Spectral Analysis

Nik Bear Brown

In this lesson we'll learn the how to implement Spectral Analysis in R.

Additional packages needed

To run the code you may need additional packages.

- If necessary install the followings packages.

```
install.packages("TSA");
```

```
library(TSA)

## Loading required package: leaps

## Loading required package: locfit

## locfit 1.5-9.1    2013-03-22

## Loading required package: mgcv

## Loading required package: nlme

## This is mgcv 1.8-15. For overview type 'help("mgcv-package")'.

## Loading required package: tseries

##
## Attaching package: 'TSA'

## The following objects are masked from 'package:stats':
##
##     acf, arima

## The following object is masked from 'package:utils':
##
##     tar
```

Data

We will be using two data sets:

A. Monthly Deaths from Lung Diseases in the UK. It consists of three time series giving the monthly deaths from bronchitis, emphysema and asthma in the UK, 1974-1979, both sexes (ldeaths), males (mdeaths) and females (fdeaths).

B. A regular time series giving the luteinizing hormone in blood samples at 10 mins intervals from a human female, 48 samples.

```
data(lh)
lh

## Time Series:
## Start = 1
## End = 48
## Frequency = 1
##  [1] 2.4 2.4 2.4 2.2 2.1 1.5 2.3 2.3 2.5 2.0 1.9 1.7 2.2 1.8 3.2 3.2
2.7
## [18] 2.2 2.2 1.9 1.9 1.8 2.7 3.0 2.3 2.0 2.0 2.9 2.9 2.7 2.7 2.3 2.6
2.4
## [35] 1.8 1.7 1.5 1.4 2.1 3.3 3.5 3.5 3.1 2.6 2.1 3.4 3.0 2.9

head(lh)

## [1] 2.4 2.4 2.4 2.2 2.1 1.5

names(lh)

## NULL

data(UKLungDeaths)
deaths <- ts.union(mdeaths, fdeaths)
head(deaths)

##      mdeaths fdeaths
## [1,]    2134     901
## [2,]    1863     689
## [3,]    1877     827
## [4,]    1877     677
## [5,]    1492     522
## [6,]    1249     406

names(deaths)

## NULL
```

Spectral Analysis

Spectral analysis or Spectrum analysis is based on the analysis of frequencies rather than fluctuations of numbers. In statistical signal processing, the goal of spectral density estimation (SDE) is to estimate the spectral density (also known as the power spectral density) of a random signal from a sequence of time samples of the signal. Intuitively speaking, the spectral density characterizes the frequency content of the signal. One purpose of estimating the spectral density is to detect any periodicities in the data, by observing peaks at the frequencies corresponding to these periodicities.

statistics and signal processing, an algorithm that estimates the strength of different frequency components (the power spectrum) of a time-domain signal. This may also be called frequency domain analysis. There are many approaches to spectral density. Below is a partial list of parametric and non-parametric spectral density estimation techniques.

Non-parametric spectral density estimation techniques:

- Periodogram, the basic modulus-squared of the discrete Fourier transform

- Bartlett's method is the average of the periodograms taken of multiple segments of the signal to reduce variance of the spectral density estimate Welch's method a windowed version of Bartlett's method that uses overlapping segments

- Multitaper is a periodogram-based method that uses multiple tapers, or windows, to form independent estimates of the spectral density to reduce variance of the spectral density estimate

- Least-squares spectral analysis, based on least squares fitting to known frequencies

- Non-uniform discrete Fourier transform is used when the signal samples are unevenly spaced in time

- Singular spectrum analysis is a nonparametric method that uses a singular value decomposition of the covariance matrix to estimate the spectral density Short-time Fourier transform

Parametric spectral density estimation techniques:

- Autoregressive model (AR) estimation, which assumes that the nth sample is correlated with the previous p samples.

- Moving-average model (MA) estimation, which assumes that the nth sample is correlated with noise terms in the previous p samples.

- Autoregressive moving average (ARMA) estimation, which generalizes the AR and MA models.

- Maximum entropy spectral estimation is an all-poles method useful for SDE when singular spectral features, such as sharp peaks, are expected.

Periodogram

Gvien a signal that is sampled at N different times, with the samples uniformly spaced by Δt, giving values x_n. Since the power spectral density of a continuous function defined on the entire real line is the modulus squared of its Fourier transform, the simplest technique to estimate the spectrum is the periodogram, given by the modulus squared of the discrete Fourier transform,

$$S(f) = \frac{\Delta t}{N} \left| \sum_{n=0}^{N-1} x_n \, e^{-i2\pi nf} \right|^2, \qquad -\frac{1}{2\Delta t} < f \leq \frac{1}{2\Delta t}$$

where $1/(2\Delta t)$ is the Nyquist frequency. The name "periodogram" was coined by Arthur Schuster in 1898. Despite the simplicity of the periodogram, the method suffers from severe deficiencies. It is an inconsistent estimator because it does not converge to the true spectral density as $N \to \infty$. It exhibits very high spectral leakage although this can be reduced by multiplying x_n by a window function. In the presence of additive noise, the estimate has a positive bias.

