Estátistica II

Lista 2

Fazer a todos os testes estatísticos e gráficos necessários e a predição para os próximos 6 meses do índice de produção de bebidas para os seguintes modelos:

i. ETS;

ii. ARIMA OU SARIMA (verificar se existe sazonalidade ou não e decidir qual modelo é mais adequado)

Obs: separe os últimos 12 meses da série para testar o modelo.

Lendo e organizando

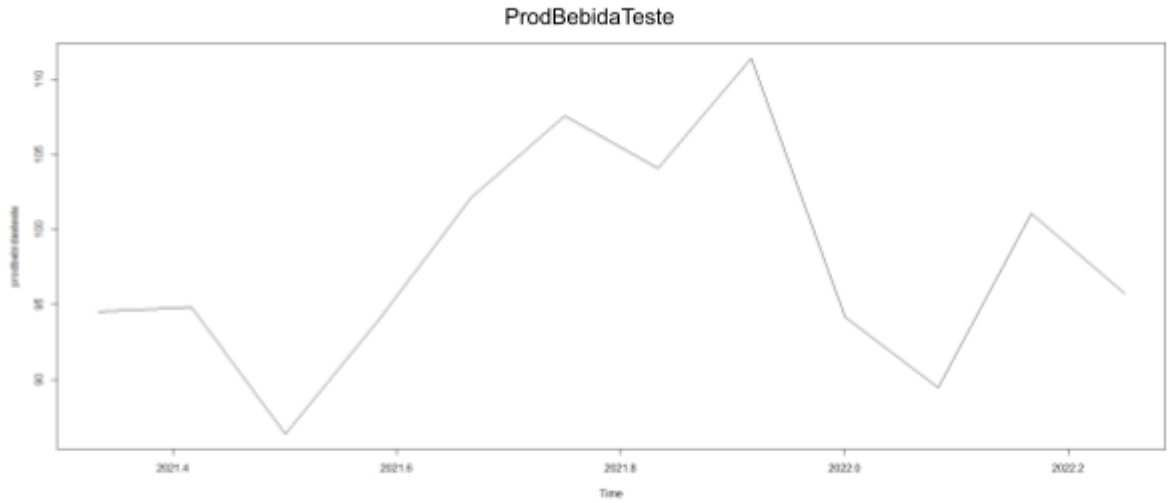
```
> prodbebidas <- read_excel("C:/Users/bruno/Documents/Facul/Listas/Arquivos_para_R/prodbebidas.xls")
>
> prodbebidas <- prodbebidas[2]
> prodbebidas
# A tibble: 244 × 1
   Prodbebidas
         <dbl>
 1        62.6
 2        57.8
 3        60.7
 4        62.7
 5        62.3
 6        60.5
 7        60.2
 8        67.3
 9        66.6
10        83.4
# … with 234 more rows
# ℹ Use `print(n = ...)` to see more rows
> prodbebidas_ts <- ts(data = prodbebidas,
+                      start = c(2002, 1),
+                      end = c(2022, 4),
+                      frequency = 12)
> prodbebidas_ts
          Jan       Feb       Mar       Apr       May       Jun       Jul       Aug       Sep       Oct       Nov       Dec
2002  62.55626  57.84550  60.69372  62.65435  62.31734  60.49615  60.16045  67.33102  66.58131  83.37684  82.94314  88.43612
2003  62.56067  60.76872  63.89176  56.67937  58.01843  56.00493  62.18952  59.96014  64.55340  74.25945  73.53011  89.46737
2004  64.98702  61.60299  63.94879  64.76350  60.49053  55.94533  62.98432  64.14763  76.73094  81.73803  81.12081  88.85680
2005  72.34144  62.85888  69.02932  71.94410  66.01501  66.99988  62.51204  70.91745  76.58744  82.16382  85.04216  93.53437
2006  74.93546  73.24877  75.90443  69.65612  71.03721  66.88274  70.67572  77.28594  82.61556  85.93397  92.50947 101.55033
2007  84.58096  73.88488  79.93531  77.96069  76.84217  69.34325  70.84053  81.14352  82.44995  92.75016  96.49290 106.67008
2008  90.77849  74.45360  74.39670  74.87630  75.97903  73.19335  73.75979  79.37945  85.26304  92.47269  94.22112 106.76814
2009  88.31490  77.76599  86.68112  79.92901  80.37831  74.77481  78.34275  85.56901  91.28898 103.91359 102.94418 116.14504
2010 103.04385  89.78474 103.23820  88.96776  90.48966  90.24111  88.98598  90.07683 100.45606 108.18073 109.28844 122.89162
2011  99.22271  94.29694  93.23749  86.52706  86.43170  80.64185  88.62650  96.58327 105.64314 113.91694 112.75390 126.67227
2012  96.28838  92.58111 101.34431  88.81832  89.67958  85.85891  85.06311 101.58247 103.70791 110.84586 118.23289 125.99715
2013 108.73152  91.10860  95.46076  88.22379  90.24233  81.92203  91.58699  96.99718 100.17055 103.09622 107.75285 119.52384
2014 102.63978  97.57184 102.66134  91.07341  93.08294  89.25355  92.64526  90.95589  96.70269 102.49691 106.82068 124.02974
2015 104.69755  92.04724  91.46396  79.64809  83.19867  84.83558  83.45016  92.24364  98.67980 104.27626 106.69583 112.18983
2016  94.20398  87.41853  87.75295  86.38086  85.75426  81.75488  83.92279  91.00071  96.52690  96.81942 100.89925 105.92023
2017  93.25303  86.77665  92.37962  76.62109  84.51705  82.02118  84.23977  91.74115  96.70937 104.56042 104.51038 109.64334
2018 102.54607  88.93495  87.77399  85.60336  73.29208  94.39913  95.45803  93.03829  86.49107 102.33571 100.04859 106.71818
2019 103.70168  93.51690  96.61046  89.90138  91.18451  89.10353  87.64335  89.11508  95.66114 106.90414 107.33023 113.37001
2020 105.18485  97.68750  77.95396  44.20794  75.98773  96.44583 102.01063 100.41817 109.47033 118.00079 118.37337 116.13132
2021 104.11460  91.99985  89.60123  84.58705  94.50620  94.81354  86.37121  93.96004 102.17477 107.55287 104.06301 111.40675
2022  94.16480  89.46395 101.07021  95.75946
```

