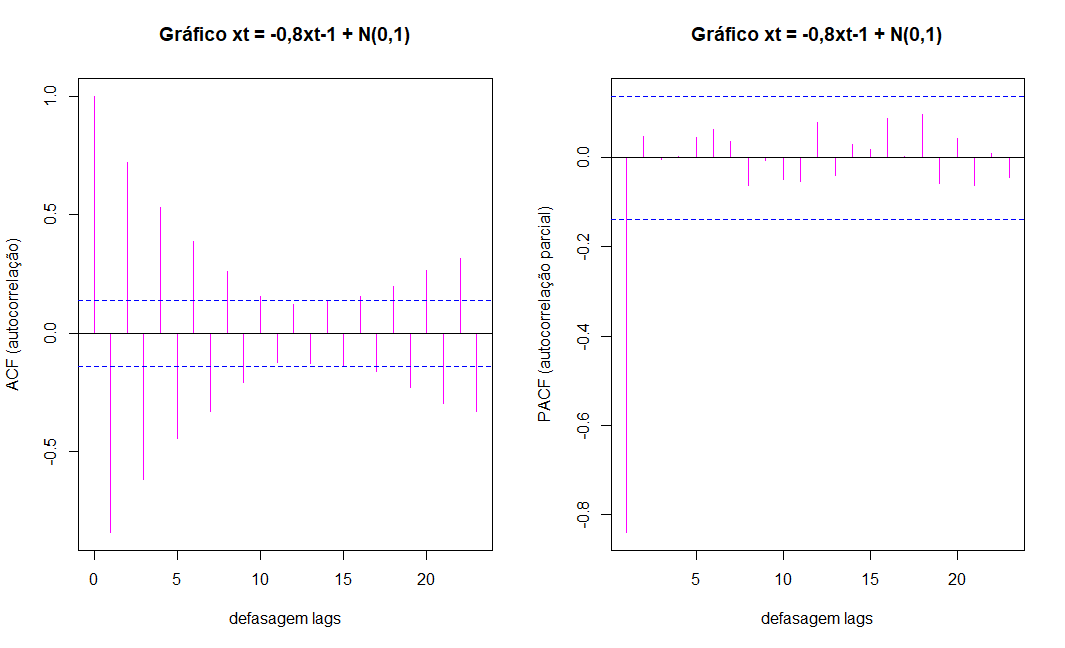
*> plot.ts(x4, main="Série Temporal X4", xlab='tempo',ylab='observações', col=6)*

*> par(mfrow=c(1,2))*

*> acf(x4, main="Gráfico xt = -0,8xt-1 + N(0,1)", xlab="defasagem lags", ylab="ACF (autocorrelação)", col=6)*

*> pacf(x4, main="Gráfico xt = -0,8xt-1 + N(0,1)", xlab="defasagem lags", ylab="PACF (autocorrelação parcial)", col=6)*





* *Gráfico ACF indica que se trata de uma série estacionária, pois ela converge.*
* *Gráfico PACF sugere lag 1, como a defasagem ideal para o modelo.*

**2.e. Médias móveis: xt = t + 0,6t-1 , t ~ N(0,1)**

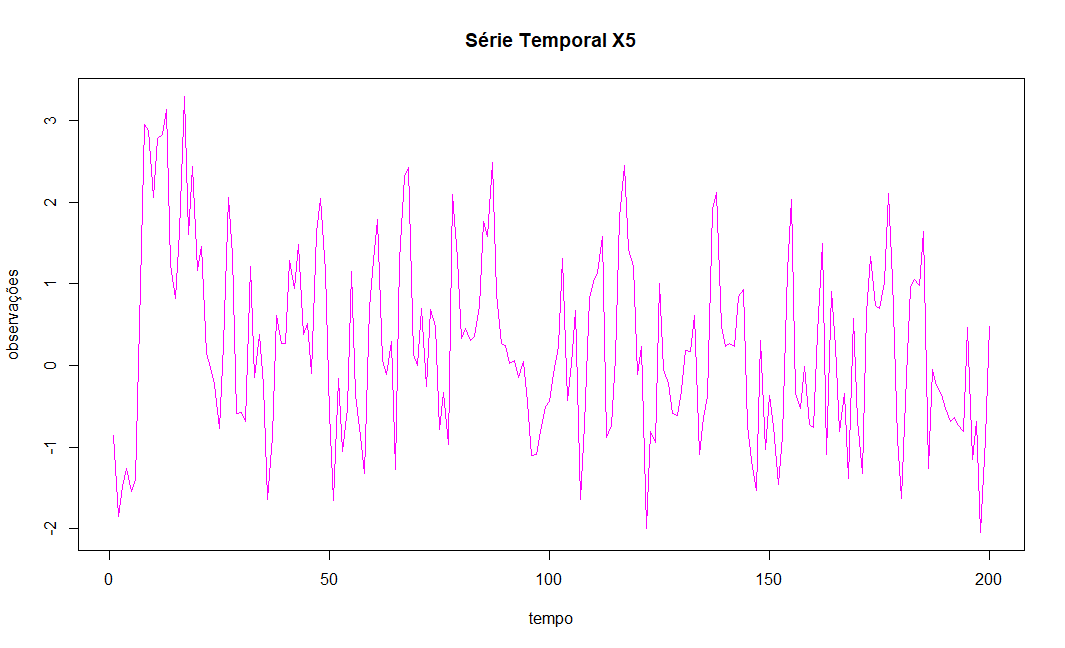
*> x5<-arima.sim(list(ar=0.6), n=200)*

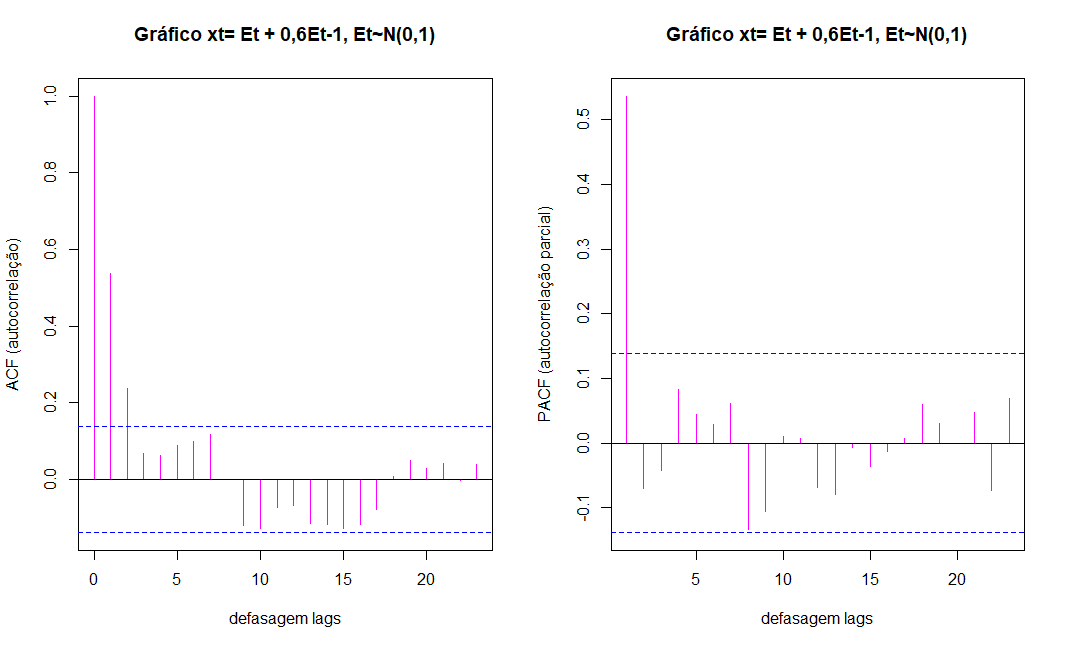
*> plot.ts(x5, main="Série Temporal X5", xlab='tempo',ylab='observações', col=6)*

*> par(mfrow=c(1,2))*

*> acf(x5, main="Gráfico xt= Et + 0,6Et-1, Et~N(0,1)", xlab="defasagem lags", ylab="ACF (autocorrelação)", col=6)*

*> pacf(x5, main="Gráfico xt= Et + 0,6Et-1, Et~N(0,1)", xlab="defasagem lags", ylab="PACF (autocorrelação parcial)", col=6)*





* *Gráfico ACF indica que se trata de uma série estacionária, pois ela converge.*
* *Gráfico PACF sugere a lag 1, como a defasagem ideal para o modelo.*

1. **Utilize a série abaixo para resolver cada item:**

An example of a time series that can probably be described using an additive model with a trend and no seasonality is the time series of the annual diameter of women’s skirts at the hem, from 1866 to 1911. The data is available in the file **http://robjhyndman.com/tsdldata/roberts/skirts.dat** (original data from Hipel and McLeod, 1994).

