

# Predicción de Series Temporales.

Nos han facilitado la información de la producción de leche en el periodo Enero 1962 a Diciembre 1975.

Fuente:

[https://datamarket.com/data/set/22ox/monthly-milk-productionpounds-per-cow-jan-62-dec-75#!  
ds=22ox&display=line](https://datamarket.com/data/set/22ox/monthly-milk-productionpounds-per-cow-jan-62-dec-75#!ds=22ox&display=line)

Se pide calcular la predicción de producción de leche por vaca para todos los meses de 1976 y 1977.

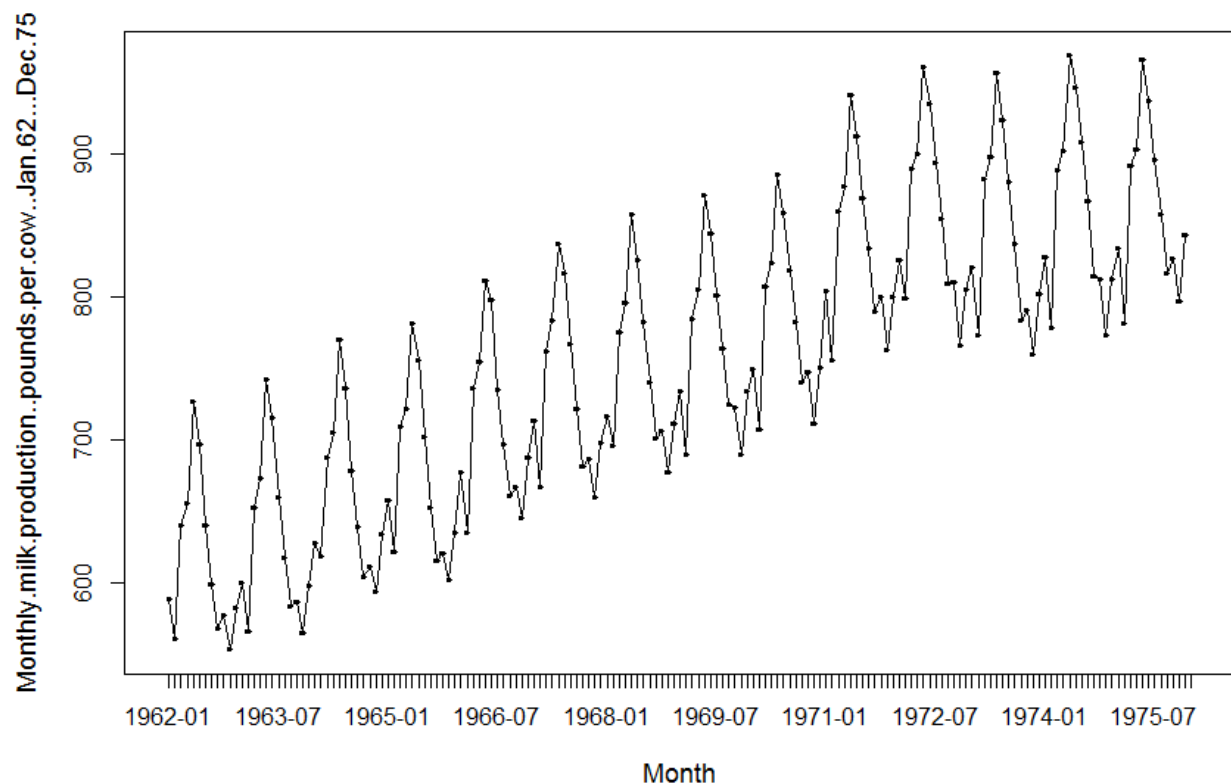
## 1. Carga de datos

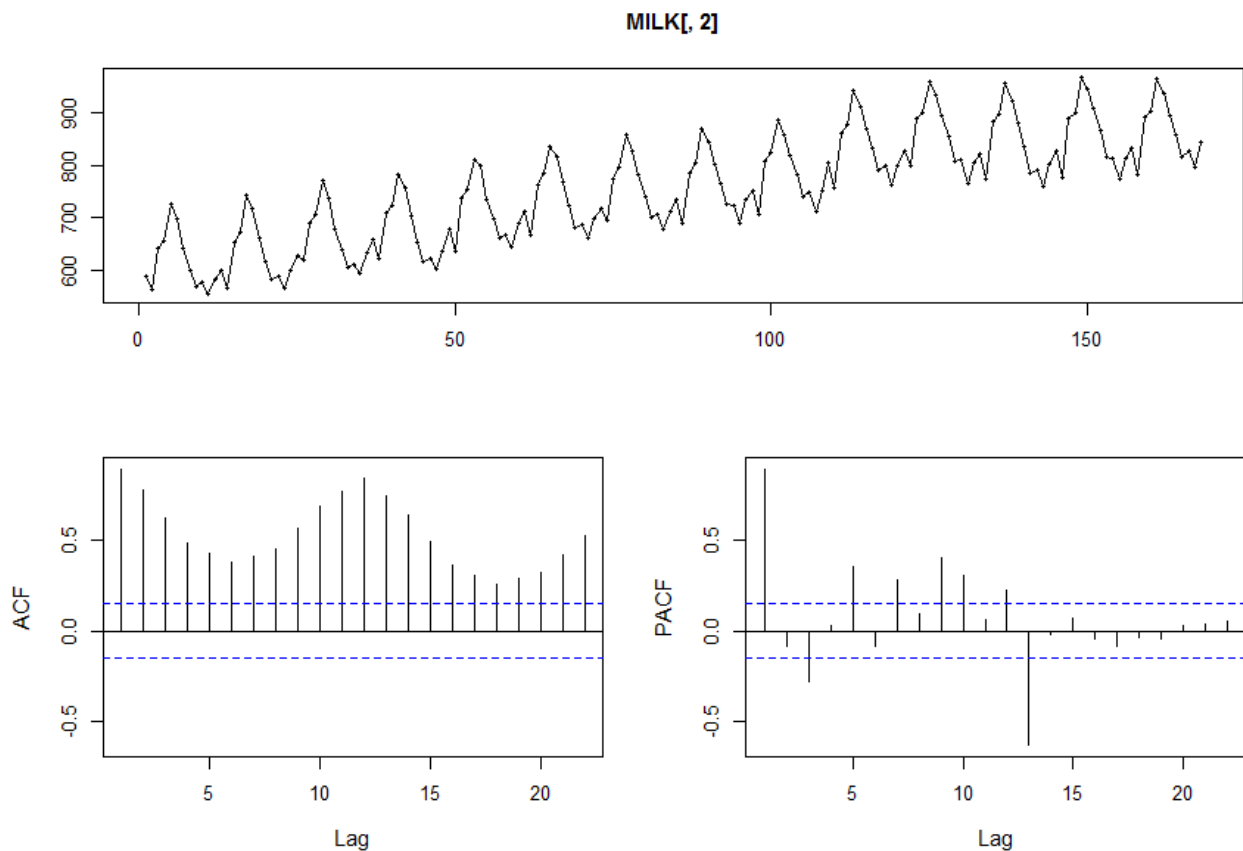
Los datos se descargan de la fuente en formato csv y se cargan a R-studio. En un análisis de calidad se comprueba que los datos están ordenados y se observa un valor nulo que es la última observación. El valor nulo queda eliminado.

```
summary(MILK)
  Month Monthly.milk.production..pounds.per.cow..Jan.62...Dec.75
1962-01: 1 Min. :553.0
1962-02: 1 1st Qu.:677.8
1962-03: 1 Median :761.0
1962-04: 1 Mean :754.7
1962-05: 1 3rd Qu.:824.5
1962-06: 1 Max. :969.0
(other):162
```

## 2. Análisis de datos

En un análisis gráfico de los datos observamos que la serie tiene estacionalidad y tendencia.





Para analizar la serie de datos y seleccionar un modelo utilizaré la metodología Box – Jenkins.

Podemos observar que la serie no es estacionaria. La función de autocorrelación (ACF) que mide la correlación entre variables separadas por  $k$  periodos no decrece hasta cero.

También observamos saltos periódicos, aproximadamente cada año, que indican estacionalidad.

