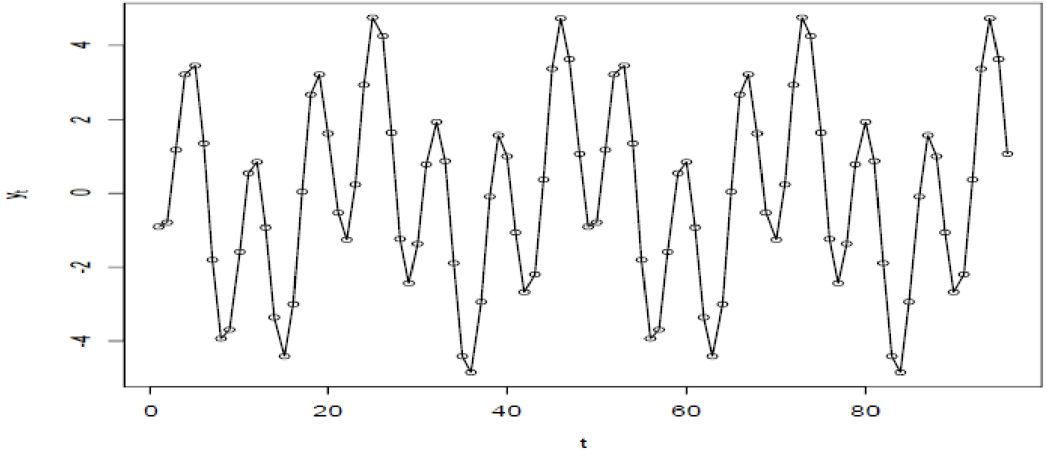
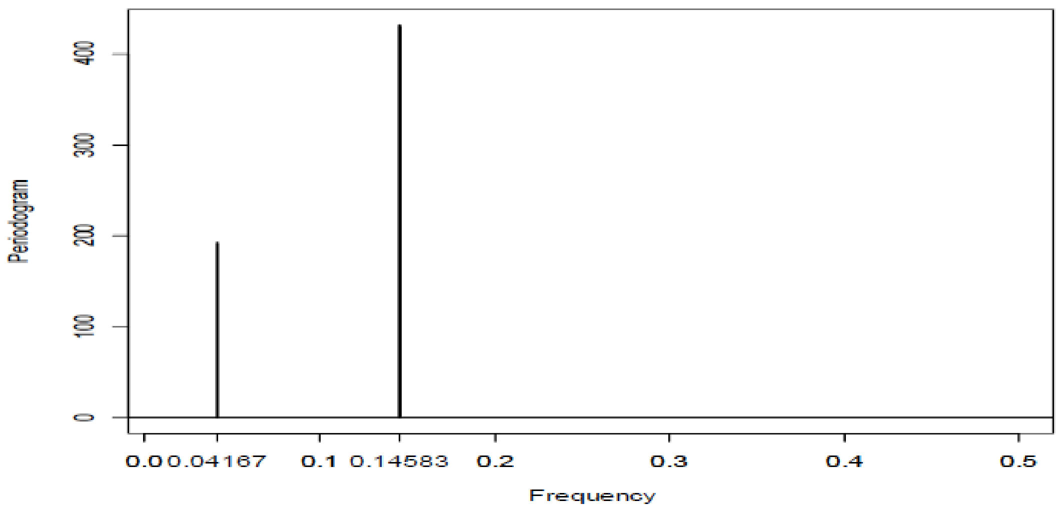
9장 Spectral 분석

Time Domain vs. Frequency Domain

- Time domain: $\{x, y\} = \{time,\, y_t\}$



- Frequency domain: $\{x, y\} = \{frequency, periodogram\}$



Fourier Transforms and Fourier Series

- Let $A(\lambda), -\infty < \lambda < \infty$ be a complex-valued function of period 2π with $\int_{-\pi}^{\pi} |A(\lambda)| d\lambda < \infty$, then

$$A(\lambda) = \sum_{u=-\infty}^{\infty} a_u \exp\{-i\lambda u\}, \qquad (1)$$

where $a_u = \dfrac{1}{2\pi} \displaystyle\int_{-\pi}^{\pi} \exp\{-i\lambda u\} A(\lambda) d\lambda$

