



Published on *STAT 510* (<https://onlinecourses.science.psu.edu/stat510>)

[Home](#) > Lesson 7: The Periodogram

Lesson 7: The Periodogram

Assignments:

- Read Examples 2.8-2.9 (pages 67-70) and Sections 4.1-4.2 (pages 173-180) of your text.
- Read through the Lesson 7 online notes that follow.
- Complete Lesson 7 Assignment.

Overview:

This week we'll study the periodogram, a tool for describing and identifying the dominant cycles in a time series.

Learning Objectives:

After successfully completing this lesson, you should be able to:

- Create a periodogram in R
- Identify the dominant periods (or frequencies) of a time series

7.1 The Periodogram

Any time series can be expressed as a combination of cosine (or sine) waves with differing periods (how long it takes to complete a full cycle) and amplitudes (maximum/minimum value during the cycle). This fact can be utilized to examine the periodic (cyclical) behavior in a time series.

A **periodogram** is used to identify the dominant periods (or frequencies) of a time series. This can be a helpful tool for identifying the dominant cyclical behavior in a series, particularly when the cycles are not related to the commonly encountered monthly or quarterly seasonality.

Properties of a Cosine Function

For discrete time (meaning time t = integer values), these definitions are useful for a cosine (or sine) wave:

- The **period** (T) is the number of time periods required to complete a single cycle of the cosine function.

- The **frequency** is $\omega = 1/T$. It is the fraction of the complete cycle that's completed in a single time period.

Imagine fitting a single cosine wave to a time series observed in discrete time. Suppose that we write this cosine wave as

$$x_t = A \cos(2\pi\omega t + \phi)$$

A is the amplitude. It determines the maximum absolute height of the curve.

ω is the frequency. It controls how rapidly the curve oscillates.

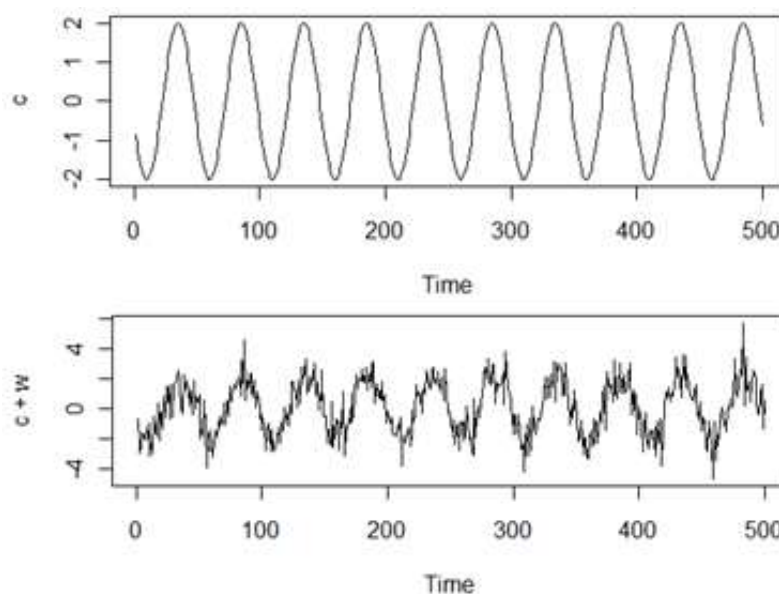
ϕ is the phase. It determines the starting point, in angle degrees, for the cosine wave.

To temporarily simplify things, suppose that $\phi=0$ and think about the quantity $2\pi\omega t$. Recall that T = number of time periods for a full cycle and that $\omega = 1/T$. As we move through time from $t = 0$ to $t = T$, the value of $2\pi\omega t = 2\pi t/T$ ranges from 0 at $t = 0$ to 2π at $t = T$. In angle degrees, this represents a full cycle of a cosine wave.

