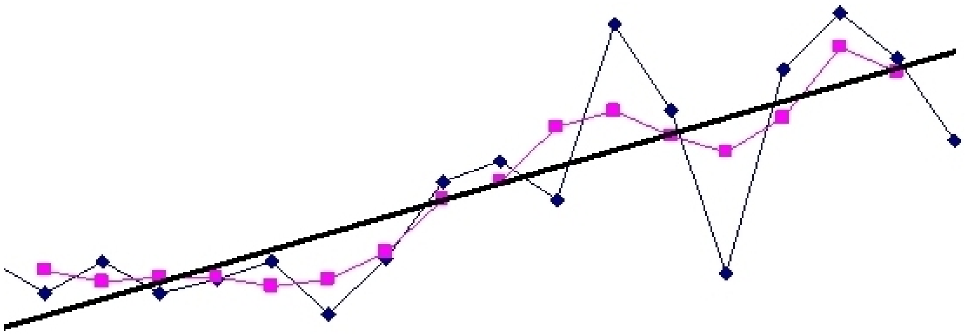
Author: **Tibor Nyári** PhD

University of Szeged
Department of Medical Physics and Informatics

www.model.u-szeged.hu
www.szote.u-szeged.hu/dmi

Tests for cyclic variation Sinusoidal curve fitting



Phase Shift

- The general formula for sinusoidal function with one cycle is :
- $Y=A\sin(\omega x-\Phi)=A\sin[\omega(x-\Phi/\omega)]$, where $\omega>0$ and $\Phi\in R$(real numbers)
- The number of Φ/ω is called the phase shift of the graph $Y=A\sin(\omega x-\Phi)$.

Period of this function

- A period begin when
  - $\omega x - \Phi = 0$
- and will end when
  - $\omega x - \Phi = 2\pi$

- For the graph $Y = A\sin(\omega x - \Phi)$ the period is

$$Period\ \ T = \frac{2\pi}{\omega}, \qquad Phase\ \ shift = \frac{\Phi}{\omega},\ \ Amplitude\ = |A|$$

Characteristics of sine function

- Domain: real numbers ($x \in R$)
- Range: $-1 \leq y \leq +1 \mid y \in R$

Finding sinusoidal function from data

- To fit the data to a sine function of the form:
  - $y=A\,sin(\omega x-\Phi)+B$
- where A, B, ω and Φ are constants.

$$Amplitude \quad A = \frac{largest\ data\ value - smallest\ data\ value}{2}$$

$$Vertical\ shift \quad B = \frac{largest\ data\ value + smallest\ data\ value}{2}$$

Example

- The data given in this rable represent the average monthly temperatures in Denver, Colorado.

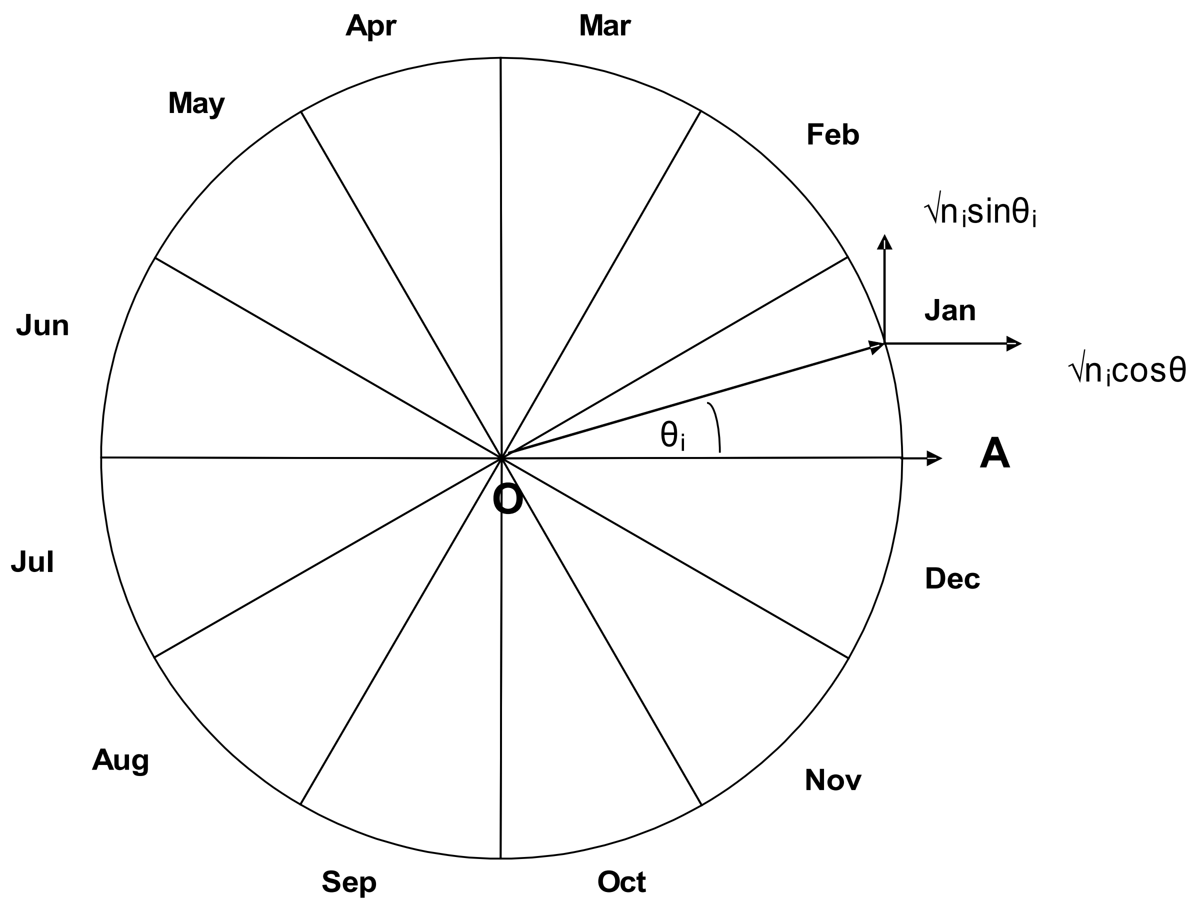
Statistical methods

- Four methods employed for the detection of seasonal variations in epidemiological data are described:
  - Edwards' method,
  - Walter-Elwood's method,
  - Logistic regression (Stolwijk et al) including periodic functions (a sine and a cosine function, simultaneously).
  - Cosinor (linear regression) model

