14_lung_diseases.R

felixreichel

2021-11-14

```
# Course: Time series analysis
# Exercise: 14th / Lung diseases
#           monthly deaths from bronchitis, emphysema and asthma in the UK, 1974–1979
# Author: Felix Reichel
```

```
require(astsa)
```

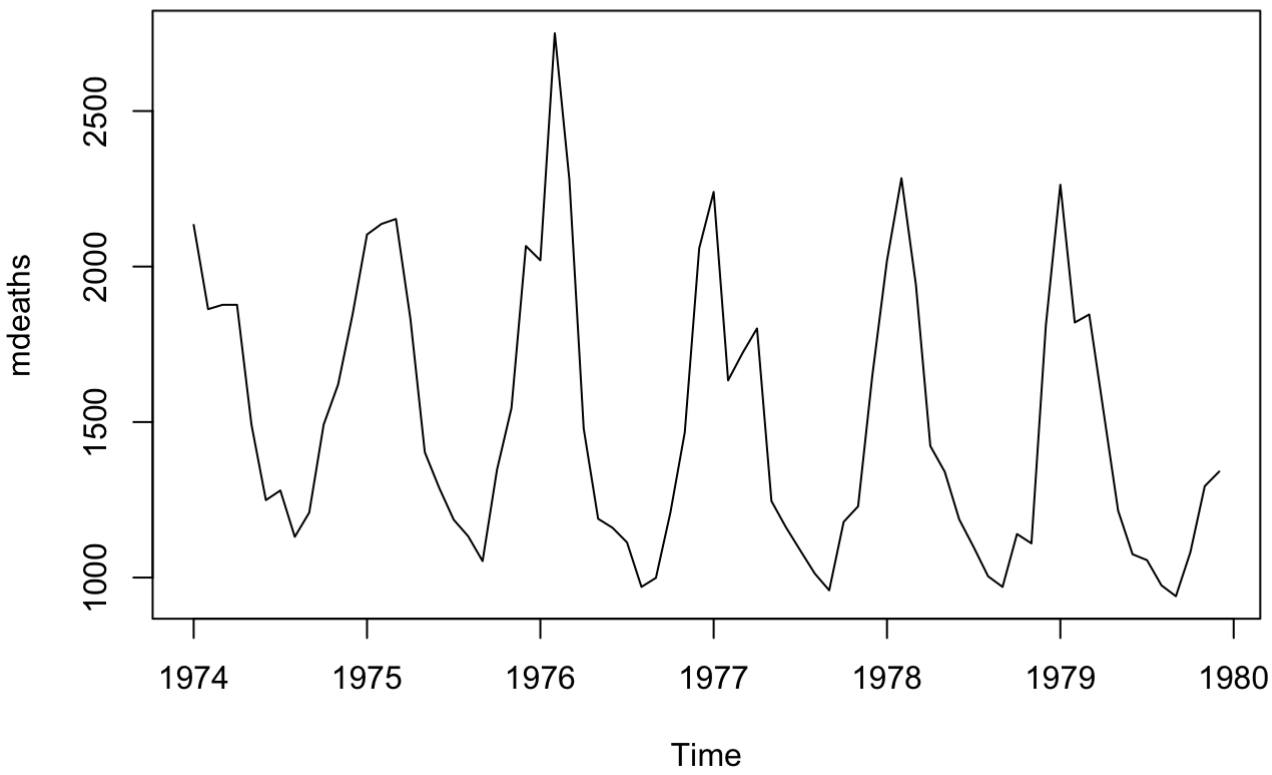
```
## Loading required package: astsa
```

```
require(tseries)
```