Fourier analysis

Fourier analysis is the study of the way general functions may be represented or approximated by sums of simpler trigonometric functions. Fourier analysis grew from the study of Fourier series. A Fourier series is a way to represent a (wave-like) function as the sum of simple sine waves

If, s(x) denotes a function of the real variable x, and s is integrable on an interval [x0, x0 + P], for real numbers x0 and P. We will attempt to represent s in that interval as an infinite sum, or series, of harmonically related sinusoidal functions. Outside the interval, the series is periodic with period P (frequency 1/P). It follows that if s also has that property, the approximation is valid on the entire real line. We can begin with a finite summation (or partial sum):

$$ s_N(x) = \frac{A_0}{2} + \sum_{n=1}^N A_n\cdot \sin(\tfrac{2\pi nx}{P}+\phi_n), \quad \scriptstyle \text{for integer}\ N\ \ge\ 1. $$

$s_N(x)$ is a periodic function with period P. Using the identities:

$$\sin(2\pi nx/P + \phi_n) \equiv \sin(\phi_n)\cos(2\pi nx/P) + \cos(\phi_n)\sin(2\pi nx/P)$$

$$\sin(2\pi nx/P + \phi_n) \equiv \mathrm{Re}\left\{\frac{1}{i} \cdot e^{i(2\pi nx/P+\phi_n)}\right\}$$

$$= \frac{1}{2i} \cdot e^{i(2\pi nx/P+\phi_n)} + \left(\frac{1}{2i} \cdot e^{i(2\pi nx/P+\phi_n)}\right)^*,$$

Function s(x) (in red) is a sum of six sine functions of different amplitudes and harmonically related frequencies. Their summation is called a Fourier series. The

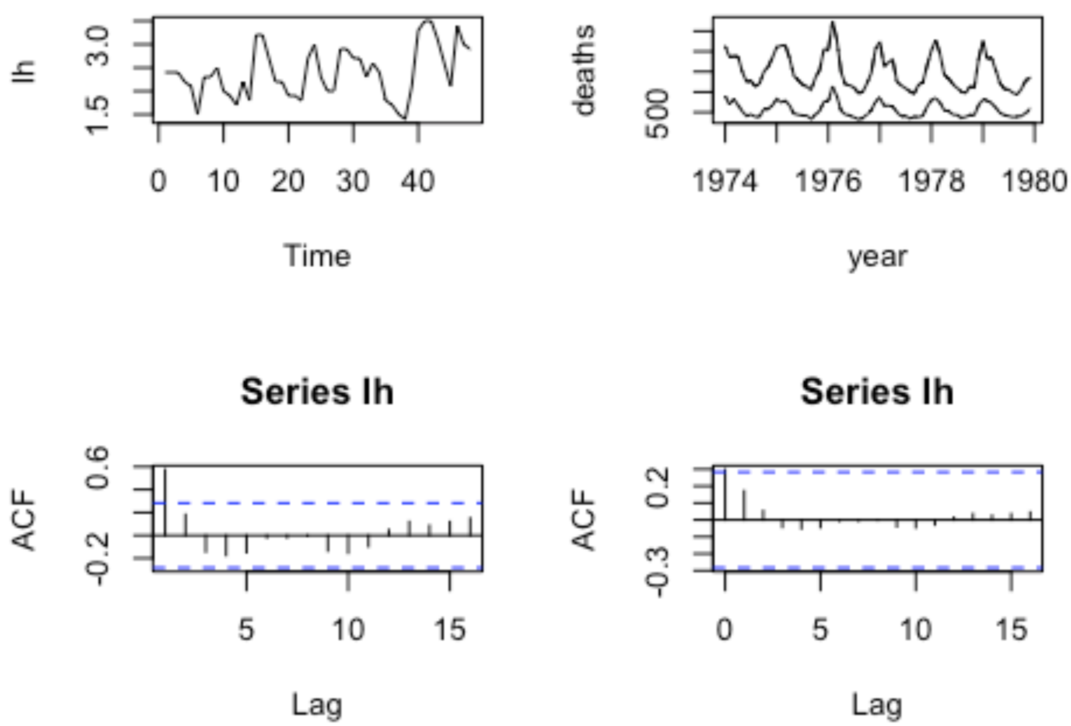
Fourier transform, S(f) (in blue), which depicts amplitude vs frequency, reveals the 6 frequencies and their amplitudes.