Note: $\exp\{ix\} = \cos x + i \sin x$

Spectral Density

- If a process $\{X_t\}_{t=-\infty}^{\infty}$ is weakly stationary with zero mean, the key quantity in the time domain analysis is the autocovariance sequence

$$\gamma_h = \mathrm{cov}(X_t, X_{t+h}) = E(X_t, X_{t+h}) = E(X_t X_{t-h}) = \gamma_{-h},$$

$$\text{where } \; h = 0, \pm 1, \pm 2, \ldots.$$

- Assuming that our data, $x_0, \ldots, x_{n-1}$, form a piece of realization of this process, the traditional estimate of γ_h is the sample autocovariance,

$$\hat{\gamma}_h = \frac{1}{n} \sum_{t=|h|}^{n-1} x_i x_{t-|h|}$$

- In frequency domain analysis, the key quantity is the spectrum (or the spectral density),

$$f(w) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma_h e^{ihw}$$

Following from (1) $\quad \gamma_h = \int_{-\pi}^{\pi} e^{ihw} f(w) dw$

Spectral Density & Periodogram

- An estimate of $f(w)$ is the periodogram,

$$I(w) = \frac{1}{2\pi} \sum_{h=-(n-1)}^{n-1} \hat{\gamma}_h \exp(-iwh) = \frac{1}{2\pi n} \left| \sum_{t=0}^{n-1} x_t \exp(-iwt) \right|^2$$

- It can be shown that

$$\hat{\gamma}_h = \int_{-\pi}^{\pi} e^{ihw} I(w) dw$$

Periodicity

- $\{x_i\}_{t=0}^{n-1}$, a time series, is any sequence of data values observed at particular values of time, t. In some cases we know a collection of periods that may be present, and we have to find the associated amplitudes and phases.

- **Eg.** Pure cosine wave, $x_i = R\cos(wt+\phi)$

  Define

  Period = # Time Units/Complete Cycle= $2\pi / w$

  Frequency= # Cycles/Unit Time= $w/2\pi$ =1/Period

  Angular frequency= # Radians/Unit Time

  $= 2\pi \cdot$frequency= w

  R is amplitude and ϕ is the phase.

  Determine Period, Frequency, Angular frequency, R of the following cosine wave.

Periodicity

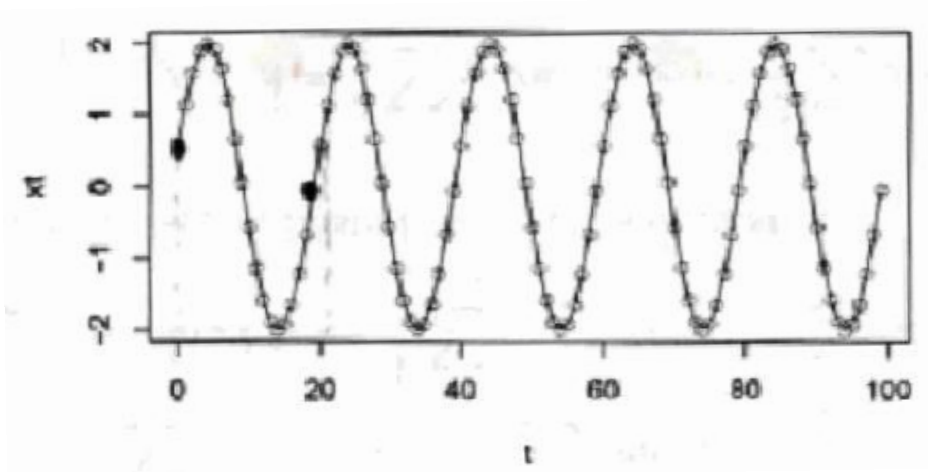


Figure 1: $x_t = R\cos(\omega t + \phi)$

- The general $x_t = R\cos(wt+\phi)$ can be written as

$$x_t = A\cos wt + B\sin wt$$

where $A = R\cos\phi$ $\quad B = -R\sin\phi$

$R = (A^2 + B^2)^{1/2}$ $\quad \phi = \tan^{-1}(-B/A)$

- **Eg.** Cosine wave +Noise

$$y_t = \mu + R\cos(wt+\phi) + \varepsilon_t = \mu + A\cos wt + B\sin wt + \varepsilon_t$$