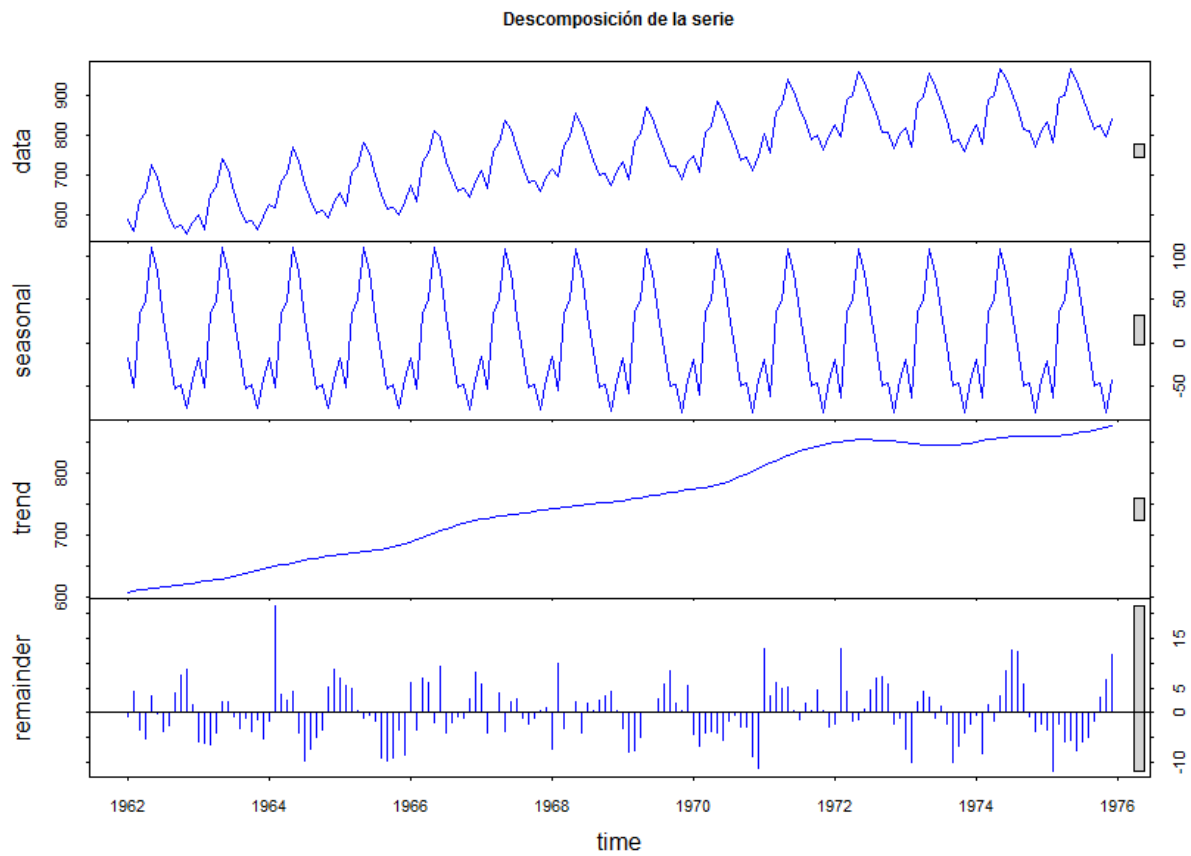
La función de correlación parcial (PACF), que mide la correlación entre dos variables separadas por  $k$  periodos cuando no se considera la dependencia creada por los retardos intermedios existentes entre ambas, muestra varios valores no nulos y tampoco decrece rápidamente.

