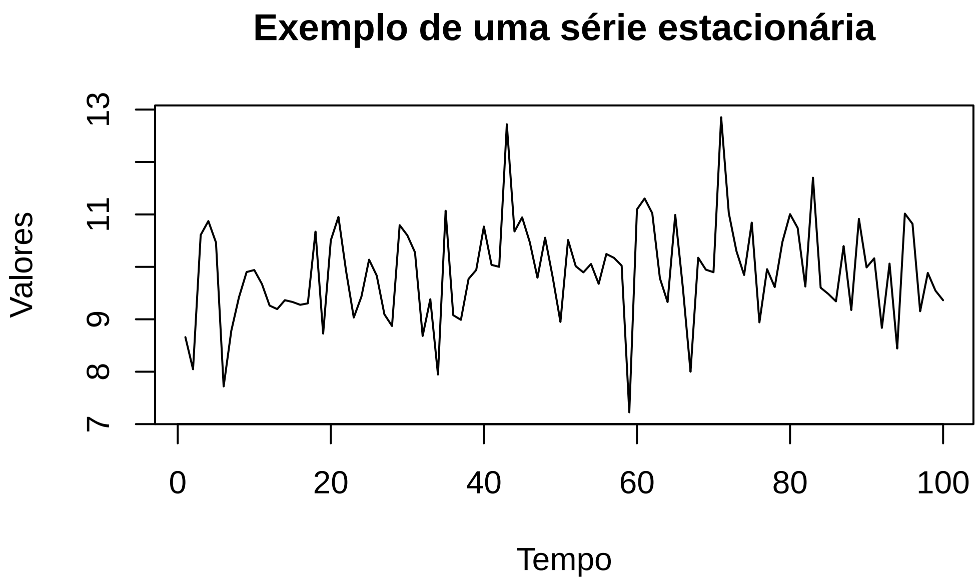
Modelos Autoregressivos de médias móveis (ARMA)

Os modelos ARMA fornecem outra abordagem para a previsão de séries temporais. Os modelos de suavização exponencial e ARMA são as duas abordagens mais utilizadas para a previsão de séries temporais. Enquanto os modelos de suavização exponencial são baseados na descrição da tendência e sazonalidade dos dados, os modelos ARMA visam descrever as autocorrelações nos dados.

Antes de introduzirmos os modelos ARMA, devemos primeiro discutir o conceito de estacionariedade e a técnica de diferenciação de séries temporais.

Estacionariedade

Uma das suposições que podemos fazer em uma série temporal é a de que ela é **estacionária**. Isto é, ao longo do tempo seus valores estão aleatoriamente próximos de uma média constante, refletindo de alguma forma um equilíbrio estável. No gráfico a seguir, tempos um exemplo de uma série temporal estacionária.



Processos estocásticos estacionários

Definição: Um processo estocástico $Z = \{Z(t), t \in \tau\}$$ diz-se estritamente estacionário se todas as distribuições finito-dimensionais permanecem as mesmas sob translações no tempo, ou seja,

$$F(z_1, \dots, z_n; t_1 + \tau, \dots, t_n + \tau) = F(z_1, \dots, z_n; t_1, \dots, t_n)$$

para quaisquer $t_1, \dots, t_n, \tau$.

Isto significa, em particular, que todas as distribuições unidimensionais são invariantes sob translações do tempo, logo a média $\mu(t)$ e a variância $V(t)$ são constantes, isto é,

$$\mu(t) = \mu, \quad V(t) = \sigma^2,$$

para todo $t \in \tau$. Sem perda de generalidade, podemos supor que $\mu = 0$; caso contrário, considere o processo $\{Z(t) - \mu\}$.

Contudo, em várias situações práticas, vamos ter que lidar com **séries não estacionárias**. É bastante comum, por exemplo, em séries econômicas e financeiras a presença de tendências de longos períodos ou curtos períodos (o que geralmente caracteriza uma mudança de nível). Os modelos ARMA assumem como pressupostos que o processo é estacionário. Assim, vamos ter que "tratar" a não estacionariedade das séries através de transformações. A transformação mais comum seria tomar diferenças sucessivas da série original, até obter uma série estacionária.

Operador Diferença

o operador diferença (Δ) de primeira ordem pode ser definido como

$$\Delta Z_t = Z_t - Z_{t-1}$$

a segunda diferença é

$$\Delta^2 Z_t = \Delta[\Delta Z_t] = \Delta[Z_t - Z_{t-1}]$$

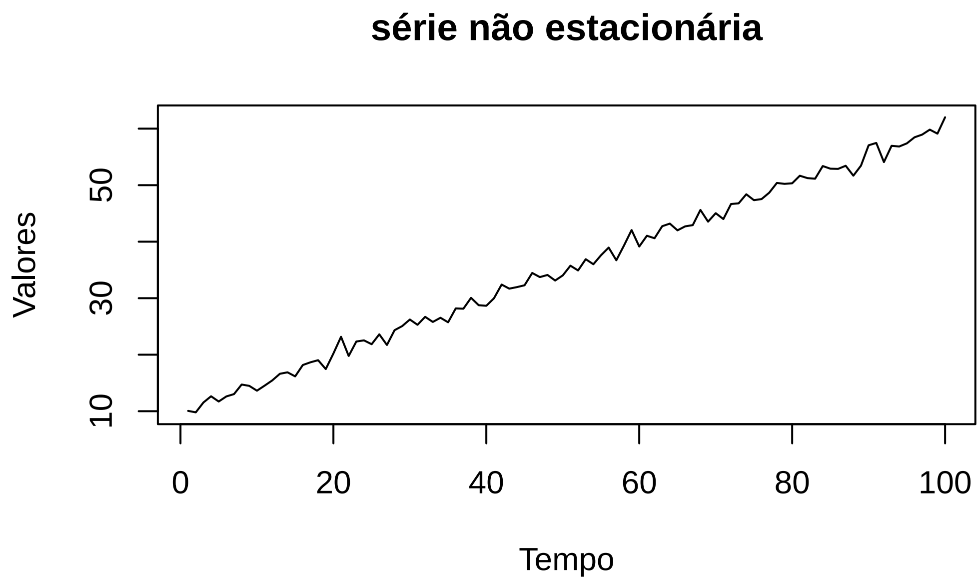
De modo geral, a n-ésima diferença de Z_t é

$$\Delta^n Z_t = \Delta[\Delta^{n-1} Z_t]$$

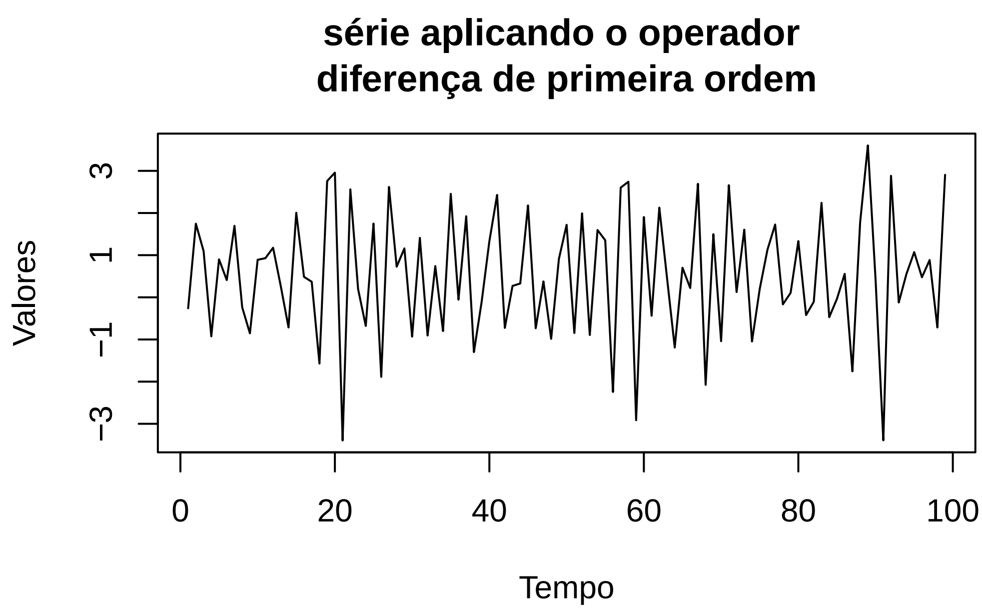
Na prática, é comum tomar uma ou duas diferenças para tornar a série temporal estacionária. O exemplo a seguir, ilustra como o operador diferença pode tornar uma série claramente não estacionária em uma série estacionária.

Exemplo simulado no R:

```
set.seed(1000)
x1 = c()
for (i in 1:100) x1[i] = rnorm(1, mean = 10+(i/2), sd = 1)
x2 = diff(x1, 1)

plot.ts(x1, xlab = "Tempo", ylab = "Valores", main = "série não estacionária" )
```




```
plot.ts(x2, xlab = "Tempo", ylab = "Valores", main = "série aplicando o operador \ndiferen
```



Teste da Raíz unitária

Uma maneira de determinar mais objetivamente se a diferenciação é necessária é usar um *teste de raiz unitária* . Estes são testes estatísticos de hipóteses de estacionariedade que são projetados para determinar se a diferenciação é necessária.

Para verificar se uma série é estacionária, vamos usar o *teste de Kwiatkowski-Phillips-Schmidt-Shin (KPSS)* (Kwiatkowski et al., 1992) . Neste teste, a hipótese nula é que os dados são estacionários, e procuramos evidências de que a hipótese nula é falsa. Consequentemente, pequenos valores de *p-value* (por exemplo, menos de 0,05 ou 0,01) sugerem que a diferenciação é necessária. O teste pode ser calculado usando a função do R `unitroot_kpss()`.

