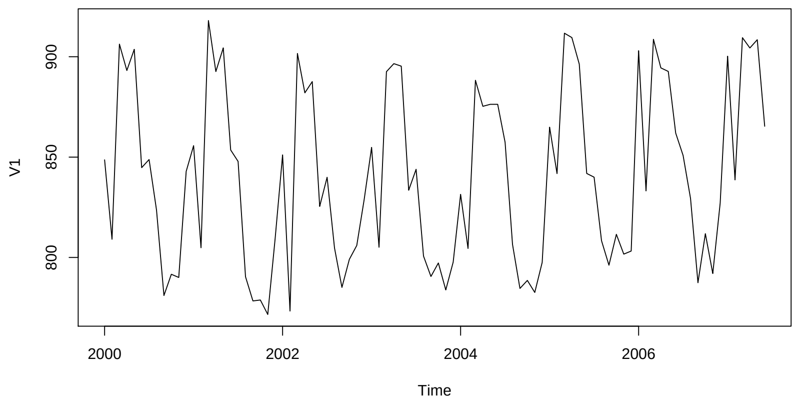
Diciembre-2021

1. Justifica si se puede asumir homogeneidad en la varianza de los datos, y en caso negativo explica si se podría aplicar alguna transformación para conseguirla.

```
library(TSA); library(tseries); library(forecast)
y=read.table("qlecheItalia.txt")
y=ts(y, frequency=12, start=2000)
plot(y)
```



Del gráfico de la serie parece que la varianza es constante

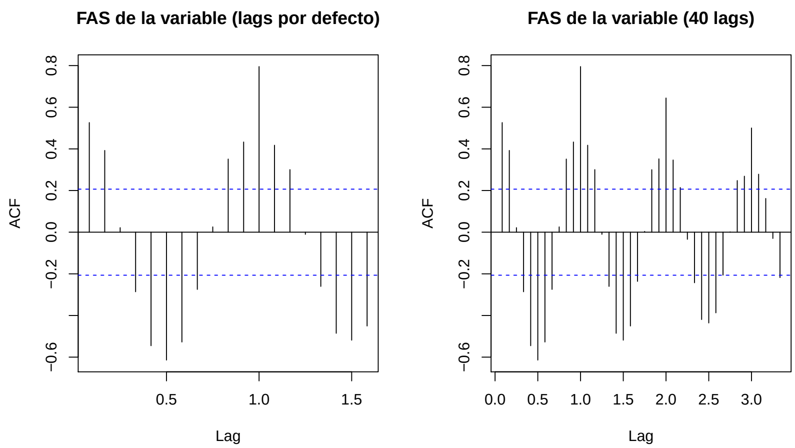
```
bc=BoxCox.ar(y); bc
#----------------------------------------------------------------------
$lambda
 [1] -2.0 -1.9 -1.8 -1.7 -1.6 -1.5 -1.4 -1.3 -1.2 -1.1 -1.0 -0.9 -0.8 -0.7 -0.6 -0.5 -0.4
[18] -0.3 -0.2 -0.1  0.0  0.1  0.2  0.3  0.4  0.5  0.6  0.7  0.8  0.9  1.0  1.1  1.2  1.3
[35]  1.4  1.5  1.6  1.7  1.8  1.9  2.0

$loglike
 [1] 360.4485 360.5736 360.6950 360.8126 360.9264 361.0364 361.1425 361.2446 361.3435
[10] 361.4379 361.5289 361.6160 361.6992 361.7788 361.8533 361.9249 361.9922 362.0556
[19] 362.1151 362.1706 362.2218 362.2693 362.3143 362.3526 362.3893 362.4219 362.4493
[28] 362.4737 362.4937 362.5094 362.5221 362.5307 362.5348 362.5356 362.5335 362.5242
[37] 362.5138 362.4360 362.4780 362.4571 362.4301

$mle
[1] 1.3

$ci
[1] -1.8  2.0
#-----------------------------------------------------
```

Como $1 \in$ `ci`, no transformamos.