### 3. Descomposición de los datos

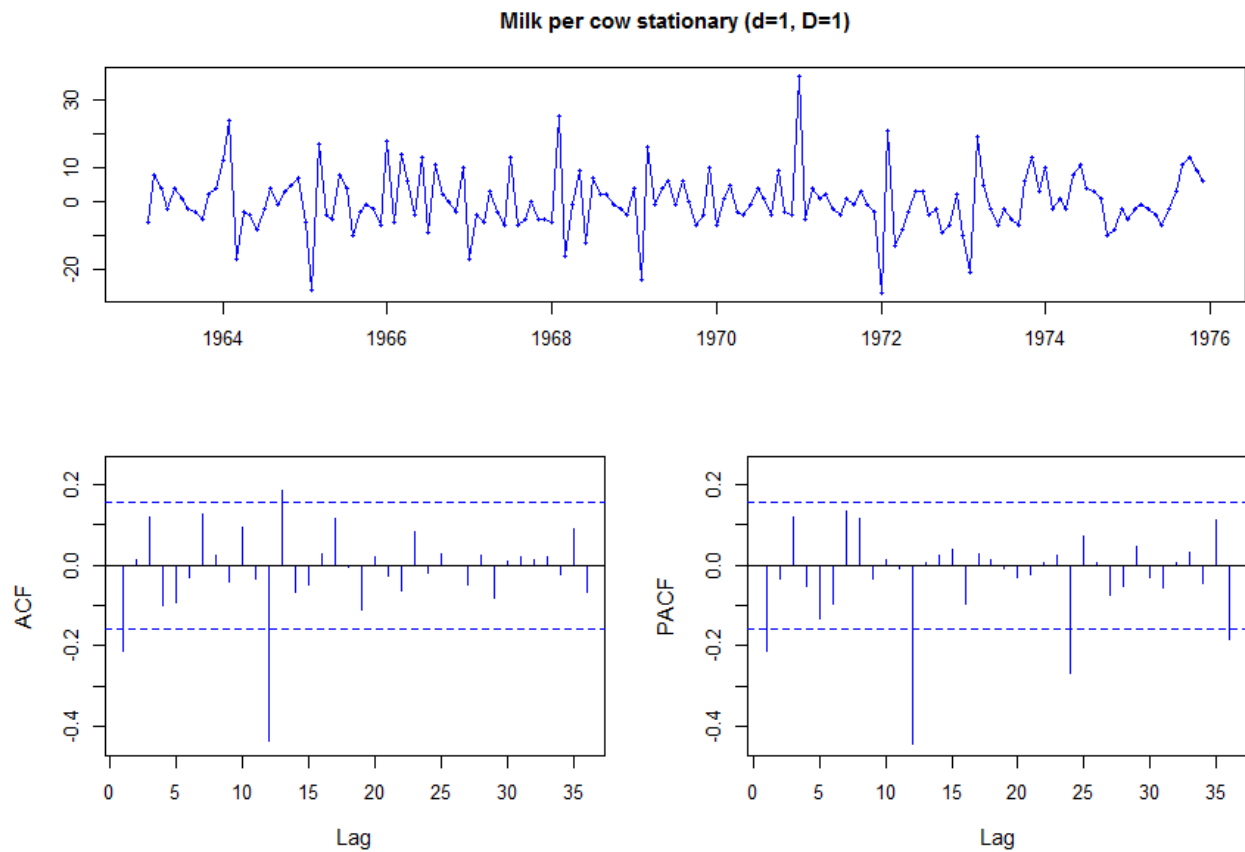
Podemos descomponer los datos en componente estacional, tendencia y el resto.

Observamos claramente el componente estacional bastante visible alrededor de la mitad de cada año y también una tendencia creciente de la cantidad media de leche por vaca obtenido a lo largo de los años.

En el siguiente paso voy a separar los componentes de estacionalidad y tendencia, dejando solamente la parte aleatoria, para poder trabajar con una serie estacionaria.



Después de transformación la serie parece estacionaria.