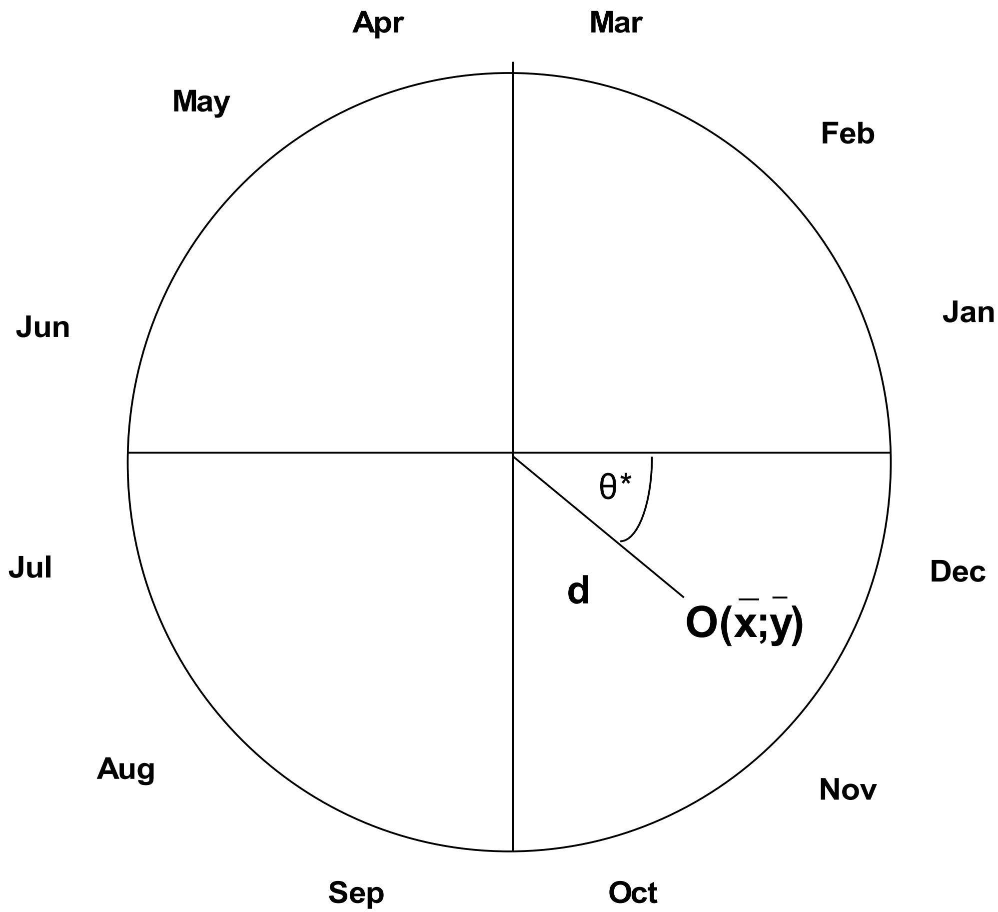
Edwards' method

- Edwards JH. The recognition and estimation of cyclic trends. Ann Hum Genet Lond 1961;25:83–87

 8


Apr
Mar
May
Feb
$\sqrt{n_i}\sin\theta_i$
Jan
Jun
$\sqrt{n_i}\cos\theta_i$
θ_i
A
O
Jul
Dec
Aug
Nov
Sep
Oct

- A simple test for cyclic trend in independent events is presented in the form of the rim of an unit circle divided into equal sector, corresponding to time intervals, and a number in each rim-sector specifying the number of events observed
- In the absence of any cyclic trend the expected centre of gravity of these masses will be at the centre of the circle. Any excess of deficit in neighbouring sectors will have consistent effect on the position of the centre of gravity, and whose distance from the centre will have a probability distribution on the null hypothesis, and whose direction will indicate the position of maximum or minimum liability, or both.
- Consider N events distributed over k equal sectors (eg. 12 months);
- Let the number of events in sector ith n_i. Take square root of ni , then any sector contribution to the moment about any arbitrary diameter making an angle θ_i with $\sqrt{n_i} \sin \theta_i$



 11

$$\bar{x} = \frac{\sum_{i=1}^{k} \sqrt{n_i} \cos\theta_i}{k} \qquad \bar{y} = \frac{\sum_{i=1}^{k} \sqrt{n_i} \sin\theta_i}{k}$$