where ε_t are uncorrelated random variables with mean zero and variance σ^2

Periodicity

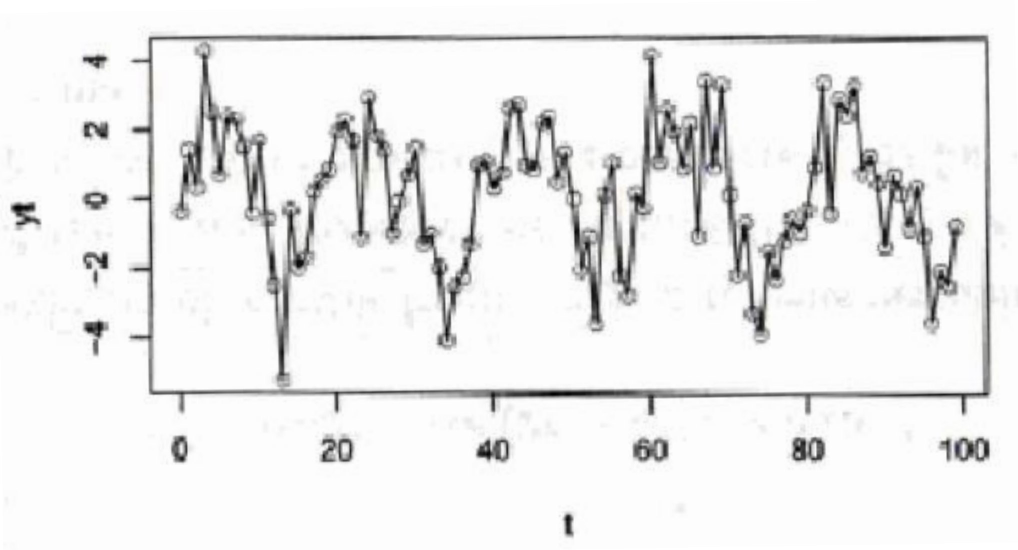


Figure 2: $y_t = \mu + R\cos(\omega t + \phi) + \epsilon_t$

- Suppose the frequency w is regarded as *known,* the least squares estimates of the parameters μ , A, B are solutions to the normal equations

$$\sum (y_t - \mu - A\cos wt - B\sin wt) = 0$$

$$\sum \cos wt (y_t - \mu - A\cos wt - B\sin wt) = 0$$

$$\sum \sin wt (y_t - \mu - A\cos wt - B\sin wt) = 0$$

Periodicity

- Using the following approximations

$$\sum(\cos wt)^2 \approx \frac{n}{2}, \qquad \sum(\sin wt)^2 \approx \frac{n}{2}$$

$$\frac{1}{n}\sum \cos wt \sin wt \approx 0, \qquad \frac{1}{n}\sum \cos wt \approx 0, \qquad \frac{1}{n}\sum \sin wt \approx 0$$

- We obtain the approximate least squares estimations

$$\hat{\mu} = \frac{1}{n}\sum y_t, \qquad \hat{A} = \frac{2}{n}\sum y_t \cos wt, \qquad \hat{B} = \frac{2}{n}\sum y_t \sin wt$$

Multiple Periodicity

- **Eg.** Multiple periodicities a more general model is

$$z_t = \mu - \sum_{u=1}^{m}(A_u \cos w_u t + B_u \sin w_u t) + \varepsilon_t$$

Alias

- Every frequency, w not in the range $[0,\pi]$ has an alias $w' \in [0,\pi]$.
- The frequencies w and w' are said to be aliases of each other, and w' is the principal alias.
- **Eg.** Suppose that $x_t = \cos wt$, $w = 2\pi - w'$ where $w' \in [0,\pi]$

$$x_t = \cos wt = \cos(2\pi - w')t = \cos w't$$

- For example, let us assume we have two cosine curves, one with frequency 1/4 and the other with 3/4. If we only observe the series at the discrete-time points 1,2,…,5, two series seem to be identical. With discrete-time observations, we could never distinguish between these two curves. We say that the two frequencies are aliased with on another.
- The maximum observable frequency is π radians per unit time. It is called the folding frequency or Nyquist frequency. A sinusoid at the Nyquist frequency executes 1 complete cycle for every two units of time.