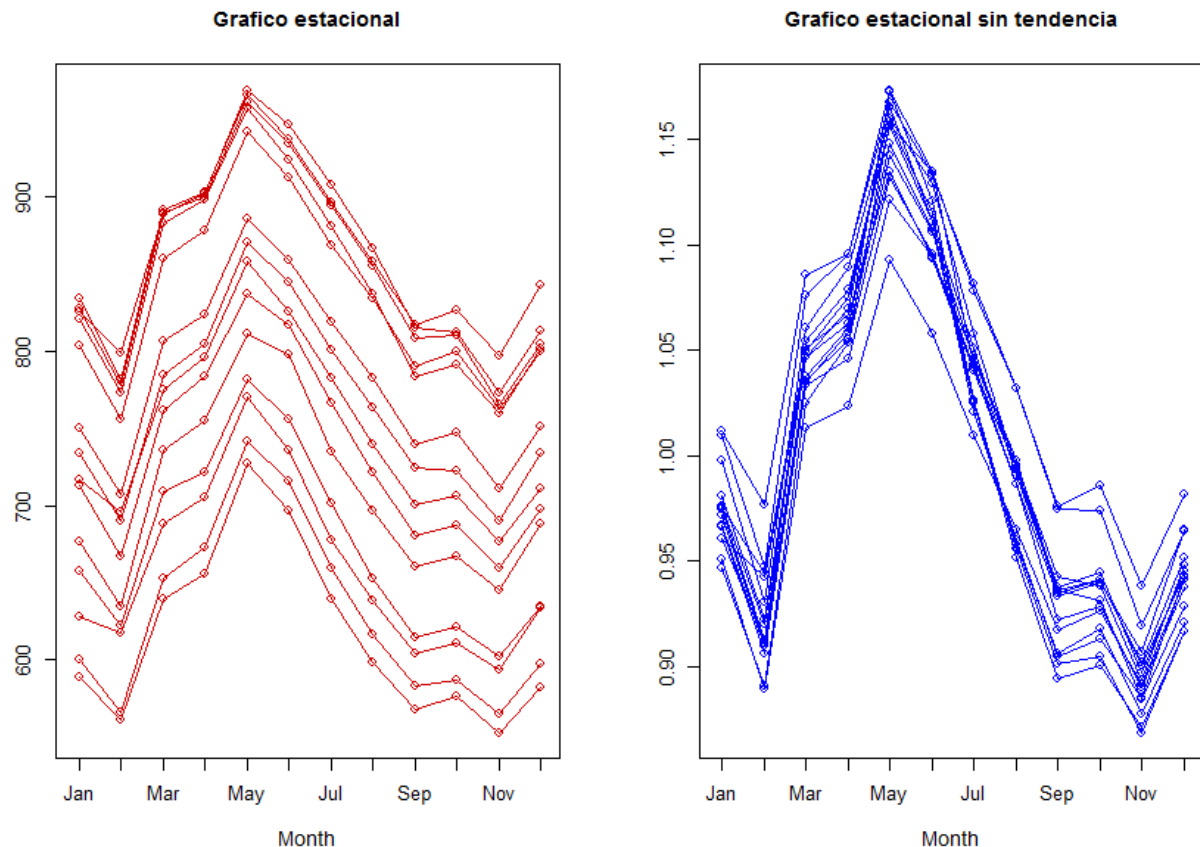


#### 4. Construcción del modelo

En el primer paso he observado que la serie muestra una tendencia creciente, lo cual excluye utilización del modelo MA únicamente. Este argumento apoya también un lento decrecimiento de ACF.

PACF tampoco decrece rápidamente en la serie estacionaria, por lo cual tampoco será suficiente un modelo auto regresivo AR y voy a tener que elegir un modelo ARIMA.

Analizando el componente estacional obtenemos:



Podemos observar un patrón bastante claro: un ascenso de la cantidad de leche obtenida en los meses de verano, sobre todo en mayo y junio.

De los gráficos de la función ACF y PACF se ha podido concluir que hace falta utilizar un parámetro de media móvil y un parámetro estacional [ARIMA(0,1,1)(0,1,1)].

De todos modos voy a utilizar la optimización Auto Arima para elegir el mejor modelo.

```
autoarimamilk <- auto.arima(MILKSERIE,seasonal=TRUE,trace=TRUE)
```

```
->Best model: ARIMA(0,1,1)(0,1,1)[12]
```