In Example 1.12 of the text, on pages 16-17, the authors give a plot of the function

$$x_t = 2 \cos\left(2\pi \frac{1}{50}t + 0.6\pi\right)$$

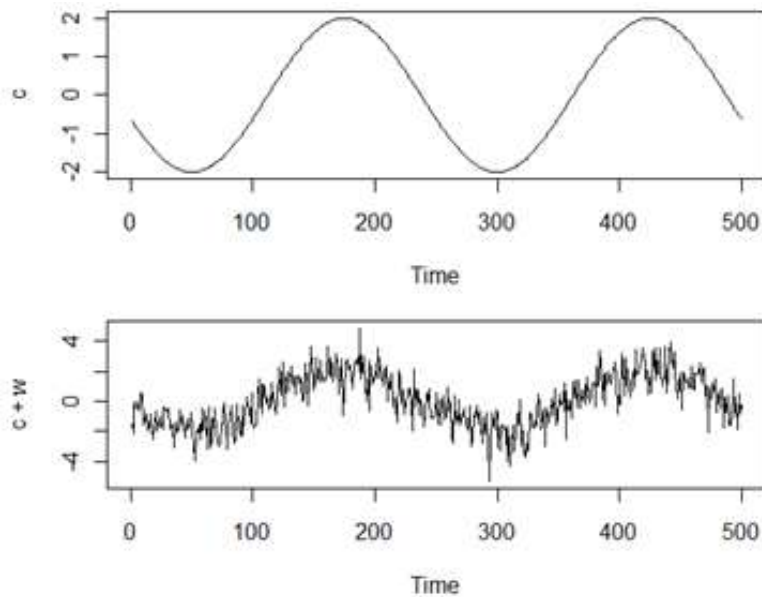
for $t = 1, 2, \dots, 500$. In addition they add normally distributed errors with mean 0 and variance 1 to this function in a second plot, and add normally distributed errors with mean 0 and variance 25 in a third plot. The R code is given on page 16. Following are the first two plots, the basic cosine function and the function plus errors with variance 1. In the basic plot, the period = 50 and the frequency is $1/50$. Thus it takes 50 time periods to cycle through the cosine function. Before errors are added, the maximum and minimum values are +2 and -2, respectively.



Following is what the plots look like when we change the period to 250 so that the frequency is $1/250 = 0.004$. The function is

$$x_t = 2 \cos\left(2\pi \frac{1}{250}t + 0.6\pi\right)$$

for $t = 1, 500$.



Notice above that longer period (250 for the second set of plots versus 50 in the first set of plots) leads to fewer cycles.

A Useful Identity

A useful trigonometric identity is

$$A \cos(2\pi\omega t + \phi) = \beta_1 \cos(2\pi\omega t) + \beta_2 \sin(2\pi\omega t),$$

with $\beta_1 = A \cos(\phi)$ and $\beta_2 = -A \sin(\phi)$.

This identity is used when we determine the *periodogram* of a series.

The Periodogram

In the area of time series called spectral analysis, we view a time series as a sum of cosine waves with varying amplitudes and frequencies. One goal of an analysis is to identify the important frequencies (or periods) in the observed series. A starting tool for doing this is the periodogram. The periodogram graphs a measure of the relative importance of possible frequency values that might explain the oscillation pattern of the observed data.

Suppose that we have observed data at n distinct time points, and for convenience we assume that n is even. Our goal is to identify important frequencies in the data. To pursue the investigation, we consider the set of possible *frequencies* $\omega_j = j/n$ for $j = 1, 2, \dots, n/2$. These are called the harmonic frequencies.

We will represent the time series as

$$x_t = \sum_{j=1}^{n/2} \left[\beta_1 \left(\frac{j}{n} \right) \cos(2\pi\omega_j t) + \beta_2 \left(\frac{j}{n} \right) \sin(2\pi\omega_j t) \right].$$

This is a sum of sin and cosine functions at the harmonic frequencies. The form of the equation comes from the identity given above in the section entitled “A Useful Identity”).

