PracticaSeriesTemporales

AlbertoMarino

11 de diciembre de 2016

## Analisis de la produccion de leche

Analizaremos la produccion de leche de un fichero "monthly-milk-production-pounds-p.csv" el cual reside en el mismo directorio del .R que estamos ejecutando.

Es necesario transformar el dataset, ya que los formatos de las fechas no son optimos para tratar con series temporales

setwd("C:\\Desarrollo\\repos\\Master-BigData-CIFF-2016\\analisis estadístico\\Practica2")  
  
data = read.csv('monthly-milk-production-pounds-p.csv',header = TRUE, skip=0, sep = ";")  
class(data)

## [1] "data.frame"

data$Year <- format(as.Date(paste0(as.character(data[,1]),"-01"),format = "%Y-%m-%d"),"%Y")  
data$Month <- format(as.Date(paste0(as.character(data[,1]),"-01"),format = "%Y-%m-%d"),"%b")  
colnames(data) <- c("month","milk","year")  
data <- na.omit(data)  
head(data)

## month milk year  
## 1 ene. 589 1962  
## 2 feb. 561 1962  
## 3 mar. 640 1962  
## 4 abr. 656 1962  
## 5 may. 727 1962  
## 6 jun. 697 1962

min(data$year)

## [1] "1962"

max(data$year)

## [1] "1975"

milk\_ts=ts(data$milk,start=1962,frequency=12)  
milk\_ts

## Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec  
## 1962 589 561 640 656 727 697 640 599 568 577 553 582  
## 1963 600 566 653 673 742 716 660 617 583 587 565 598  
## 1964 628 618 688 705 770 736 678 639 604 611 594 634  
## 1965 658 622 709 722 782 756 702 653 615 621 602 635  
## 1966 677 635 736 755 811 798 735 697 661 667 645 688  
## 1967 713 667 762 784 837 817 767 722 681 687 660 698  
## 1968 717 696 775 796 858 826 783 740 701 706 677 711  
## 1969 734 690 785 805 871 845 801 764 725 723 690 734  
## 1970 750 707 807 824 886 859 819 783 740 747 711 751  
## 1971 804 756 860 878 942 913 869 834 790 800 763 800  
## 1972 826 799 890 900 961 935 894 855 809 810 766 805  
## 1973 821 773 883 898 957 924 881 837 784 791 760 802  
## 1974 828 778 889 902 969 947 908 867 815 812 773 813  
## 1975 834 782 892 903 966 937 896 858 817 827 797 843

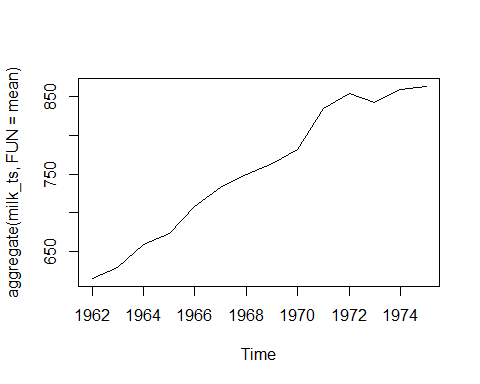
Ploteamos y vemos que es una serie creciente en el tiempo, agregando las medias

cycle(milk\_ts)