**3.a. Faça a leitura da série de dados e os tratamentos necessários para considerar a mesma como uma série temporal.**

*> library(readxl)*

*> skirt <- read\_excel("C:/Users/suely/2 FGV Big Data/João Chela/Mat Complementar/skirt.xlsx")*

*> View(skirt)*

*>*

*> diam.ts<-ts(skirt, frequency=1, start=c(1866))*

*> diam.ts*

*Time Series:*

*Start = 1866*

*End = 1911*

*Frequency = 1*

*DIAM*

*[1,] 608*

*[2,] 617*

*[3,] 625*

*[4,] 636*

*[5,] 657*

*[6,] 691*

*[7,] 728*

*[8,] 784*

*[9,] 816*

*[10,] 876*

*[11,] 949*

*[12,] 997*

*[13,] 1027*

*[14,] 1047*

*[15,] 1049*

*[16,] 1018*

*[17,] 1021*

*[18,] 1012*

*[19,] 1018*

*[20,] 991*

*[21,] 962*

*[22,] 921*

*[23,] 871*

*[24,] 829*

*[25,] 822*

*[26,] 820*

*[27,] 802*

*[28,] 821*

*[29,] 819*

*[30,] 791*

*[31,] 746*

*[32,] 726*

*[33,] 661*

*[34,] 620*

*[35,] 588*

*[36,] 568*

*[37,] 542*

*[38,] 551*

*[39,] 541*

*[40,] 557*

*[41,] 556*

*[42,] 534*

*[43,] 528*

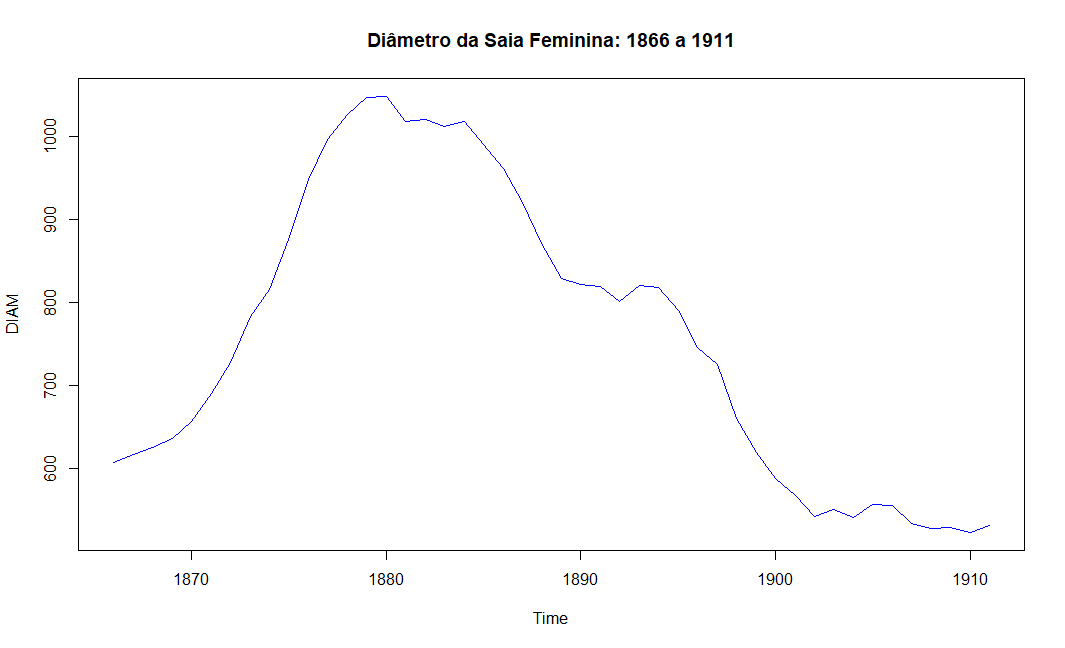
*[44,] 529*

*[45,] 523*

*[46,] 531*

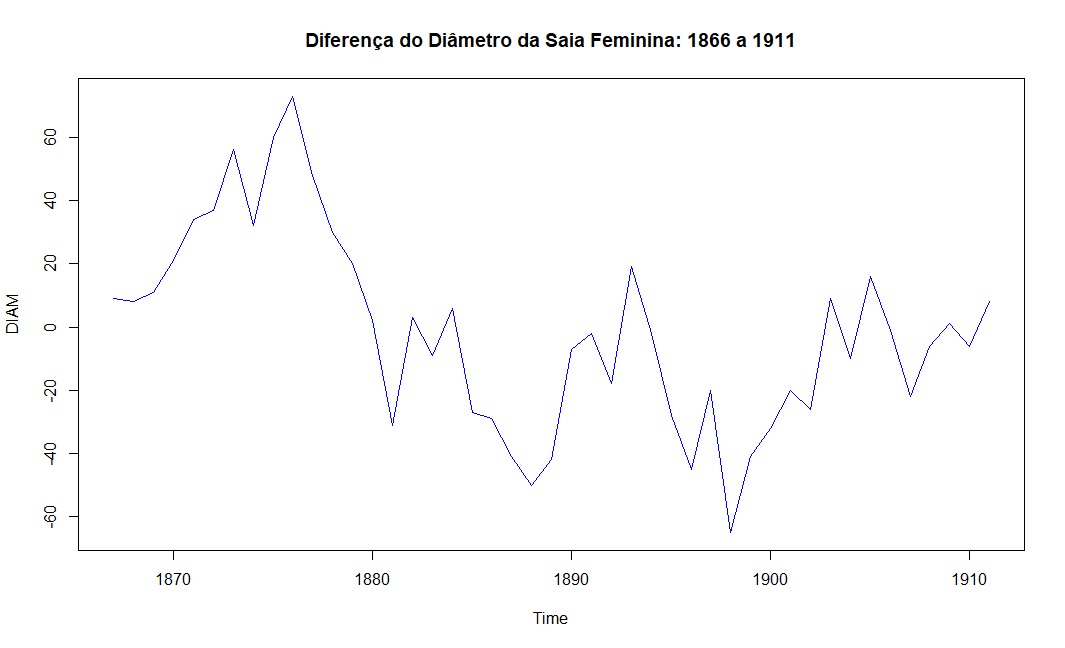
*>*

*> plot(diam.ts, col=4, main="Diâmetro da Saia Feminina: 1866 a 1911")*



*> d.diam.ts<-diff(diam.ts)*

*> plot(d.diam.ts, col=4, main="Diferença do Diâmetro da Saia Feminina: 1866 a 1911")*