2. Justifica si se puede asumir homogeneidad en la media de los datos, o bien es necesasio alguna(s) diferenciación(es) para conseguirla, indicando las diferencias y el orden de las mismas

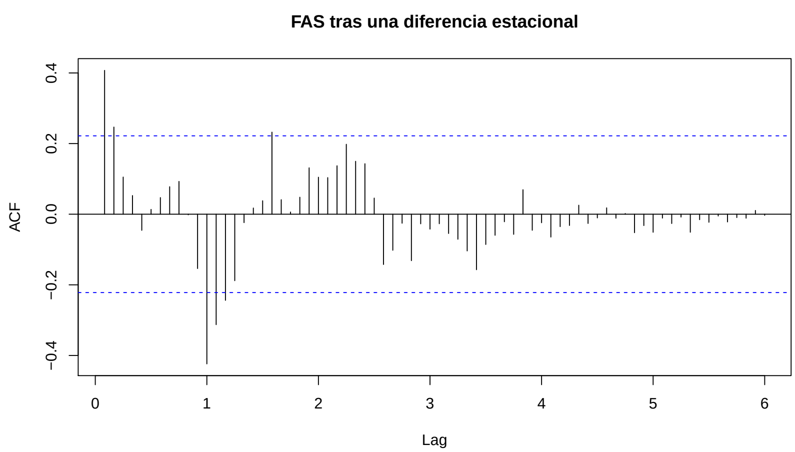
Del gráfico parece que la serie tiene una fuerte componente estacional. Representamos la FAS

```
layout(matrix(1:2, ncol=2))
acf(y, main="FAS de la variable (lags por defecto)")
acf(y, main="FAS de la variable (40 lags)", lag.max=40)
layout(matrix(1, ncol=1))
```



Las autocorrelaciones estacionales desciende muy lentamente Aplicamos una diferencia estacional y representamos la FAS

```
y1=diff(y,lag=12, diff=1)
acf(y1, main="FAS tras una  diferencia estacional", lag.max=72)
```



Tiene el aspecto de la FAS de un proceso estacionario.

3. Aplica algún contraste adecuado para concluir si la serie transformada (o no transformada, en caso que no sea necesario) puede considerarse estacionaria
Aplicamos test de la raíz unitaria

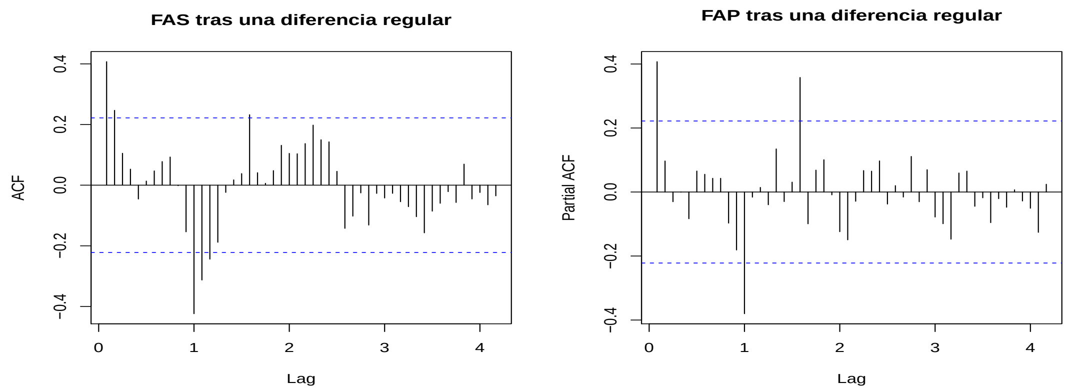
```
library(tseries)
adf.test(y1)
#----------------------------------------------------------
        Augmented Dickey-Fuller Test

data:  y1
Dickey-Fuller = -3.8367, Lag order = 4, p-value = 0.0216
alternative hypothesis: stationary
#----------------------------------------------------
```

Se rechaza la hipótesis de que el polinomio autorregresivo contenga una raíz unitaria, concluimos que la serie es estacionaria

4. Identifica razonadamente la estructura ARIMA de los datos, detallando y justificando los pasos efectuados (órdenes considerados, decisiones tomadas, etc). Escribe el modelo seleccionado

Representamos las autocorrelaciones simples (FAS) y parciales (FAP)

```
layout(matrix(1:2, ncol=2))
acf(y2, main="FAS tras una diferencia  estacional")
pacf(y2, main="FAP tras una diferencia estacional")
layout(matrix(1, ncol=1))
```