$$d = \sqrt{\bar{x}^2 + \bar{y}^2}$$

$$1 + \alpha \sin(\theta_i - \theta^*)$$

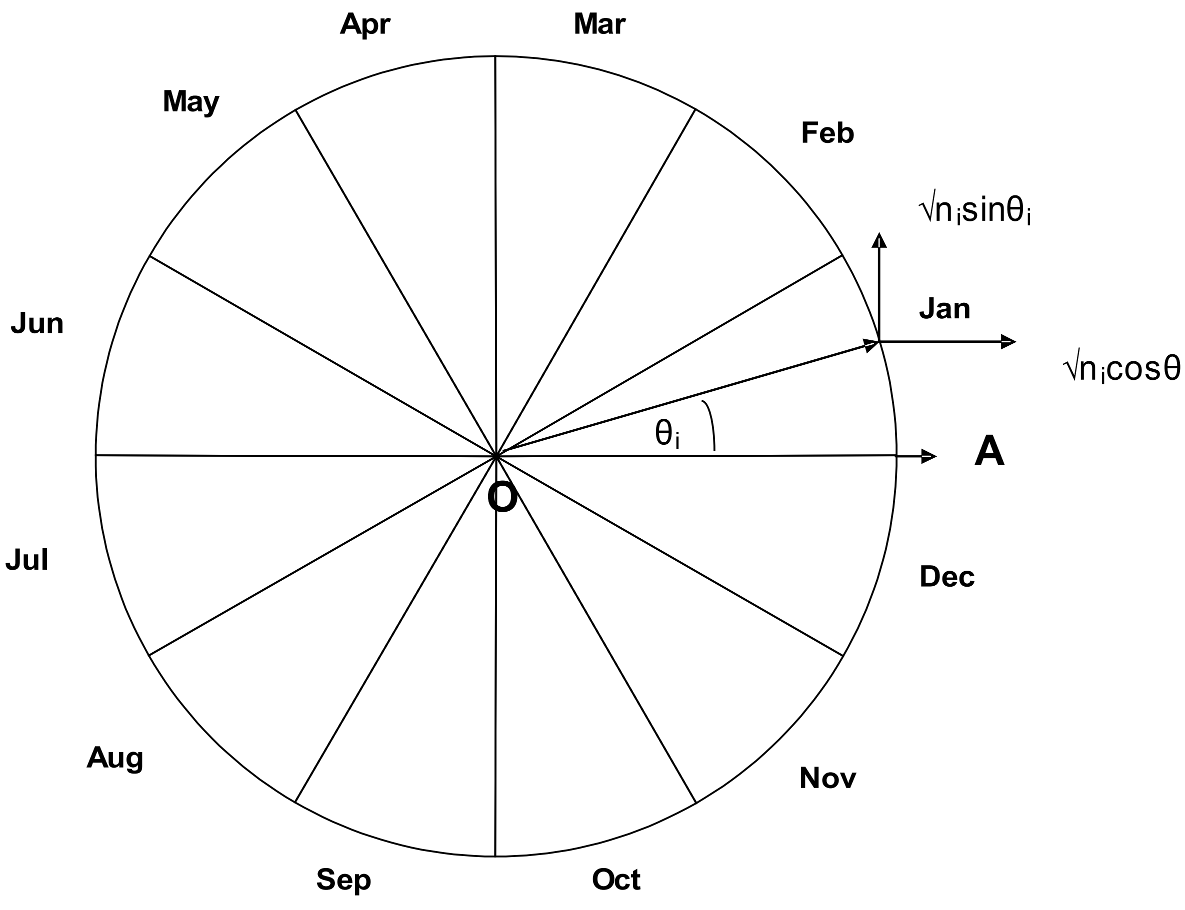
$$\alpha = 4d$$

$$\theta^* = \arctan\left(\frac{\bar{y}}{\bar{x}}\right)$$

Walter-Elwood's method

- Walter SD, Elwood JM. A test for seasonality of events with a variable population at risk. Br J Prev Soc Med 1975;29:18–21

13


Apr
Mar
May
Feb
√nᵢsinθᵢ
Jan
Jun
√nᵢcosθᵢ
θᵢ
A
O
Jul
Dec
Aug
Nov
Sep
Oct

14

Walter-Elwood

- We suppose that within certain time span (e.g.: a year) there are k sectors (e.g.: 12 months) and that in sector I there are ni events from a population at risk of size mi (e.g.: total births during that month). The total number of events is $N=\sum n_i$ and the total number of population at risk is $M=\sum m_i$.
- H0: The excepted number of events in a sector is proportional to the population at risk in that sector, i.e.: $E(n_i)=Nm_i/M$
- The data by weights $\sqrt{} n_i$ placed around a unit circle at points corresponding to the sector midpoints at angles θ i to an arbitrary diameter (e.g.: the diameter through 1 January).

 15

$$N = \sum_{i=1}^{k} n_i ; M = \sum_{i=1}^{k} m_i$$

$$\bar{x} = \frac{\sum_{i=1}^{k} \sqrt{n_i} \cos\theta_i}{\sum_{i=1}^{k} \sqrt{n_i}}$$

$$\bar{y} = \frac{\sum_{i=1}^{k} \sqrt{n_i} \sin\theta_i}{\sum_{i=1}^{k} \sqrt{n_i}}$$

$E(n_i)=Nm_i/M$

$$\mu_{\bar{x}} = \frac{\sum_{i=1}^{k} \sqrt{Nm_i / M} \cos\theta_i}{\sum_{i=1}^{k} \sqrt{Nm_i / M}} = \frac{\sum_{i=1}^{k} \sqrt{m_i} \cos\theta_i}{\sum_{i=1}^{k} \sqrt{m_i}}$$

$$\mu_{\bar{y}} = \frac{\sum_{i=1}^{k} \sqrt{m_i} \sin\theta_i}{\sum_{i=1}^{k} \sqrt{m_i}}$$

$$\sigma^2_{\bar{x}} = \frac{\sum_{i=1}^{k} \frac{1}{4} \cos^2\theta_i}{[\sum_{i=1}^{k} \sqrt{Nm_i / M}]^2}$$

$$\sigma^2_{\bar{y}} = \frac{\sum_{i=1}^{k} \frac{1}{4} \sin^2\theta_i}{[\sum_{i=1}^{k} \sqrt{Nm_i / M}]^2}$$

Test statistics

$$\left(\frac{\bar{x}-\mu_{\bar{x}}}{\sigma_{\bar{x}}}\right)^2+\left(\frac{\bar{y}-\mu_{\bar{y}}}{\sigma_{\bar{y}}}\right)^2$$

Which on the null hypothesis is distributed as χ2 with 2 d.f.

The distance d of the sample centre of gravity from its null expectation is given by

$$d=\sqrt{(\bar{x}-\mu_{\bar{x}})^2+(\bar{y}-\mu_{\bar{y}})^2}$$

 17

If it is required to fit a simple harmonic trend to the data, we may suppose that the expected frequency in sector i is proportional to

$$c_i = m_i\left[1 + \alpha \cos(\theta_i - \theta^*)\right]$$

$$\theta^* = \arctan\left(\frac{\bar{y} - \mu_{\bar{y}}}{\bar{x} - \mu_{\bar{x}}}\right)$$

$$\alpha = \frac{2[d\sqrt{(kM)} - \sum_{i=1}^{k}\sqrt{m_i}\cos(\theta_i - \theta^*)]}{\sum_{i=1}^{k}\sqrt{m_i}\cos^2(\theta_i - \theta^*)]}$$

 18

In the case $\theta_i = 2\pi i/k$, $m_i = M/k$, $i = 1, 2, ..k$ it may be shown that

$$\mu_{\bar{x}} = 0 \qquad \mu_{\bar{y}} = 0$$

The adequacy of the simple harmonic curve may be evaluated by a goodness-of-fit test using a further χ2 statistics (k-1 df):

$$\sum_{i=1}^{k} \left(n_i - n_i^{'}\right)^2 / n_i^{'}$$

$$n_i^{'} = N c_i / \sum_{i=1}^{k} c_i$$

20

Logistic regression

- Stolwijk AM, Straatman H and Zielhuis GA. Studying seasonality by using sine and cosine functions in regression analysis J Epidemiol Community Health 1999;53:235-38