Think of the $\beta_1(j/n)$ and $\beta_2(j/n)$ as regression parameters. Then there are a total of n parameters because we let j move from 1 to $n/2$. This means that we have n data points and n parameters, so the fit of this regression model will be exact.

The first step in the creation of the periodogram is the estimation of the $\beta_1(j/n)$ and $\beta_2(j/n)$ parameters. It's actually not necessary to carry out this regression to estimate the $\beta_1(j/n)$ and $\beta_2(j/n)$ parameters. Instead a mathematical device called the Fast Fourier Transform (FFT) is used. We'll skip the details of that – it's advanced Calculus, something not required for this course.

After the parameters have been estimated, we define

$$P \left(\frac{j}{n} \right) = \hat{\beta}_1^2 \left(\frac{j}{n} \right) + \hat{\beta}_2^2 \left(\frac{j}{n} \right)$$

This is the value of the sum of squared “regression” coefficients at the frequency j/n .

This is the periodogram value at the frequency j/n , although the authors of our textbook (on page 69) say they will call this the scaled periodogram value. Thus, for them the scaled periodogram is a plot of $P(j/n)$ versus j/n for $j = 1, 2, \dots, n/2$.

Interpretation and Use:

A relatively large value of $P(j/n)$ indicates relatively more importance for the frequency j/n (or near j/n) in explaining the oscillation in the observed series. $P(j/n)$ is proportional to the squared correlation between the observed series and a cosine wave with frequency j/n . The dominant frequencies might be used to fit cosine (or sine) waves to the data, or might be used simply to describe the important periodicities in the series.

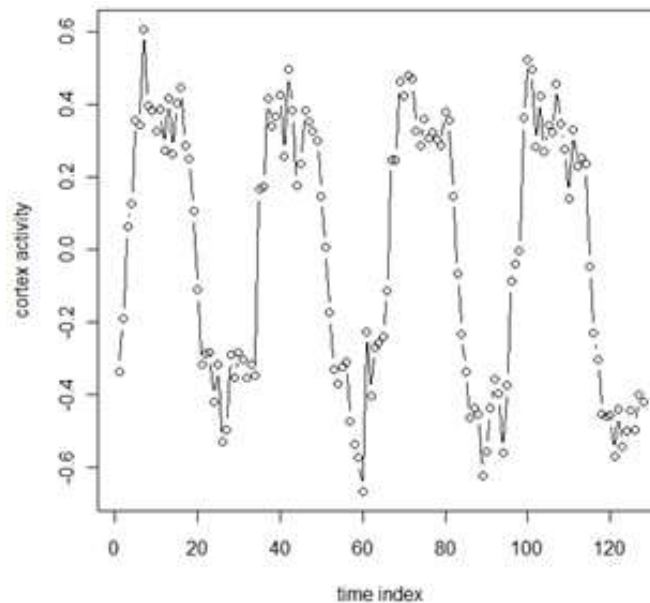
Some R Issues

The Fast Fourier Transform in R doesn't quite give a direct estimate of the scaled periodogram. A small bit of scaling has to be done (and the FFT produces estimates at more frequencies than we need). These things are easy to fix. Some code is given on page 70 of the book for simulated data. The first two lines of the code on page 70 generate simulated data, so those lines won't be there when you analyze an actual dataset.