Código R para rodar o teste da raiz unitaria no exemplo anterior

```
## Teste raiz unitaria para a serie com tendencia
feasts::unitroot_kpss( x1  )
```

```
 kpss_stat kpss_pvalue
  2.102721    0.010000
```

```
## Teste raiz unitaria para a serie diferenciada
feasts::unitroot_kpss( x2  )
```

```
 kpss_stat kpss_pvalue
0.03973563  0.10000000
```

Exemplo no R (Preço das ações do Google)

Carregando os pacotes:

```
library(dplyr)
library(forecast)
library(fpp3)
library(tsibble)
```

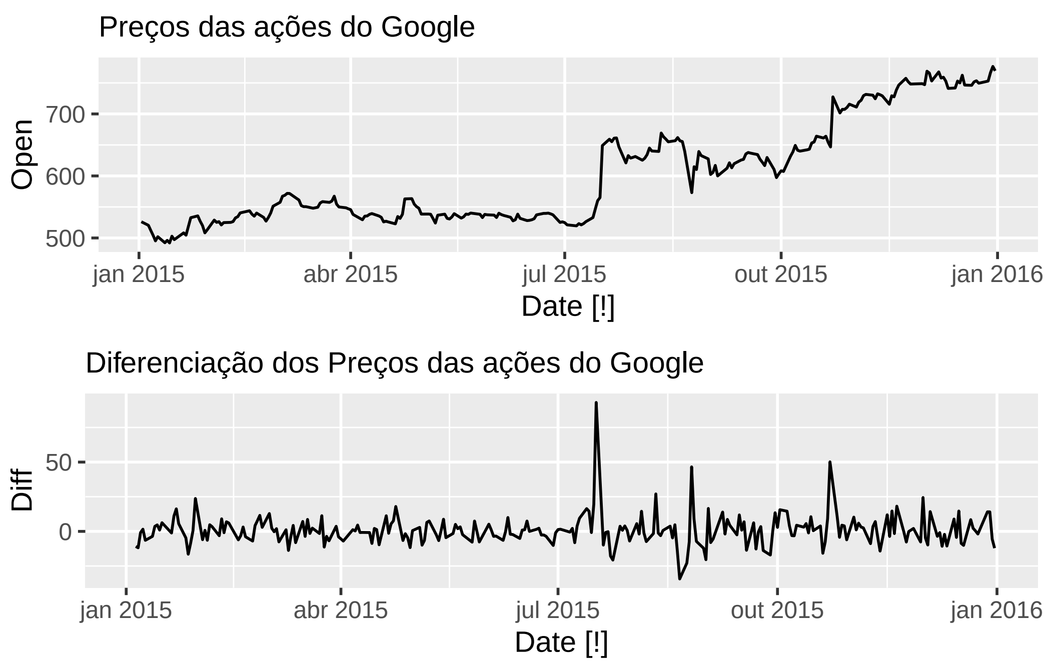
Obtendo os dados

```
google_2015 <- gafa_stock %>%
  filter(Symbol == "GOOG", year(Date) == 2015)

p1 = google_2015 %>%
```

```
  autoplot(  ) + labs(subtitle = "Preços das ações do Google")

p2 = google_2015 %>%
  mutate( Diff = difference(Close) )  %>%
  select( Date, Diff ) %>%
  autoplot(   ) + labs(subtitle = "Diferenciação dos Preços das ações do Google")
```

Séries dos preços das ações do google no ano de 2015 e sua respectiva diferenciação.

```
gridExtra::grid.arrange(p1, p2, ncol = 1)
```



Usando a função `unitroot_kpss` para testar se existe estacionariedade na serie

```
google_2015 %>%
  features(Close, unitroot_kpss)
```

```
# A tibble: 1 x 3
  Symbol kpss_stat kpss_pvalue
  <chr>      <dbl>       <dbl>
1 GOOG        3.56        0.01
```

```
google_2015 %>%
   mutate( Diff = difference(Close) )  %>%
  features(Diff, unitroot_kpss)
```

```
# A tibble: 1 x 3
  Symbol kpss_stat kpss_pvalue
  <chr>      <dbl>       <dbl>
1 GOOG      0.0989         0.1
```

Portanto, através do teste da raiz unitária, podemos concluir que a propriedade de estacionariedade da série é alcançada com apenas uma diferenciação.

Função de autocovariância

Seja $\{Z_t, t \in \mathbb{Z}\}$ um processo estocástico real com tempo discreto, de média zero e função de autocovariância (FACV) $\gamma_r = \mathbf{E}(Z_t Z_{t-r})$.

Proposição: A FACV γ_r satisfaz as seguintes propriedades:

(i) $\gamma_0 > 0$

(ii) $\gamma_{-r} = \gamma_r$

(iii) $\mid \gamma_r \mid \leq \gamma_0$

Autocorrelação

Assim como a correlação mede a extensão de uma relação linear entre duas variáveis, a autocorrelação mede a relação linear entre os valores defasados de uma série temporal.

Existem vários coeficientes de autocorrelação, correspondentes a cada linha vertical no gráfico dos lags. Por exemplo, r_1 mede a relação entre y_t e y_{t-1}, r_2 mede a relação entre y_t e y_{t-2}, e assim por diante.

O valor de k pode ser escrito como

$$r_k = \frac{\sum_{t=k+1}^{T} (y_t - \bar{y}) (y_{t-k} - \bar{y})}{\sum_{t=1}^{T} (y_t - \bar{y})^2},$$

Onde T é o comprimento da série temporal. Os coeficientes de autocorrelação compõem a função de autocorrelação ou ACF.

Tendência e sazonalidade nas parcelas ACF

Quando os dados apresentam um comportamento de tendência, as autocorrelações para pequenas defasagens tendem a ser grandes e positivas porque as observações próximas no tempo também estão próximas em valor. Assim, o ACF de uma série temporal de tendência tende a ter valores positivos que diminuem lentamente à medida que as defasagens aumentam.

Quando os dados possuem sazonalidade, as autocorrelações serão maiores para as defasagens sazonais (em múltiplos do período sazonal) do que para outras defasagens.

Portanto, note que a autocorrelação é uma medida extremamente importante em séries temporais, pois pode identificar padrões de tendência e sazonalidade. Além disso, vamos ver mais a frente que a autocorrelação pode definir processos ARMA.

Processos Autoregressivos

Em um modelo de regressão múltipla, predizemos a variável de interesse usando uma combinação linear de preditores (covariáveis). Em um modelo de auto-regressão, prevemos a variável de interesse usando uma combinação linear de valores passados da própria variável.

Definição: Um modelo autoregressivo de ordem p pode ser escrito como:

$$y_t = \alpha + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + w_t$$

em que w_t é um ruído branco Gaussiano com média zero e variância σ_w^2, $\phi_1, \dots, \phi_p$ são constantes, $y_{t-1}, \dots, y_{t-p}$ são os valores das séries defasadas no tempo e $\alpha = \mu(1 - \phi_1 - \cdots - \phi_p)$.

Forma usando o operador diferença

É bastante comum escrever a equação anterior usando o operador diferença da seguinte forma:

$$\left(1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p\right) y_t = w_t,$$

ou de forma mais compacta como

$$\phi(B) y_t = w_t,$$

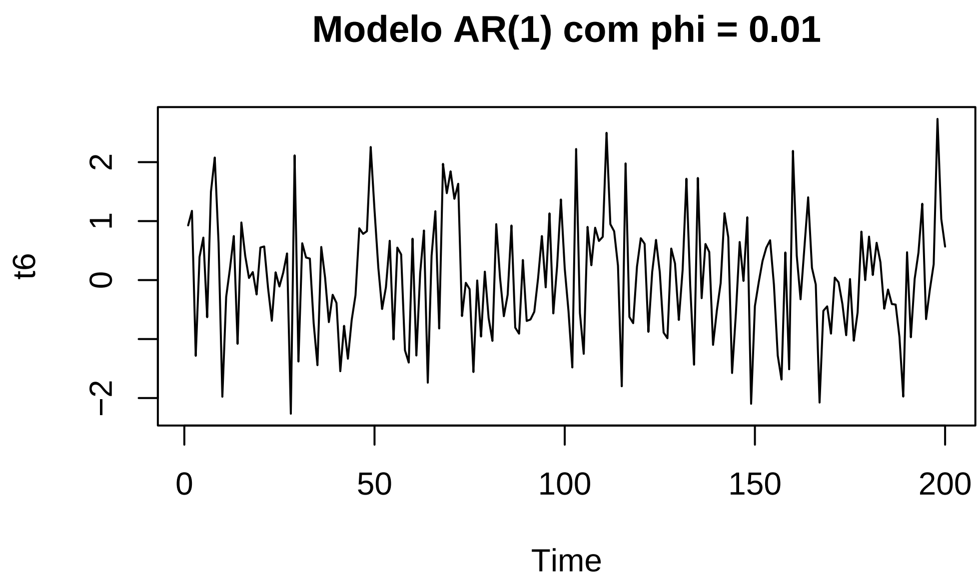
em que $\phi(B) = \left(1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p\right)$.

OBS: um modelo autorregressivo de ordem p é denominado como AR(p).