A logistic regression model was developed to analyse seasonality. Such a model will have the following form:

$$\ln\left(\frac{P}{1-P}\right) = \beta_0 + \beta_{season} \text{ x season} + \beta_{C_1} \text{ x } C_1 + \ldots + \beta_{C_N} \text{ x } C_N$$

 22

- To define the variable “season” in these models, it is hypothesised that the seasonal pattern under study follows a cosine function with variable amplitude and horizontal shift. In this cosine function, two periods must be defined:
  - (i) the time period that defines the measure of malformation, for example, “month”
  - (ii) the period described by one cosine function.
- As an example we take “month” as the time period under study, and “one year” as the period of the cosine function.

 23

The cosine function can be described as:

$$f(t) = \alpha \text{ x } \cos\left[\left(\frac{2\pi \ t}{T}\right) - \theta\right]$$

T= number of time periods described by one cosine function over (0, 2π) (for example, T = 12 months);

t = time period (for example, for January: t = 1, for February: t = 2, etc);

α· = amplitude, > 0;

θ = horizontal shift of the cosine function (in radials).

 24

A trigonometrical identities

$$\cos(\alpha - \beta) = \cos\alpha \cos\beta + \sin\alpha \sin\beta$$

$$f(t) = \beta_1 \text{ x } \sin\left(\frac{2\pi t}{T}\right) + \beta_2 \text{ x } \cos\left(\frac{2\pi t}{T}\right)$$

$$\beta_1 = \alpha \sin\theta \qquad \beta_2 = \alpha \cos\theta$$

$$\ln\left(\frac{P}{1-P}\right) = \beta_0 + \beta_1 \text{ x } \sin\left(\frac{2\pi t}{T}\right) + \beta_2 \text{ x } \cos\left(\frac{2\pi t}{T}\right) + \beta_{C_1} \text{ x } C_1 + \ldots + \beta_{C_N} \text{ x } C_N$$

$$P_t = \frac{e^{\beta_0 + \beta_1 \text{ x } \sin\left(\frac{2\pi t}{T}\right) + \beta_2 \text{ x } \cos\left(\frac{2\pi t}{T}\right) + \beta_{C_1} \text{ x } C_1 + \ldots + \beta_{C_N} \text{ x } C_N}}{1 + e^{\beta_0 + \beta_1 \text{ x } \sin\left(\frac{2\pi t}{T}\right) + \beta_2 \text{ x } \cos\left(\frac{2\pi t}{T}\right) + \beta_{C_1} \text{ x } C_1 + \ldots + \beta_{C_N} \text{ x } C_N}}$$

 26

$$f(t) = \alpha \text{ x } \cos\left[\left(\frac{2\pi t}{T}\right) - \theta\right]$$

$$f(t) = \beta_1 \text{ x } \sin\left(\frac{2\pi t}{T}\right) + \beta_2 \text{ x } \cos\left(\frac{2\pi t}{T}\right)$$

$$\beta_1 = \alpha \sin\theta \qquad \beta_2 = \alpha \cos\theta$$

$$\alpha = \sqrt{\beta_1^2 + \beta_2^2}$$

 27

Two extreme values in (0,T) can be found at the solutions of :

$$\tan\left(\frac{2\pi t}{T}\right) = \frac{\beta_1}{\beta_2}$$

$$t = \arctan\left(\frac{\beta_1}{\beta_2}\right) \times \frac{T}{2\pi}$$

If $\beta 1/\beta 2 > 0$, then $t > 0$ and indicates the first extreme; the other extreme value is found at $t + T/2$.

If $\beta 1/\beta 2 < 0$, then $t < 0$; the extreme values are found at $t + T/2$ and at $t + T$.

If $\beta 1 > 0$, the first extreme is a maximum and the second a minimum; if $\beta 1 < 0$ it is the other way around.

The maximum extreme (t_{max}) indicates the shift θ, which can be calculated by:

$$\theta = \left(\frac{2\pi \, t_{\max}}{T} \right)$$

 29

Cosinor method

- A regression model was developed by Halberg et al to analyze seasonality. Such a model will have the following form:

$$y = \beta_0 + \beta_{season} \text{ x season} + \beta_{C_1} \text{ x } C_1 + \ldots + \beta_{C_N} \text{ x } C_N$$

- To define the variable "season" in these models, it is hypothesized that the seasonal pattern under study follows a cosine function with variable amplitude and horizontal shift.
- The cosine function can be described as:

$$f(t) = \alpha \text{ x } \cos\left[\left(\frac{2\pi t}{T}\right) - \theta\right]$$

- T= number of time periods described by one cosine function over $(0, 2\pi)$ (for example, T = 12 months);
- t = time period (for example, for January: t = 1, for February: t = 2, etc);
- α = amplitude, > 0;
- θ = horizontal shift of the cosine function (in radials).
- As θ is unknown, transformation of this cosine function is required before the regression analysis can be performed. Therefore the following formula is included into a regression model:

$$f(t) = \beta_1 \ \mathrm{x} \ \sin\left(\frac{2\pi t}{T}\right) + \beta_2 \ \mathrm{x} \ \cos\left(\frac{2\pi t}{T}\right)$$