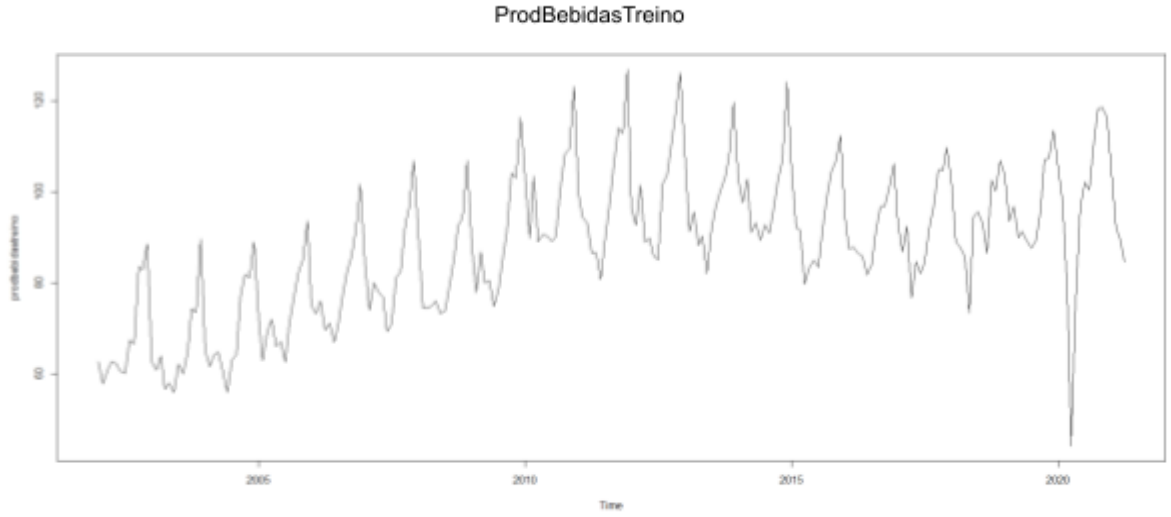
Tamanho da amostra

```
> length(prodbebidas)
[1] 244
```

Separando testes e treinos

```
> prodbebidastreino=window(prodbebidas,start=c(2002,1),
+                          end=c(2021,4))
> plot(prodbebidastreino)
> length(prodbebidastreino)
[1] 232
>
> prodbebidasteste=window(prodbebidas,start=c(2021,5),
+                         end=c(2022,4))
> plot(prodbebidasteste)
> length(prodbebidasteste)
[1] 12
```



