

Method Of Curve Fitting And Time Series Analysis Using FOSS

Indrani YRL¹, Dr. Varalaxmi T Shedole²

¹ Assistant Professor, Department of Mathematics, Nrupathunga University, Bengaluru, Karnataka

² Assistant Professor, Department of Statistics, Nrupathunga University, Bengaluru, Karnataka

Abstract: Maxima a free and open-source software is regarded as a computer algebra system. In this paper, graphical problems are easily solved on the method of curve fitting from statistical methods and Time Series Analysis using this software. It is one of the advanced methods which will be applicable for the problems on future prediction related to various fields for economic development and is definitely a new start for researchers in the field of statistics and other fields.

Keywords: Maxima, free and open-source software, curve fitting, trend series, semi average method

1. INTRODUCTION

Curve fitting is the process of constructing a curve, or mathematical function, that has the best fit to a series of data points, possibly subject to constraints. Let us consider bivariate distribution $(x_1, y_1), (x_2, y_2), \dots \dots (x_n, y_n)$ of n pairs of observations. Here x is independent variable and y is dependent variable . the method of finding the suitable relationship between the random variables for the given data in the form $y = f(x)$ is called curve fitting. In pair of observations, the values of y are given values. Using the functional relationship, obtained $f(x)$ values are called estimated values of y. The error (or difference) between actual values of y and their estimated values of y, which is $y - f(x)$ is called residual of y.

Arranging the statistical data in chronological order i.e., according to the occurrence of time is called time series[2]. Time series analysis is a specific way of analyzing a sequence of data points collected over an interval of time.

If we observe the definition and examples of time series, a mathematical relationship exists between the variable and time period. If we take time period as 't' and variable by y_t

then, the mathematical functional relationship can be expressed as $y_t = f(t)$.

Let $y_{t1}, y_{t2}, \dots \dots y_{tn}$ are the values of the variable at time periods $t_1, t_2, \dots \dots t_n$ respectively , then time series may be written as

t	t_1	t_2	t_n
y_t	y_{t1}	y_{t2}	y_{tn}

For

example, y_t may be considered as agricultural production or population or bank deposits or sales or profits or temperature etc., and t may be taken as years, months, weeks, days etc.

Maxima a Computer Algebra System (CAS) is a free and open source software and is applicable[3] for the steering of symbolic and numerical expressions, sets, lists, vectors, matrices, differentiation[4] and integration, Taylor series, Laplace transforms, ordinary differential equations, systems of linear equations, polynomials, statistics and tensors. It can plot functions and data in two and three dimensions. It even has a programming language that can be used to extend Maxima's capabilities. Using this software the method of curve fitting from statistical methods [5,6] and Time Series Analysis[1] can be easily solved.

2. EXPERIMENTAL METHODS OR METHODOLOGY

Fit a second degree parabola for the following data

x	0	1	2	3	6
y	1	1.8	1.3	2.5	6.3

Solution: Let the second degree parabola be

$$y = a + bx + cx^2$$

Normal equations are

$$\sum_{i=1}^n y_i = na + b \sum_{i=1}^n x_i + \sum_{i=1}^n x_i^2 \dots\dots\dots(1)$$

$$\sum_{i=1}^n x_i y_i = a \sum_{i=1}^n x_i + b \sum_{i=1}^n x_i^2 + \sum_{i=1}^n x_i^3 \dots\dots\dots(2)$$

$$\sum_{i=1}^n x_i^2 y_i = a \sum_{i=1}^n x_i^2 + b \sum_{i=1}^n x_i^3 + c \sum_{i=1}^n x_i^4 \dots\dots\dots(3)$$

x_i	y_i	x_i^2	x_i^3	x_i^4	$x_i y_i$	$x_i^2 y_i$
0	1	0	0	0	0	0
1	1.8	1	1	1	1.8	1.8
2	1.3	4	8	16	2.6	5.2
3	2.5	9	27	81	7.5	22.5
4	6.3	16	64	256	25.2	100.8
$\sum_{i=1}^n x_i = 10$	$\sum_{i=1}^n y_i = 12.9$	$\sum_{i=1}^n x_i^2 = 30$	$\sum_{i=1}^n x_i^3 = 100$	$\sum_{i=1}^n x_i^4 = 354$	$\sum_{i=1}^n x_i y_i = 37.1$	$\sum_{i=1}^n x_i^2 y_i = 130.3$