**3.b. Faça a decomposição da série do item (a): Sazonalidade, Tendência e Aleatória.**

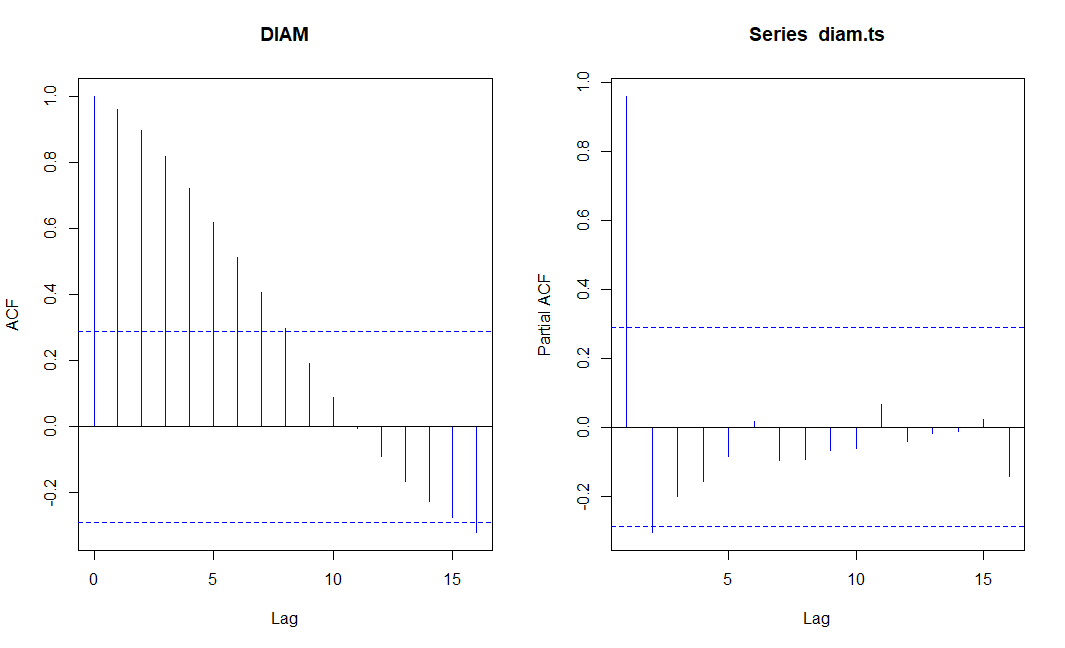
* *Não precisa decompor, pois a série é composta apenas por anos, e não há sazonalidade anual !!!!!*

**3.c. Calcule a ACF e PACF da série e da primeira diferença.**

*> par(mfrow=c(1,2))*

*> acf(diam.ts, col=4)*

*> pacf(diam.ts, col=4)*

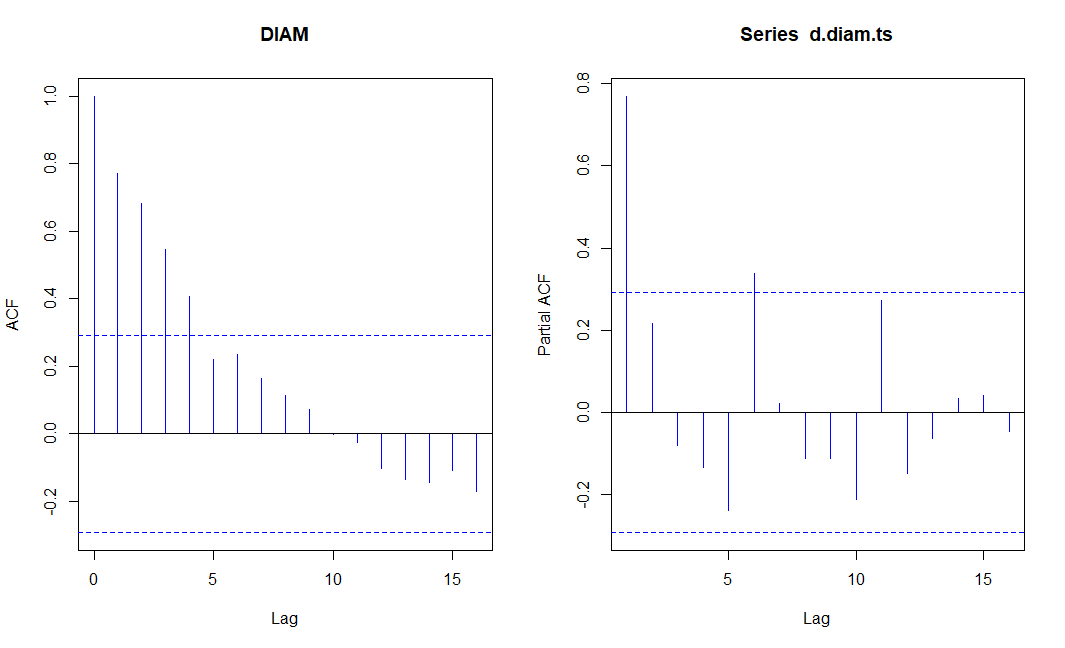


* *O gráfico ACF da série indica que a mesma não é estacionária, pois apresenta tendência e não converge.*
* *O gráfico PACF não sugere nenhuma lag, pois a série não é estacionária.*

*> d.diam.ts<-diff(diam.ts)*

*> par(mfrow=c(1,2))*

*> acf(d.diam.ts, col=4)*

*> pacf(d.diam.ts, col=4)* 

* *Após utilizar a diferença dos valores, o gráfico ACF indica que a série é estacionária, pois converge.*
* *Já o gráfico PACF, sugere a lag 1 como a defasagem ideal para o modelo.*

1. **Usando a função arima.sim gere as seguintes simulações (300 pontos):**

Para cada simulação, plote o gráfico da série, calcule o ACF e PACF. Usando estes resultados, conclua como deve ser o comportamento da ACF e PACF de um modelo autoregressivo( AR.)

**4.a. Processo AR(1) onde θ0=0, θ1=0.7**

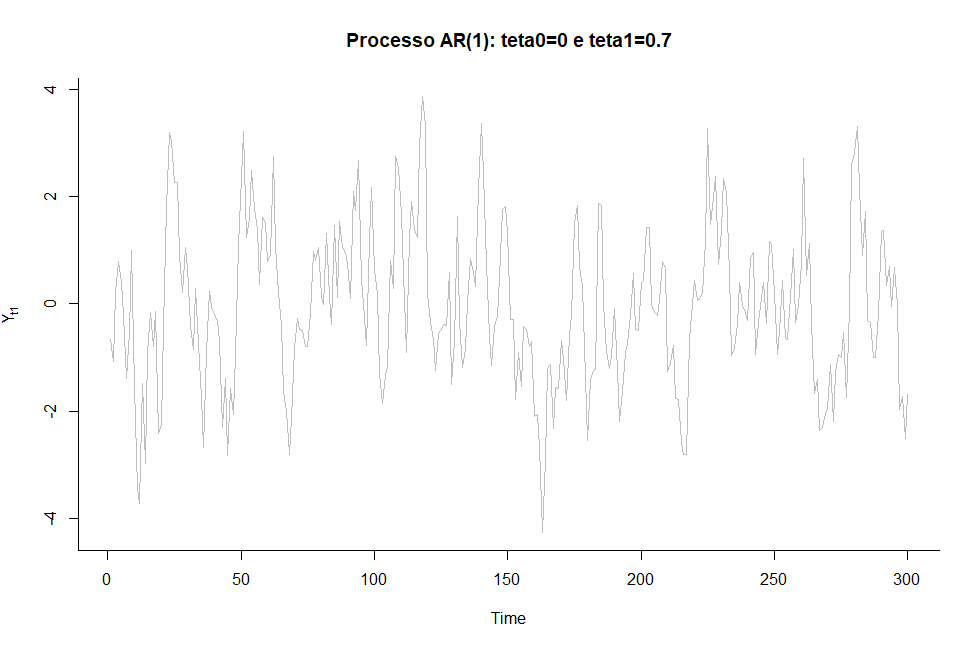
*> Yt1<-arima.sim(model=list(ar=0.7), n=300)*