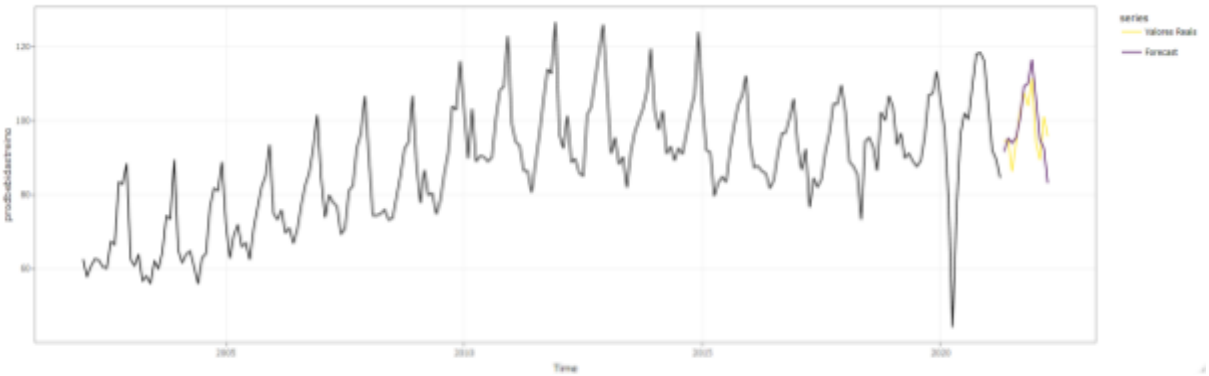


Estimando modelo ETS

```
> prodbebidastreino.ets <- ets(prodbebidastreino)
> summary(prodbebidastreino.ets)
ETS(A,N,A)

Call:
 ets(y = prodbebidastreino)

  Smoothing parameters:
    alpha = 0.5324
    gamma = 0.1855

  Initial states:
    l = 68.3188
    s = 22.2949 11.6767 10.4517 0.0984 -2.9925 -7.4788
           -10.2453 -6.8968 -6.2432 -3.005 -6.982 -0.678

  sigma:  5.6518

     AIC     AICc      BIC
2082.839 2085.061 2134.540

Training set error measures:
                    ME     RMSE      MAE          MPE     MAPE      MASE      ACF1
Training set 0.2423636 5.478619 3.752648 -0.02880651 4.494228 0.6607421 0.1047277
```

Prevendo para 12 meses

```
> prodbebidas.ets.forecasts <- forecast.ets(prodbebidastreino.ets,
+                                           h = 12)
> summary(prodbebidas.ets.forecasts)

Forecast method: ETS(A,N,A)

Model Information:
ETS(A,N,A)

Call:
 ets(y = prodbebidastreino)

  Smoothing parameters:
    alpha = 0.5324
    gamma = 0.1855

  Initial states:
    l = 68.3188
    s = 22.2949 11.6767 10.4517 0.0984 -2.9925 -7.4788
           -10.2453 -6.8968 -6.2432 -3.005 -6.982 -0.678

  sigma:  5.6518

     AIC     AICc      BIC
2082.839 2085.061 2134.540

Error measures:
                    ME     RMSE      MAE         MPE     MAPE      MASE      ACF1
Training set 0.2423636 5.478619 3.752648 -0.02880651 4.494228 0.6607421 0.1047277

Forecasts:
         Point Forecast     Lo 80     Hi 80    Lo 95    Hi 95
May 2021       91.68192  84.43884  98.92499 80.60459 102.7592
Jun 2021       95.22489  87.01919 103.43060 82.67535 107.7744
Jul 2021       94.00474  84.93804 103.07144 80.13842 107.8711
Aug 2021       95.38278  85.53005 105.23552 80.31432 110.4512
Sep 2021      100.58180  90.00126 111.16235 84.40026 116.7633
Oct 2021      109.23093  97.96952 120.49234 92.00809 126.4538
Nov 2021      110.04264  98.13924 121.94603 91.83797 128.2473
Dec 2021      116.56363 104.05114 129.07611 97.42744 135.6998
Jan 2022      106.04728  92.95401 119.14055 86.02286 126.0717
Feb 2022       94.99386  81.34450 108.64322 74.11896 115.8688
Mar 2022       92.56915  78.38548 106.75283 70.87710 114.2612
Apr 2022       83.17693  68.07010  98.28376 60.07303 106.2808
```