 31

where

$$\beta_1 = \alpha \sin\theta \qquad \beta_2 = \alpha \cos\theta$$

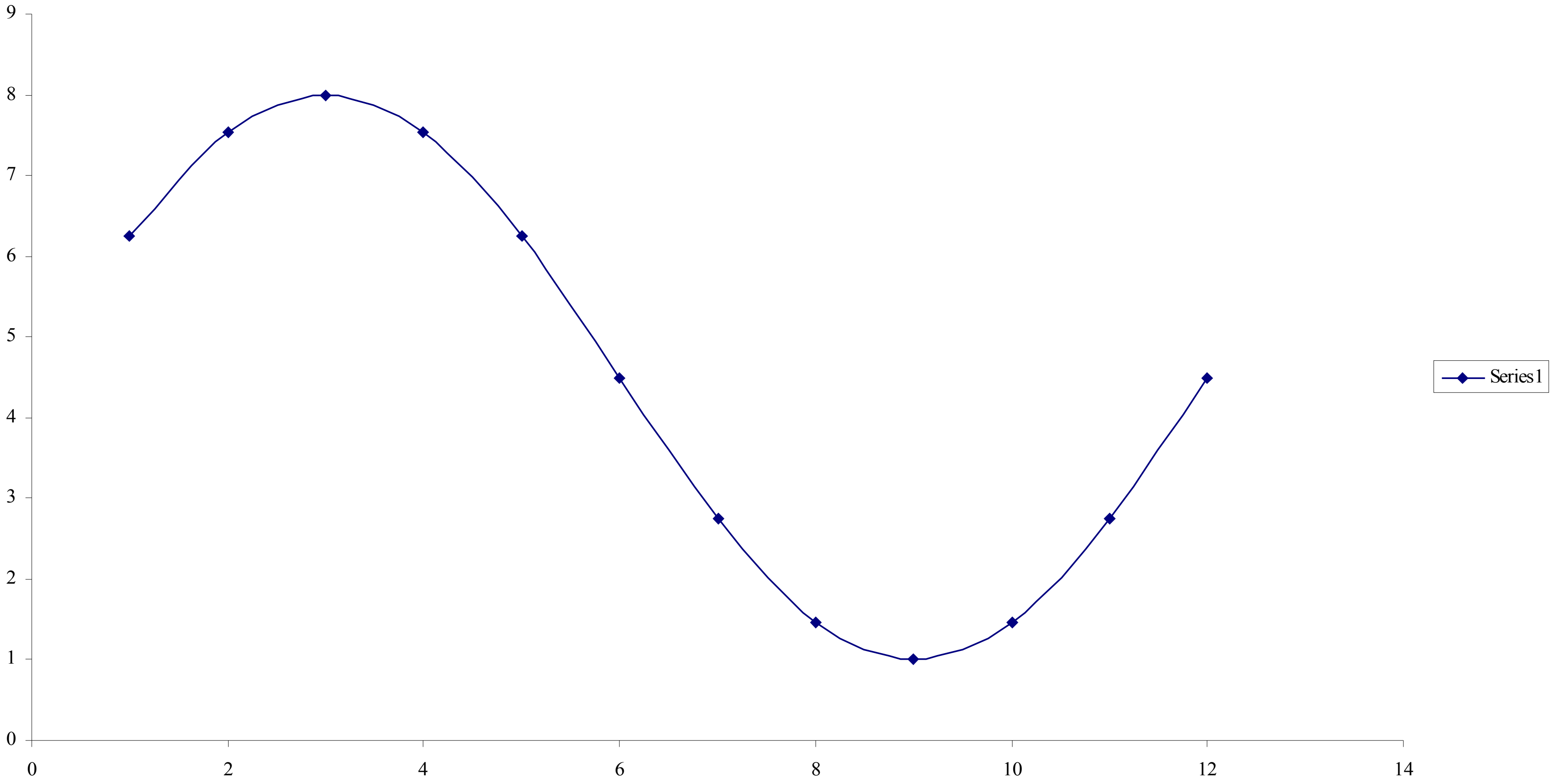
- The amplitude is: $\alpha = \sqrt{\beta_1^2 + \beta_2^2}$

- and two extreme values in (0,T) can be found at

$$t = \arctan\left(\frac{\beta_1}{\beta_2}\right) \text{ x } \frac{T}{2\pi}$$

32

Example



Example

- In the following table are given monthly frequencies of cases of anencephalus and total births for Canada in the period 1954-62 (Elwood, 1975). A cursory inspection of the data reveals a distinct general excess of total births during the summer months whereas the anenchephalus cases demonstrate no consistent seasonal patern.

Data table

Month	Anecephalus	Total birth
1	468	340797
2	399	318319
3	471	363626
4	437	359689
5	376	373878
6	410	361290
7	399	368867
8	472	358531
9	418	363551
10	448	352173
11	409	331964
12	397	336894
Total	5104	4229579