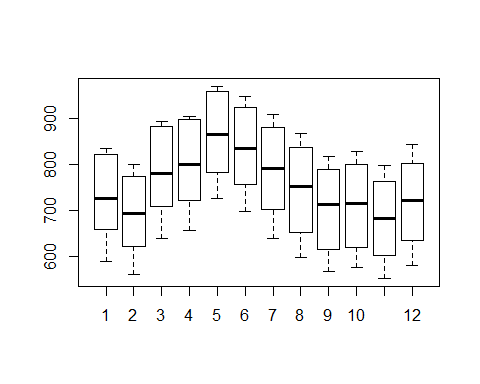
## Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec  
## 1962 1 2 3 4 5 6 7 8 9 10 11 12  
## 1963 1 2 3 4 5 6 7 8 9 10 11 12  
## 1964 1 2 3 4 5 6 7 8 9 10 11 12  
## 1965 1 2 3 4 5 6 7 8 9 10 11 12  
## 1966 1 2 3 4 5 6 7 8 9 10 11 12  
## 1967 1 2 3 4 5 6 7 8 9 10 11 12  
## 1968 1 2 3 4 5 6 7 8 9 10 11 12  
## 1969 1 2 3 4 5 6 7 8 9 10 11 12  
## 1970 1 2 3 4 5 6 7 8 9 10 11 12  
## 1971 1 2 3 4 5 6 7 8 9 10 11 12  
## 1972 1 2 3 4 5 6 7 8 9 10 11 12  
## 1973 1 2 3 4 5 6 7 8 9 10 11 12  
## 1974 1 2 3 4 5 6 7 8 9 10 11 12  
## 1975 1 2 3 4 5 6 7 8 9 10 11 12

vemos que tiene cierta estacionalidad en los meses de Mayo y Junio donde son mayores las producciones de leche

plot(aggregate(milk\_ts,FUN = mean))

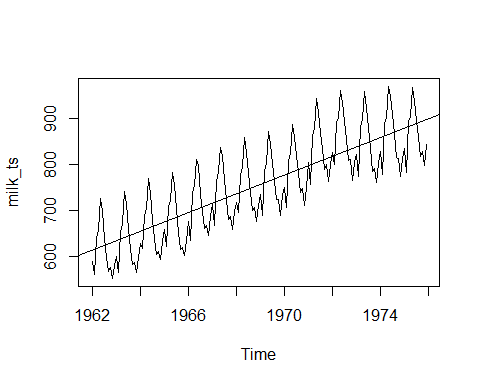


boxplot(milk\_ts~cycle(milk\_ts))



Ploteamos la serie y vemos que tiene tendencia, lo que indica que no es estacionaria y da pistas de que tendremos que hacel al menos una diferenciacion para eliminar la tendencia

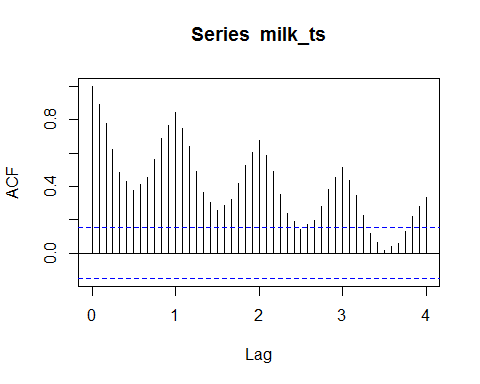
plot(milk\_ts)  
abline(reg=lm(milk\_ts~time(milk\_ts)))



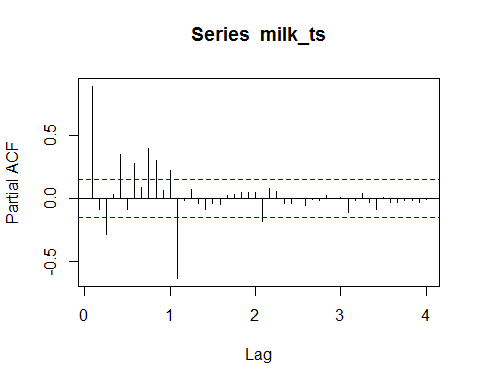
#adf.test(diff(log(milk\_ts)), alternative="stationary", k=0)

Ploteamos 4 años la serie normal

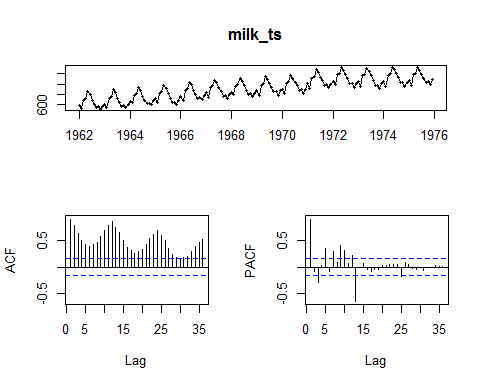
acf(milk\_ts,lag.max=48)



pacf(milk\_ts,lag.max=48)