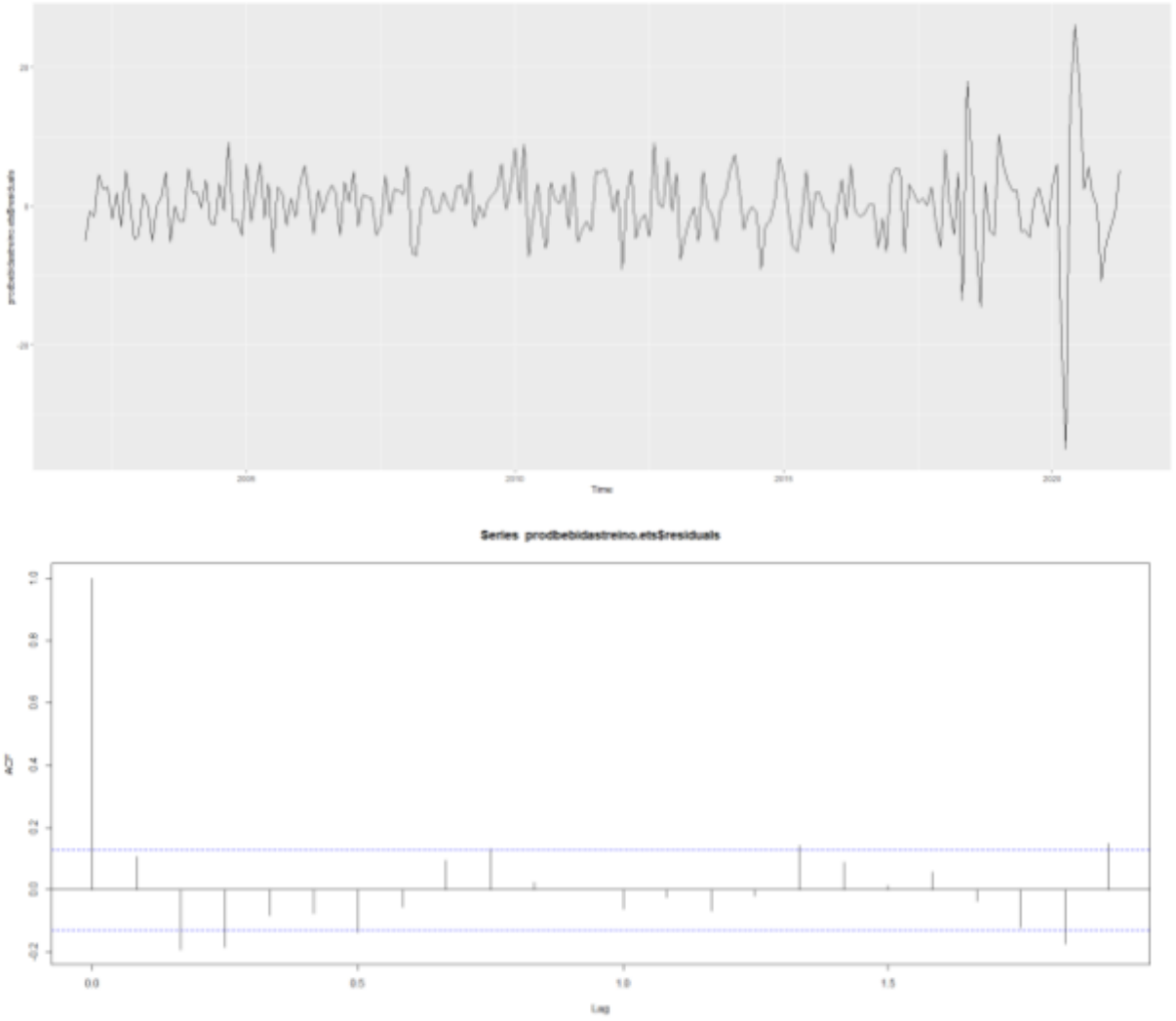


Previsões

Acurácia

```
> accuracy(prodbebidas.ets.forecasts$mean,prodbebidasteste)
                ME     RMSE      MAE       MPE     MAPE      ACF1 Theil's U
Test set -1.182813 6.69891 5.432951 -1.285254 5.634968 0.3820066 0.8175471
```

Resíduo



Autocorrelação da estimativa

Analisando os resíduos (erros) das previsões:

Não podem ser correlacionados; se forem correlacionados ficaram informações nos resíduos que deveriam estar no modelo, devem possuir média zero, caso não possua então as previsões são enviesadas