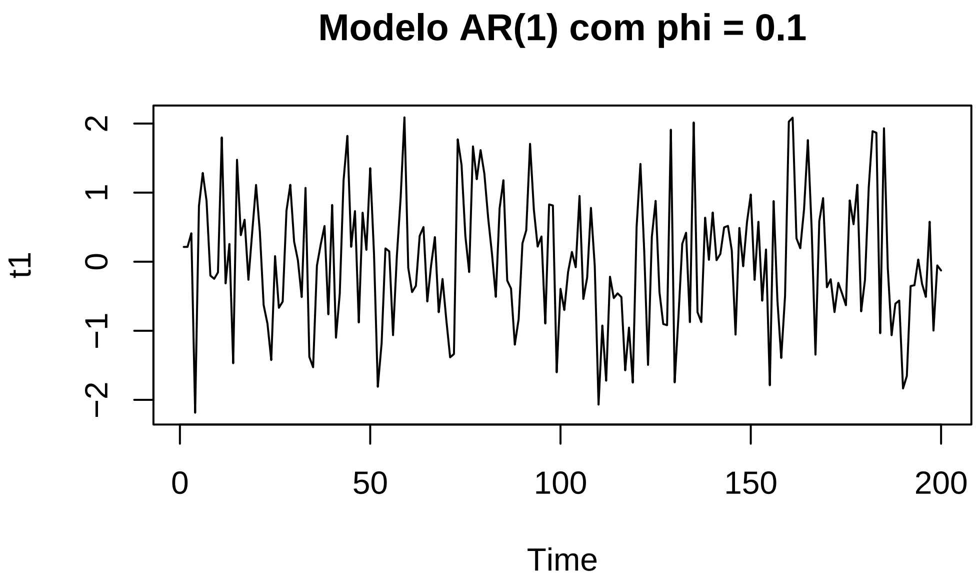
Os modelos autoregressivos podem modelar vários comportamento de séries temporais. Na Figura a seguir, podemos ver que para valores diferentes de ϕ, podemos ter comportamentos diferentes da série.

```
t1 <- arima.sim(list(order = c(1,0,0), ar = 0.1), n = 200)
t2 <- arima.sim(list(order = c(1,0,0), ar = 0.5), n = 200)
t3 <- arima.sim(list(order = c(1,0,0), ar = -0.5), n = 200)
t4 <- arima.sim(list(order = c(1,0,0), ar = 0.8), n = 200)
t5 <- arima.sim(list(order = c(1,0,0), ar = -0.8), n = 200)
t6 <- arima.sim(list(order = c(1,0,0), ar = 0.01), n = 200)
```

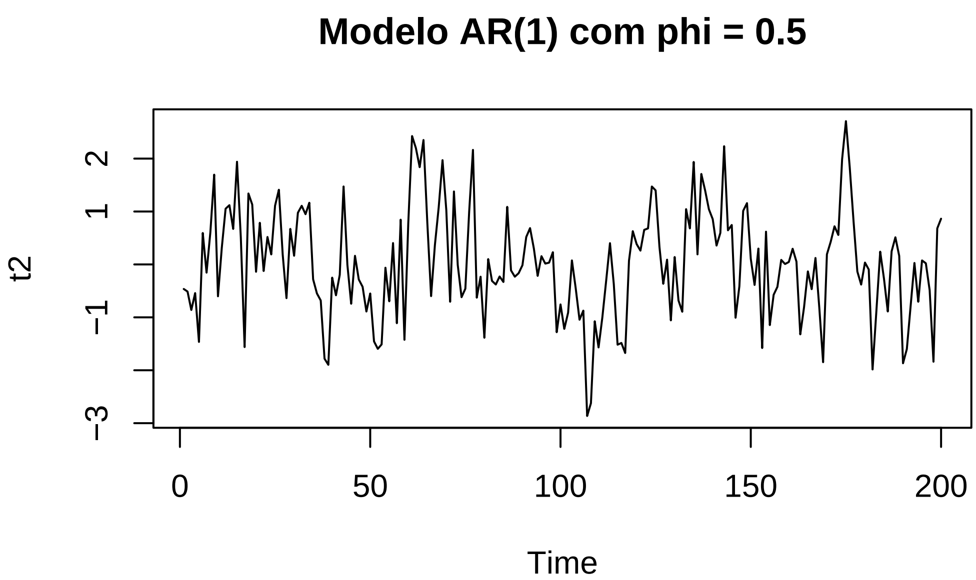
```
ts.plot(t6, main = " Modelo AR(1) com phi = 0.01 ")
```



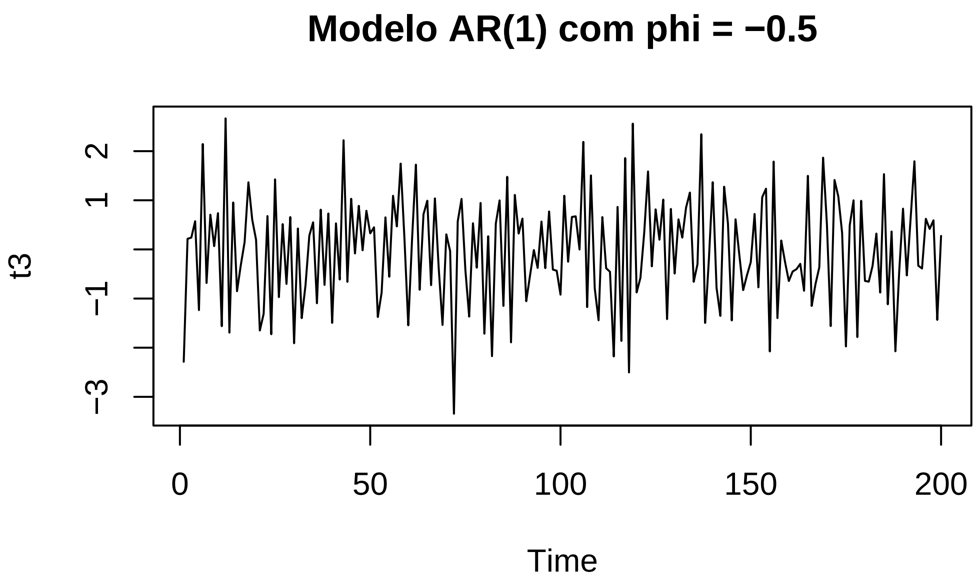

```
ts.plot(t1, main = " Modelo AR(1) com phi = 0.1 ")
```



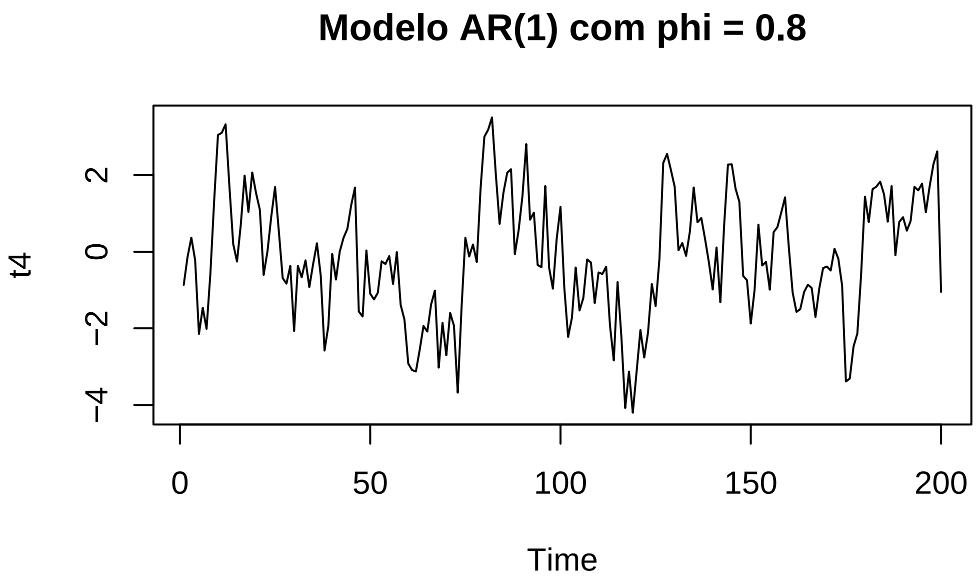

```r
ts.plot(t2, main = " Modelo AR(1) com phi = 0.5 ")
```



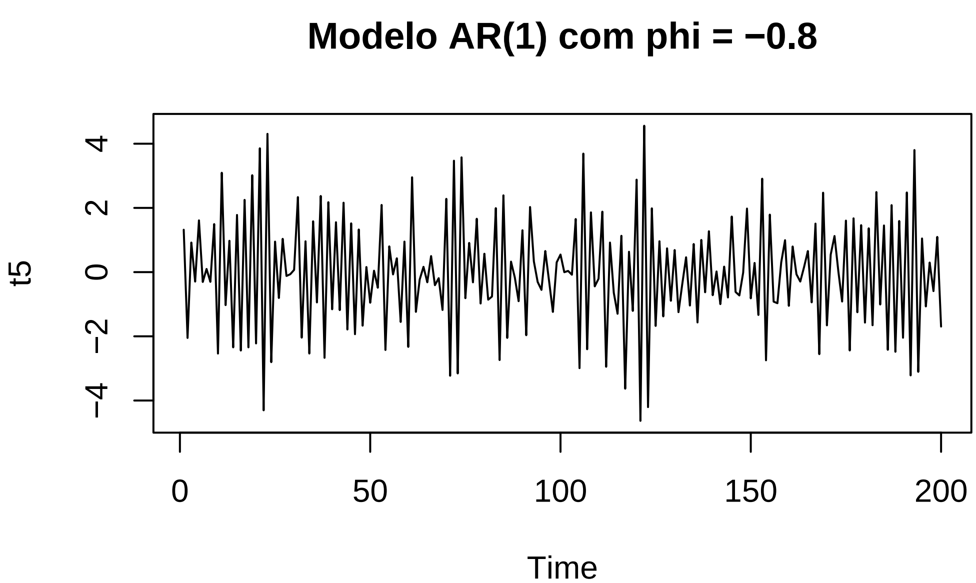
```r
ts.plot(t3, main = " Modelo AR(1) com phi = -0.5 ")
```




```r
ts.plot(t4, main = " Modelo AR(1) com phi = 0.8 ")
```




```
ts.plot(t5, main = " Modelo AR(1) com phi = -0.8 ")
```



Região de estabilidade

Os modelos autorregressivos são restritos a dados estacionários, caso em que algumas restrições nos valores dos parâmetros são necessárias.

- Para um modelo AR(1): $-1 < \phi_1 < 1$.
- Para um modelo AR(2): $-1 < \phi_2 < 1$, $\phi_1 + \phi_2 < 1$ e $\phi_2 - \phi_1 < 1$

Quando $p \geq 3$, as restrições são muito mais complicadas.