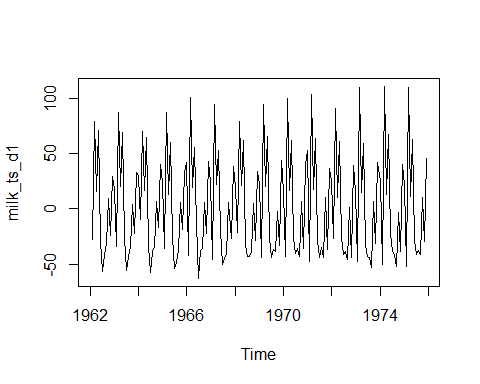
tsdisplay(milk\_ts)



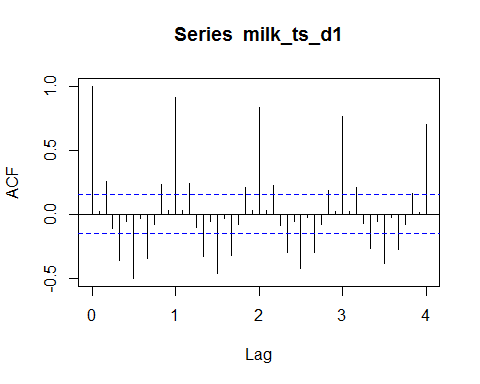
calculamos las primeras diferencias y ploteamos.

Vemos que en la primera diferencia,tanto las correlaciones como las correlaciones parciales están fuera del marge, por lo que siguen siendo no estacionarias

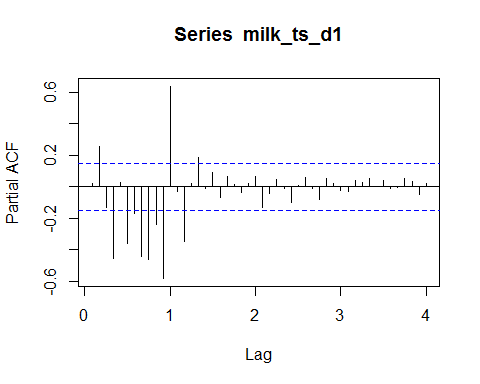
milk\_ts\_d1=diff(milk\_ts,lag=1,differences=1)  
plot(milk\_ts\_d1)



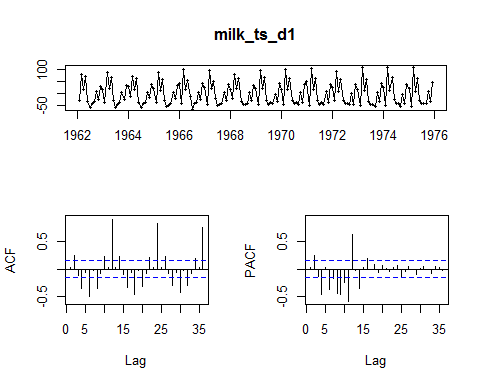
acf(milk\_ts\_d1,lag.max=48)



pacf(milk\_ts\_d1,lag.max=48)

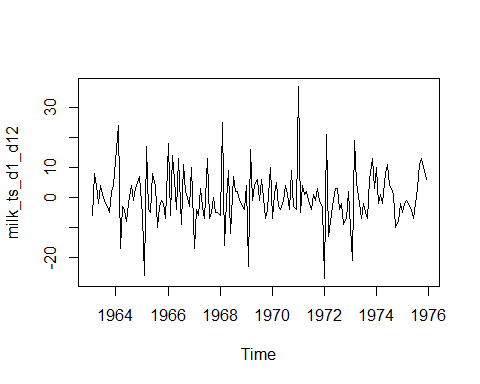


tsdisplay(milk\_ts\_d1)