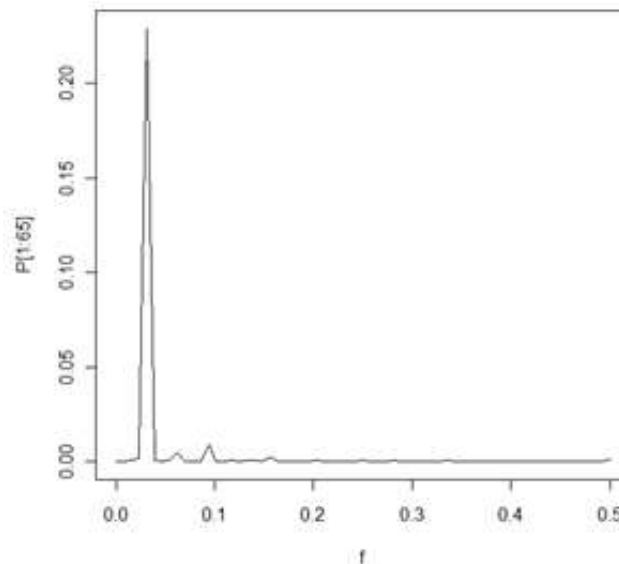
Example 1: The series is $n = 128$ values of brain cortex activity, measured every 2 seconds for 256 seconds. A stimulus, brushing of the back of the hand, was applied for 32 seconds and then was stopped for 32 seconds. This pattern was repeated for a total of 256 seconds. The series is actually the average of this process for five different subjects.

A time series plot follows. We see a regularly repeating pattern that seems to repeat about

every 30 or so time periods. This may not be surprising. The stimulus was applied for 16 time periods (of 2 seconds) and not applied for another 16 time periods (of 2 seconds). So, we might expect a repeating pattern every $16+16 = 32$ time periods.



The periodogram shows a dominant spike at a low frequency –



It's hard to judge the exact location of the peak. It would help to print out the first few values of the periodogram and the frequencies. The first 16 scaled periodogram values and frequencies follow. The peak value of periodogram is the fifth value, and that corresponds to a frequency of 0.0312500. The period for this value = $1/0.0312500 = 32$. That is, it takes 32 time periods for a complete cycle. This is what we expected!

```
> f
```

```
[1] 0.0000000 0.0078125 0.0156250 0.0234375 0.0312500 0.0390625
0.0468750 0.0546875
```

```
[9] 0.0625000 0.0703125 0.0781250 0.0859375 0.0937500 0.1015625
```

```
0.1093750 0.1171875
```

```
> round(P, 7)
```

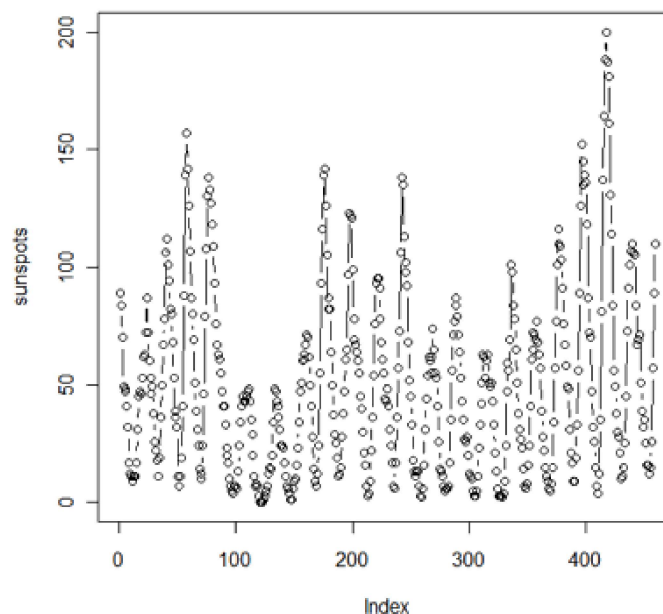
```
[1] 0.0002098 0.0001267 0.0006303 0.0019079 0.2290052 0.0008480
0.0002474 0.0006253
```

```
[9] 0.0046971 0.0004434 0.0000723 0.0004260 0.0092948 0.0002298
0.0003149 0.0007845
```

Here's the code. The value 128 in a few of the lines is the sample size. The fast Fourier Transform (fft) includes a frequency value of 0 and goes all the way to 1. The second line, creates the basis for what our textbook authors call the unscaled periodogram. We only need to go to 0.5 for the periodogram so in the third line below we pick off the first $(n/2)+1 = 65$ elements of FF (the object created in the second line). Also in the third line, the multiplication by $4/128 (=4/n)$ creates the scaled periodogram as described on pages 68-70 of the book. In a sense, this scaling is optional because we would see the same pattern even if we did not use this constant.