Spectral Analysis in R

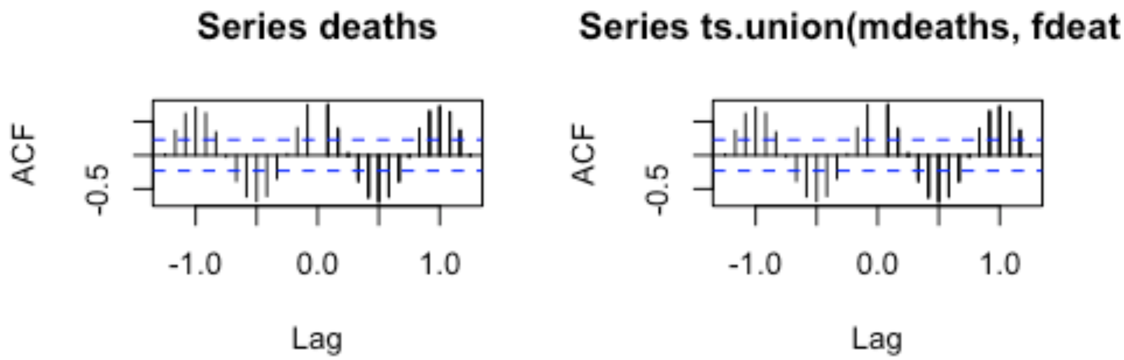
```
tsp(mdeaths)

## [1] 1974.000 1979.917   12.000

start(mdeaths)

## [1] 1974    1

end(mdeaths)

## [1] 1979   12

frequency(mdeaths)

## [1] 12

cycle(deaths)

##      Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
## 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
## 1976   1   2   3   4   5   6   7   8   9  10  11  12
## 1977   1   2   3   4   5   6   7   8   9  10  11  12
## 1978   1   2   3   4   5   6   7   8   9  10  11  12
## 1979   1   2   3   4   5   6   7   8   9  10  11  12

# Plot time series
par(mfrow = c(2, 2))
ts.plot(lh)
ts.plot(deaths, mdeaths, fdeaths,
          lty = c(1, 3, 4), xlab = "year", ylab = "deaths")
#obtain quarterly sums or annual means of deaths
aggregate(deaths, 1,sum)

## Time Series:
## Start = 1974
## End = 1979
## Frequency = 1
##      mdeaths fdeaths
## 1974   19071    7069
## 1975   19247    6854
## 1976   18697    7021
## 1977   16927    6302
## 1978   17329    6622
## 1979   16437    6501

aggregate(deaths, 1, mean)
```

```
## Time Series:
## Start = 1974
## End = 1979
## Frequency = 1
##       mdeaths  fdeaths
## 1974 1589.250 589.0833
## 1975 1603.917 571.1667
## 1976 1558.083 585.0833
## 1977 1410.583 525.1667
## 1978 1444.083 551.8333
## 1979 1369.750 541.7500
```

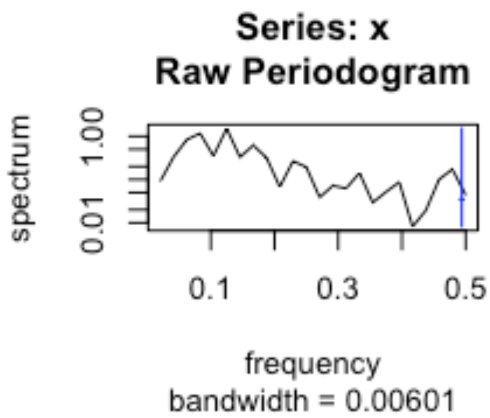
```
#Second-Order Summaries
#acf for multiple time series:
acf(lh)
acf(lh, type = "covariance")
```

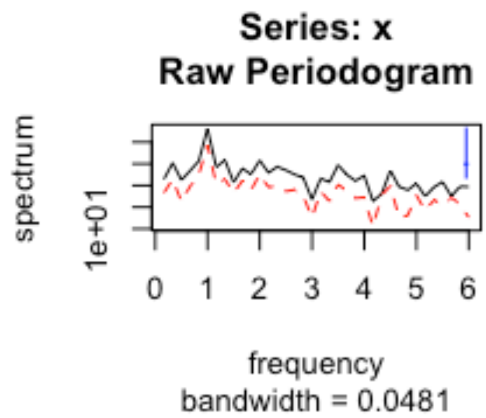



```
acf(deaths)
acf(ts.union(mdeaths, fdeaths))
#Note: spectrum by de-fault removes a linear trend from the series
before estimating the spectral density
par(mfrow = c(2, 2))
```