From the normal equations

$$12.9 = 5a + 10b + 30c \dots\dots\dots(1)$$

$$37.1 = 5a + 10b + 30c \dots\dots\dots(2)$$

$$130.3 = 30a + 100b + 354c \dots\dots\dots(3)$$

By solving these equations , we get

$$a = 1.42, b = -1.07, c = 0.55$$

The best fit of parabola for the given data is

$$y = 1.42 - 1.07x + 0.55x^2$$

Program for the above calculations and graphical representation:

```
kill(all)$d1:[0,1,2,3,4];d2:[1,1.8,1.3,2.5,6.3];
d3:d1^2;d4:d1^3;d5:d1^4;d6:d1*d2;d7:d3*d2;d:[d1,d2,d3,d4,d5,d6,d7];
s:[0,0,0,0,0,0,0]$n:length(d1)$m:7$
for j:1 thru m do(for i:1 thru n do(s[j]:s[j]+d[j][i]),disp(s[j]))$
e1:s[2]=n*a+s[1]*b+s[3]*c;e2:s[6]=s[1]*a+s[3]*b+s[4]*c;e3:s[7]=s[3]*a+s[4]*b+s[5]*c;
s1:linsolve([e1, e2,e3], [a,b,c]);s2:float(s1);fprintfprec:5$e:y=a+b*x+c*x^2;
wxplot2d([0.55*x^2-1.07*x+1.42], [x,-8,10]);
```

Output:

```
(%o1) [0,1,2,3,4](%o2)[1,1.8,1.3,2.5,6.3](%o3)[0,1,4,9,16](%o4)[0,1,8,27,64]
(%o5) [0,1,16,81,256](%o6) [0,1.8,2.6,7.5,25.2](%o7)[0,1.8,5.2,22.5,100.8]
(%o8) [[0,1,2,3,4],[1,1.8,1.3,2.5,6.3],[0,1,4,9,16],[0,1,8,27,64],[0,1,16,81,256],
[0,1.8,2.6,7.5,25.2],[0,1.8,5.2,22.5,100.8]]
10 12.9 30 100 354 37.1 130.3
(%o13) 12.9=30*c+10*b+5*a (%o14)37.1=100*c+30*b+10*a
(%o15) 130.3=354*c+100*b+30*a (%o16)[a=71/50,b=-107/100,c=11/20]
(%o17) [a=1.42,b=-1.07,c=0.55] (%o19)y=c*x^2+b*x+a (%o20)y=0.55*x^2-1.07*x+1.42
(%t21) (Graphics)
```