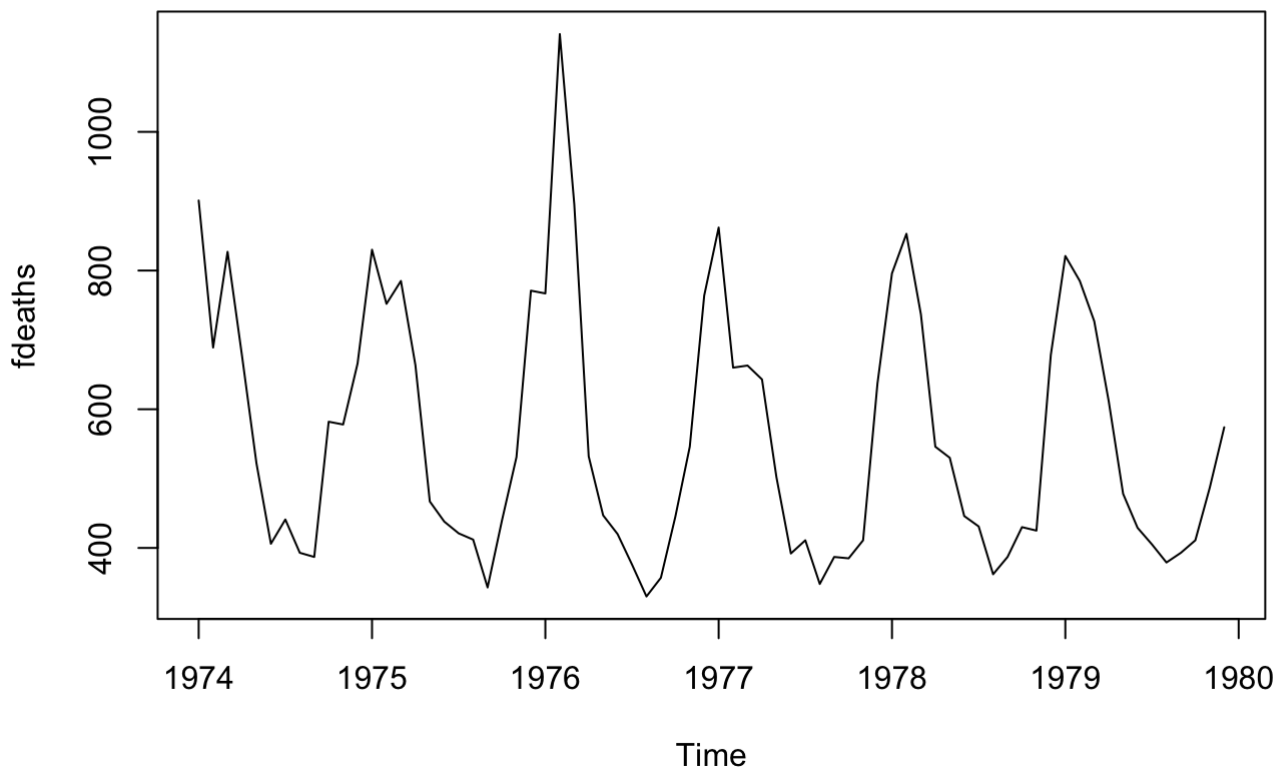
```
## Loading required package: tseries
```

```
## Registered S3 method overwritten by 'quantmod':
##   method            from
##   as.zoo.data.frame zoo
```

```
plot(mdeaths)
```



```
plot(fdeaths)
```




```
freq <- frequency(mdeaths)  # 12
start <- start(mdeaths)     # num[1:2]1974 1
end <- end(mdeaths)         # num[1:2]1979 12
delatat <- deltat(mdeaths)  # 1/freq
Tau <- length(mdeaths)      # Tau
years <- Tau/freq           # 6
t <- rep(1:72)

# Normalized seasonal vector
plot(window(mdeaths, start=1974, end=1975))
```


```
seasonal_decomp <- matrix(decompose(mdeaths)$seasonal)[1:12,]
normalize <- function(x){(x - min(x)) / (max(x) - min(x))}
seasonal_norm <- normalize(seasonal_decomp)
seasonal_norm
```

```
##  [1] 1.00000000 0.99924681 0.87970837 0.54517244 0.24813021 0.16021812
##  [7] 0.10384954 0.01046931 0.00000000 0.21541926 0.32965527 0.77881885
```