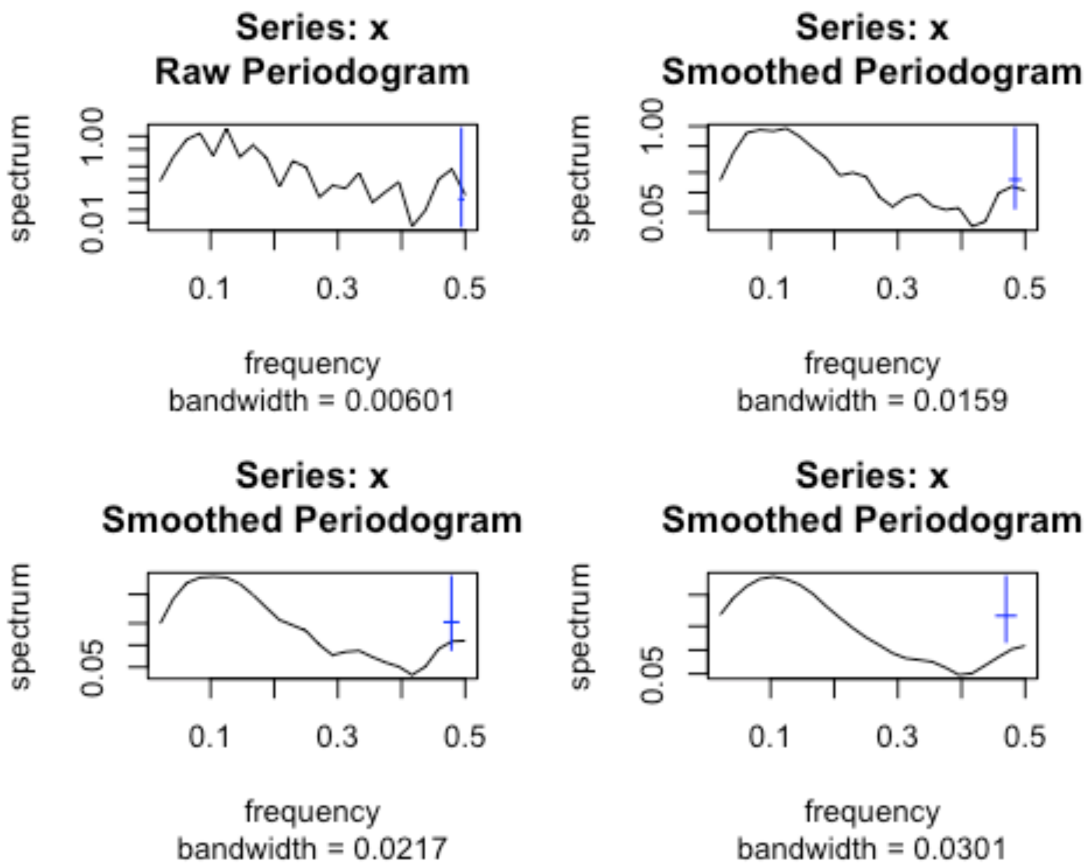



```
spectrum(lh)
spectrum(deaths)
#deaths.spec <- spectrum(deaths, plot=FALSE)
#deaths.spec$spec     # these are the periodogram ordinates for both
series
#deaths.spec$freq     # these are the frequencies omega
#The function spectrum also produces smoothed plots, using repeated
#smoothing with modified Daniell smoothers (Bloomfield, 2000, p. 157),
which
#are moving averages giving half weight to the end values of the span.
Trial-and-
#error is needed to choose the spans
# The function spectrum estimates of the spectral density at
frequencies
par(mfrow = c(2, 2))
```