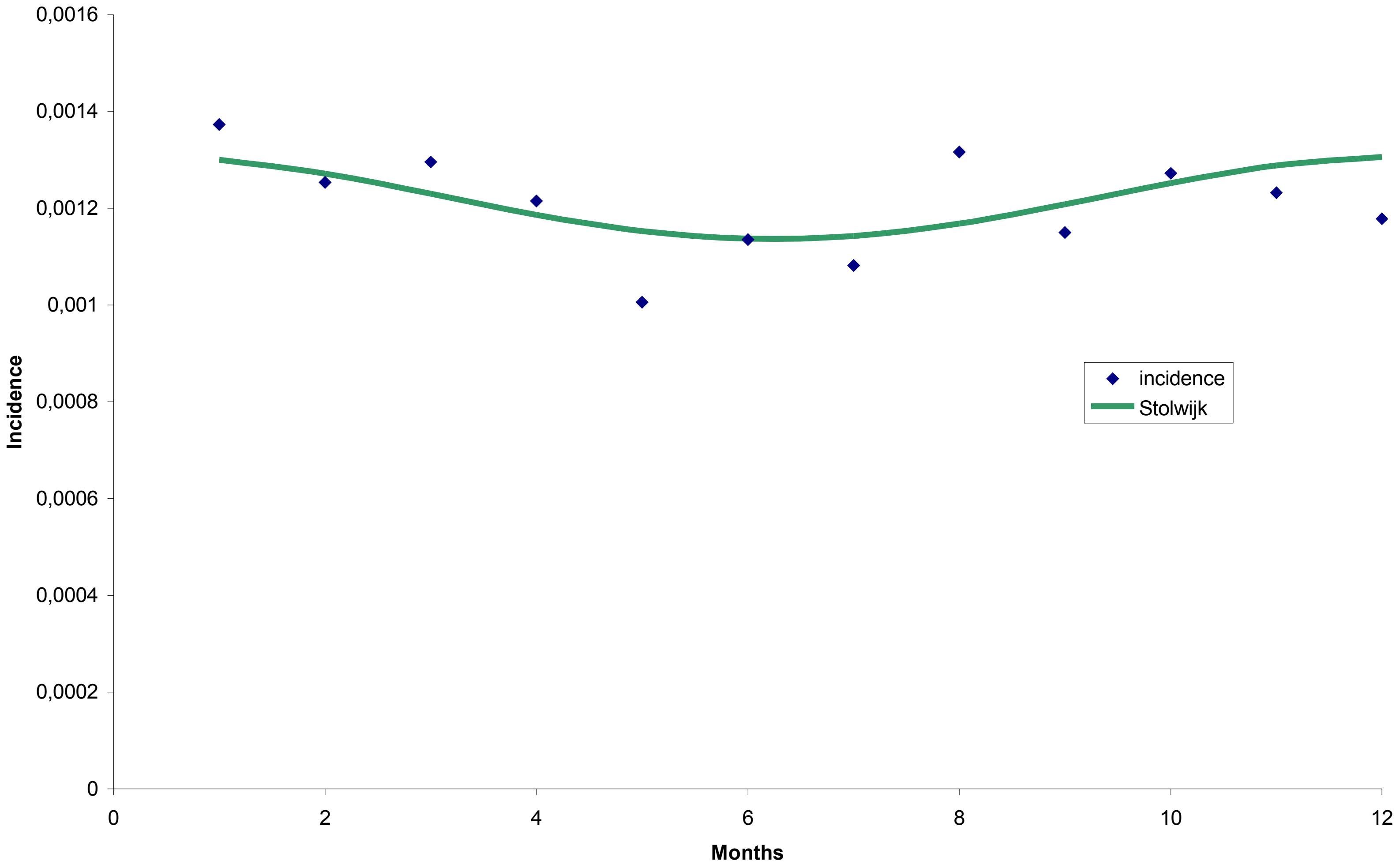
 35

Results

All children	Edwards'	Walter-Elwood	Stolwijk's
Amplitude	0.017	0.07	0.069
angle of maximum rate	-18.8	-7.4	2.73
p-value	0.67	0.002	0.029

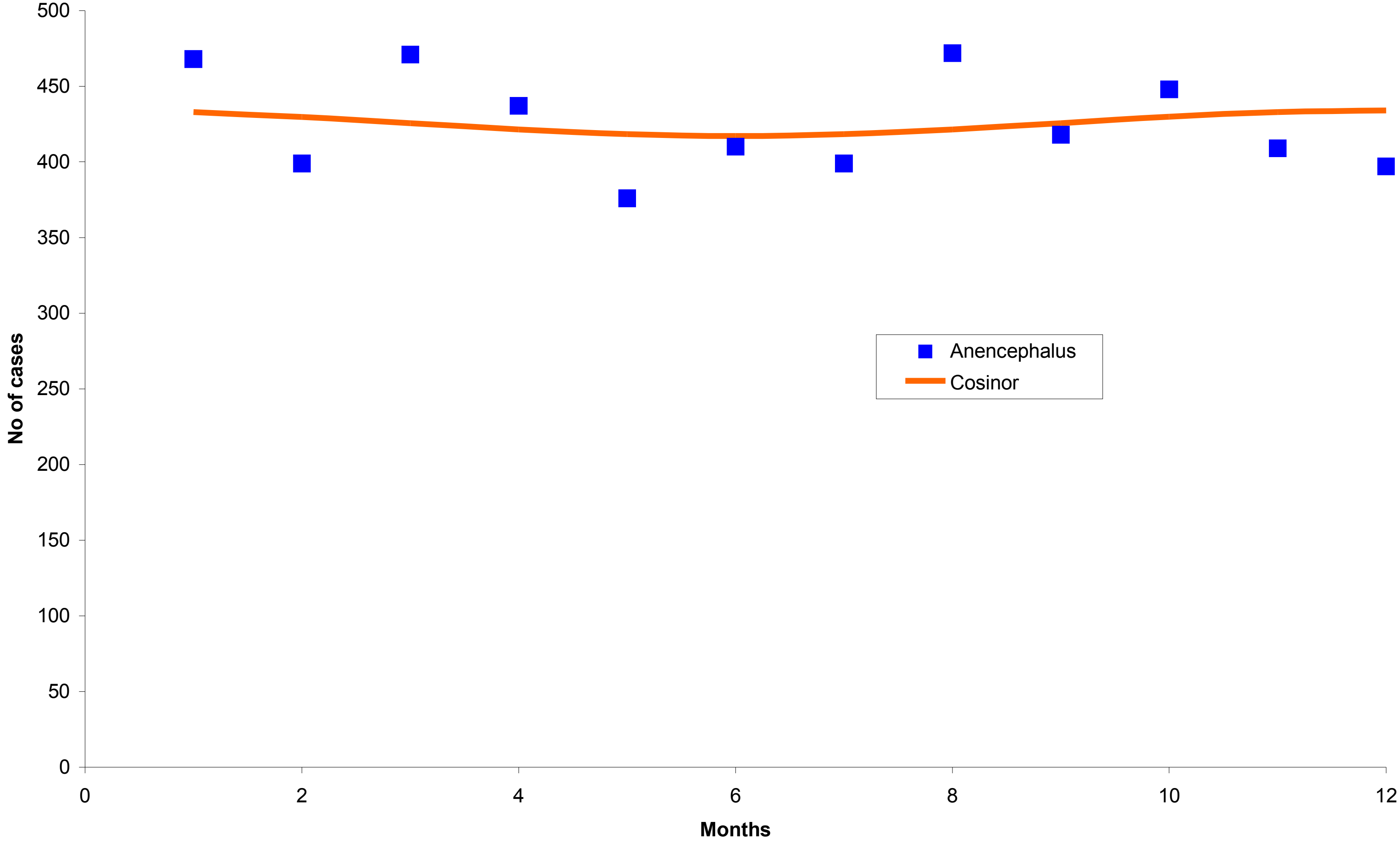
 36

Logistic regression fit



37

Cosinor method



 38

Example

- The mean temperature values of Cape Town. Fit a cyclic trend if possible.