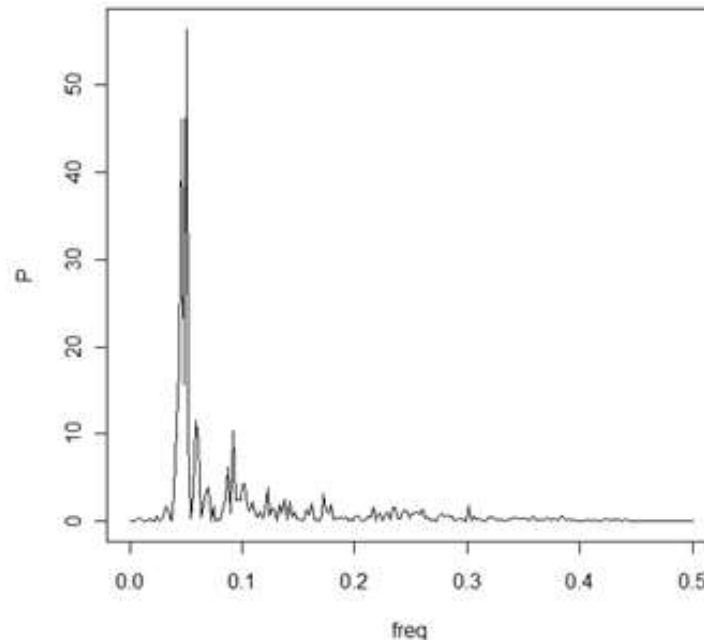
```
x = scan("cortex.dat")
FF = abs(fft(x)/sqrt(128))^2
P = (4/128)*FF[1:65] # Only need the first (n/2)+1 values of the FFT result.
f = (0:64)/128 # this creates harmonic frequencies from 0 to .5 in steps of
1/128.
plot(f, P, type="l") # This plots the periodogram; type = "l" creates a line
plot. Note: l is lowercase L, not number 1.
```

Example 2: The series is semi-annual sunspot activity (smoothed number of sunspots) for $n = 459$ time periods. This is semi-annual data, so this is $459/2 = 229.5$ years worth of data. The following time series plot shows ups and downs, but it's hard to judge the span(s) of any regular periodicity.



The odd number of points creates a minor problem as the periodogram work was described in terms of an even number of points. There are a few possibilities for dealing with this. We might set aside a data value from one end of the series or pad the series with something like the mean value for the series. All tries I made along this path resulted in a periodogram that was overwhelmingly dominated by the lowest possible frequency. Generally this is a sign that trend may be present. It's generally best to detrend the data either using differences or some sort of trend line before determining the periodogram. The differenced data (for this example) has the added benefit of creating a sample size that's even - with $n = 459$ data values there are 458 differences.

Following is the periodogram of the differenced (detrended) data.



The dominant peak area occurs somewhere around a frequency of 0.05. Investigation of the periodogram values indicates that the peak occurs at nearly exactly this frequency. This corresponds to a period of about $1/0.05 = 20$ time periods. That's 10 years, since this is semi-annual data. Thus there appears to be a dominant periodicity of about 10 years in sunspot activity.

```
sunspots=scan("sunspots.dat")
plot(sunspots,type="b")
x = diff(sunspots)
I = abs(fft(x)/sqrt(458))^2
P = (4/458)*I[1:230]
freq = (0:229)/458
plot(freq,P,type="l")
```

Spectral Analysis

Spectral analysis, described in Chapter 4 of our textbook, is the analysis of the dominant frequencies in a time series. In practice, spectral analysis imposes smoothing techniques on the periodogram. With certain assumptions, we can also create confidence intervals to estimate the peak frequency regions. Spectral analysis can also be used to examine the association between two different time series. We'll look at some of these techniques later in

the course.

Source URL: <https://onlinecourses.science.psu.edu/stat510/node/52>