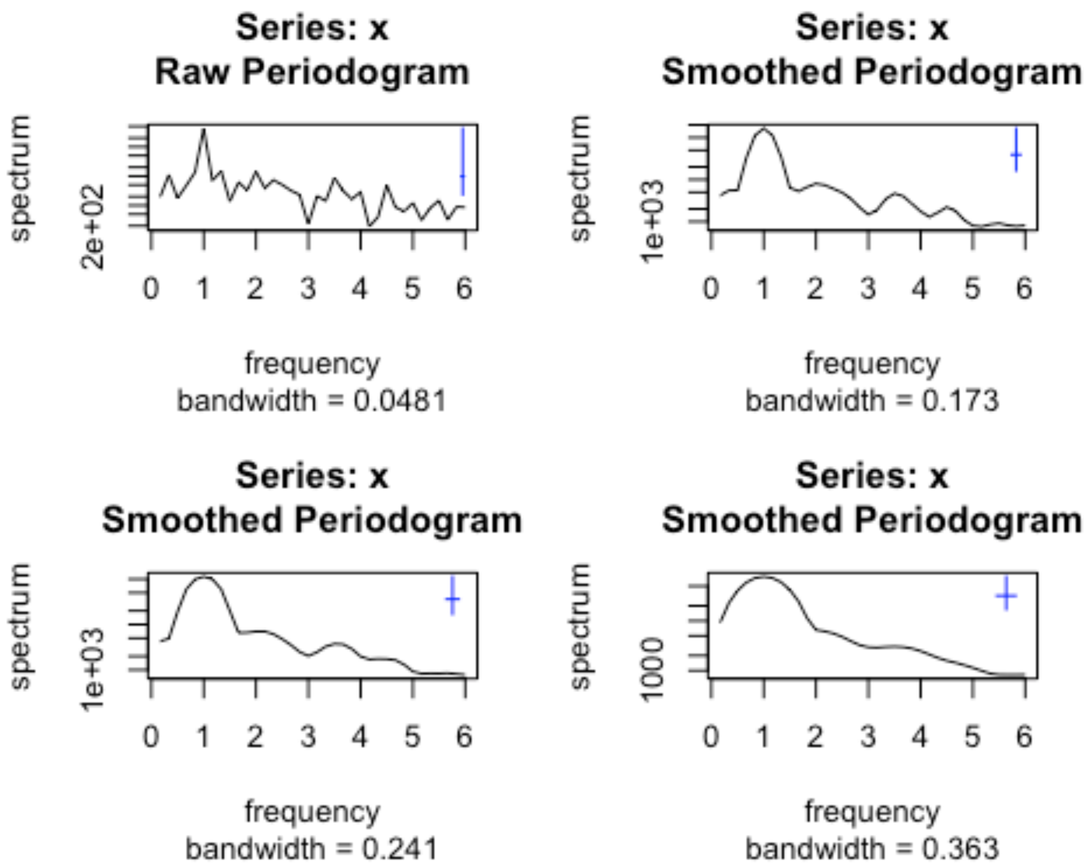




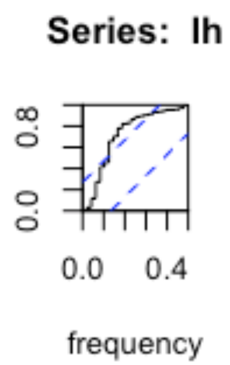

```r
spectrum(lh)
spectrum(lh, spans = 3)
spectrum(lh, spans = c(3, 3))
spectrum(lh, spans = c(3, 5))
```

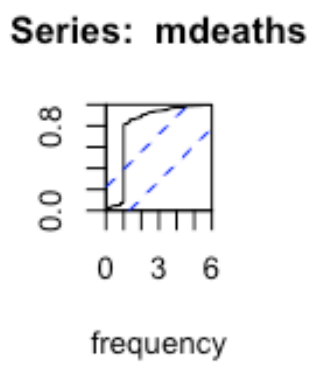


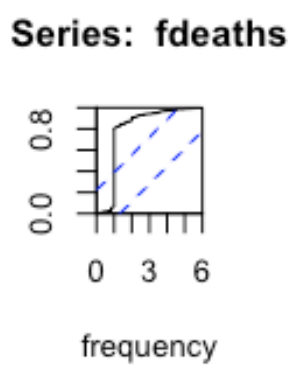

```r
spectrum(mdeaths)
spectrum(mdeaths, spans = c(3, 3))
spectrum(mdeaths, spans = c(3, 5))
spectrum(mdeaths, spans = c(5, 7))
```

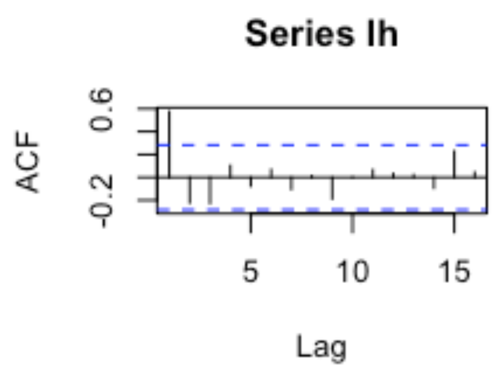



```r
# plot the cumulative periodogram
cpgram(lh)
cpgram(mdeaths)
cpgram(fdeaths)
# ARIMA models
acf(lh, type = "partial")
```