Alias

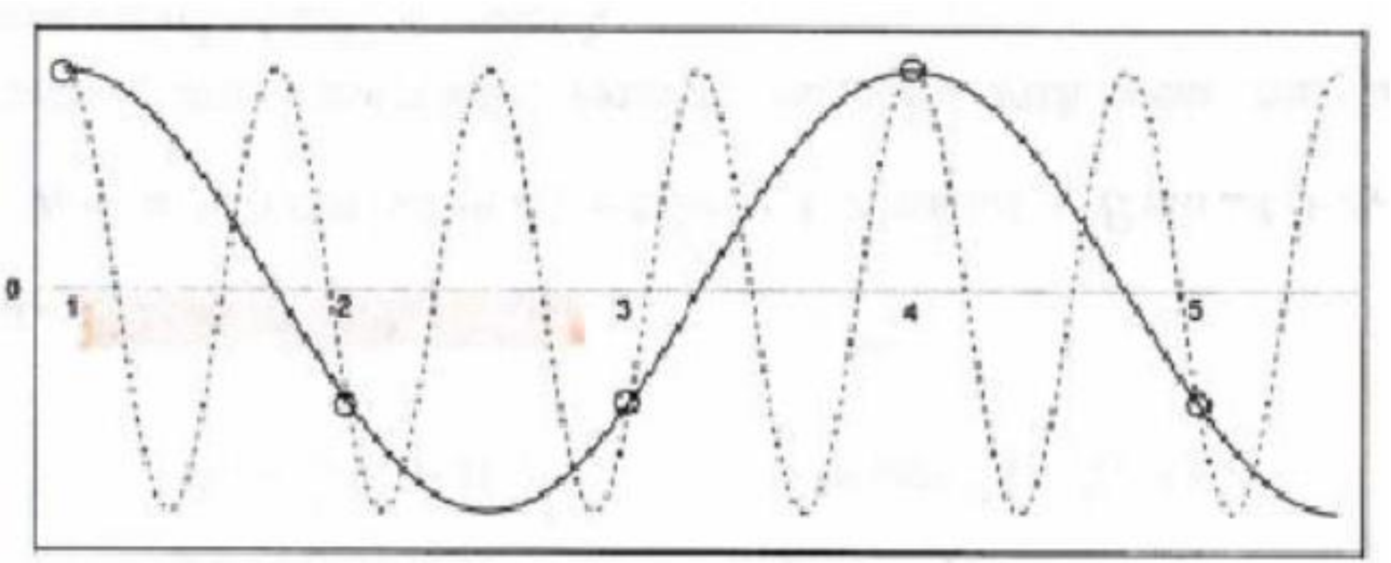


Figure 3: The aliasing effect

Discrete Fourier Transform

- $\{x_t\}_{t=0}^{n-1}$, a time series, is any sequence of data values observed at particular values of time t.