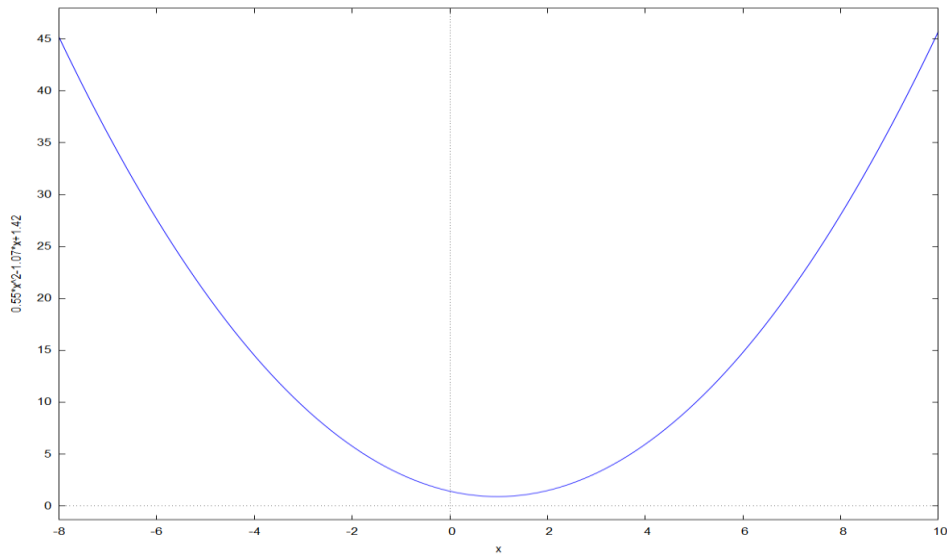


Fig. 1. Fitting a second degree parabola

3. RESULTS AND DISCUSSION

3.1 Fit a straight line of the form $y = a + bx$ to the following data

Solution: $y = 3.8 + (2.67)x$

Output:

```
(%o1) [2,4,6,8,10,12] (%o2) [10,14,19,25,31,36] (%o3)[4,16,36,64,100,144]
(%o4) [20,56,114,200,310,432](%o5)
[[2,4,6,8,10,12],[10,14,19,25,31,36],[4,16,36,64,100,144],[20,56,114,200,310,432]]
42 135 364 1132 (%o10)135=42*b+6*a (%o11)1132=364*b+42*a(%o12)[a=19/5,b=187/70]
(%o13)[a=3.8,b=2.6714](%o15)y=b*x+a(%o16)y=2.6714*x+3.8(%t17) (Graphics)
```

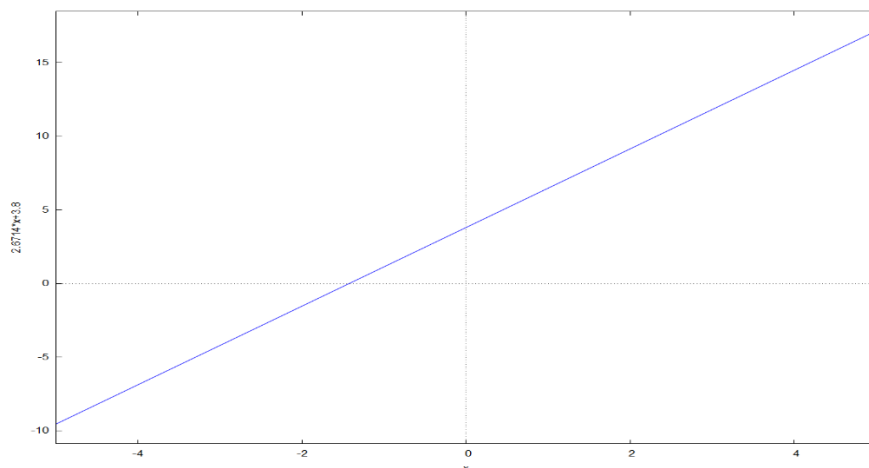


Fig. 2. Fitting a straight line

3.2 Fit a trend line to the following data by the method of semi-averages.

Year	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994
Sales (in lakhs)	68	72	75	77	82	78	80	85	90	94

Solution:

Since $n=10$ years, first we divide the data into two parts 1985-1989 and 1990-1994 by omitting the sales in 1995. Now, we calculate averages

Arithmetic Mean of first part = $\bar{y}_1 = \frac{374}{5} = 74.8$

Arithmetic Mean of second part = $\bar{y}_2 = \frac{427}{5} = 85.4$

These averages are plotted against the mid-values of the first part and second part respectively

