The Method of Least Squares

The method of least squares assumes that the best fit curve of a given type is the curve that has the minimal sum of deviations, i.e., least square error from a given set of data.

The method of least squares is a widely used method of fitting curve for a given data. It is the most popular method used to determine the position of the trend line of a given time series. The trend line is technically called the best fit. In this method a mathematical relationship is established between the time factor and the variable given. Let (t_1, y_1), (t_2, y_2), …, (t_n, y_n) denote the given time series. In this method the trend value y_c of the variable *y* are computed so as to satisfy the conditions:

1.The sum of the deviations of *y* from their corresponding trend values is zero.

i.e.,

$\sum(y-y_c)=0$

2.The sum of the square of the deviations of the values of *y* from their corresponding trend values is the least.

i.e.,

$\sum(y-y_c)^2$ = is least.

The equation of the trend line can be expressed as

$y_c=a+bx$

where *a* and *b* are constants and the trend line satisfies the conditions:

1.

$\sum(y-y_c)=0$

2.

$\sum(y-y_c)^2$ is least.

The values of *a* and *b* determined such that they satisfy the equations.

$\sum y=na+b\sum x$

$\sum xy=b\sum x+a\sum x^2$

They are called normal equations.

In the equation, $y_c=a+bx$, of the trend, *a* represents the trend of the variable when x=0 and *b* represents the slope of the trend line. If *b* is positive, the trend line will be upward and if *b* is negative the trend line will be downward.

When the origin is mentioned and the deviations from the origin is denoted by *x*, we get

$a=\frac{\sum y}{n}, b=\frac{\sum xy}{\sum x^2}$

$\because$(The sum of derivation from the origin=$\sum x=0$).

Note: If *n* is odd, we take the middle value (Middle year) as the origin. If *n* is even, there will be two middle values. In this case we take the mean of the two middle values as the origin.

Merits

1. The method is mathematically sound.
2. The estimates *a* and *b* are unbiased.
3. The least square method gives trend values for all the years and the method is devoid of all kinds of subjectivity.
4. The algebraic sum of deviations of actual values from trend values is zero and the sum of the deviations $\sum(y-yc)2$ is minimum.

Demerits

1. The least square method is highly mathematical, therefore, it is difficult for a layman to understand it.
2. The method is not flexible. If certain new values are included in the given, time series, the values of *n*, $\sum x$, $\sum y$,$\sum x2$, and $\sum xy$ would change. Which affects the trend values.
3. It has been assumed that y is only a linear function of time period x. Which may not be true in many situations.