Months	Temperature
Jan	28
Febr	27
Mar	25,5
Apr	22
May	18,5
Jun	16
Jul	15
Aug	16
Sep	18
Oct	21,5
Nov	24
Dec	26

 39

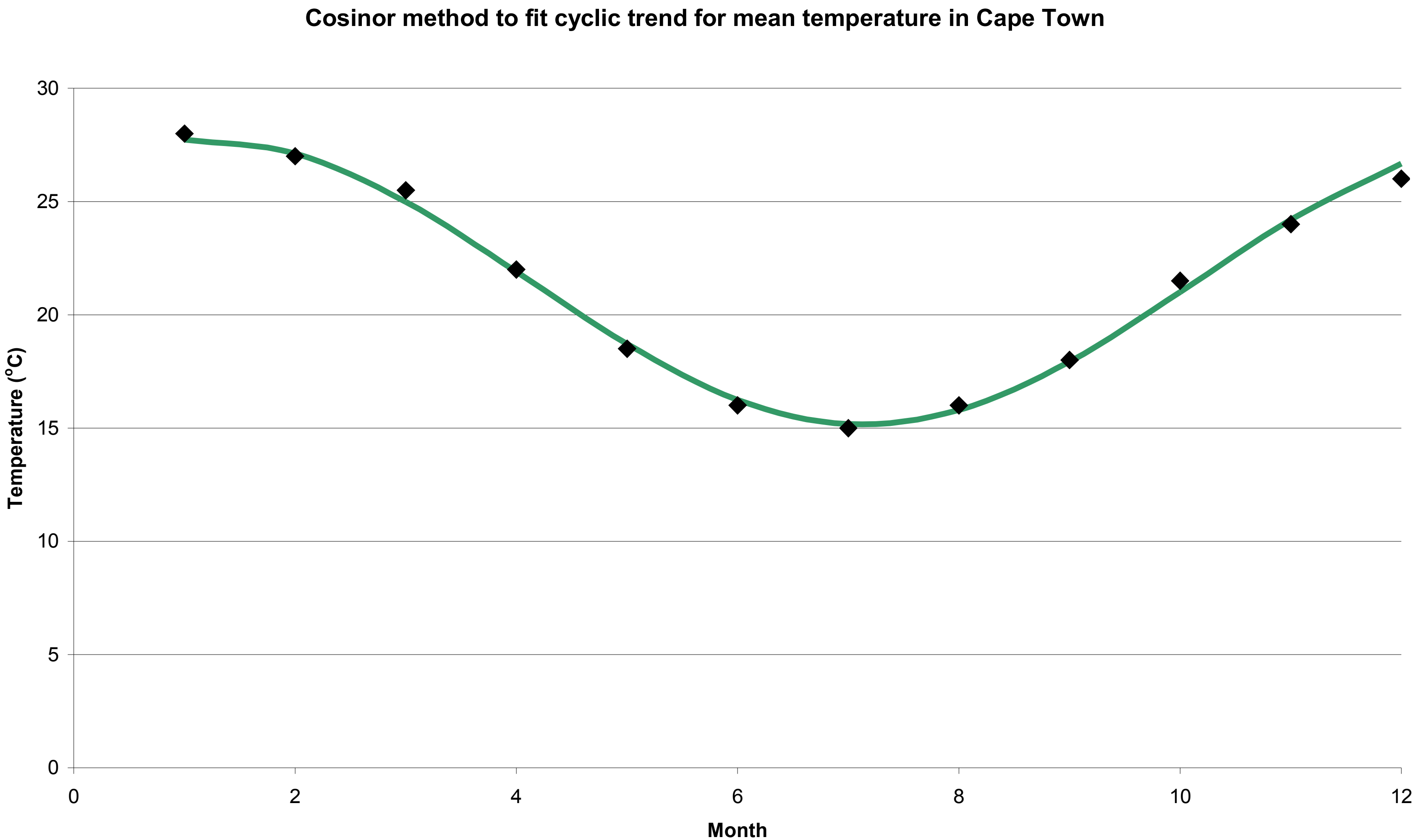
Cosinor analysis (Temperature)

```
Regression with robust standard errors                 Number of obs =      12
                                                       F(  2,      9) =  927.21
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.9947
                                                       Root MSE      =  .37604

------------------------------------------------------------------------------
         |               Robust
    temp |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
   sin_x |   3.534882   .1420283     24.889   0.000       3.213592    3.856172
   cos_x |   5.211912   .1642083     31.740   0.000       4.840447    5.583377
   _cons |   21.45833   .1085546    197.673   0.000       21.21277     21.7039
------------------------------------------------------------------------------
```

 40

Cosinor method to fit harmonic trend for monthly mean temperature



41

Example

- The possibility that Crohn' disease has an infective aetiology has been postulatae don various occasions since 1932. Cave and Freedman (1975) investigated the case histories of patients with an acute onset of symptoms leading to a diagnosis of inflammatory bowel disease to ascertain the presence of environmental or other factors which might have a bearing on aetiology in relation to transmissibility, and compared the findings among patients with Crohn's disease and ulcerative colitis.