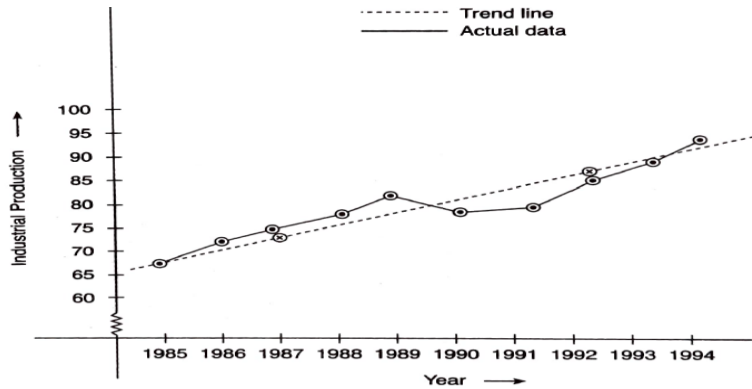


Fig. 3. Trend Analysis for 3.2 drawn roughly

Output:

```
(%o2)[[1985,68],[1986,72],[1987,75],[1988,77],[1989,82],[1990,78],[1991,80],[1992,85],
[1993,90],[1994,94]] (%o6) 74.8 (%o8) 85.4 (%o9)10 (%o10) 2.12 (%o11) 2.12*(x-1987.0)+74.8 4.0*(x-1985.0)+68.0
3.0*(x-1986.0)+72.0 2.0*(x-1987.0)+75.0 5.0*(x-1988.0)+77.0
82.0-4.0*(x-1989.0) 2.0*(x-1990.0)+78.0 5.0*(x-1991.0)+80.0 5.0*(x-1992.0)+85.0
4.0*(x-1993.0)+90.0 (%t14) (Graphics)
```

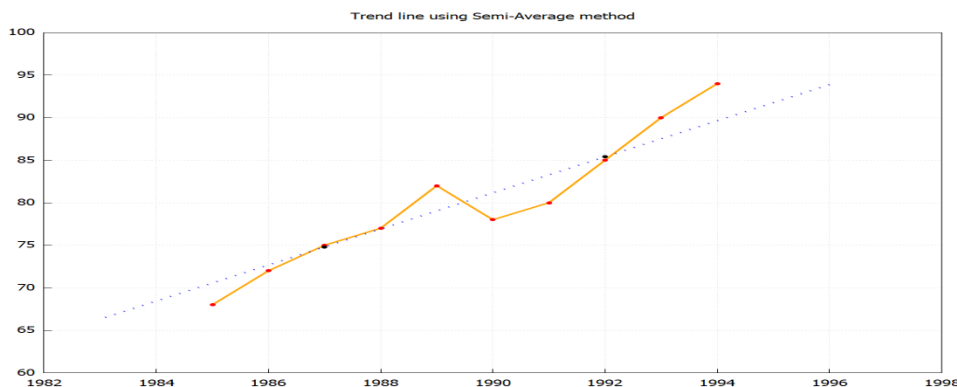


Fig. 4. Trend Analysis for 3.2 using maxima program

3.3 Fit a trend line to the following data by the method of semi-averages

Year	1991	1992	1993	1994	1995	1996	1997	1998	1999
Sales (in lakhs)	20	30	24	36	40	45	38	44	40