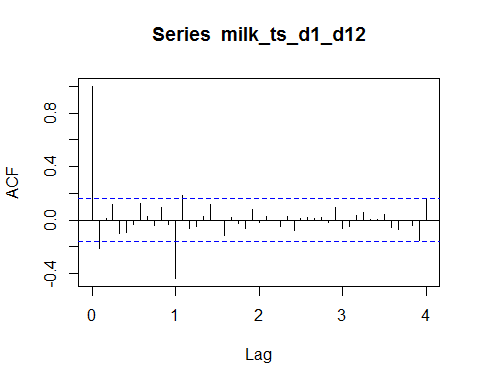


Calculamos las segundas diferencias, y vemos que las autocorrelaciones tienden a cero despues de el primer año y que las correlaciones parciales, todavía no

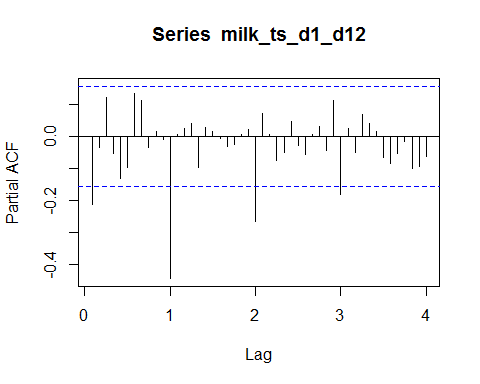
milk\_ts\_d1\_d12=diff(diff(milk\_ts, lag=1,differences=1),lag=12,differences=1)  
plot(milk\_ts\_d1\_d12)



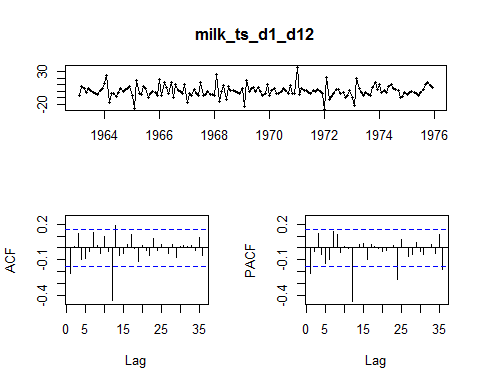
acf(milk\_ts\_d1\_d12,lag.max=48)



pacf(milk\_ts\_d1\_d12,lag.max=48)



tsdisplay(milk\_ts\_d1\_d12)



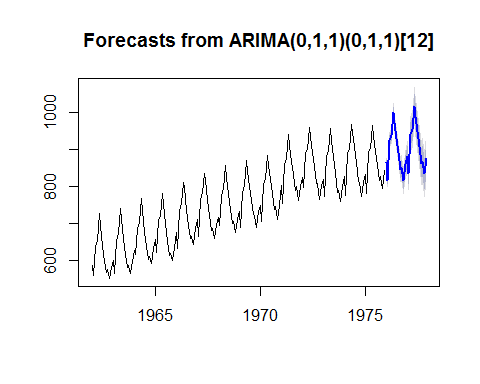
Aplicamos el autoarima para ver que nos indica.

Parece que la serie tiene media movil, y nos da que el mejor modelo es aquel que se realiza con una sola diferencia. la gráficas muestran como es estacionaria

Auto\_milk\_model=auto.arima(milk\_ts,trace=TRUE)

##   
## ARIMA(2,1,2)(1,1,1)[12] : 1019.845  
## ARIMA(0,1,0)(0,1,0)[12] : 1063.122  
## ARIMA(1,1,0)(1,1,0)[12] : 1020.115  
## ARIMA(0,1,1)(0,1,1)[12] : 1007.968  
## ARIMA(0,1,1)(1,1,1)[12] : 1011.709  
## ARIMA(0,1,1)(0,1,0)[12] : 1058.167  
## ARIMA(0,1,1)(0,1,2)[12] : 1009.592  
## ARIMA(0,1,1)(1,1,2)[12] : 1009.233  
## ARIMA(1,1,1)(0,1,1)[12] : 1011.808  
## ARIMA(0,1,0)(0,1,1)[12] : 1013.309  
## ARIMA(0,1,2)(0,1,1)[12] : 1009.573  
## ARIMA(1,1,2)(0,1,1)[12] : 1013.943  
##   
## Best model: ARIMA(0,1,1)(0,1,1)[12]