Processo de Médias Móveis

Uma alternativa aos processos autoregressivos, é o modelo Médias Móveis de ordem q (Moving Average - MA(q)). Seja w_t um ruído branco gaussiano ($w_t \sim RB(0, \sigma_w^2)$), então definimos um processo de médias movéis de ordem q como:

$$y_t = w_t + \theta_1 w_{t-1} + \theta_2 w_{t-2} + \cdots + \theta_q w_{t-q},$$

em que q são os lags no processo MA e $\theta_1, \dots, \theta_q$ são os parâmetros do modelo.

Usando o operador de médias móveis

$$\theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^q$$

Podemos escrever o processo MA(q) da seguinte forma:

$$y_t = \theta(B) w_t.$$

Processo Autoregressivo de médias móveis (ARMA)

Uma série temporal y_t é um processo ARMA(p, q) se é estacionário e

$$y_t = \alpha + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + w_t + \theta_1 w_{t-1} + \theta_2 w_{t-2} + \cdots + \theta_q w_{t-q},$$

em que $\phi_p \neq 0$, $\theta_q \neq 0$ e $\sigma_w^2 > 0$. Os parâmetros p e q são as ordens dos modelos autoregressivo e médias móveis, respectivamente.

Em particular, o modelo ARMA(p, q) pode ser escrito de forma concisa como:

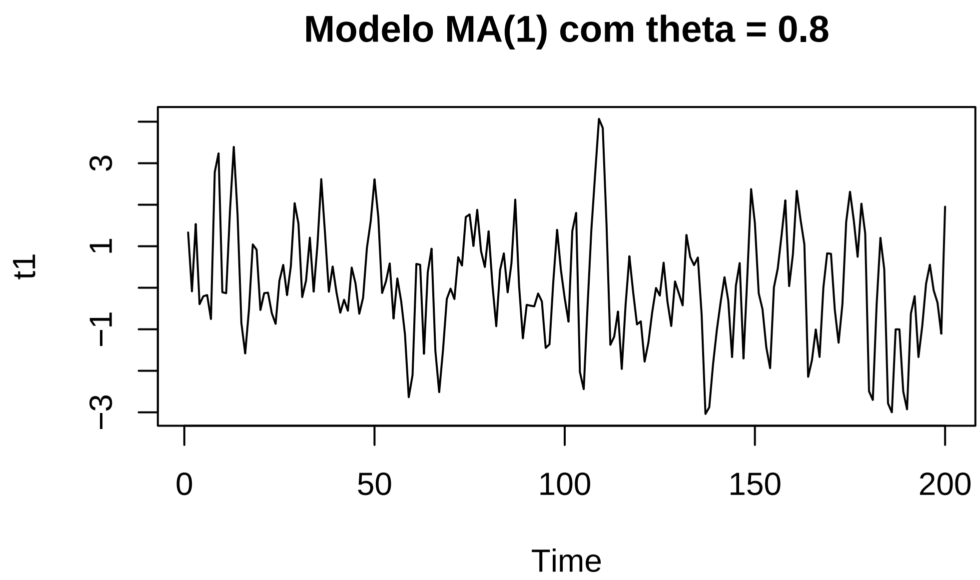
$$\phi(B) y_t = \theta(B) w_t$$

esse é um processo causal e invertível.

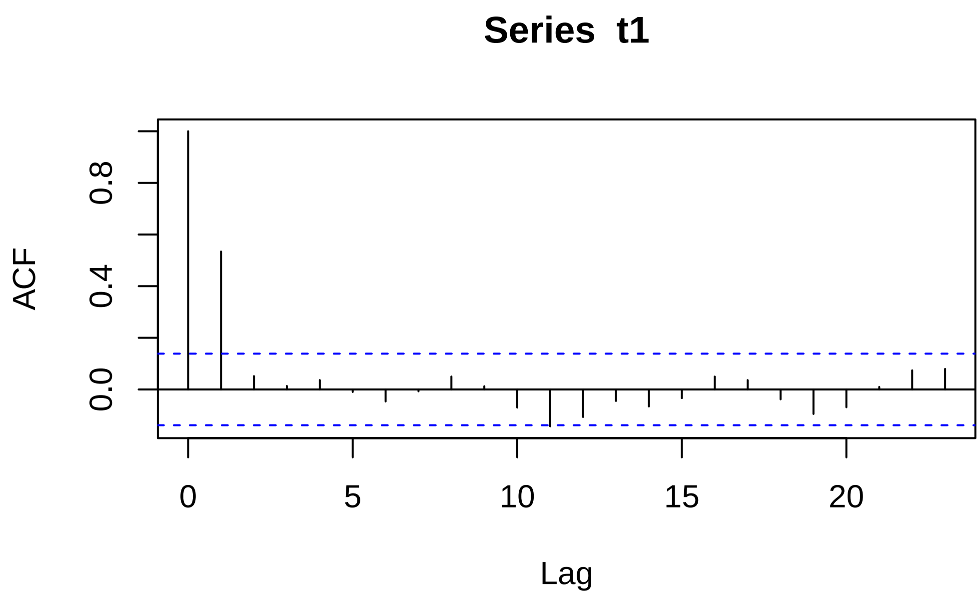
Autocorrelação e autocorrelação parcial nos processos AR e MA

- A autocorrelação define um processo MA(q):

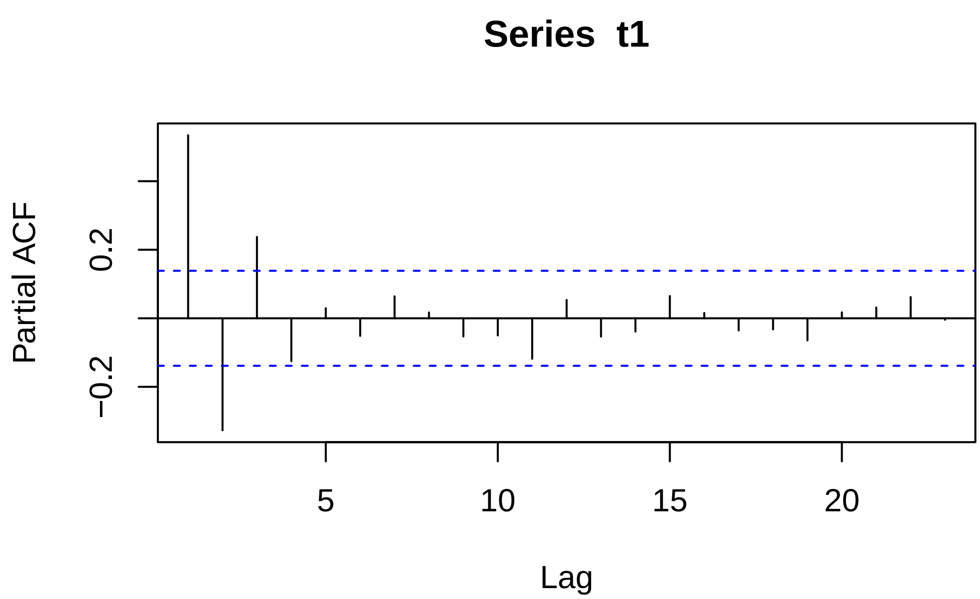
```
### Modelo MA(1)
t1 <- arima.sim(list(order = c(0,0,1), ma = 0.8), n = 200)
ts.plot(t1, main = " Modelo MA(1) com theta = 0.8 ")
```