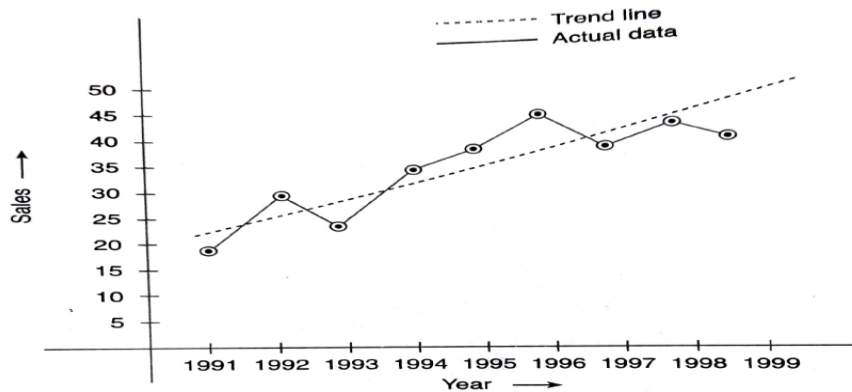


Fig. 4. Trend Analysis for 3.3 drawn roughly

Output:

```
(%o2)[[1991,20],[1992,30],[1993,24],[1994,36],[1995,40],[1996,45],[1997,38],[1998,44],
[1999,40]] (%o9)9 (%o10)57/20 (%o11)(57*(x-3985/2))/20+55/2
10*(x-1991)+20 30-6*(x-1992) 12*(x-1993)+24 4*(x-1994)+36 5*(x-1995)+40
45-7*(x-1996) 6*(x-1997)+38 44-4*(x-1998) (%t14) (Graphics)
```

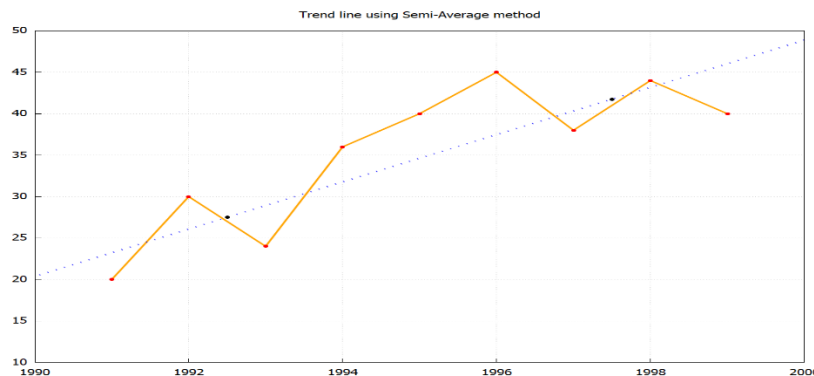


Fig. 4. Trend Analysis for 3.3 using maxima program

CONCLUSION

Curve fitting is useful - to estimate the value of one variable corresponding to a specified value of other variable, in correlation and regression analysis. Time series is useful - in determining nature of variations in the data, to forecast or predict the behavior of phenomenon in future, for planning and execution of business activities, decision making of various kind of government policies. Like Every software even Maxima has its own advantages and disadvantages, taking up the advantages as long as it supports the requirement, it's the best tool to know and implement.

REFERENCES

1. S.C. Gupta and V.K.Kapoor (Reprint 2010). "A Text Book of Fundamentals of Applied Statistics". Published by Sultan Chand & Sons, ISBN:9758-81-8054-705-8.
2. Pre Jonsson and Lars Eklundh. "TIMESAT – a program for analyzing time-series of satellite sensor." In Computers & Geosciences, 30, pp. 833 – 845. Elsevier, 2004.
3. Bassam Bou – Fakhreddine. "Computer Algebra System a Symbolic Approach Using Maxima". by author content, 2018.
4. H V Geetha, T G Sudha and Harshini Srinivas. "Solution of wave equation by method of separation of variables using Foss Tools Maxima." In International Journal of Pure and Applied Mathematics, 117, pp.167 – 174. ISSN: 1311-8080, 2017
5. Natanael Karjanto and Husty Serviana Husain. "Not Another Computer Algebra System: Highlighting wxMaxima in Calculus." In mathematics, 9, 1317. MDPI, 2021.



6. D Afriza, C M Zubainur and A Veloo. "The use of the wxMaxima linear algebra module on Gauss elimination lesson for mathematics education students." In IOP Conf. Series: Journal of Physics, 1088(),012088. IOP Publishing, 2018.
7. Dwi Taufik Hidayat, Robby Kurniawan Budhi and Fanny Lindawati. "Analysis of blook stock on red cross Surabaya as decision support using semi average method."In International Seminar on Application for Technology of Information and Communication(iSemantic), pp.268-272. IEEE, 2017.