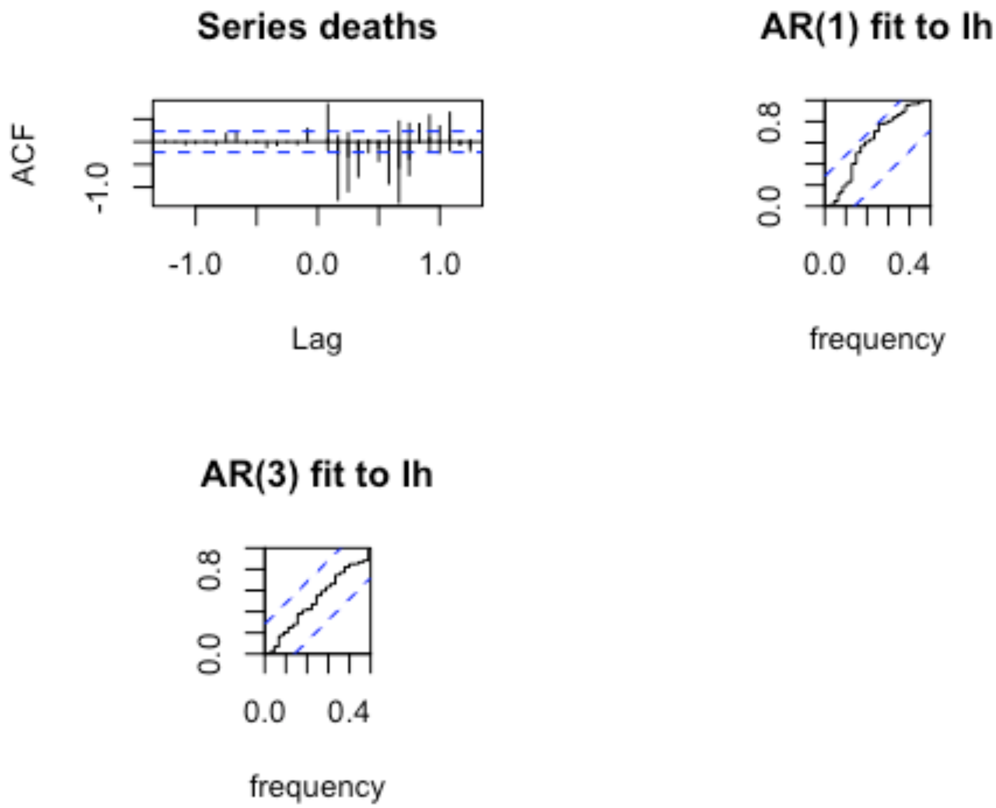


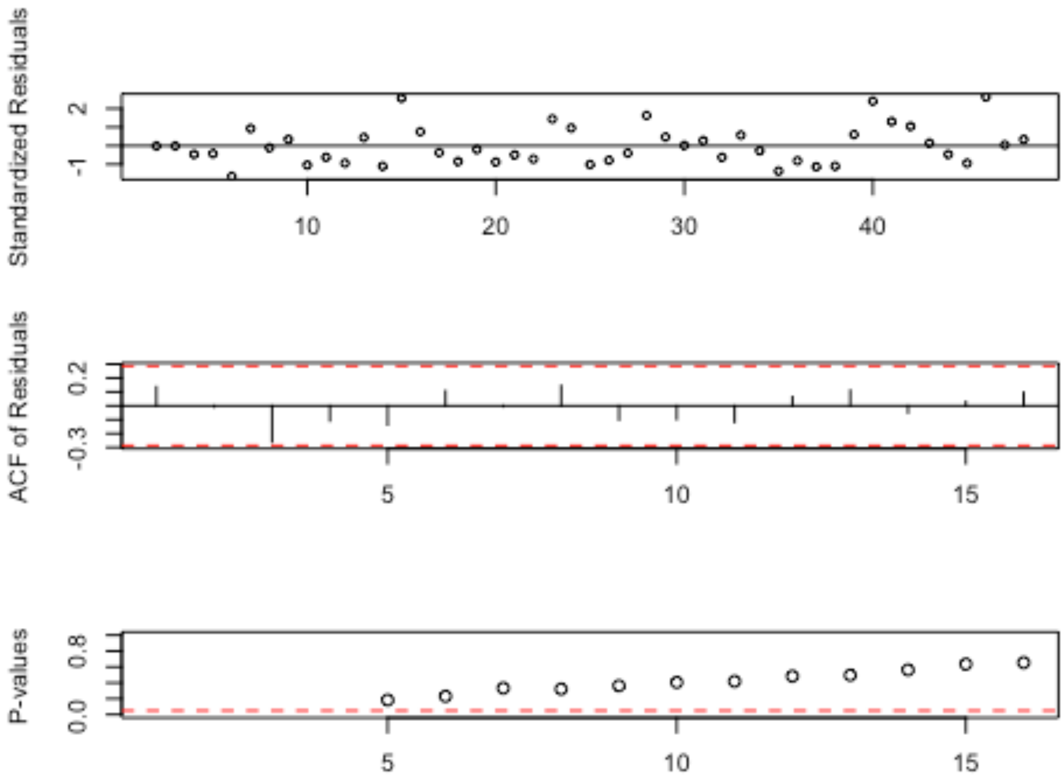



```
acf(deaths, type = "partial")
# Model Fitting
lh.ar1 <- ar(lh, aic = F, order.max = 1)
cpgram(lh.ar1$resid, main = "AR(1) fit to lh")
lh.ar <- ar(lh, order.max = 9)
lh.ar$aic
```