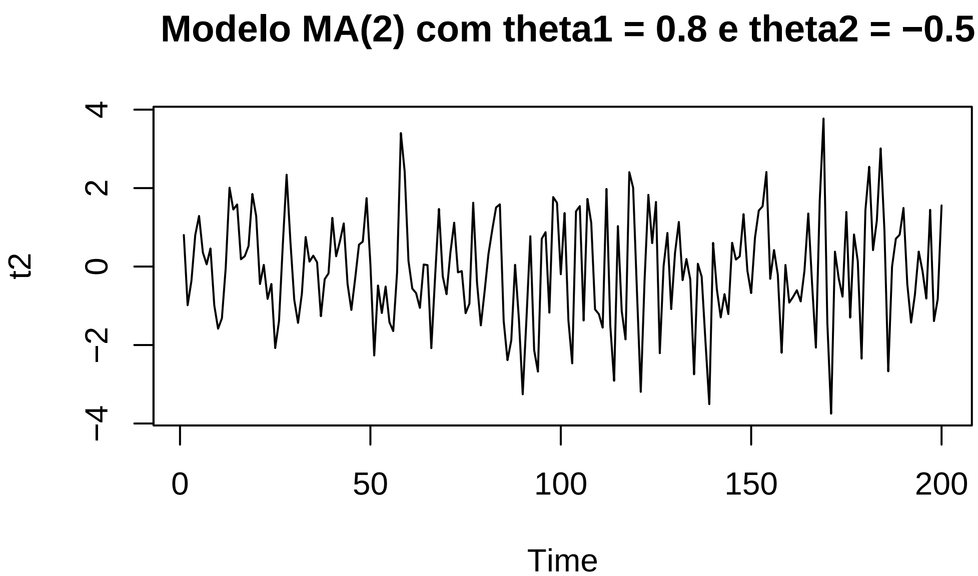

```
acf(t1)
```


Series t1
ACF
0.0
0.4
0.8
0
5
10
15
20
Lag

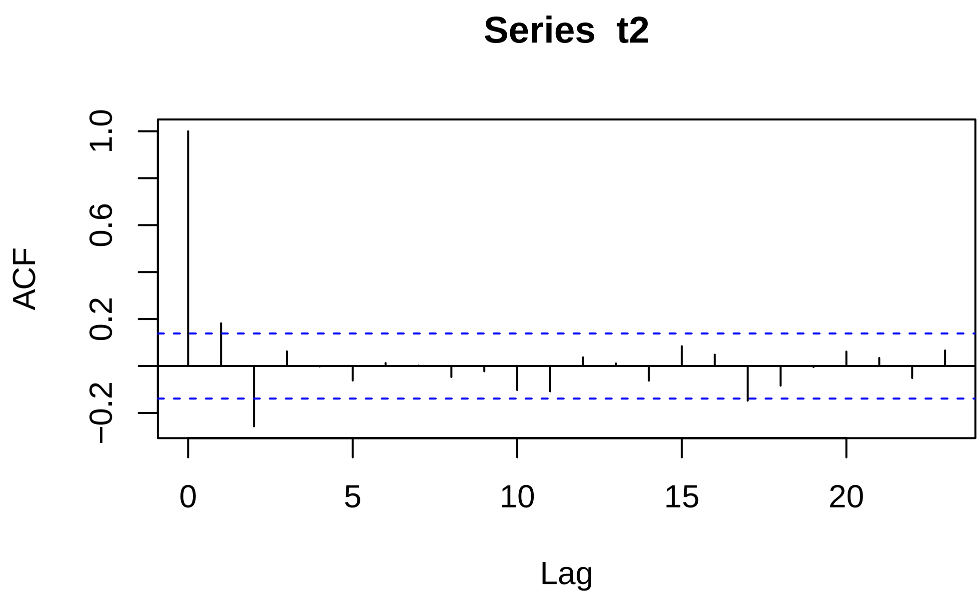
```r
pacf(t1)
```


Series t1
Partial ACF
−0.2
0.2
5
10
15
20
Lag

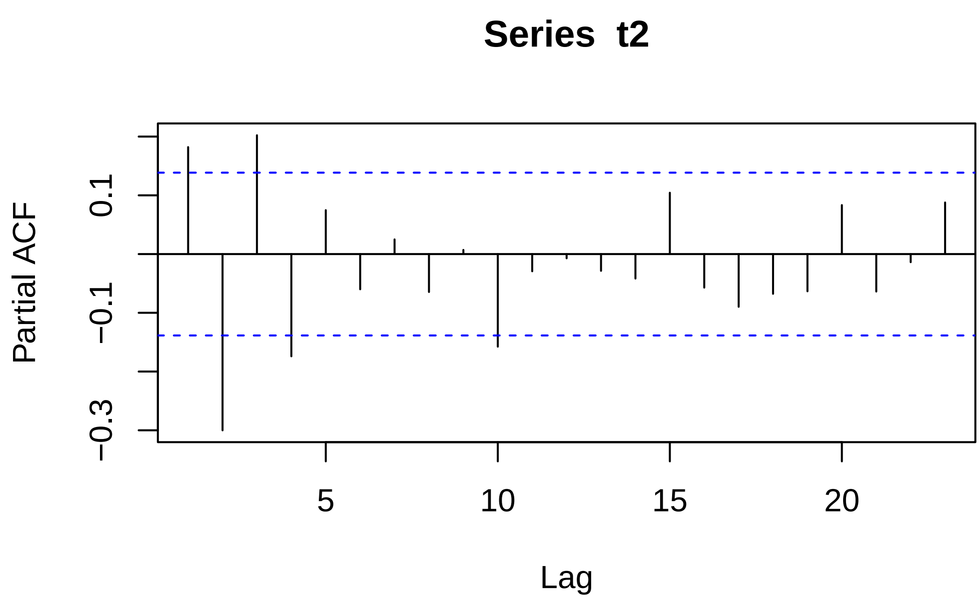
```r
### Modelo MA(2)
t2 <- arima.sim(list(order = c(0,0,2), ma = c(0.8,-0.5) ), n = 200)
ts.plot(t2, main = " Modelo MA(2) com theta1 = 0.8 e theta2 = -0.5")
```




```r
acf(t2)
```

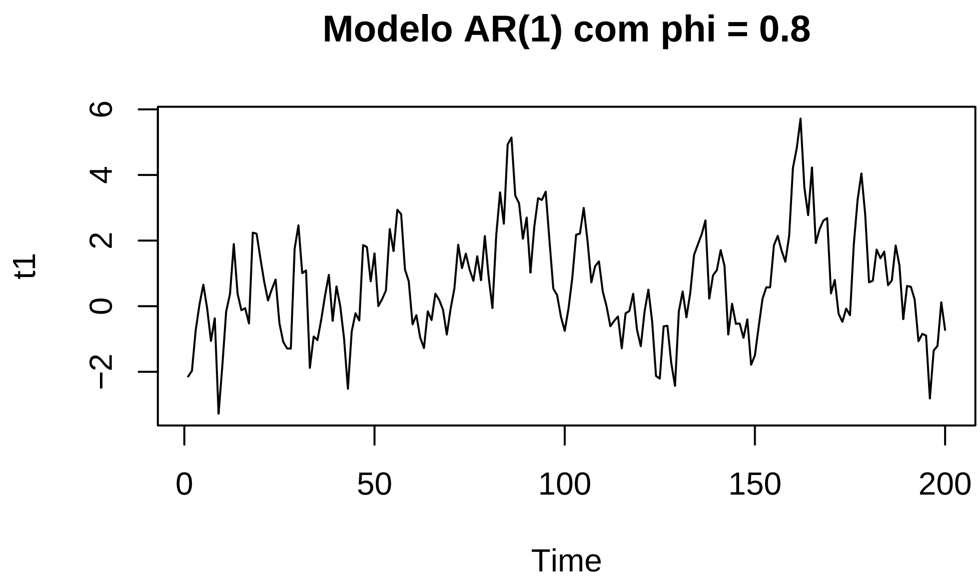

Series t2
ACF
1.0
0.6
0.2
−0.2
0
5
10
15
20
Lag

```
pacf(t2)
```

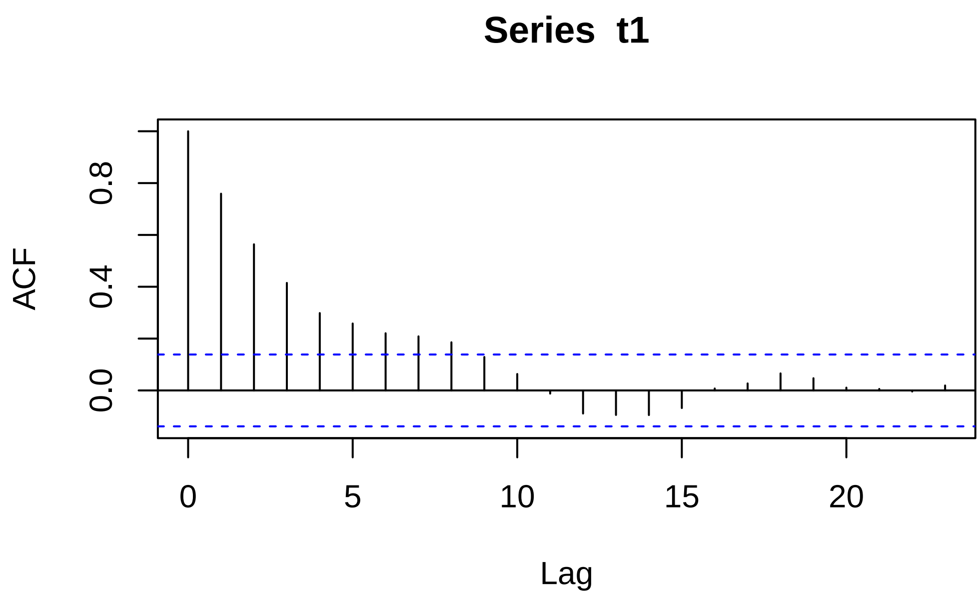

Series t2
Partial ACF
0.1
−0.1
−0.3
5
10
15
20
Lag

- A autocorrelação parcial define um processo AR(p)

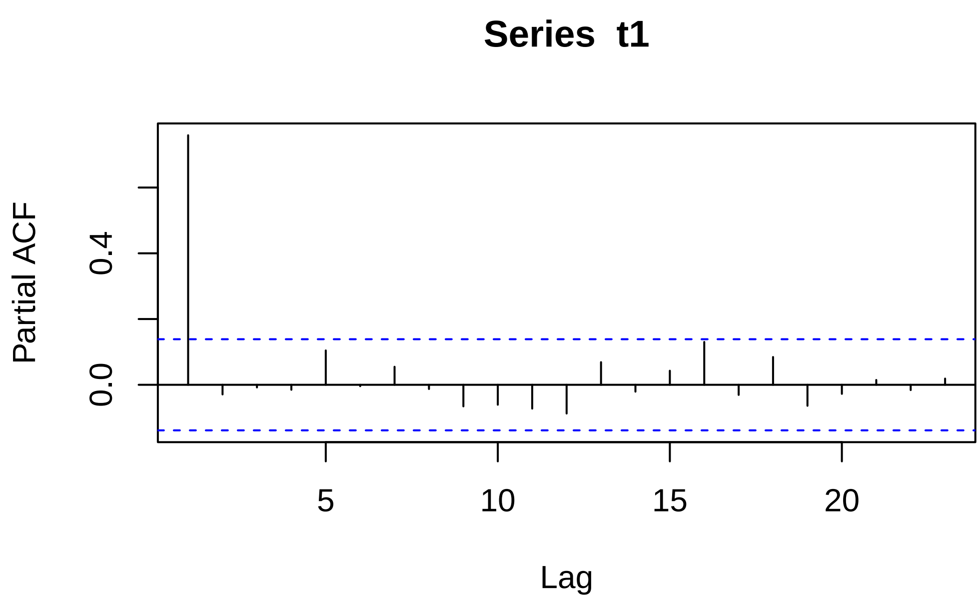
```
### Modelo AR(1)
t1 <- arima.sim(list(order = c(1,0,0), ar = 0.8), n = 200)
ts.plot(t1, main = " Modelo AR(1) com phi = 0.8 ")
```