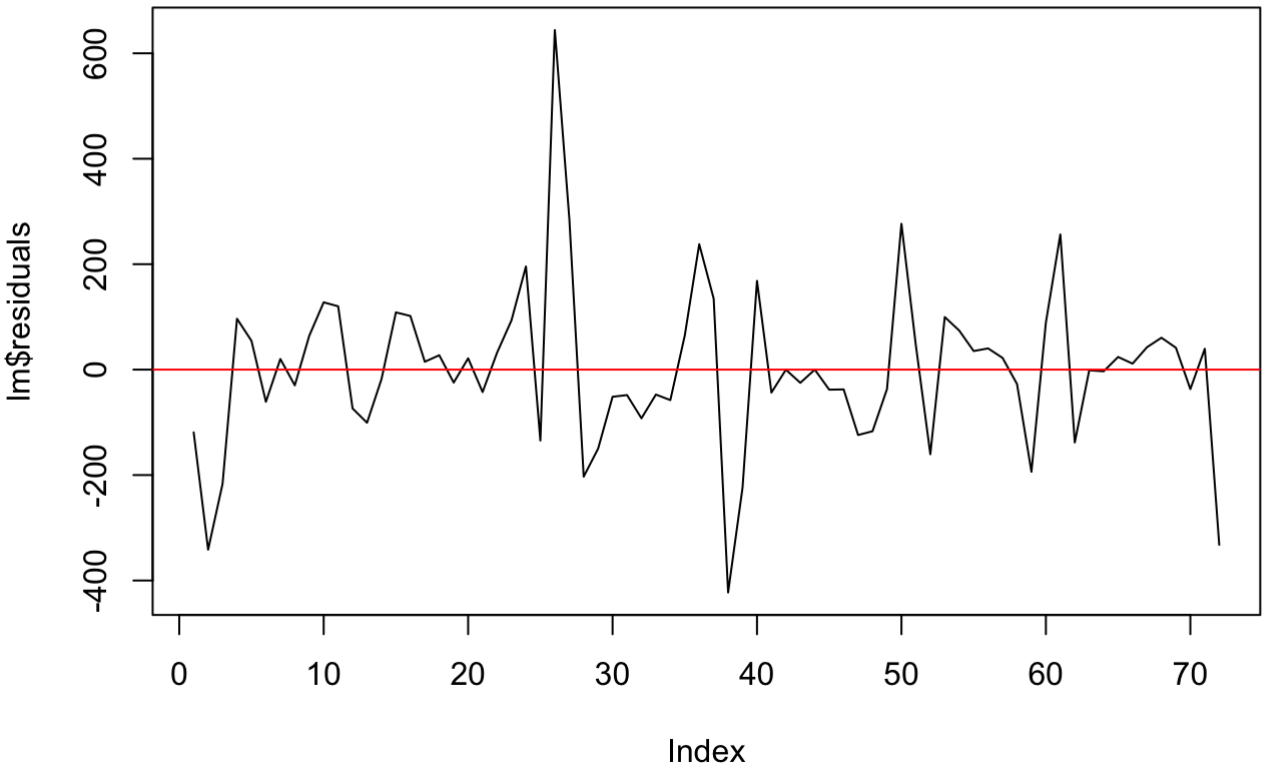
```
# Fit an appropriate regression model
ltr <- 1:Tau
seas <- C(as.factor(rep(1:freq, years)), contr.sum)

lm = lm(mdeaths ~ ltr + seas)
summary(lm)
```

```
##
## Call:
## lm(formula = mdeaths ~ ltr + seas)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -422.70  -58.60   -0.11   64.29  644.03
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1645.8117    40.4442  40.693  < 2e-16 ***
## ltr           -4.1060     0.9662  -4.250 7.72e-05 ***
## seas1        611.3062    65.8865   9.278 3.94e-13 ***
## seas2        566.9121    65.8156   8.614 5.06e-12 ***
## seas3        460.1847    65.7589   6.998 2.73e-09 ***
## seas4        151.1240    65.7163   2.300 0.025024 *
## seas5       -187.9367    65.6879  -2.861 0.005832 **
## seas6       -311.1641    65.6736  -4.738 1.40e-05 ***
## seas7       -357.2248    65.6736  -5.439 1.08e-06 ***
## seas8       -452.1188    65.6879  -6.883 4.27e-09 ***
## seas9       -464.0129    65.7163  -7.061 2.13e-09 ***
## seas10      -240.4069    65.7589  -3.656 0.000547 ***
## seas11       -99.8010    65.8156  -1.516 0.134765
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 168 on 59 degrees of freedom
## Multiple R-squared:  0.875,  Adjusted R-squared:  0.8495
## F-statistic: 34.41 on 12 and 59 DF,  p-value: < 2.2e-16
```

```
# Residual analysis
lm$residuals
```

```
##             1             2             3             4             5
## -119.01190476 -341.51190476 -216.67857143   96.48809524   54.65476190
##             6             7             8             9            10
##  -61.01190476   20.15476190  -29.84523810   64.15476190  127.65476190
##            11            12            13            14            15
##  120.15476190  -73.67857143 -100.74047619  -18.24047619  108.59285714
##            16            17            