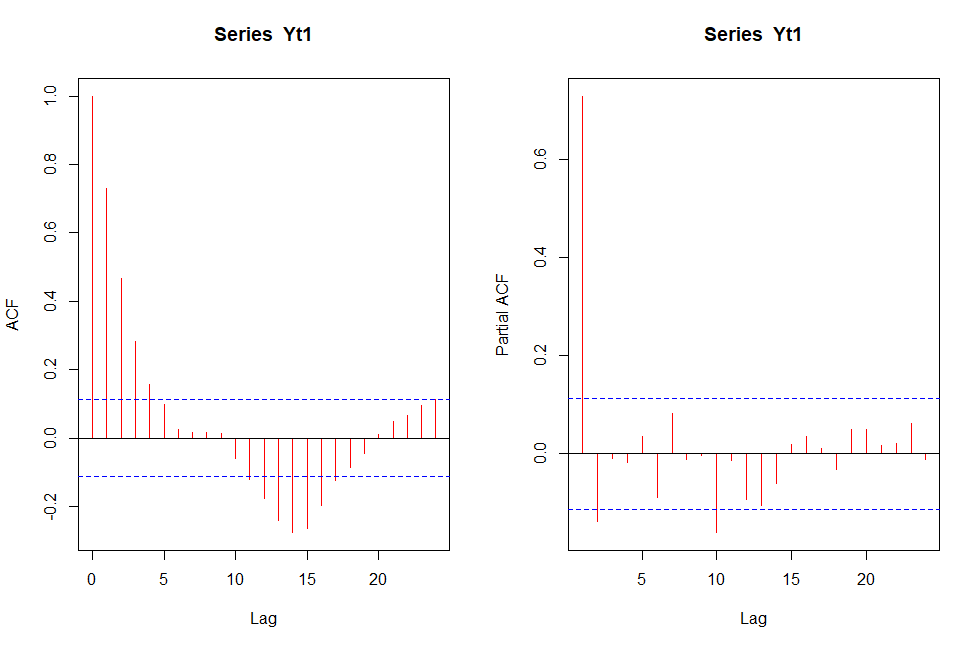
*> plot(Yt1, lty=1, bty='l', col=8, main="Processo AR(1): teta0=0 e teta1=0.7", ylab=expression(paste(Y[t1])))*

*> par(mfrow=c(1,2))*

*> acf(Yt1, col=2)*

*> pacf(Yt1, col=2)*





* *O gráfico ACF indica que a série é estacionária, pois converge.*
* *O gráfico PACF sugere a lag 1 como a defasagem ideal para o modelo.*

**4.b. Processo AR(1) onde θ0=0, θ1= -0.7**

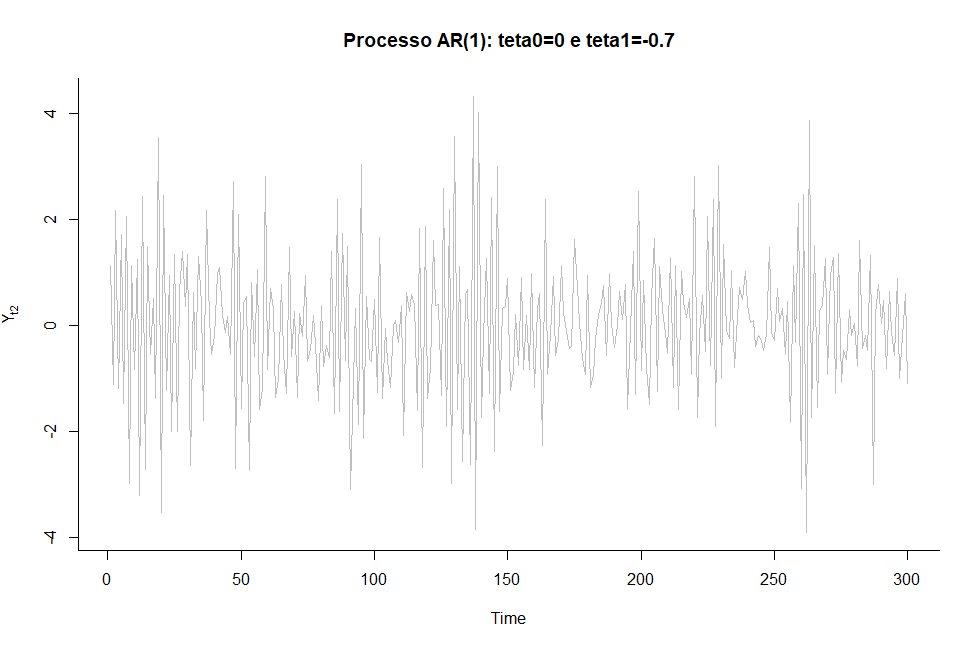
*> Yt2<-arima.sim(model=list(ar=-0.7), n=300)*

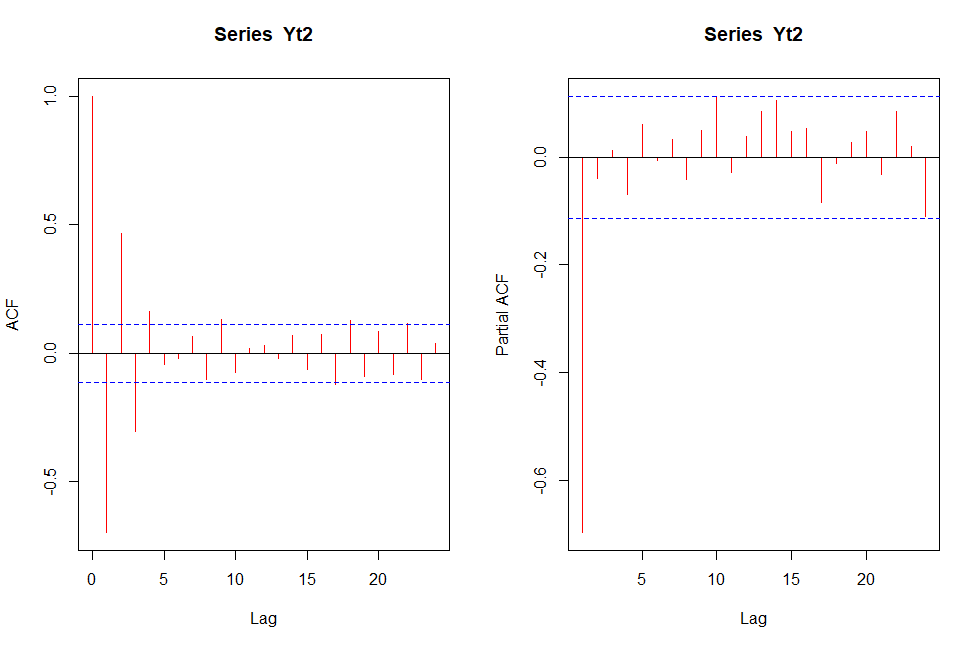
*> plot(Yt2, lty=1, bty='l',col=8, main="Processo AR(1): teta0=0 e teta1=-0.7", ylab=expression(paste(Y[t2])))*

*> par(mfrow=c(1,2))*

*> acf(Yt2, col=2)*

*> pacf(Yt2, col=2)*





* *O gráfico ACF indica que a série é estacionária, pois converge.*
* *O gráfico PACF sugere a lag 1 como a defasagem ideal para o modelo.*