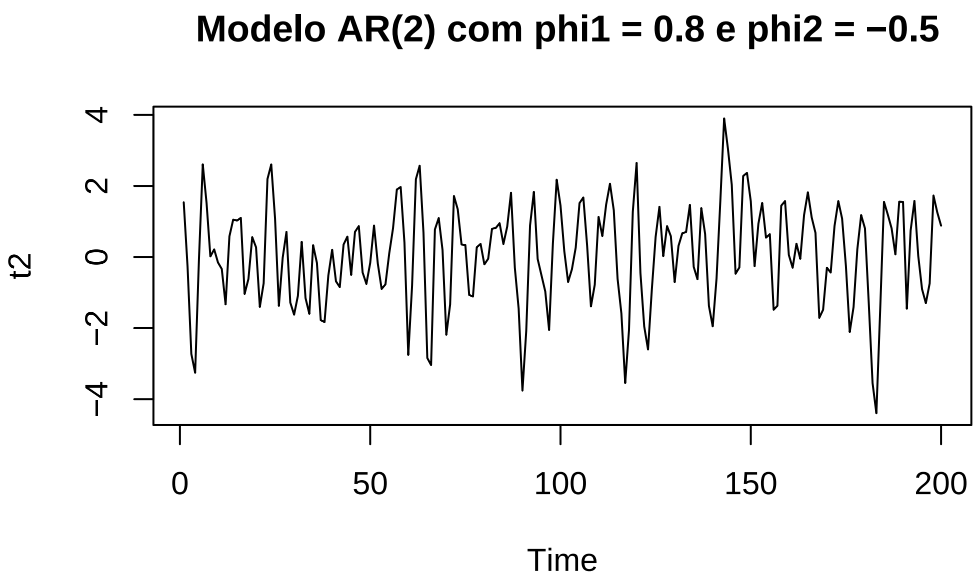

```
acf(t1)
```


Series t1
ACF
0.0
0.4
0.8
0
5
10
15
20
Lag

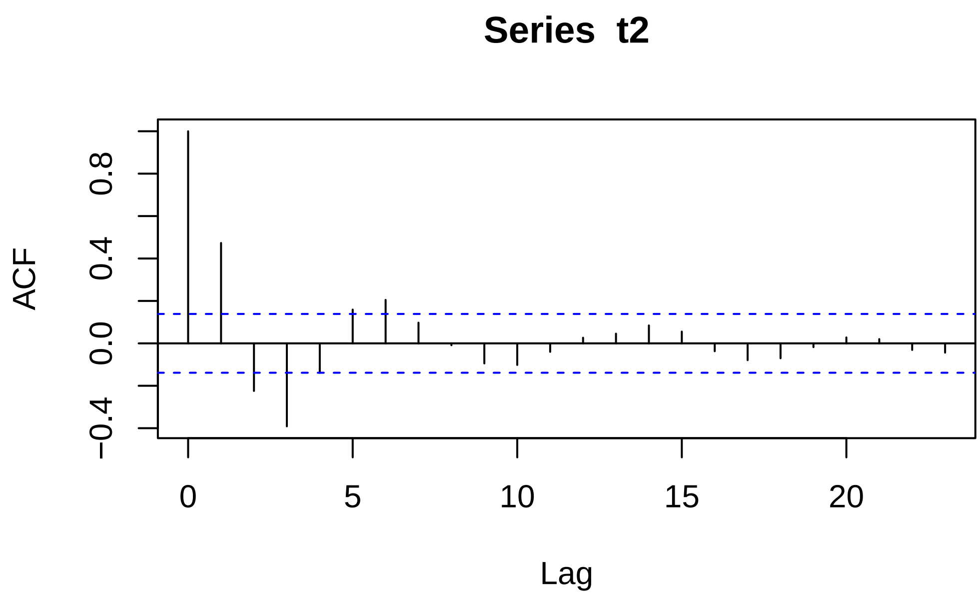
```
pacf(t1)
```


Series t1
Partial ACF
0.0
0.4
5
10
15
20
Lag

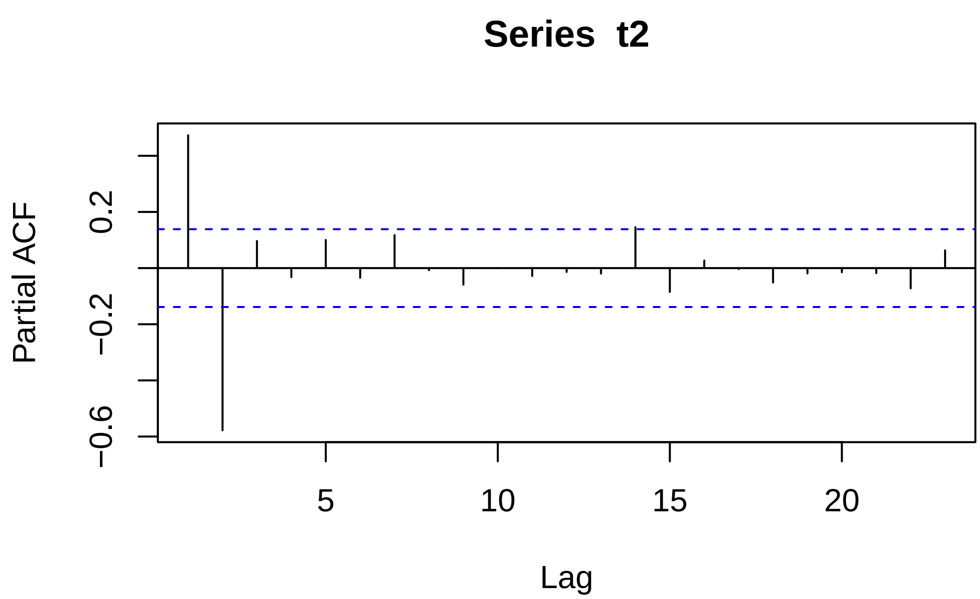
```r
### Modelo AR(2)
t2 <- arima.sim(list(order = c(2,0,0), ar = c(0.8,-0.5) ), n = 200)
ts.plot(t2, main = " Modelo AR(2) com phi1 = 0.8 e phi2 = -0.5")
```




```r
acf(t2)
```


Series t2
ACF
0.8
0.4
0.0
−0.4
0
5
10
15
20
Lag

```r
pacf(t2)
```


Series t2
Partial ACF
0.2
−0.2
−0.6
5
10
15
20
Lag

A autocorrelação de um processo MA(q) é igual a zero para os lags $> q$ e a autocorrelação parcial de um processo AR(p) é zero para lags $> p$. Portanto, os gráficos da ACF e PACF podem auxiliar a escolha das ordens dos modelos AR e MA. Entretanto, adicionalmente deve ser usado critérios de seleção de modelos para a escolha dos modelos.

Observação sobre o inverso do operador diferença

```
set.seed(100)
y=rpois(10, 5)    # serie original
dy=diff(y)     # diff ordem 1
invdy= cumsum(c(y[1],dy)) # serie diff inversa
cbind(y, dy, invdy)
```

```
      y dy invdy
 [1,] 4 -1     4
 [2,] 3  2     3
 [3,] 5 -3     5
 [4,] 2  3     2
 [5,] 5  0     5
 [6,] 5  2     5
 [7,] 7 -3     7
 [8,] 4  1     4
 [9,] 5 -2     5
[10,] 3 -1     3
```

Entretanto, não precisamos nos preocupar com isso. Pois, o próprio R já faz isso. A seguir, vamos apresentar o modelo que incorpora as diferenciações no modelo.

Modelos Autoregressivos integrados de médias móveis (ARIMA)

Se combinarmos diferenciação com autorregressão e um modelo de média móvel, obtemos um modelo ARIMA não sazonal. ARIMA é um acrônimo para *AutoRegressive Integrated Moving Average*. O modelo completo pode ser escrito como

$$y'_t = c + \phi_1 y'_{t-1} + \cdots + \phi_p y'_{t-p} + \theta_1 w_{t-1} + \cdots + \theta_q w_{t-q} + w_t$$

em que y'_t é a série diferenciada (pode ter sido diferenciada mais de uma vez). Os "preditores" no lado direito incluem ambos os valores defasados de y_t e w_t. Chamamos isso de ARIMA(p, d, q) modelo, onde

- $p =$ ordem da parte autorregressiva;
- $d =$ grau de primeira diferenciação envolvido;
- $q =$ ordem da parte média móvel.

Observação: As mesmas condições de estacionaridade e invertibilidade usadas para modelos autorregressivos e de média móvel também se aplicam a um modelo ARIMA.

Muitos dos modelos que já discutimos são casos especiais do modelo ARIMA, conforme a seguir:

- Ruído branco: ARIMA$(0, 0, 0)$
- Autoregressivo (AR): ARIMA$(p, 0, 0)$
- Médias Móveis (MA): ARIMA$(0, 0, q)$
- Autoregressivo de Médias Móveis (ARMA): ARIMA$(p, 0, q)$

Aplicando o operador diferença, podemos representar o modelo ARIMA(p, d, q) como

$$\left(1 - \phi_1 B - \cdots - \phi_p B^p\right) \quad (1 - B)^d y_t = c + \left(1 + \theta_1 B + \cdots + \theta_q B^q\right) w_t.$$

Critérios de informação

Akaike's Information Criterion (AIC) é usado para determinar a ordem de um modelo ARIMA. Pode ser escrito como

$$\text{AIC} = -2\log(L) + 2(p + q + k + 1),$$

em que L é a verossimilhança dos dados, $k = 1$ se $c \neq 0$ e $k = 0$ se $c = 0$. Note que o último termo em parênteses é o número de parâmetros do modelo (incluindo a variância dos resíduos, σ_w^2).

Para modelos ARIMA, o AIC corrigido pode ser escrito como

$$\text{AICc} = \text{AIC} + \frac{2(p + q + k + 1)(p + q + k + 2)}{T - p - q - k - 2}$$

em que T é o tamanho da amostra.