- Define DFT of $\{x_t\}_{t=0}^{n-1}$

$$J(w) = \frac{1}{n}\sum_{t=0}^{n-1} x_t \exp(-iwt) \qquad (2)$$

Periodogram

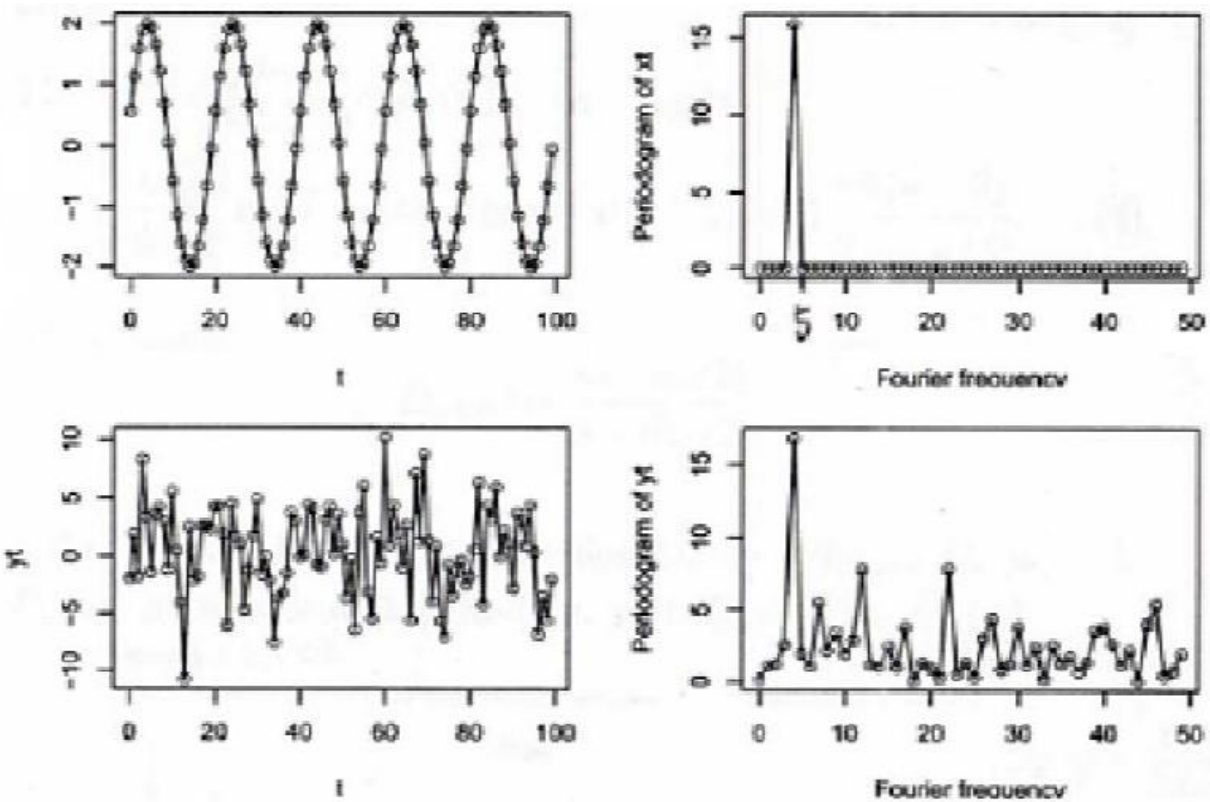
- The Periodogram is defined by $$I(w) = \frac{n}{2\pi}|J(w)|^2$$

- Note that $$\sum_{t=0}^{n-1}|x_t|^2 = 2\pi\sum_{j=0}^{n-1} I(w_j)$$

- In general, a peak in the periodogram at a given frequency indicates a strong harmonic component in the data $\{x_t\}$ at that frequency. In Handout 1, if w in the model for y_t is unknown, it can be estimated by maximizing I (w) over all frequencies. We need the distribution of the periodogram under the null where there are no underlying periodic components to declare we have found a "true" cycle. We will discuss the test for the periodicity later.

Periodogram

Graphical Representation



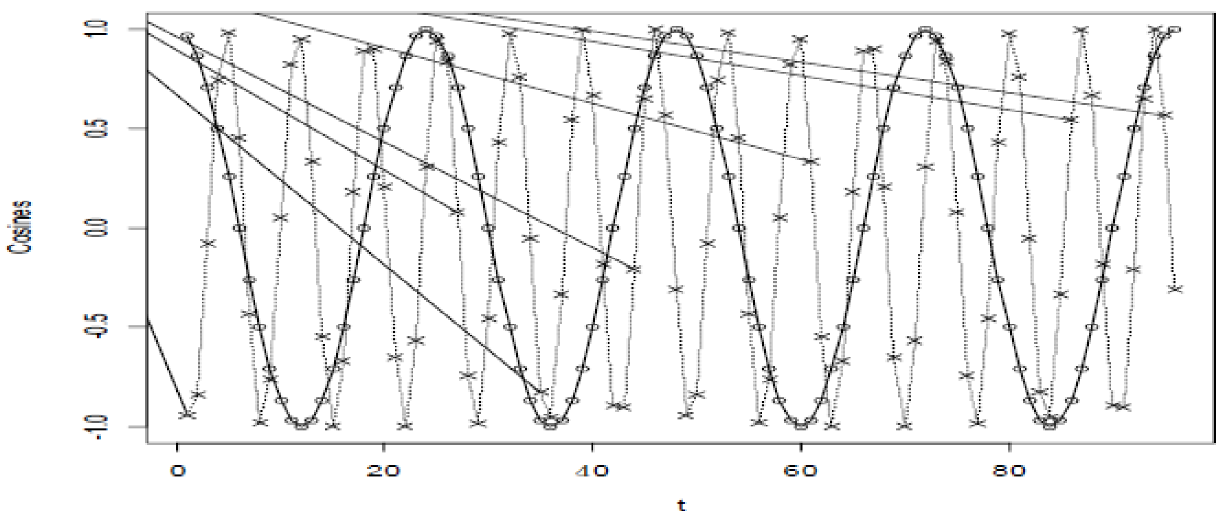
$$x_t = R\cos(wt+\phi) = A\cos wt + B\sin wt$$