18            19            20
##  101.75952381   14.92619048   27.25952381  -24.57380952   21.42619048
##            21            22            23            24            25
##  -42.57380952   31.92619048   93.42619048  195.59285714 -134.46904762
##            26            27            28            29            30
##  644.03095238  287.86428571 -202.96904762 -149.80238095  -51.46904762
##            31            32            33            34            35
##  -48.30238095  -92.30238095  -47.30238095  -57.80238095   64.69761905
##            36            37            38            39            40
##  237.86428571  134.80238095 -422.69761905 -223.86428571  168.30238095
##            41            42            43            44            45
##  -43.53095238   -0.19761905  -25.03095238   -0.03095238  -38.03095238
##            46            47            48            49            50
##  -37.53095238 -124.03095238 -116.86428571  -36.92619048  276.57380952
##            51            52            53            54            55
##   45.40714286 -160.42619048   99.74047619   74.07380952   35.24047619
##            56            57            58            59            60
##   40.24047619   22.24047619  -27.25952381 -193.75952381   89.40714286
##            61            62            63            64            65
##  256.34523810 -138.15476190   -1.32142857   -3.15476190   24.01190476
##            66            67            68            69            70
##   11.34523810   42.51190476   60.51190476   41.51190476  -36.98809524
##            71            72
##   39.51190476 -332.32142857
```

```
# Plot residuals
plot(lm$residuals, type="l")
abline(h = mean(lm$residuals), col="red")
```




```
require(lmtest)
```

```
## Loading required package: lmtest
```

```
## Loading required package: zoo
```