**4.c. Processo AR(2) onde θ0=0, θ1=0.3 e θ2=0.5**

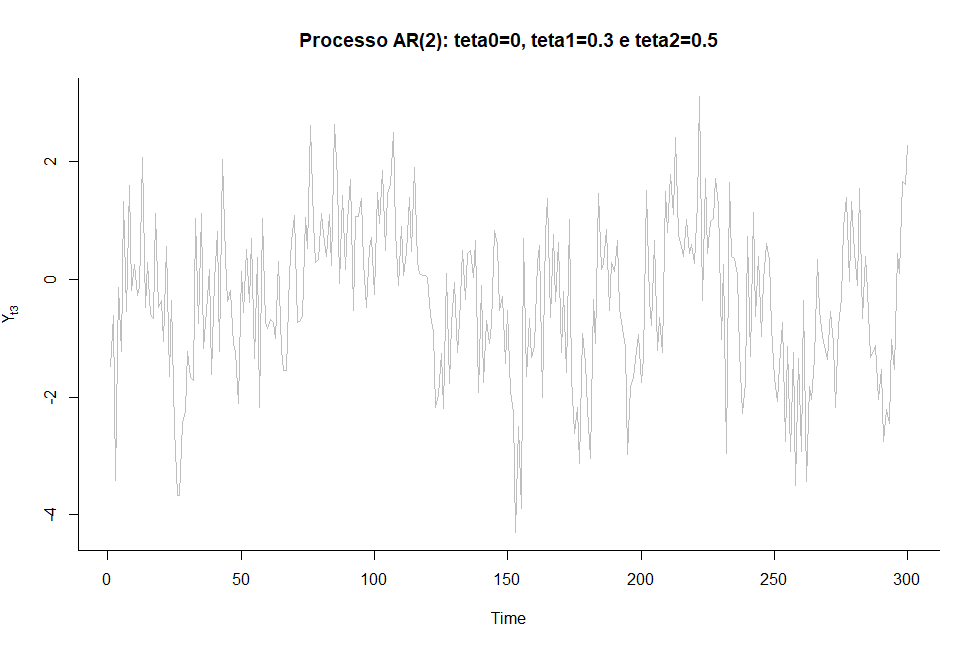
*> Yt3<-arima.sim(model=list(ar=c(0.3,0.5)), n=300)*

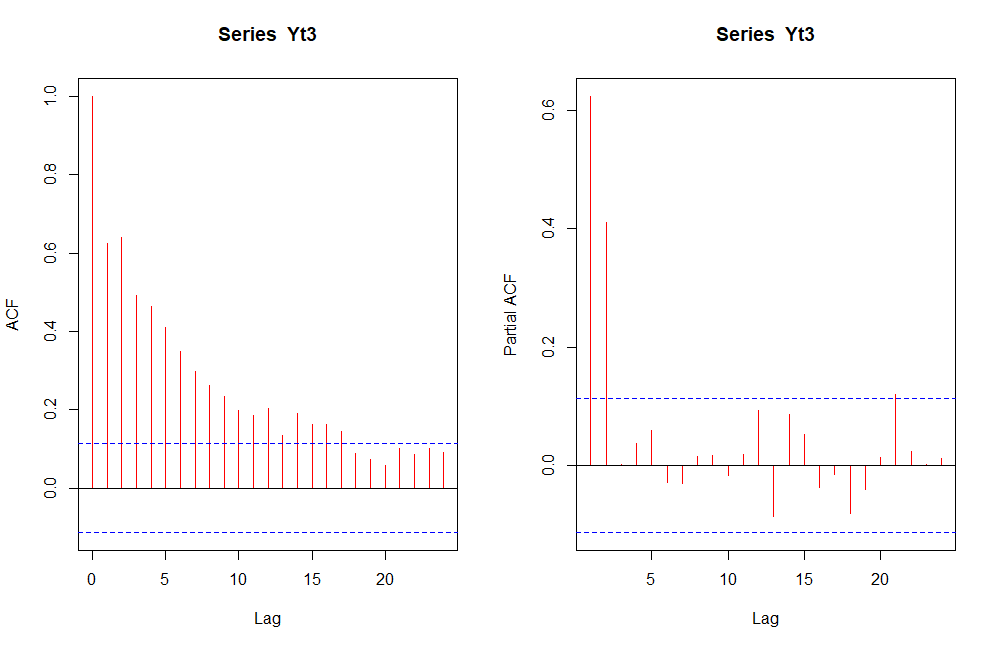
*> plot(Yt3, lty=1, bty='l',col=8, main="Processo AR(2): teta0=0, teta1=0.3 e teta2=0.5", ylab=expression(paste(Y[t3])))*

*> par(mfrow=c(1,2))*

*> acf(Yt3, col=2)*

*> pacf(Yt3, col=2)*





* *O gráfico ACF indica que a série é estacionária, pois converge.*
* *O gráfico PACF sugere as lags 1 e 2 como a defasagem ideal para o modelo.*

**4.d. Processo MA(1) onde θ0=0, θ1=0.6**

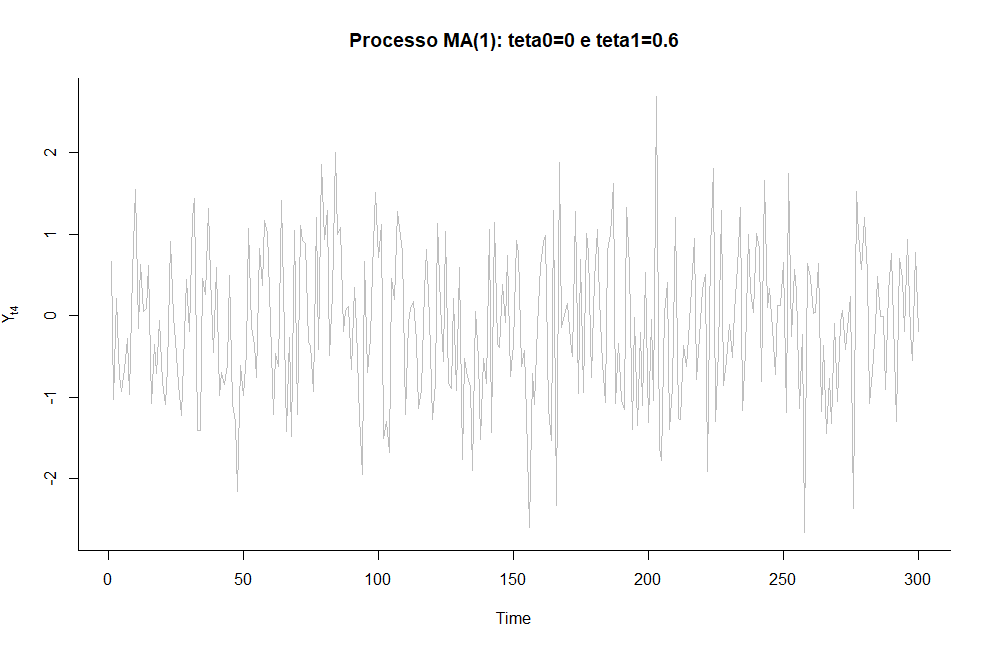
*> Yt4<-arima.sim(model=list(am=0.6), n=300)*

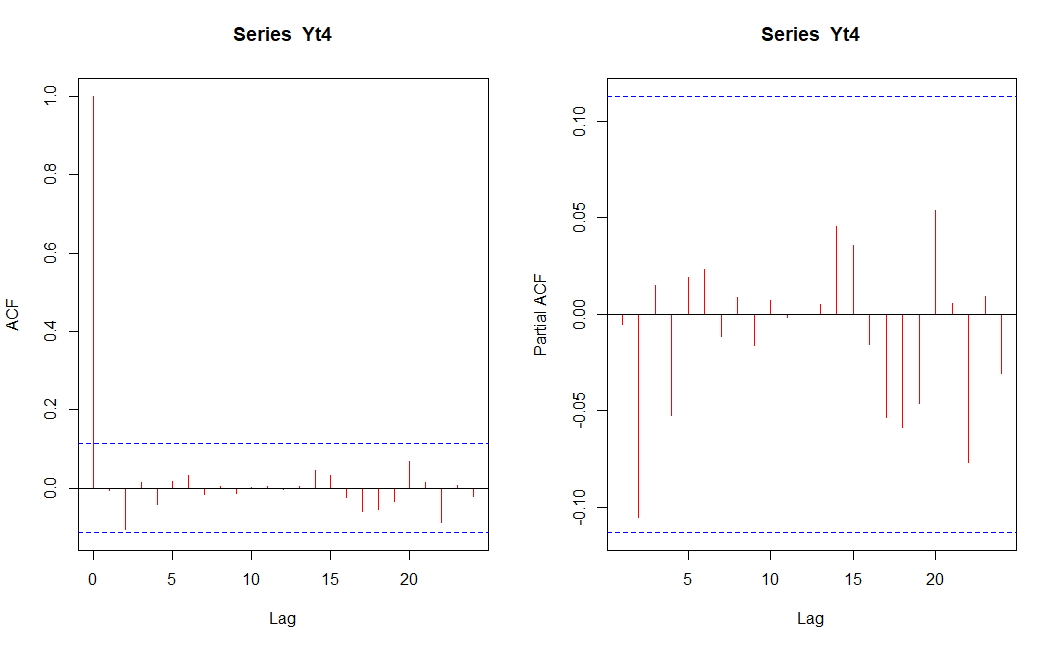
*> plot(Yt4, lty=1, bty='l',col=8, main="Processo MA(1): teta0=0 e teta1=0.6", ylab=expression(paste(Y[t4])))*