Modelo para la parte estacional: AR(1) o MA(1)
Modelo para la parte regular: AR(1)

```
#-------------------------------------------------------------------------
ajuste1=arima(y, order=c(1,0,0), seasonal=list(order=c(1,1,0), period=12)); ajuste1
#----------------------------------------------------------------
Call:
arima(x = y, order = c(1, 0, 0), seasonal = list(order = c(1, 1, 0), period = 12))

Coefficients:
         ar1     sar1
      0.3696  -0.3891
s.e.  0.1050   0.1048

sigma^2 estimated as 225.3:  log likelihood = -323.02,  aic = 650.04
#----------------------------------------------------------------
checkresiduals(ajuste1)
#-----------------------------------------------------------------
        Ljung-Box test

data:  Residuals from ARIMA(1,0,0)(1,1,0)[12]
Q* = 12.499, df = 16, p-value = 0.709

Model df: 2.   Total lags used: 18
#---------------------------------------------
# modelo  v\'alido
#---------------------------------------------
ajuste2=arima(y, order=c(1,0,0), seasonal=list(order=c(0,1,1), period=12)); ajuste2
#----------------------------------------------------------------
Call:
arima(x = y, order = c(1, 0, 0), seasonal = list(order = c(0, 1, 1), period = 12))

Coefficients:
         ar1     sma1
      0.3902  -0.4620
s.e.  0.1034   0.1201

sigma^2 estimated as 218.5:  log likelihood = -322.28,  aic = 648.57
#----------------------------------------------------------------
checkresiduals(ajuste2)
#-----------------------------------------------------------------
data:  Residuals from ARIMA(1,0,0)(0,1,1)[12]
Q* = 11.145, df = 16, p-value = 0.8005

Model df: 2.   Total lags used: 18
#---------------------------------------------
# modelo  v\'alido, y con menor AIC que el ajuste 1.
#---------------------------------------------
```

Veamos qué hace auto.arima:

```
(u=auto.arima(y, d=0, D=1, step=FALSE))
#----------------------------------------------
Series: y
ARIMA(1,0,0)(0,1,1)[12]

Coefficients:
         ar1     sma1
      0.3902  -0.4620
s.e.  0.1034   0.1201

sigma^2 estimated as 224.3:  log likelihood=-322.28
AIC=650.57   AICc=650.89   BIC=657.64
```

Coincide con ajuste2. Nos quedamos con ajuste2 porque tiene menor AIC. Los coefientes ar de la parte estacional y regular son significativamente no nulos

5. Con el modelo ajustado, escribe la predicción, así como un intervalo de confianza al 90% para los 24 meses siguientes.

```
pred=predict(ajuste2,n.ahead=24)
lo=pred$pred-qnorm(0.95)*pred$se; up=pred$pred+qnorm(0.95)*pred$se
cbind(pred$pred, lo,up)
#----------------------------------------------------
         pred$pred       lo       up
Jul 2007  851.8496 827.5363 876.1628
Aug 2007  820.2941 794.1952 846.3930
Sep 2007  789.7632 763.4030 816.1234
Oct 2007  807.6334 781.2336 834.0332
Nov 2007  793.0523 766.6464 819.4581
Dec 2007  816.2161 789.8094 842.6229
Jan 2008  891.1131 864.7064 917.5198
Feb 2008  833.9736 807.5669 860.3803
Mar 2008  907.9522 881.5455 934.3589
Apr 2008  900.3960 873.9892 926.8027
May 2008  900.8333 874.4265 927.2400
Jun 2008  861.0124 834.6056 887.4191
Jul 2008  850.1530 820.6838 879.6221
Aug 2008  819.6321 789.7240 849.5401
Sep 2008  789.5048 759.5306 819.4791
Oct 2008  807.5326 777.5482 837.5169
Nov 2008  793.0129 763.0270 822.9988
Dec 2008  816.2008 786.2147 846.1869
Jan 2009  891.1071 861.1211 921.0931
Feb 2009  833.9712 803.9852 863.9573
Mar 2009  907.9513 877.9653 937.9373
Apr 2009  900.3956 870.4096 930.3816
May 2009  900.8331 870.8471 930.8192
Jun 2009  861.0123 831.0263 890.9983
```

6. Calcula la predicción de la serie para los 24 meses siguientes suponiendo que los datos se pueden describir mediante un modelo aditivo para la tendencia y la componente estacional, utilizando un método basado en el uso de medias móviles.