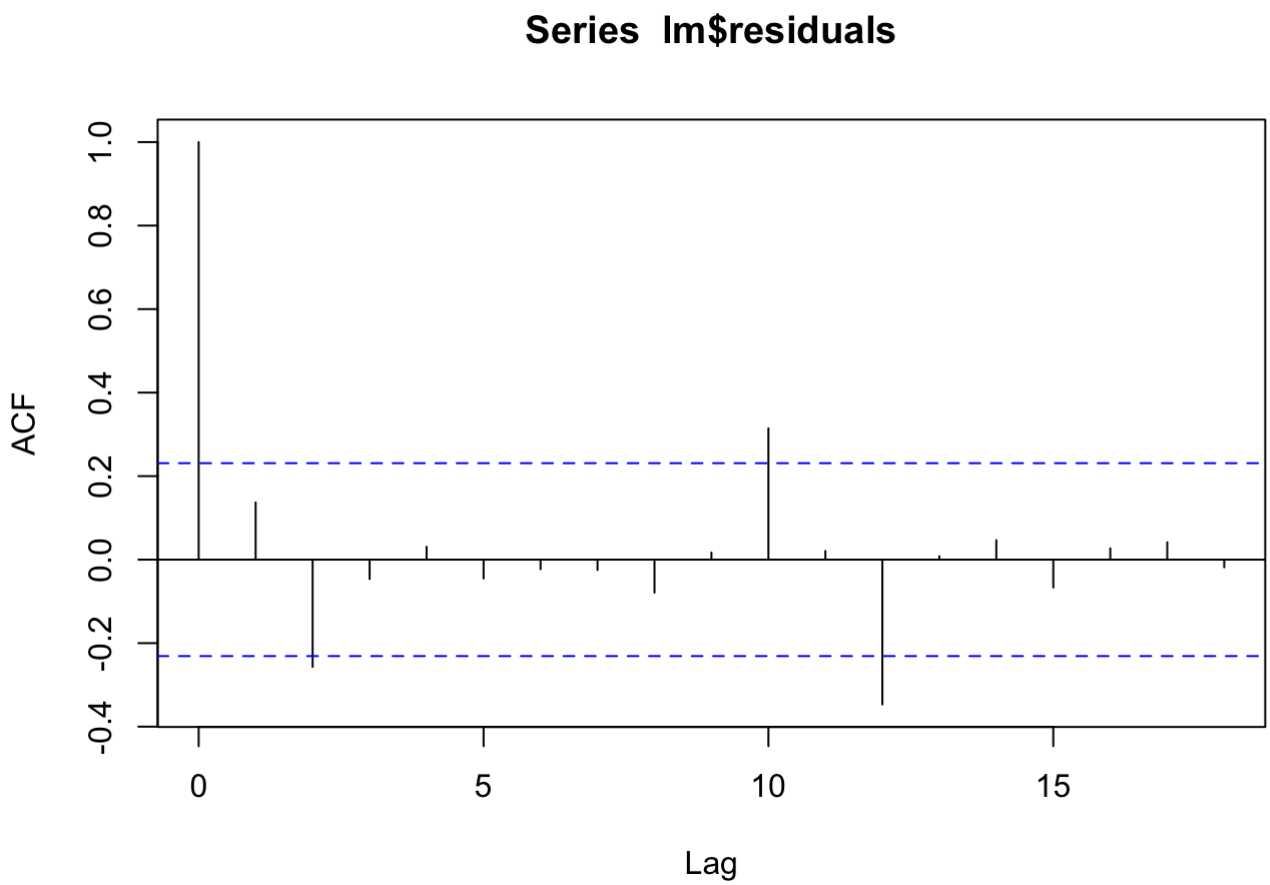
```
##
## Attaching package: 'zoo'
```

```
## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