$$y_t = R\cos(wt+\phi) + \varepsilon_t = A\cos wt + B\sin wt + \varepsilon_t$$

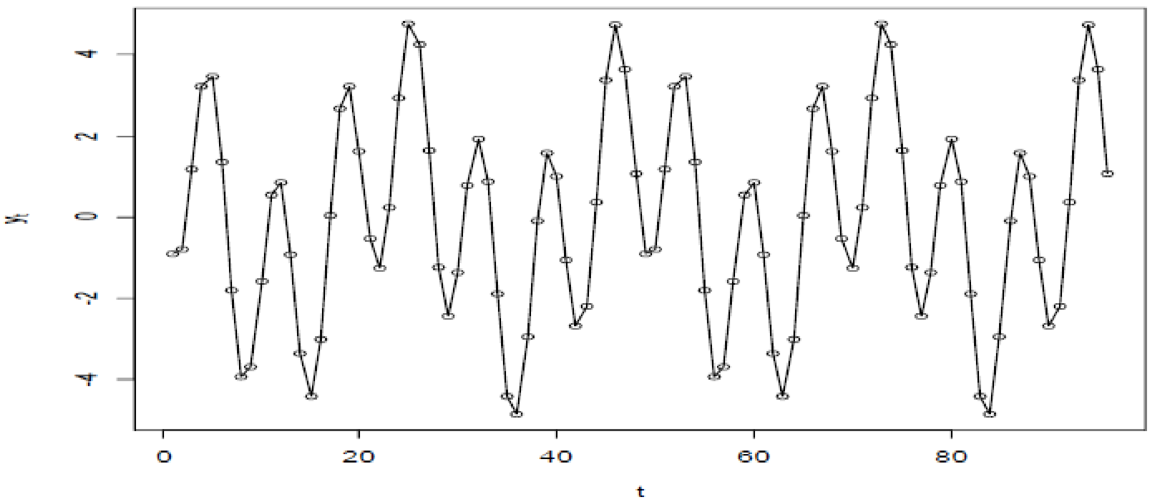
Examples using R-code

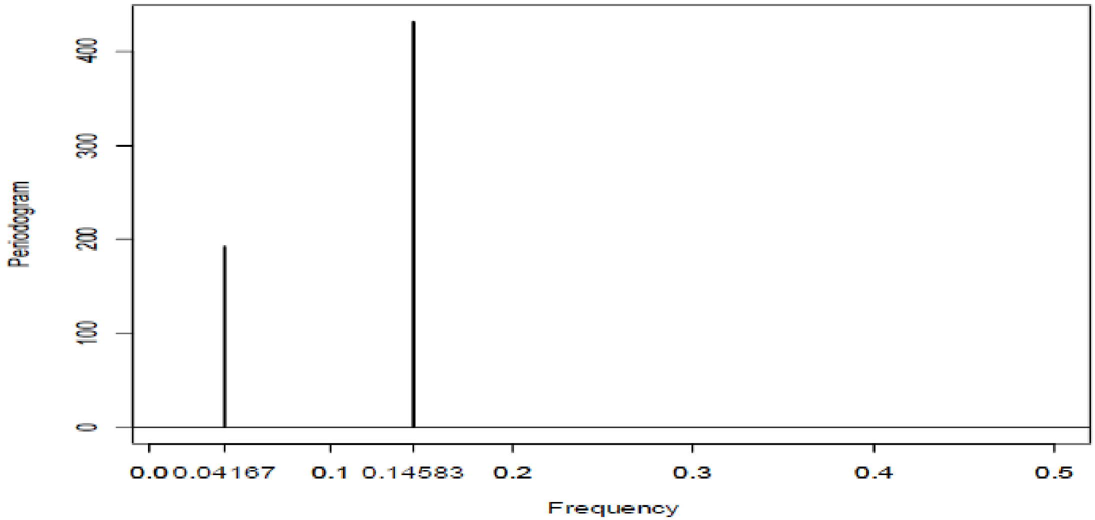
- **Example**: Multiple periodicity: linear combination of two cosine curves

```
t=1:96;cos1=cos(2*pi*t*4/96); cos2=cos(2*pi*(t*14/96+.3))
plot(t,cos1, type='o', ylab='Cosines');lines(t,cos2, lty='dotted',type='o',pch=4)
y=2*cos1+3*cos2;
plot(t,y,type='o',ylab=expression(y[t]))
periodogram(y); abline(h=0);axis(1,at=c(0.04167,.14583))
```