Por fim, o *Bayesian Information Criterion* (BIC) pode ser definido como

$$\text{BIC} = \text{AIC} + [\log(T) - 2](p + q + k + 1)$$

Bons modelos são obtidos minimizando o AIC, AICc ou BIC. É preferível usar o AICc.

É importante notar que esses critérios de informação tendem a não ser bons guias para selecionar a ordem apropriada de diferenciação (d) de um modelo, mas apenas para selecionar os valores de p e q. Isso ocorre porque a diferenciação altera os dados nos quais a verossimilhança é calculada, tornando os valores de AIC entre modelos com diferentes ordens de diferenciação não comparáveis. Então, precisamos usar alguma outra abordagem para escolher d , e então podemos usar o AICc para selecionar p e q.

Modelos SARIMA

Até agora, restringimos nossa atenção a dados não sazonais e modelos ARIMA não sazonais. No entanto, os modelos ARIMA também são capazes de modelar uma ampla gama de dados sazonais.

Um modelo ARIMA sazonal é formado pela inclusão de termos sazonais adicionais nos modelos ARIMA que vimos até agora. Está escrito da seguinte forma:

$$\text{ARIMA} \quad \underbrace{(p,d,q)}_{\substack{\uparrow \\ \text{parte não sazonal} \\ \text{do modelo}}} \quad \underbrace{(P,D,Q)_m}_{\substack{\uparrow \\ \text{Parte sazonal} \\ \text{do modelo}}}$$

em que $m =$ o período sazonal (por exemplo, se a série for mensal o m é 12, ou seja, em um ano temos 12 meses, De forma similar, se temos uma série trimestral em um ano temos 4 trimestres ou seja, $m = 4$). Usamos notação em maiúsculo para as partes sazonais do modelo e notação em minúsculo para as partes não sazonais do modelo.

Definição: O modelo multiplicativo *Seasonal Autoregressive Integrated Moving Average* (SARIMA) é definido por:

$$(1 - \phi_1 B)(1 - \Phi_1 B^m)(1 - B)(1 - B^m)\, y_t = (1 + \theta_1 B)(1 + \Theta_1 B^m)\, w_t$$

em que w_t é o usual processo ruído branco gaussiano, m é o período sazonal, (p, d, q) são as ordens para a estrutura não sazonal e (P, D, Q) são as ordens para a estrutura sazonal.

A notação geral do modelo SARIMA é

$$ARIMA(p, q, q) \times (P, D, Q)_m.$$

Por exemplo, Considere o seguinte modelo SARIMA denotado por $ARIMA(0,0,1)\times(1,0,0)_{12}$. Dessa forma assumindo que a série é mensal, o período $m = 12$ corresponde a quantidade de meses em um ano. Além disso, o modelo acima corresponde a um modelo MA(1) para a parte não sazonal e AR(1) para a parte sazonal.

Comportamento da ACF e PACF de modelos SARIMA

A parte sazonal de um modelo AR ou MA será vista nas defasagens sazonais do PACF e ACF. Por exemplo, um ARIMA $(0,0,0)(0,0,1)_{12}$ o modelo mostrará:

- um pico no lag 12 no ACF, mas nenhum outro pico significativo;
- Decaimento exponencial nas defasagens sazonais do PACF (isto é, nas defasagens $12, 24, 36, \ldots$).

Da mesma forma, um ARIMA $(0,0,0)(1,0,0)_{12}$ o modelo mostrará:

- decaimento exponencial nas defasagens sazonais da ACF;
- um único pico significativo no lag 12 no PACF.

Ao considerar as ordens sazonais apropriadas para um modelo ARIMA sazonal, restrinja a atenção aos *lags* sazonais.

O procedimento de modelagem é quase o mesmo que para dados não sazonais, exceto que precisamos selecionar termos sazonais AR e MA, bem como os componentes não sazonais do modelo. O processo é melhor ilustrado por meio de exemplos.

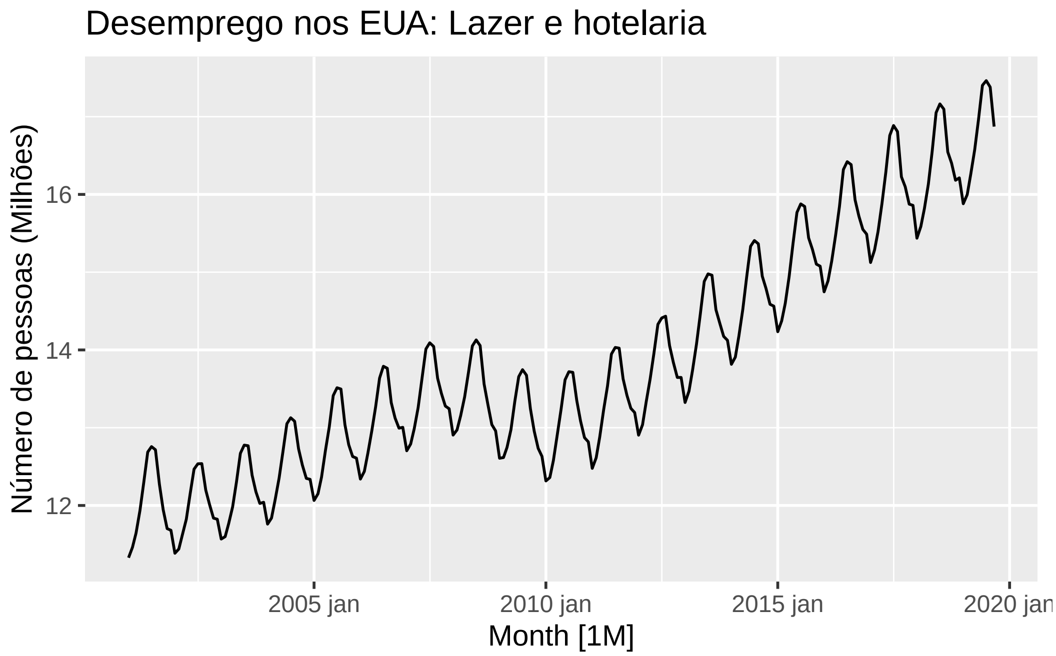
Exemplo: emprego mensal nas áreas de lazer e hotelaria nos EUA

Descreveremos a modelagem ARIMA sazonal usando dados mensais de emprego nos EUA para empregos nas áreas de lazer e hotelaria de janeiro de 2001 a setembro de 2019, como podemos ver na figura a seguir.

```r
## Bibliotecas de séries temporais
library(TSA)
library(forecast)
library(fpp3)
```

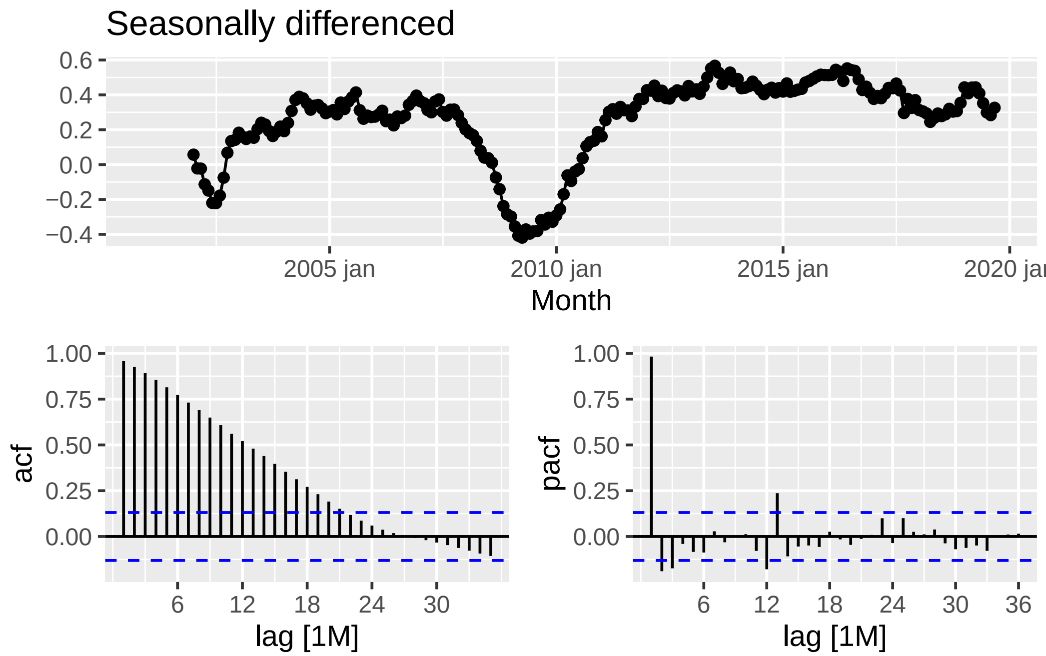
```r
leisure <- us_employment %>%
  filter(Title == "Leisure and Hospitality",
         year(Month) > 2000) %>%
  mutate(Employed = Employed/1000) %>%
```

```
  select(Month, Employed)
autoplot(leisure, Employed) +
  labs(title = "Desemprego nos EUA: Lazer e hotelaria",
       y="Número de pessoas (Milhões)")
```