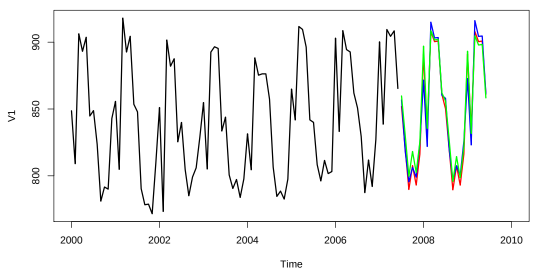
```
dec=decompose(y, type="additive")
y_des=y-dec$seasonal #serie desestacionalizada
#--------------------------------------------
# ajustamos una recta a la serie desestacionalizada
#----------------------------------------
Time=time(y_des)
tendencia=lm(y_des~Time)
summary(tendencia)
#------------------------
Call:
lm(formula = y_des ~ Time)

Residuals:
    Min      1Q  Median      3Q     Max
-35.542  -7.033  -0.443   6.736  35.330

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -3787.6072  1290.9645  -2.934 0.004267 **
Time            2.3089     0.6443   3.584 0.000555 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 13.23 on 88 degrees of freedom
Multiple R-squared:  0.1274,    Adjusted R-squared:  0.1174
F-statistic: 12.84 on 1 and 88 DF,  p-value: 0.0005551
#-----------------------------------------------------
str(Time)
#-----------------------------------------------------
# Time-Series [1:90]
#-----------------------------------------------------
cc=1:24/24; cc=Time[90]+cc; cc=data.frame(Time=cc)
pred_tend=predict(tendencia, cc); pred_tend
#-------------------------------------------------------
#        1        2        3        4        5        6        7        8        9       10       11       12
# 847.5006 847.5968 847.6931 847.7893 847.8855 847.9817 848.0779 848.1741 848.2703 848.3665 848.4627 848.5589
#       13       14       15       16       17       18       19       20       21       22       23       24
# 848.6551 848.7513 848.8475 848.9437 849.0399 849.1361 849.2324 849.3286 849.4248 849.5210 849.6172 849.7134
#-------------------------------------------------------
 dec$figure
#-------------------------------------------------------
# [1]  23.536221 -26.217876  66.717888  54.806777  55.001152  11.766429   9.289515 -29.042450 -52.170664
#[10] -41.470842 -48.843223 -23.372926
#-------------------------------------------------------
season=c(dec$figure[7:12], dec$figure, dec$figure[1:6])
(p2=ts(pred_tend+season, freq=12, start=c(2007,7)))
#-------------------------------------------------------
#           Jan      Feb      Mar      Apr      May      Jun      Jul      Aug      Sep      Oct      Nov      Dec
# 2007                                                       856.7902 818.5544 795.5224 806.3184 799.0422 824.6087
# 2008 871.6141 821.9562 914.9882 903.1733 903.4639 860.3253 857.9446 819.7089 796.6769 807.4729 800.1967 825.7632
# 2009 872.7686 823.1107 916.1427 904.3277 904.6183 861.4798
#-------------------------------------------------------
```

7. Calcula la predicción de la serie para los 24 meses siguientes suponiendo que los datos se pueden describir mediante un modelo aditivo para la tendencia y la componente estacional, utilizando un método basado en el uso de suavización exponencial.

```
y_suav=HoltWinters(y)
(p3=predict(y_suav, n.ahead=24))
#-----------------------------------------------------
#          Jan      Feb      Mar      Apr      May      Jun      Jul      Aug      Sep      Oct      Nov      Dec
# 2007                                                      859.7589 830.4215 799.6630 818.2278 802.4865 825.3202
# 2008 897.0332 835.5079 908.6137 901.6799 902.0980 862.1310 855.9459 826.6085 795.8501 814.4149 798.6735 821.5073
# 2009 893.2202 831.6949 904.8008 897.8670 898.2851 858.3180
#-----------------------------------------------------
```

8. Representa la serie junto con las tres predicciones realizadas para los 24 meses siguientes y comenta.

```
plot(y, xlim=c(2000, 2010), lwd=2)
p1=pred$pred
lines(p1, col="red", lwd=2) #graficoIT5.pdf
lines(p2,  col="blue", lwd=2)
lines(p3,  col="green", lwd=2)
```



Comentario: prácticamente coinciden