```
# Durbin Watson
dwtest(lm) # one sided test: rho(1)>0
```

```
##
##  Durbin-Watson test
##
## data:  lm
## DW = 1.6512, p-value = 0.08205
## alternative hypothesis: true autocorrelation is greater than 0
```

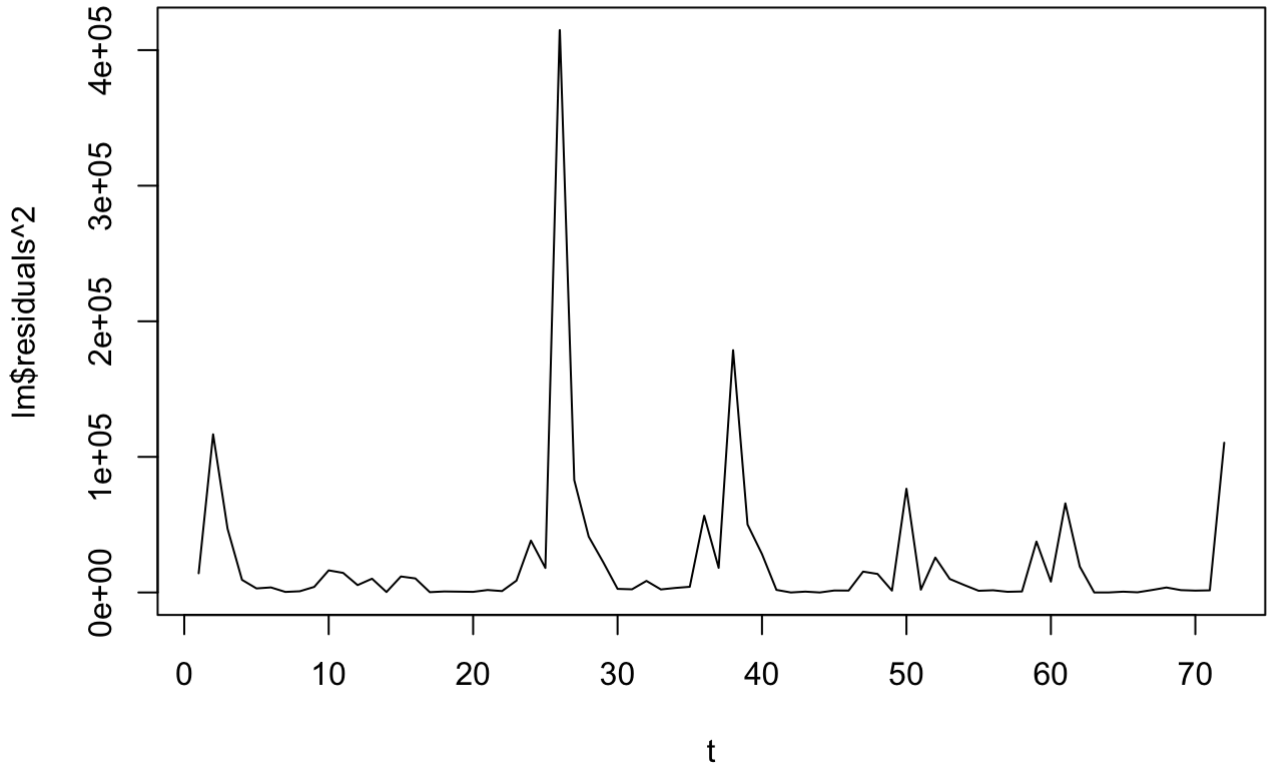
```
# 0 < DW =1.6512 < 2  => reject H0 => positive autocorrelation of errors