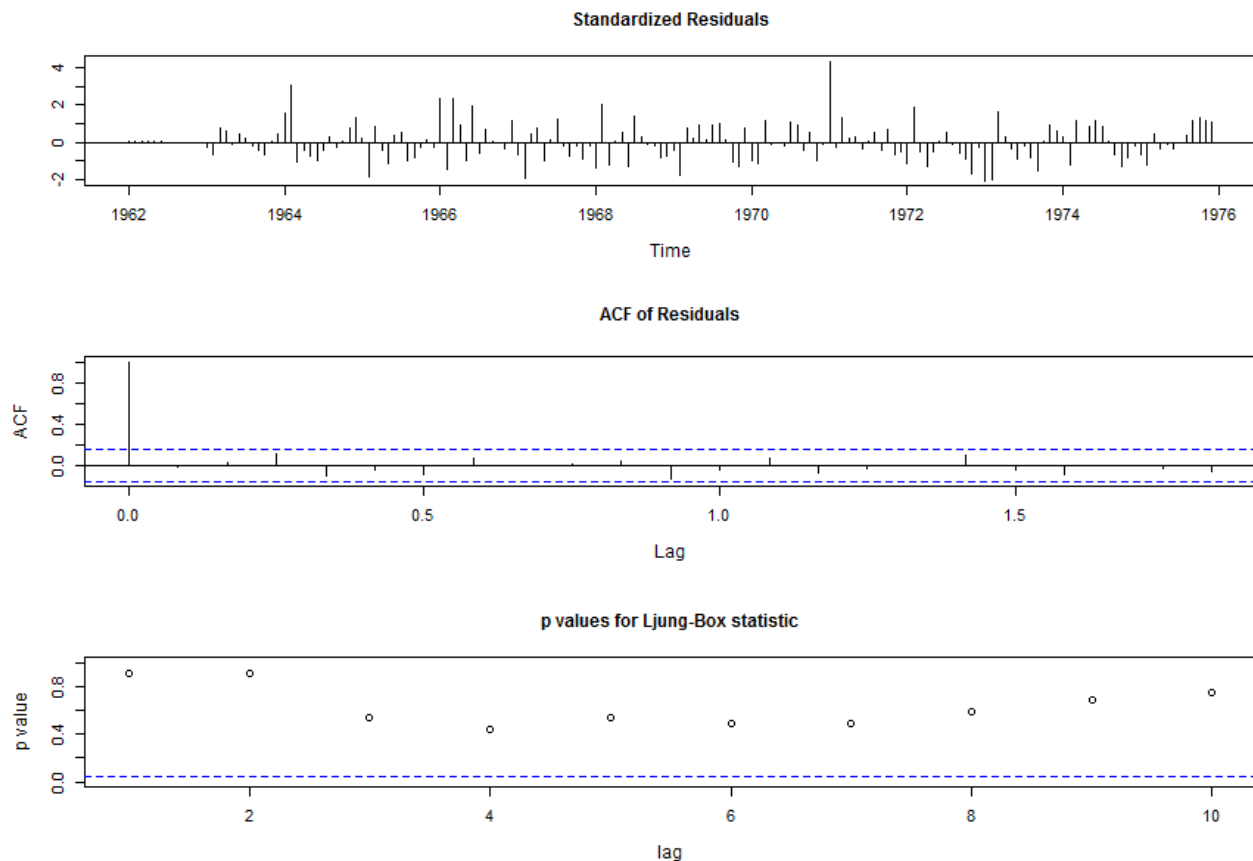
He obtenido modelo SARIMA con los siguientes parametros:

```
Series: MILKSERIE  
ARIMA(0,1,1)(0,1,1)[12]
```

```
Coefficients:  
      ma1      sma1  
    -0.2204  -0.6214  
s.e.   0.0748   0.0627
```

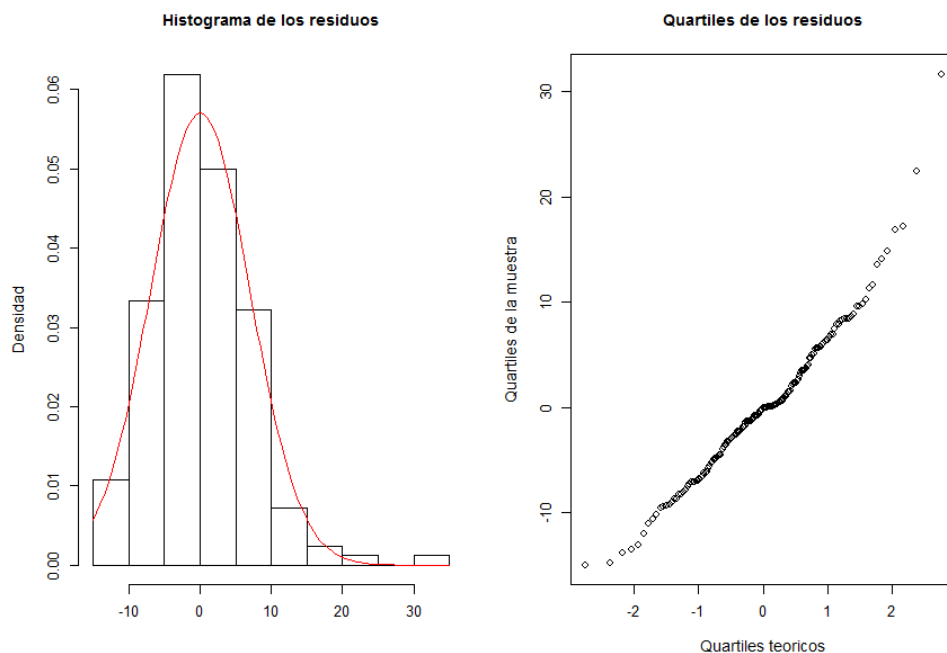
```
sigma^2 estimated as 53.42: log likelihood=-530.15  
AIC=1066.3  AICC=1066.46  BIC=1075.43
```

## 5. Diagnóstico del modelo



Si el modelo está bien ajustado, los residuos deberían tener una distribución parecida a normal con media cero y desviación típica 1 y no correlacionados. Analizando los residuos obtenemos:

En el primer gráfico observamos que los residuos tienden en media a cero y que la desviación típica no se muy alta, aunque se ven valores mayores de 1. La función de auto-correlación en el segundo gráfico también es aceptable. En el tercer gráfico vemos los p valores del test de Ljung-Box para comprobar la hipótesis de correlaciones iguales a cero. Los p valores son altos y no rechazamos la hipótesis de no correlación entre los residuos.