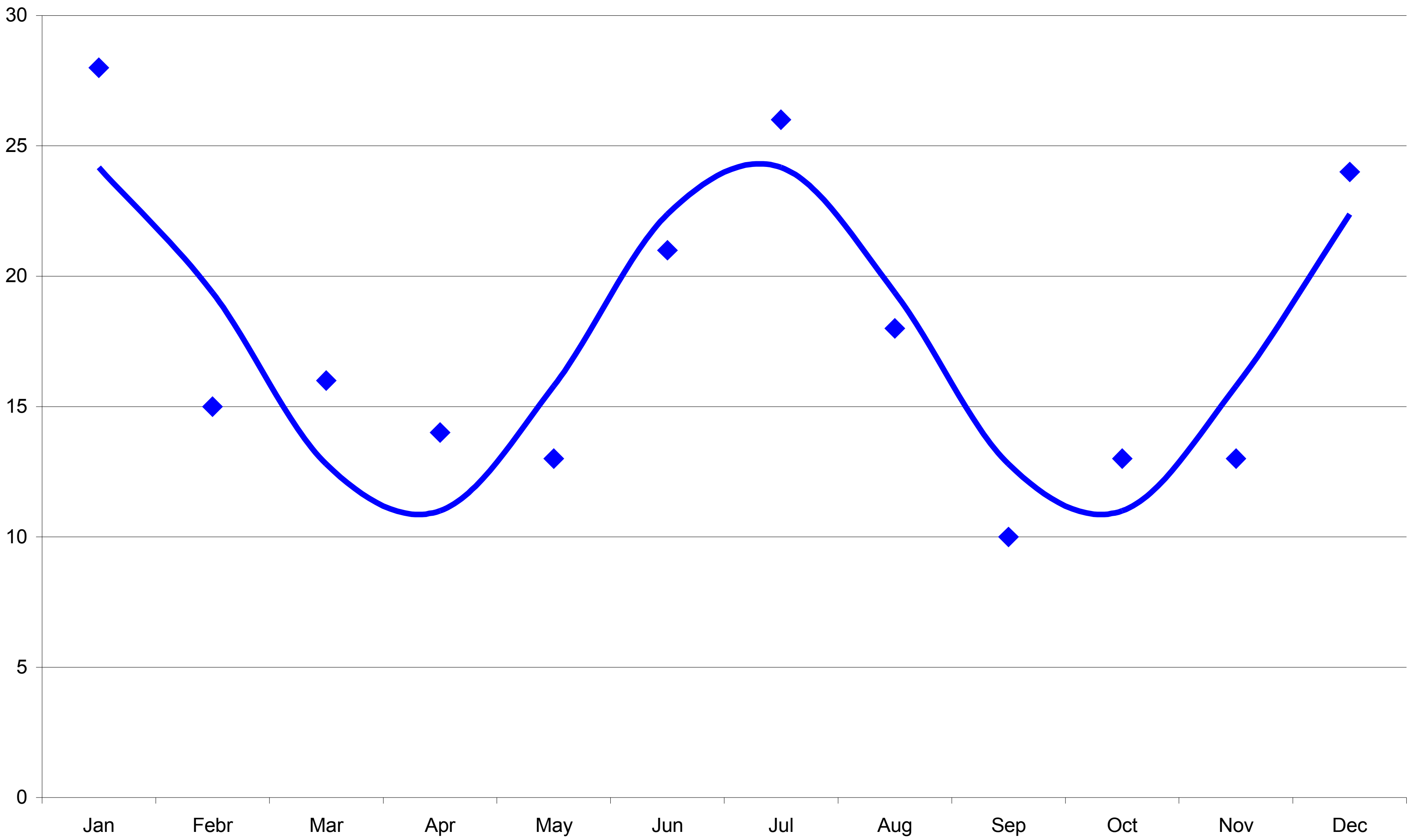
 42

Seasonality of Crohn's disease

- Month Observed Expected
- January 28 24.170095
- February 15 19.375391
- March 16 12.78863
- April 14 10.996571
- May 13 15.791274
- June 21 22.378036
- July 26 24.170095
- August 18 19.375391
- September 10 12.78863
- October 13 10.996571
- November 13 15.791272
- December 24 22.378038
- Edwards Test
- Total Number of Cases = 211
- Seasonality Test Goodness of fit test
- chi2(2) = 15.8300 chi2(2) = 5.6201
- Prob > chi2 = 0.0004 Prob > chi2 = 0.8975
- ---
- Parameter | Estimate
- --+------------
- Amplitude of cyclic variation | .38735966
- Angle of maximum rate | 15.254715
- ---

 43

Seasonal variation in month of Crohn's disease



 44

Summary I.

- **Edward's method**
  - No population at risk
  - It is being sensitive for extreme values in the data

45

Summary II

- **Walter-Elwood's method**
- Use population at risk
- Generalization of Edwards' method
- Exact month length

46

Summary III

- **Logistic regression method**
  - Use population at risk
  - Allowing confounder.

47

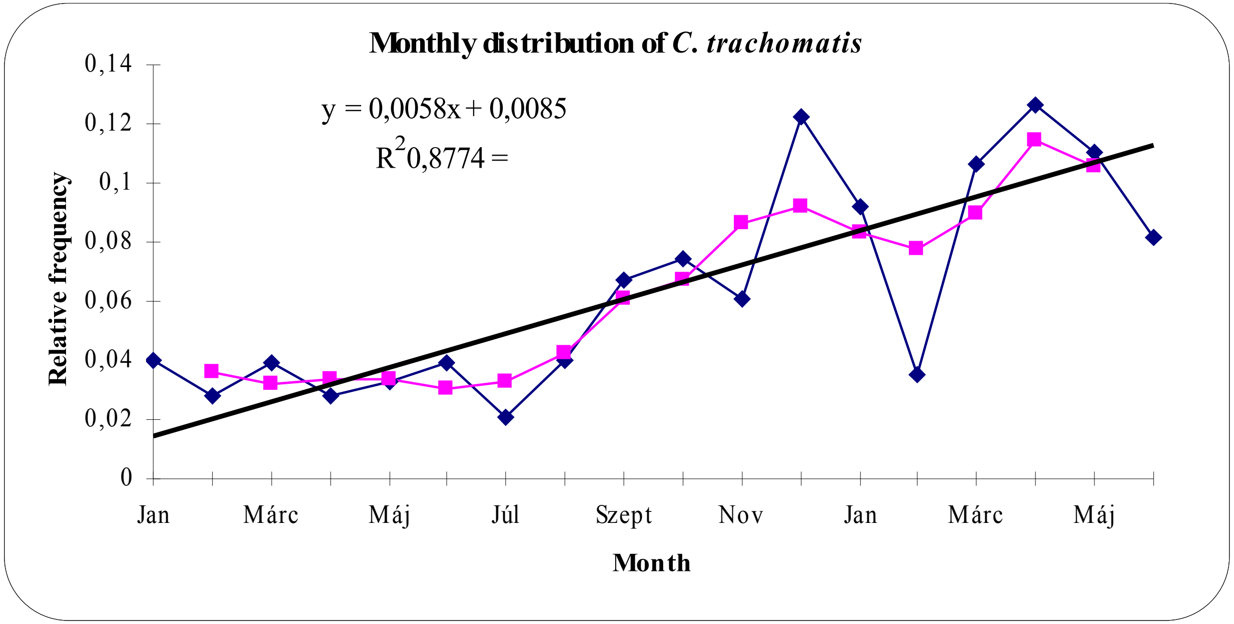
Summary IV

- **Cosinor (linear regression) method**
  - Allowing confounder.
  - No population at risk

48

Forecasting methods

- Moving average



 49