*> par(mfrow=c(1,2))*

*> acf(Yt4, col=2)*

*> pacf(Yt4, col=2)*





* *O gráfico ACF indica que a variável só tem correlação com ela mesma, portanto, trata-se de ruído branco.*
* *O gráfico PACF confirma a informação, pois não sugere nenhuma lag como defasagem ideal para o modelo.*

**4.e. Processo MA(1) onde θ0=0, θ1=-0.6**

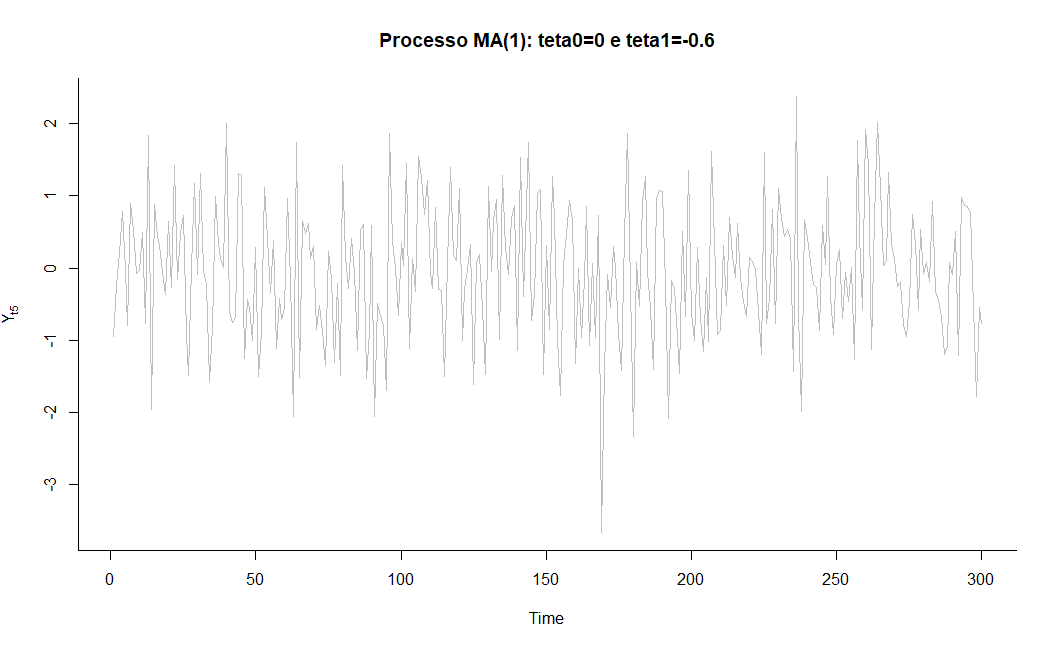
*> Yt5<-arima.sim(model=list(am=-0.6), n=300)*

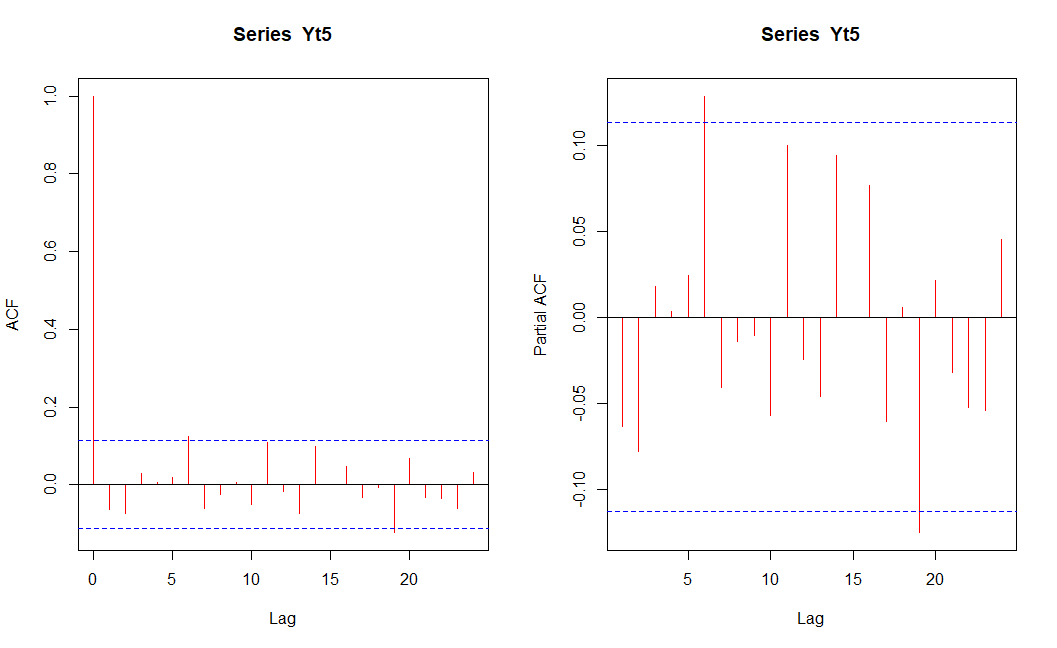
*> plot(Yt5, lty=1, bty='l',col=8, main="Processo MA(1): teta0=0 e teta1=-0.6", ylab=expression(paste(Y[t5])))*

*> par(mfrow=c(1,2))*

*> acf(Yt5, col=2)*

*> pacf(Yt5, col=2)*





* *O gráfico ACF indica que a variável só tem correlação com ela mesma, portanto, trata-se de ruído branco.*
* *O gráfico PACF confirma a informação, pois não sugere nenhuma lag como defasagem ideal para o modelo.*

1. **Obtenha a série histórica do PIB Brasil no seguinte site e responda as questões abaixo:**

[**http://www.bcb.gov.br/pre/portalCidadao/cadsis/series.asp?idpai=PORTALBCB**](http://www.bcb.gov.br/pre/portalCidadao/cadsis/series.asp?idpai=PORTALBCB)