El histograma y QQ plot indican distribución de los residuos aproximada a la normal con cola derecha.

La normalidad no se confirma en el test de Shapiro (rechazo  $H_0$  de normalidad):

```
Shapiro-wilk normality test
data:  residuos
W = 0.96515, p-value = 0.0003192
```

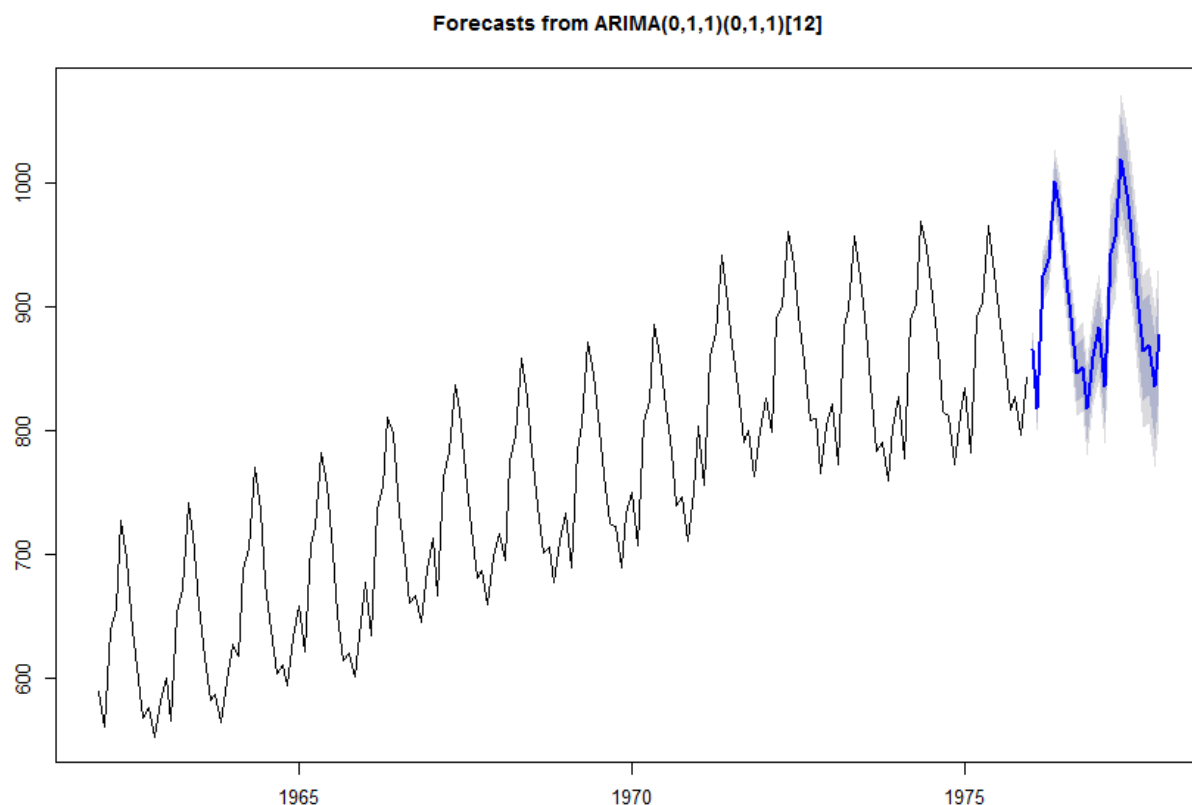
En una prueba con modelo log no mejora el resultado.

Sin embargo los coeficientes del modelo son significativamente distintos de cero:

```
p.value      ma1      sma1
0.00320582  0.00000000
```

## 6. Predicción de la cantidad de leche por vaca a meses

Aplicando el modelo predictivo obtenemos el siguiente resultado:



prog # valores pronosticados

	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