acf(lm$residuals) # ci: (- 1.96/sqrt(TT),+ 1.96/sqrt(TT))
```




```
Box.test (lm$residuals, lag = 3, type = "Ljung")
```

```
##
##  Box-Ljung test
##
## data:  lm$residuals
## X-squared = 6.598, df = 3, p-value = 0.08588
```

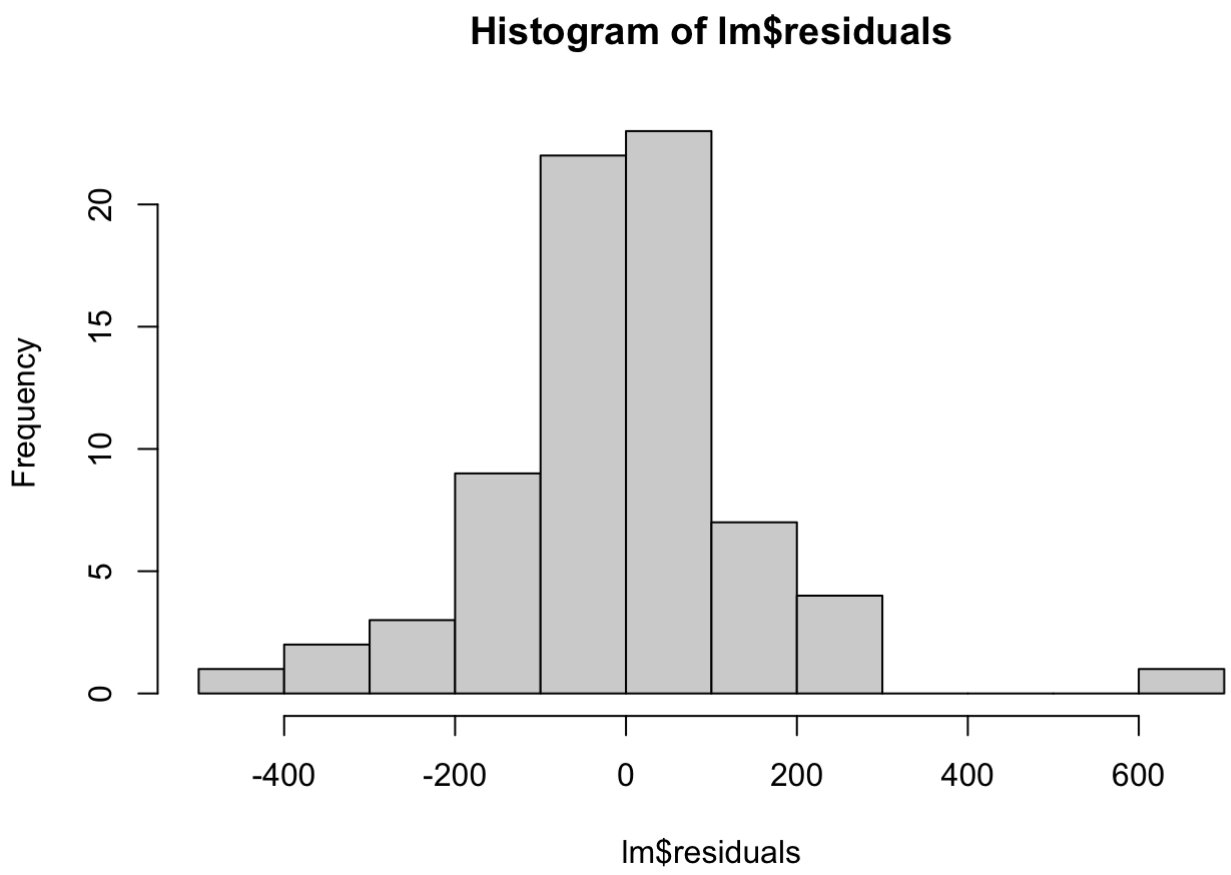
```
# p-value = 0.08588 > 0.05 => not significant => H0 (model does not show lack of fit)