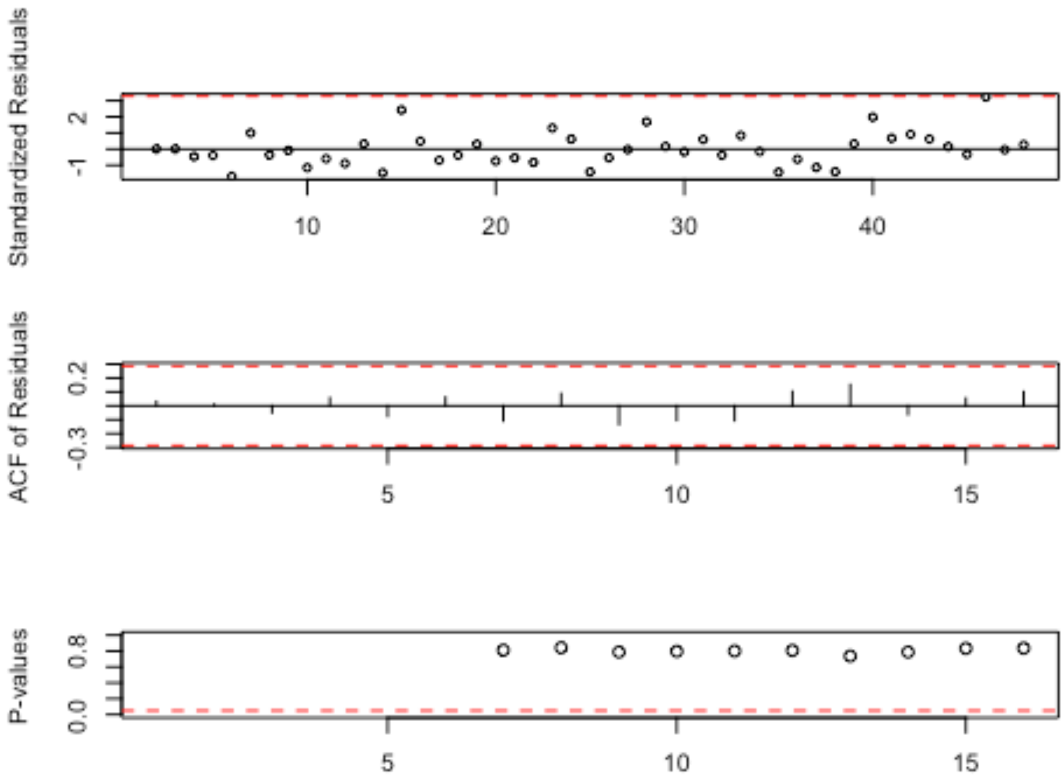
```
##          0          1          2          3          4          5
## 18.3066645  0.9956542  0.5380214  0.0000000  1.4903597  3.2127890
##          6          7          8          9
##  4.9932119  6.4694960  8.4625678  8.7411958
```

```
cpgram(lh.ar$resid, main = "AR(3) fit to lh")
lh.arima1 <- arima(lh, order = c(1,0,0))
tsdiag(lh.arima1)
```