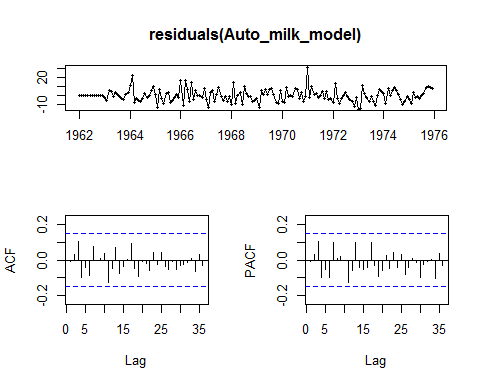
plot(forecast(Auto\_milk\_model,h=24))



summary(Auto\_milk\_model)

## Series: milk\_ts   
## ARIMA(0,1,1)(0,1,1)[12]   
##   
## Coefficients:  
## ma1 sma1  
## -0.2204 -0.6214  
## s.e. 0.0748 0.0627  
##   
## sigma^2 estimated as 53.42: log likelihood=-530.15  
## AIC=1066.3 AICc=1066.46 BIC=1075.43  
##   
## Training set error measures:  
## ME RMSE MAE MPE MAPE MASE  
## Training set 0.0356593 6.975173 5.168896 -0.003681841 0.6897827 0.2428027  
## ACF1  
## Training set -0.00843703

tsdisplay(residuals(Auto\_milk\_model))



Los resultados para los dos años posteriores a la series pueden predecise de la siguiente manera, con intervalos de confianza.

f1 = forecast(Auto\_milk\_model,24)  
f1

## Point Forecast Lo 80 Hi 80 Lo 95 Hi 95  
## Jan 1976 864.9773 855.6103 874.3443 850.6517 879.3029  
## Feb 1976 817.7493 805.8719 829.6267 799.5843 835.9142  
## Mar 1976 924.4056 910.4626 938.3485 903.0817 945.7295  
## Apr 1976 937.4836 921.7439 953.2233 913.4118 961.5554  
## May 1976 1000.6235 983.2721 1017.9749 974.0868 1027.1601  
## Jun 1976 973.2165 954.3909 992.0420 944.4252 1002.0077  
## Jul 1976 931.8501 911.6576 952.0426 900.9684 962.7318  
## Aug 1976 892.2597 870.7873 913.7322 859.4204 925.0991  
## Sep 1976 846.3679 823.6875 869.0483 811.6812 881.0545  
## Oct 1976 851.5326 827.7055 875.3597 815.0921 887.9731  
## Nov 1976 817.4931 792.5719 842.4143 779.3795 855.6068  
## Dec 1976 859.7534 833.7842 885.7225 820.0370 899.4698  
## Jan 1977 882.8150 854.6706 910.9593 839.7719 925.8581  
## Feb 1977 835.5870 805.6961 865.4779 789.8728 881.3012  
## Mar 1977 942.2433 910.7024 973.7842 894.0057 990.4809  
## Apr 1977 955.3213 922.2126 988.4301 904.6859 1005.9568  
## May 1977 1018.4612 983.8555 1053.0668 965.5364 1071.3859  
## Jun 1977 991.0542 955.0138 1027.0946 935.9351 1046.1732  
## Jul 1977 949.6878 912.2676 987.1080 892.4585 1006.9171  
## Aug 1977 910.0975 871.3465 948.8484 850.8330 969.3619  
## Sep 1977 864.2056 824.1682 904.2430 802.9736 925.4375  
## Oct 1977 869.3703 828.0865 910.6541 806.2321 932.5085  
## Nov 1977 835.3308 792.8371 877.8245 770.3423 900.3194  
## Dec 1977 877.5911 833.9210 921.2612 810.8034 944.3787