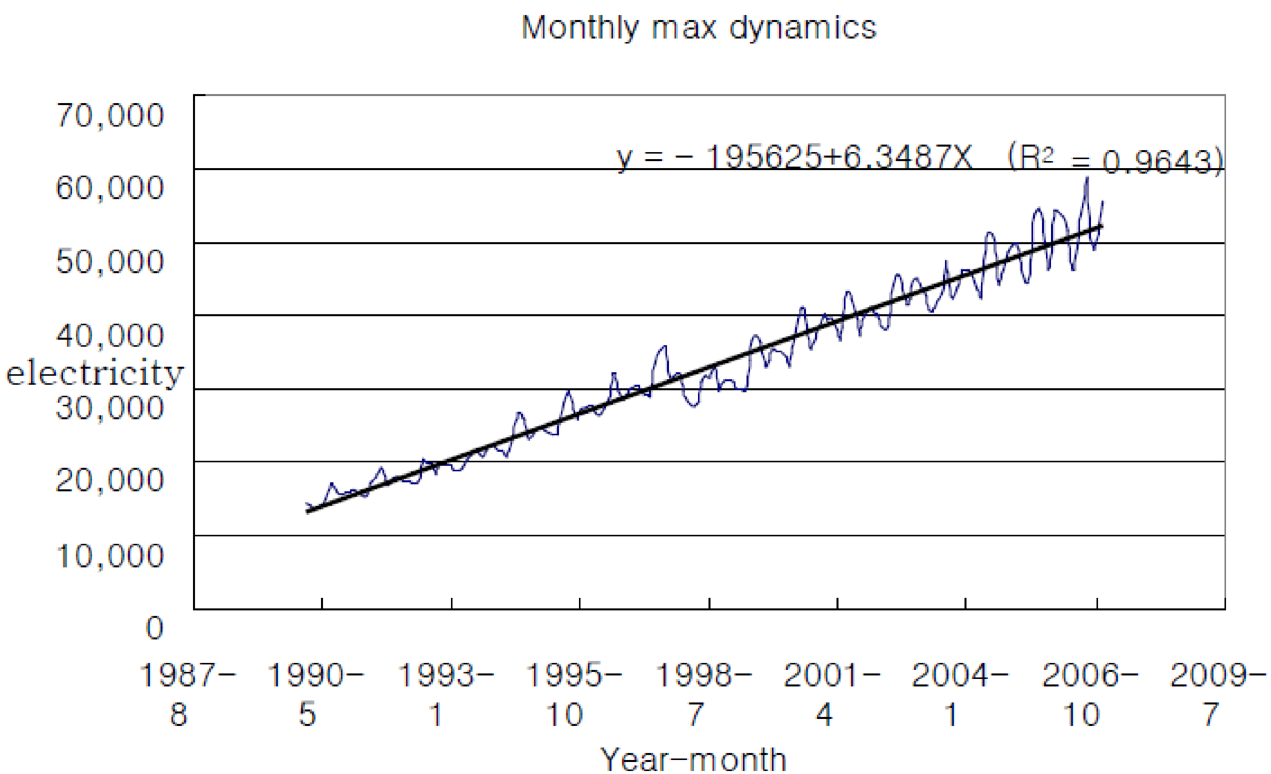
Examples using R-code





Case study

- Data can be downloaded from http://www.kpx.or.kr/. Data are monthly maximum electricity dynamics ranged from 1990.1 to 2006. 12. Total number of observation is 204.



<Fig. 1> monthly max dynamics (thin line) and estimated trend line (thick line).

Case study

- **Step 1:** Remove a trend if it exists as follows:

$$e_i = y_i - \hat{y}_i, \qquad\qquad i = 1, 2, \ldots, n$$

where y_i is observed dynamics, $\hat{y}_i = b_0 + b_1 x_i$ is estimated trend (regression). According to <Fig. 1>, ordinary least square estimators are computed: $b_0 = 195625$, $b_1 = 6.3487$. Obtained R^2 value is 96.43%.

- **Step 2:** Using the residual, carry out the spectral analysis.