Teste de Ljung-Box

```
> Box.test(prodbebidastreino.ets$residuals, lag=1,
+          type=c("Ljung-Box"))

        Box-Ljung test

data:  prodbebidastreino.ets$residuals
X-squared = 2.5776, df = 1, p-value = 0.1084
```

Teste de Ljung-box
H0: os resíduos são iid (modelo não exibe falhas de ajustes)
HA: os resíduos não são iid (modelo exibe falhas de ajustes)

Não desejamos rejeitar H0 (desejo um p-value>0.05)

```
> library(seastests)
>
> combined_test(prodbebidastreino)
Test used:  WO

Test statistic:  1
P-value:  0 0 0
```

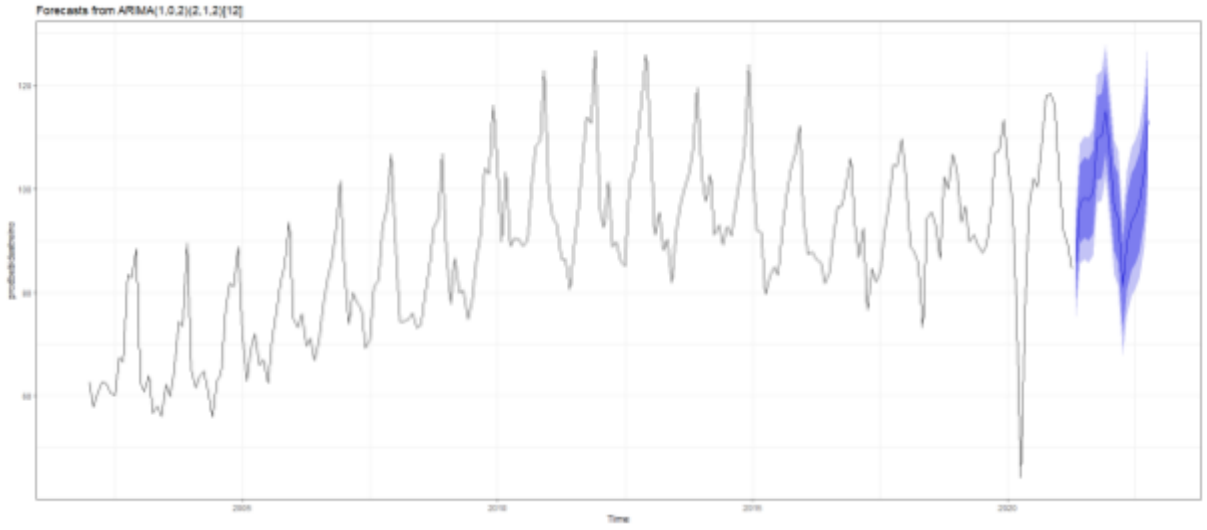
p-value é = 0 (ou seja, <0.5) então a série tem sazonalidade, então podemos utilizar o SARIMA, caso contrário teríamos de usar o ARIMA