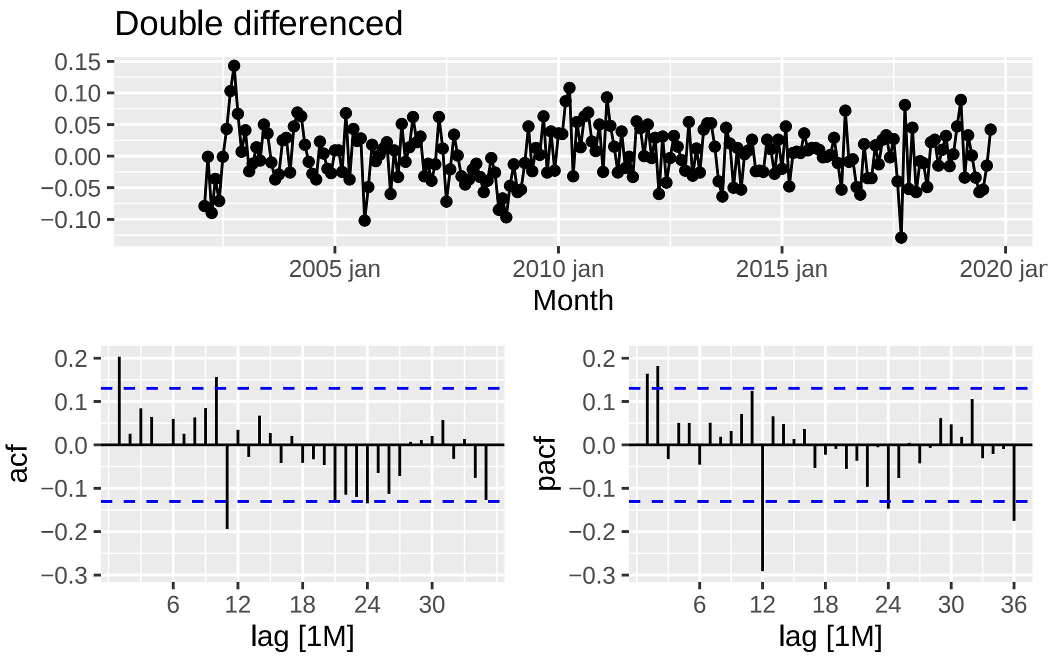
Como a série é claramente não estacionária, vamos aplicar o operador diferença:

```
leisure %>%
  gg_tsdisplay(difference(Employed, 12),
               plot_type='partial', lag=36) +
  labs(title="Seasonally differenced", y="")
```

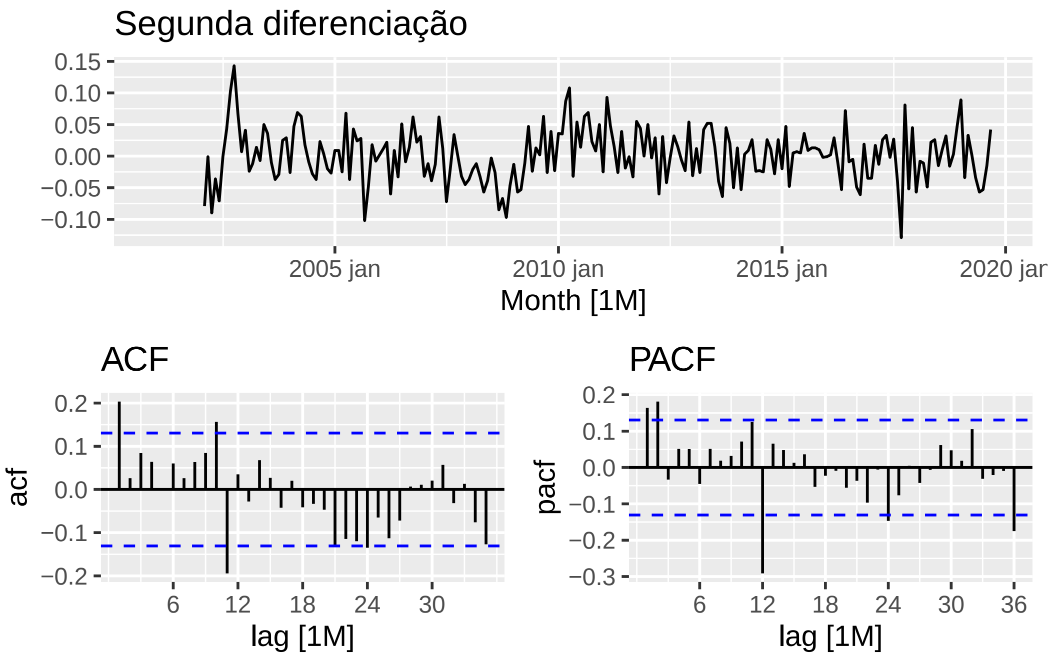


Note ainda que a primeira diferenciação não é o bastante para tornar a série estacionária. Dessa forma, é importante diferenciá-la novamente.

```
leisure %>%
  gg_tsdisplay(difference(Employed, 12) %>% difference(),
               plot_type='partial', lag=36) +
  labs(title = "Double differenced", y="")
```




```
#### Comportamento da serie
plot_serie = leisure %>%
  autoplot(difference(Employed, 12) %>% difference()) +
  labs(title="Segunda diferenciação",
       y="")


### ACF e PACF das series
acf = leisure %>%
  ACF(difference(Employed, 12) %>% difference(), lag_max = 36 ) %>%
  autoplot() + labs(title="ACF")

pacf = leisure %>%
  PACF(difference(Employed, 12) %>% difference(), lag_max = 36) %>%
  autoplot() + labs(title="PACF")


gridExtra::grid.arrange(plot_serie, acf, pacf,layout_matrix=rbind(c(1,1),c(2,3) ))
```



Analisando a Figura acima, vamos analisar as funções ACF e PACF sob duas perspectiva, (i) Não sazonal e (ii) Sazonais.

- Para a parte não sazonal:
  - Note que a ACF sai forte no lag 1, sugerindo então um processo MA(1)
  - Note que a PACF sai forte nos dois primeiros lags, sugerindo então um processo AR(2)
- Para a parte sazonal:
  - Note que a ACF sai forte apenas lag 12, sugerindo um modelo $\text{MA}(1)_{12}$
  - Note que a PACF sai forte no lag 12 e 36 (além do lag 24), sugerindo um processo AR(3)

Portanto como tivemos uma diferenciação sazonal e outra diferenciação. Processos como ARIMA $(0,1,1)(0,1,1)_{12}$ e ARIMA $(2,1,0)(3,1,0)_{12}$ são possíveis candidatos de melhor modelo.

Para realizar o ajuste do modelo SARIMA vamos usar a função `model` da seguinte maneira:

```
fit <- leisure %>%
  model(
```

```
    arima012011 = ARIMA(Employed ~ pdq(0,1,1) + PDQ(0,1,1)),
    arima210011 = ARIMA(Employed ~ pdq(2,1,0) + PDQ(3,1,0)),
    auto = ARIMA(Employed, stepwise = FALSE, approx = FALSE)
  )
fit
```

```
# A mable: 1 x 3
                arima012011                 arima210011                        auto
                    <model>                     <model>                     <model>
1 <ARIMA(0,1,1)(0,1,1)[12]> <ARIMA(2,1,0)(3,1,0)[12]> <ARIMA(2,1,0)(1,1,1)[12]>
```

Para a escolha do melhor modelo vamos usar o critério de seleção AICc ou BIC uando a função `glance`.

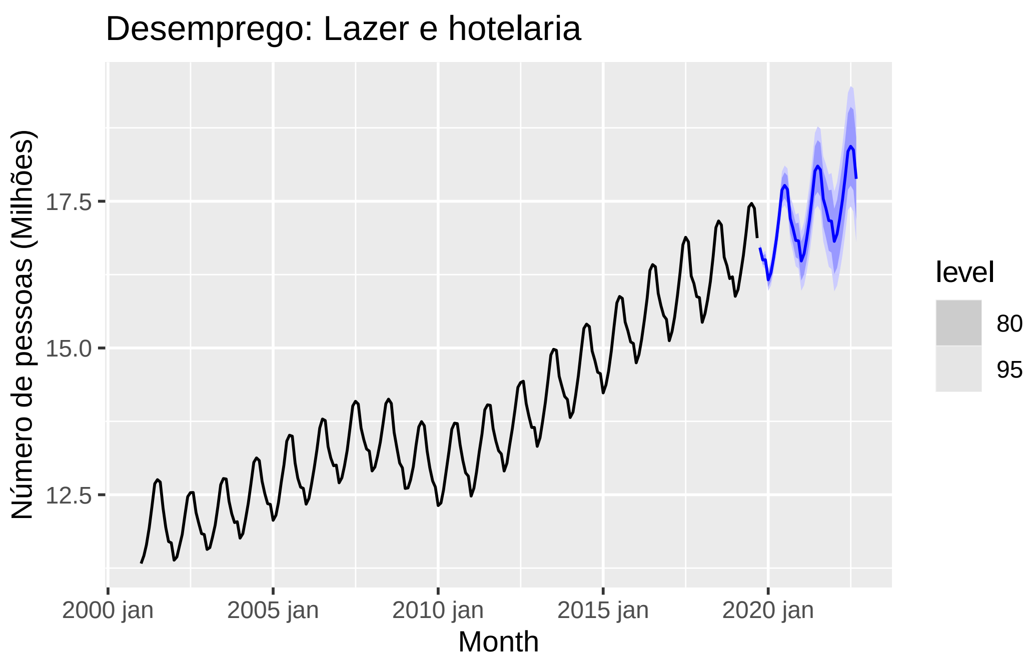
```
glance(fit) %>% arrange(AICc) %>% select(.model:BIC)
```

```
# A tibble: 3 x 6
  .model       sigma2 log_lik   AIC  AICc   BIC
  <chr>         <dbl>   <dbl> <dbl> <dbl> <dbl>
1 auto        0.00142    395. -780. -780. -763.
2 arima210011 0.00144    394. -776. -776. -756.
3 arima012011 0.00154    386. -766. -765. -756.
```

Portanto, o modelo `auto`, que representa uma ARIMA $(2,1,0)(1,1,1)_{12}$ foi o melhor modelo pois apresenta o menor AICc.

Previsão no R

```
forecast(fit, h=36) %>%
  filter(.model=='auto') %>%
  autoplot(leisure) +
  labs(title = "Desemprego: Lazer e hotelaria",
       y="Número de pessoas (Milhões)")
```


Desemprego: Lazer e hotelaria
Número de pessoas (Milhões)
17.5
15.0
12.5
2000 jan
2005 jan
2010 jan
2015 jan
2020 jan
Month
level
80
95