**Código da série: 1232**

**5.a. Plote o gráfico da série usando o R.**

*> attach(PIB.Brasil)*

*> pib.ts<-ts(pib, frequency=4, start=c(1991,1))*

*> pib.ts*

*Qtr1 Qtr2 Qtr3 Qtr4*

*1991 83.45 87.71 90.98 86.96*

*1992 80.82 85.85 90.16 90.64*

*1993 84.66 89.66 94.70 94.66*

*1994 87.63 91.63 99.90 103.92*

*1995 96.48 99.76 101.69 102.07*

*1996 95.64 100.82 107.77 104.37*

*1997 99.18 105.82 109.47 107.92*

*1998 99.97 107.36 109.39 105.82*

*1999 100.54 106.53 108.34 108.21*

*2000 105.30 110.70 112.91 112.95*

*2001 109.01 113.26 113.23 112.17*

*2002 109.11 115.39 117.44 117.62*

*2003 111.66 116.40 118.12 118.65*

*2004 116.38 123.61 125.54 125.86*

*2005 121.22 128.96 128.18 128.54*

*2006 126.47 131.45 134.26 134.77*

*2007 133.00 139.90 142.40 143.77*

*2008 141.37 148.93 152.53 145.14*

*2009 137.54 145.36 150.29 152.85*

*2010 150.39 158.10 160.70 161.00*

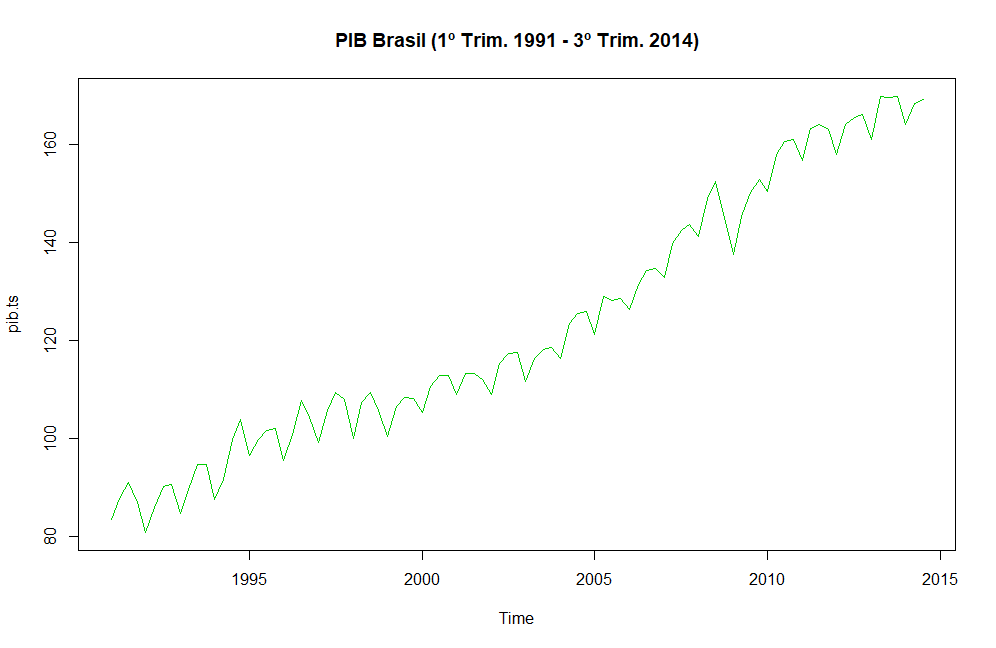
*2011 156.76 163.34 164.11 163.20*

*2012 158.01 164.24 165.66 166.18*

*2013 161.00 169.92 169.68 169.79*

*2014 164.08 168.45 169.27*

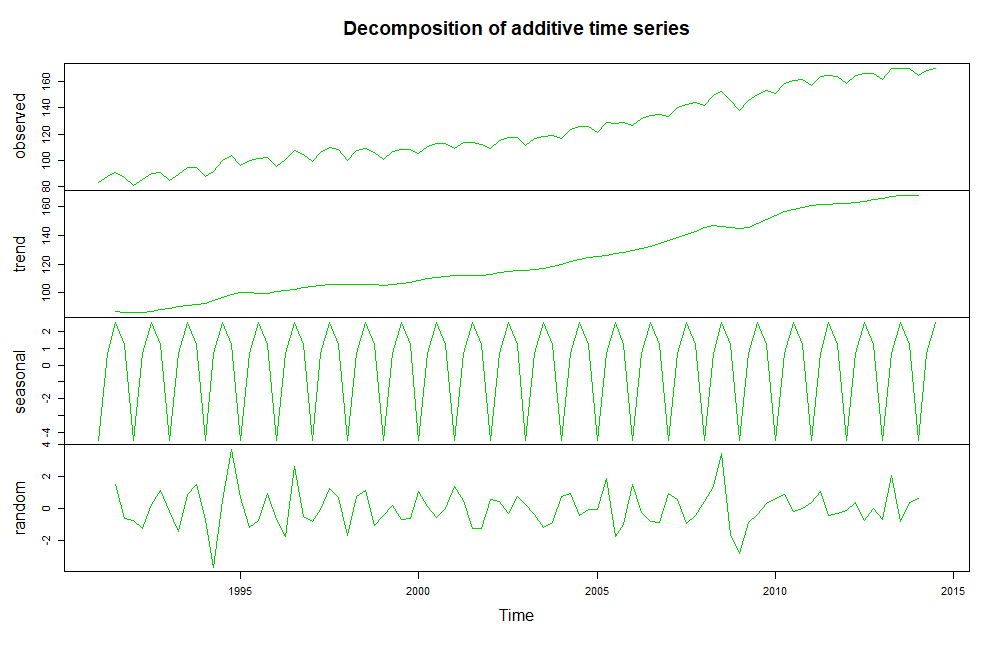
*> plot(pib.ts, col=3, main="PIB Brasil (1º Trim. 1991 - 3º Trim. 2014)")*



**5.b. Faça a decomposição da série em: Sazonalidade, Tendência e Aleatória.**

*> decomp.pib.ts<-decompose(pib.ts)*

*> plot(decomp.pib.ts, col=3)*



**5.c. Usando o índice dos últimos 12 anos, encontre uma projeção para o PIB(índice) do próximo semestre usando um modelo AR(1). Neste caso use a série das diferenças.**