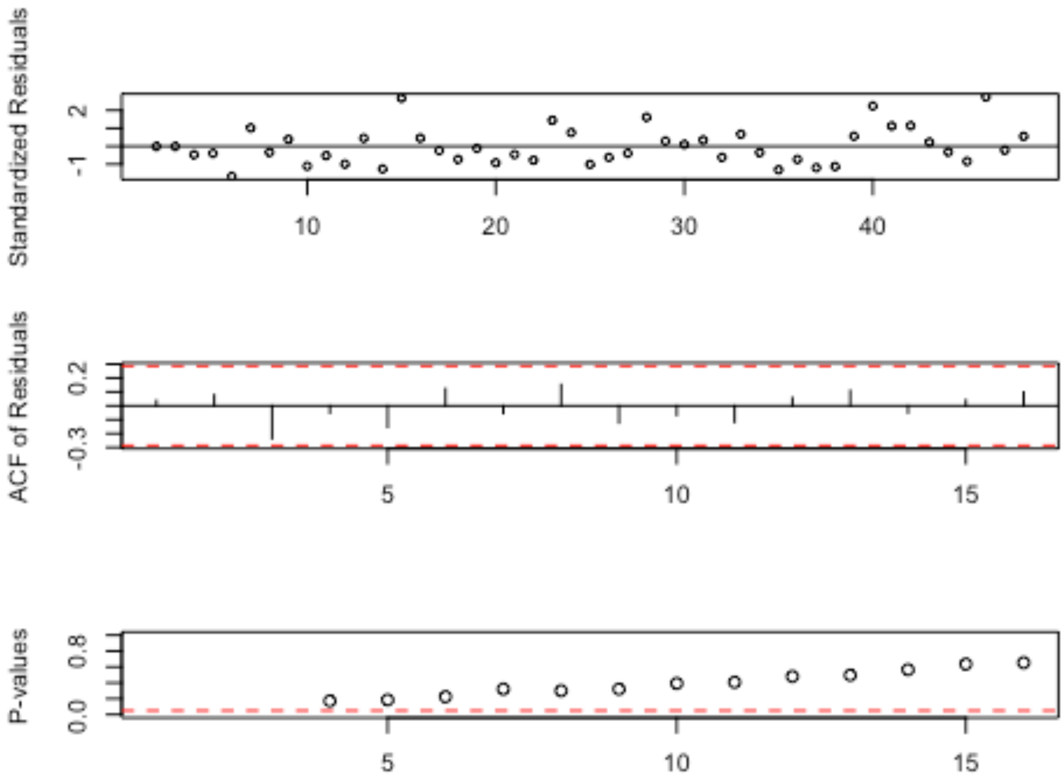

Series deaths
ACF
-1.0
-1.0
0.0
1.0
Lag
AR(1) fit to lh
0.8
0.0
0.0
0.4
frequency
AR(3) fit to lh
0.8
0.0
0.0
0.4
frequency



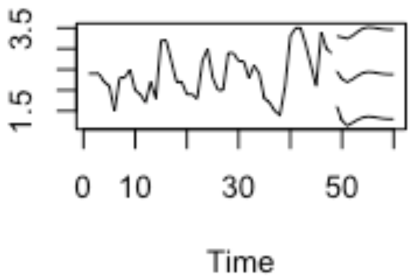

```
lh.arima3 <- arima(lh, order = c(3,0,0))
tsdiag(lh.arima3)
```




```
lh.arima11 <- arima(lh, order = c(1,0,1))
tsdiag(lh.arima11)
```




```
##-----------------Forecasting -----------------------
lh.fore <- predict(lh.arima3, 12)
ts.plot(lh, lh.fore$pred, lh.fore$pred + 2*lh.fore$se,
        lh.fore$pred - 2*lh.fore$se)
```



Resources

- Spectral Analysis
- R Time Series Tutorial
- Time Series Analysis with R
- Spectral Analysis - The R Book - Safari Books Online

References

The data, R code and lessons are based upon:

1. Time Series Analysis :

Data Source: http://www.geophysics.geol.uoa.gr/catalog/catgr_20002008.epi

Code References :

Book : Mastering Predictive Analytic with R
Author: Rui Miguel Forte

https://www.safaribooksonline.com/library/view/mastering-predictive-analytics/9781783982806/

Chapter 9: Time series Analysis

http://www.statoek.wiso.uni-goettingen.de/veranstaltungen/zeitreihen/sommer03/ts_r_intro.pdf

http://www.stat.pitt.edu/stoffer/tsa3/R_toot.htm

http://www.statoek.wiso.uni-goettingen.de/veranstaltungen/zeitreihen/sommer03/ts_r_intro.pdf

2. Trend Analysis

Code References :

Book : Mastering Predictive Analytic with R
Author: Rui Miguel Forte
https://www.safaribooksonline.com/library/view/mastering-predictive-analytics/9781783982806/

http://www.r-bloggers.com/seasonal-trend-decomposition-in-r/

3. Seasonal Models

Code references :

Book: Time Series Analysis and Its Applications
Author: Robert H. Shumway . David S. Stoffer
Link:
http://www.springer.com/us/book/9781441978646#otherversion=9781461427599

http://a-little-book-of-r-for-time-series.readthedocs.org/en/latest/src/timeseries.html

https://onlinecourses.science.psu.edu/stat510/?q=node/47

https://rpubs.com/ryankelly/tsa5

https://onlinecourses.science.psu.edu/stat510/node/68

Data Reference :
https://github.com/RMDK/TimeSeriesAnalysis/blob/master/colorado_river.csv

4. Spectral Analysis

Code References:
Book:
Modern Applied Statistics with S Fourth edition

Author: W. N. Venables and B. D. Ripley
Link: Modern Applied Statistics with S Fourth edition

http://www.maths.adelaide.edu.au/patty.solomon/TS2004/tsprac3_2004.pdf