Testando com ARIMA

```
> arimaprodbebidastreino=auto.arima(prodbebidastreino, trace=T

 Fitting models using approximations to speed things up...

 ARIMA(2,0,2)(1,1,1)[12] with drift         : 1333.728
 ARIMA(0,0,0)(0,1,0)[12] with drift         : 1469.595
 ARIMA(1,0,0)(1,1,0)[12] with drift         : 1361.981
 ARIMA(0,0,1)(0,1,1)[12] with drift         : 1378.039
 ARIMA(0,0,0)(0,1,0)[12]                    : 1479.418
 ARIMA(2,0,2)(0,1,1)[12] with drift         : 1332.551
 ARIMA(2,0,2)(0,1,0)[12] with drift         : 1418.309
 ARIMA(2,0,2)(0,1,2)[12] with drift         : 1332.972
 ARIMA(2,0,2)(1,1,0)[12] with drift         : 1359.787
 ARIMA(2,0,2)(1,1,2)[12] with drift         : 1335.388
 ARIMA(1,0,2)(0,1,1)[12] with drift         : 1332.054
 ARIMA(1,0,2)(0,1,0)[12] with drift         : 1415.699
 ARIMA(1,0,2)(1,1,1)[12] with drift         : 1331.24
 ARIMA(1,0,2)(1,1,0)[12] with drift         : 1359.318
 ARIMA(1,0,2)(2,1,1)[12] with drift         : 1331.14
 ARIMA(1,0,2)(2,1,0)[12] with drift         : 1333.754
 ARIMA(1,0,2)(2,1,2)[12] with drift         : 1329.806
 ARIMA(1,0,2)(1,1,2)[12] with drift         : 1333.162
 ARIMA(0,0,2)(2,1,2)[12] with drift         : 1367.643
 ARIMA(1,0,1)(2,1,2)[12] with drift         : 1348.304
 ARIMA(2,0,2)(2,1,2)[12] with drift         : 1332.38
 ARIMA(1,0,3)(2,1,2)[12] with drift         : 1330.356
 ARIMA(0,0,1)(2,1,2)[12] with drift         : 1378.058
 ARIMA(0,0,3)(2,1,2)[12] with drift         : Inf
 ARIMA(2,0,1)(2,1,2)[12] with drift         : 1336.933
 ARIMA(2,0,3)(2,1,2)[12] with drift         : 1332.79
 ARIMA(1,0,2)(2,1,2)[12]                    : 1327.801
 ARIMA(1,0,2)(1,1,2)[12]                    : Inf
 ARIMA(1,0,2)(2,1,1)[12]                    : 1329.164
 ARIMA(1,0,2)(1,1,1)[12]                    : 1329.363
 ARIMA(0,0,2)(2,1,2)[12]                    : 1378.698
 ARIMA(1,0,1)(2,1,2)[12]                    : 1353.626
 ARIMA(2,0,2)(2,1,2)[12]                    : 1330.579
 ARIMA(1,0,3)(2,1,2)[12]                    : 1328.304
 ARIMA(0,0,1)(2,1,2)[12]                    : 1391.541
 ARIMA(0,0,3)(2,1,2)[12]                    : 1372.243
 ARIMA(2,0,1)(2,1,2)[12]                    : 1335.41
 ARIMA(2,0,3)(2,1,2)[12]                    : 1330.864

 Now re-fitting the best model(s) without approximations...

 ARIMA(1,0,2)(2,1,2)[12]                    : 1375.786

 Best model: ARIMA(1,0,2)(2,1,2)[12]

>
```

```
> checkresiduals(arimaprodbebidastreino)

        Ljung-Box test

data:  Residuals from ARIMA(1,0,2)(2,1,2)[12]
Q* = 34.597, df = 17, p-value = 0.007024

Model df: 7.   Total lags used: 24

>
> ks.test(arimaprodbebidastreino$residuals, "pnorm", mean(arimaprodbebidastreino$residuals),
+         sd(arimaprodbebidastreino$residuals))

        Asymptotic one-sample Kolmogorov-Smirnov test

data:  arimaprodbebidastreino$residuals
D = 0.067978, p-value = 0.234
alternative hypothesis: two-sided

> ArchTest(arimaprodbebidastreino$residuals)

        ARCH LM-test; Null hypothesis: no ARCH effects

data:  arimaprodbebidastreino$residuals
Chi-squared = 31.514, df = 12, p-value = 0.001644


> prevprodbebidas=forecast::forecast(arimaprodbebidastreino, h=18)
>
> prevprodbebidas
         Point Forecast     Lo 80     Hi 80     Lo 95     Hi 95
May 2021       85.79982  79.06923  92.53040  75.50627  96.09336
Jun 2021       97.11409  89.48287 104.74532  85.44314 108.78505
Jul 2021       98.38603  90.63347 106.13859  86.52952 110.24255
Aug 2021       97.83107  89.96183 105.70031  85.79611 109.86603
Sep 2021       99.41268  91.43116 107.39421  87.20600 111.61937
Oct 2021      109.70548 101.61581 117.79515  97.33340 122.07756
Nov 2021      110.27721 102.08332 118.47110  97.74573 122.80869
Dec 2021      115.19485 106.90045 123.48924 102.50967 127.88003
Jan 2022      107.03256  98.64122 115.42391  94.19911 119.86602
Feb 2022       96.43855  87.95356 104.92353  83.46188 109.41521
Mar 2022       94.29273  85.71730 102.86816  81.17774 107.40772
Apr 2022       81.20666  72.54382  89.86951  67.95799  94.45534
May 2022       89.55655  80.45708  98.65602  75.64011 103.47299
Jun 2022       93.63624  84.31121 102.96127  79.37483 107.89764
Jul 2022       95.10998  85.66451 104.55545  80.66437 109.55558
Aug 2022       97.41385  87.85224 106.97547  82.79062 112.03709
Sep 2022      103.25094  93.57725 112.92463  88.45631 118.04557
Oct 2022      113.18924 103.40736 122.97113  98.22914 128.14935
>
> autoplot(prevprodbebidas) +
+     theme_bw()
```

Previsão



Acurácia