*> head(PIB\_index12)*

*# A tibble: 6 x 2*

*Data index12*

*<chr> <dbl>*

*1 1° Trim.2002 109.11*

*2 2° Trim.2002 115.39*

*3 3° Trim.2002 117.44*

*4 4° Trim.2002 117.62*

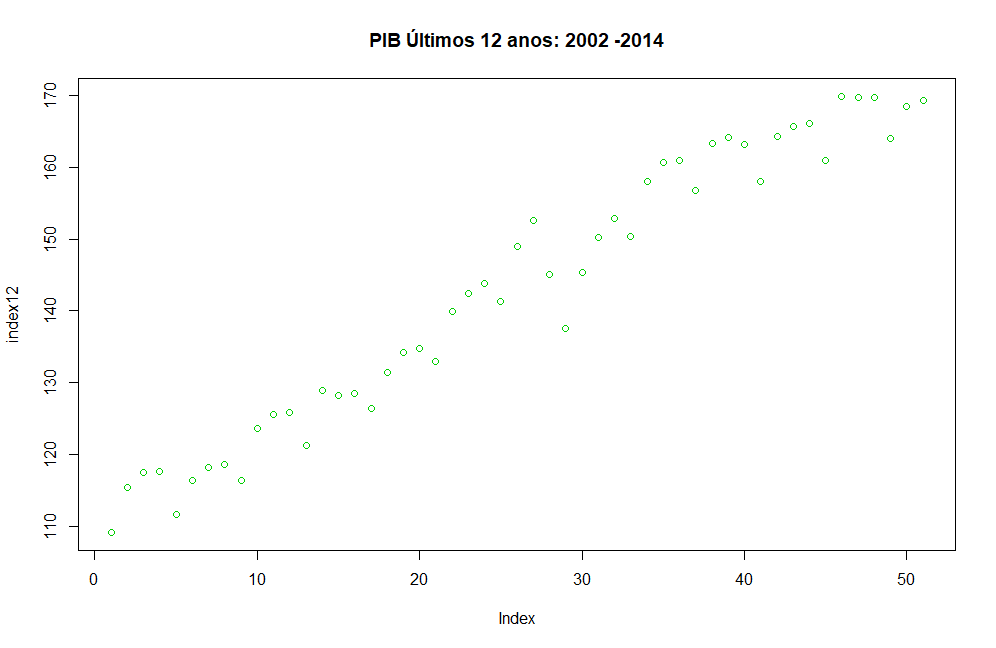
*5 1° Trim.2003 111.66*

*6 2° Trim.2003 116.40*

*>*

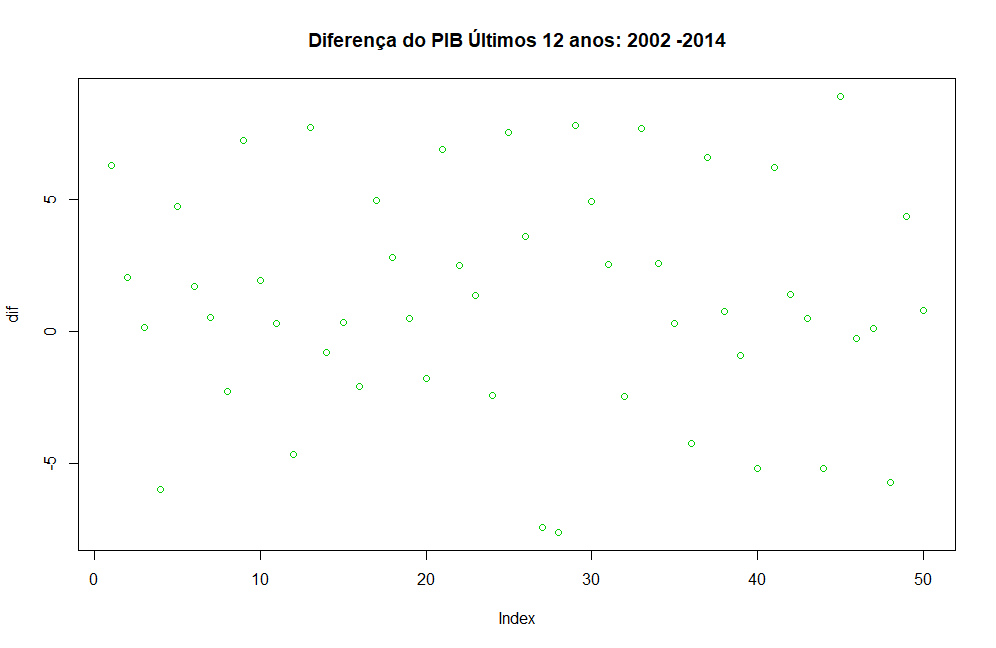
*> attach(PIB\_index12)*

*> plot(index12, col=3, main="PIB Últimos 12 anos: 2002 -2014")*



*> dif<-diff(index12)*

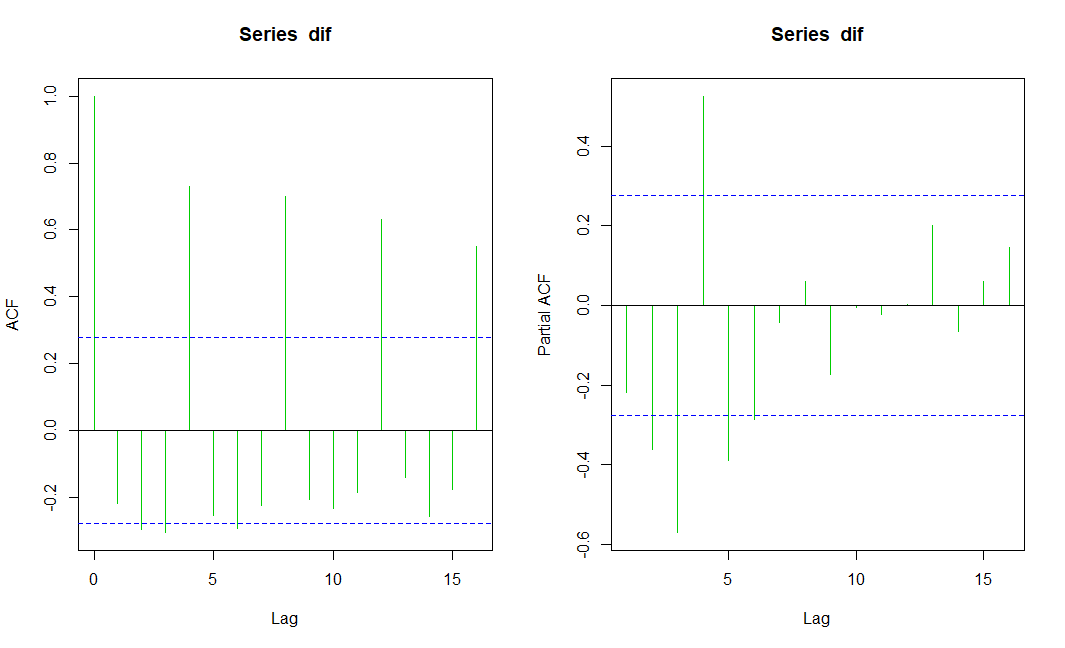
*> plot(dif, col=3, main="Diferença do PIB Últimos 12 anos: 2002 -2014")*



*> par(mfrow=c(1,2))*

*> acf(dif, col=3) # sugere que a série é estacionária, pois converge*

*> pacf(dif, col=3) # sugere 2 lags (3 e 4)*



*> pib12.arima<-arima(dif, order=c(0,1,2))*

*> pib12.arima*

*Call:*

*arima(x = dif, order = c(0, 1, 2))*

*Coefficients:*

*ma1 ma2*

*-1.6758 0.6759*

*s.e. 0.2237 0.1853*

*sigma^2 estimated as 14.68: log likelihood = -138.69, aic = 283.38*

*> pib12.pred<-predict(pib12.arima, n.ahead=2) # n.ahead=2, pois se trata de 2 trimestres adiante*

*> pib12.pred*

*$pred*

*Time Series:*

*Start = 51*

*End = 52*

*Frequency = 1*

*[1] 2.179788 1.171699*

*$se*

*Time Series:*

*Start = 51*

*End = 52*

*Frequency = 1*

*[1] 3.872926 4.628135*

*> index12[40:51]*

*[1] 163.20 158.01 164.24 165.66 166.18 161.00 169.92 169.68 169.79 164.08 168.45 169.27*

*> prev.1.trim<-index12[51]+(2.179788)*

*> prev.1.trim*

*[1] 171.4498*

*> prev.2.trim<-prev.1.trim +(1.171699)*

*> prev.2.trim*

*[1] 172.6215*

* *A previsão para o próximo semestre é:*
  + *4º trimestre 2014: 171.4498*
  + *1º trimestre 2015: 172.6215*

***Usar auto.arima para indicar ordem correta diretamente, nesse caso:***

*> auto.arima(index12)*

*Series: index12*

*ARIMA(0,1,2) with drift*

*Coefficients:*

*ma1 ma2 drift*

*-0.5663 -0.2749 1.2258*

*s.e. 0.1396 0.1322 0.1098*

*sigma^2 estimated as 13.96: log likelihood=-135.85*

*AIC=279.69 AICc=280.58 BIC=287.34*

*> pib12.arima<-arima(dif, order=c(0,1,2))*

*> pib12.arima*

*Call:*

*arima(x = dif, order = c(0, 1, 2))*

*Coefficients:*

*ma1 ma2*

*-1.6758 0.6759*

*s.e. 0.2237 0.1853*

*sigma^2 estimated as 14.68: log likelihood = -138.69, aic = 283.38*

*> pib12.pred<-predict(pib12.arima, n.ahead=2) # n.ahead=2, pois se trata de 2 trimestres adiante*

*> pib12.pred*

*$pred*

*Time Series:*

*Start = 51*

*End = 52*

*Frequency = 1*

*[1] 2.179788 1.171699*

*$se*

*Time Series:*

*Start = 51*

*End = 52*

*Frequency = 1*

*[1] 3.872926 4.628135*

*> index12[40:51]*

*[1] 163.20 158.01 164.24 165.66 166.18 161.00 169.92 169.68 169.79 164.08 168.45 169.27*

*> prev.1.trim<-index12[51]+(2.179788)*

*> prev.1.trim*

*[1] 171.4498*

*> prev.2.trim<-prev.1.trim +(1.171699)*

*> prev.2.trim*

*[1] 172.6215*