# heteroscedasticity
plot(lm$residuals^2,type="l",xlab="t")
```



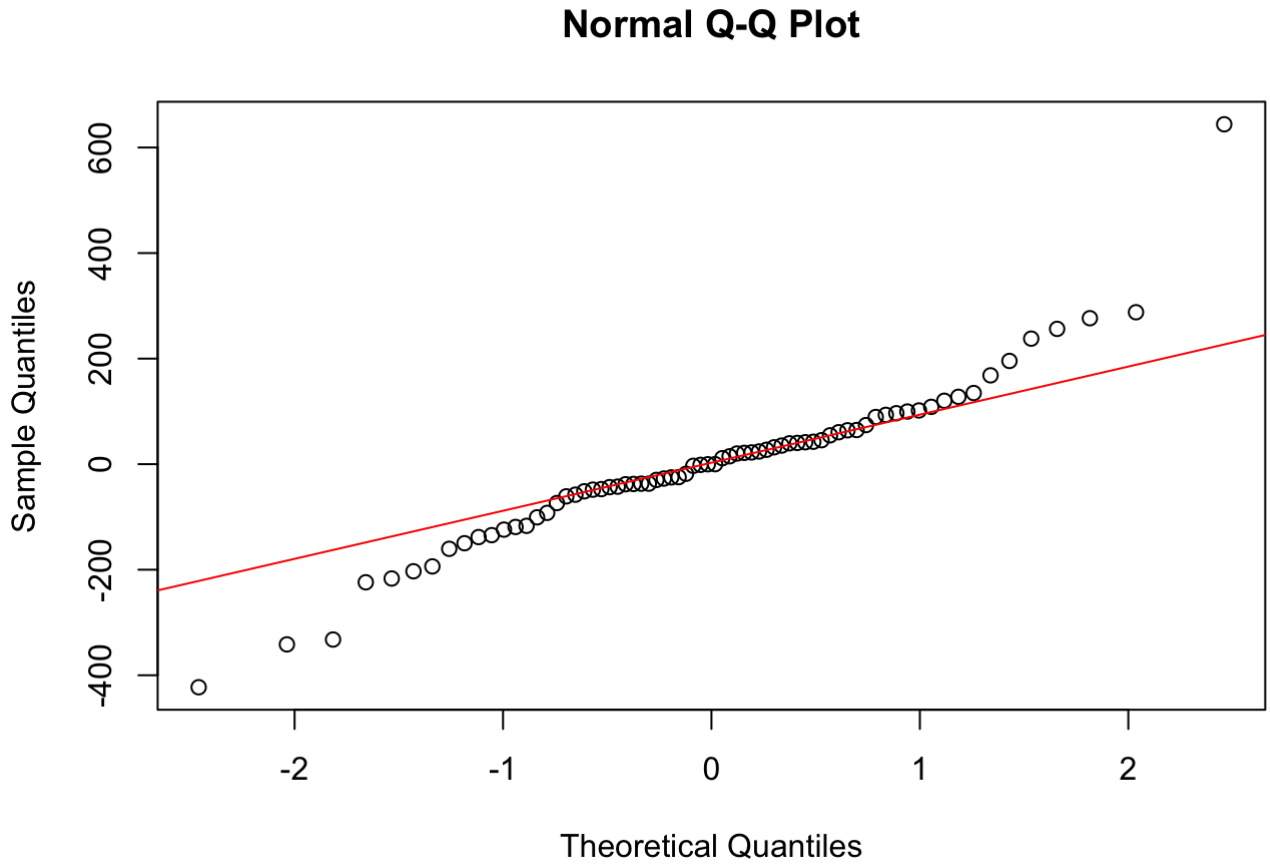

```
bptest(lm)
```

```
##
##  studentized Breusch-Pagan test
##
## data:  lm
## BP = 29.189, df = 12, p-value = 0.003693
```

```
# p-value = 0.003693 < 0.05 => H1 (heteroskedasticity)

# normal distribution
hist(lm$residuals)
```




```
qqnorm(lm$residuals)
qqline(lm$residuals,col="red")
```



```
jarque.bera.test(lm$residuals)
```

```
##
##  Jarque Bera Test
##
## data:  lm$residuals
## X-squared = 49.395, df = 2, p-value = 1.879e-11
```

```
# p-value = 1.879e-11 < 0.05 => H1 (no normal distribution)

# Seasonal effect for December
seasonal_effect_december <- seasonal_norm[1]
seasonal_effect_december
```

```
## [1] 1
```

```
# Plot the series with the fitted values
df <- data.frame(ltr = 1:Tau, s = rep(1:12, years))
pred <- predict(lm, newdata = df)
```

```
## Warning: contrasts dropped from factor seas
```

```
plot(mdeaths, col="black")
lines(ts(pred, start=1974, end=1980, deltat = 1/12),type='l',col="blue",xlab="t",ylab
="ytrseas")
```