Draw its periodogram. Find a biggest peak, then look for a corresponding period. That will be a dominant period.

```
proc spectra data=totdat out=ft0 p adjmean whitetest;var resid;run;
proc gplot data=ft0;symbol i=spline v=circle h=2;plot p_01*freq;run; proc print
data=ft0;
run;
```

Case study

- SAS/ETS provides spectral analysis. Some output will be as follows:

Test for White Noise for Variable resid

M-1	101
Max(P(*))	3.2707E8
Sum(P(*))	9.7709E8

Fisher's Kappa: (M-1)*Max(P(*))/Sum(P(*))

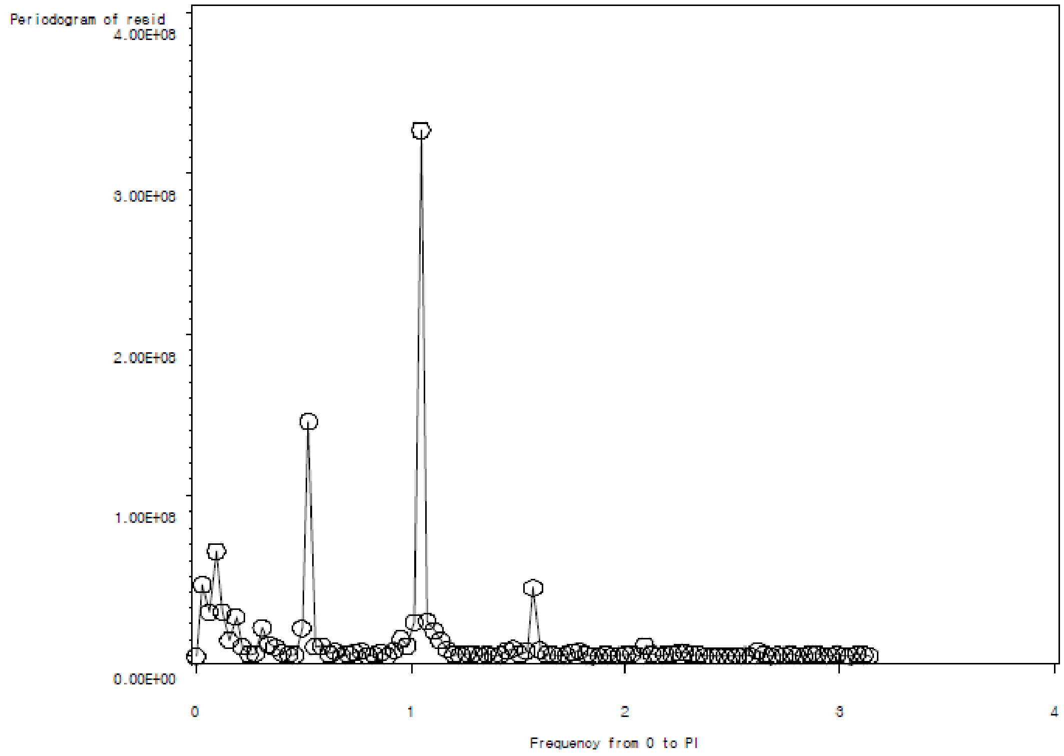
Kappa	33.80839

Bartlett's Kolmogorov-Smirnov Statistic:
Maximum absolute difference of the standardized partial sums of the periodogram and the CDF of a uniform(0,1) random variable.

Test Statistic	0.51272
Approximate P-Value	<.0001

Case study

- 'Test for White Noise' is testing sequentially uncorrelateness (autocorrelation) for 'resid'. If the test result is significant, then it is possible to have a significant period for the considered dynamics.



<Fig. 2> Periodogram.

Case study

- Y-axis is Periodogram and X-axis is Angular frequency in <Fig. 2>. Maximum periodogram seems to appear around 1 of angular frequency. Second biggest periodogram appear around angular frequency 0.5. A corresponding period for the largest peak can be obtained from <Fig. 3>, which implies the dominant period. Here period 6 implies 6 months. Therefore, the dominant period is 6 month.

OBS	FREQ	PERIOD	P_01
1	0.00000	.	0.00
2	0.03080	204.000	44519043.90
		.	
		.	
17	0.49280	12.750	17542312.75
18	0.52360	12.000	145743846.09
		.	
34	1.01640	6.182	21293455.26
35	1.04720	6.000	327067163.34
		.	
52	1.57080	4.000	42515465.32

<Fig. 3> Periodogram and corresponding period