INTERPOLATION AND EXTRAPOLATION

INTRODUCTION

Interpolation is an important statistical technique to determine the most likely value of a dependent variable (Y) on the basis of some given values of independent variables (X). For example, if we are given the population figures of a country for the years 1931, 1941, 1951, 1961, 1971, 1981, 1991, 2011 etc., then to find out the population figure for any year falling between 1931 and 2011 will be termed as *Interpolation*. If we calculate the population for any year out side the interval 1931 to 2011 it will be known as *Extrapolation*.

MEANING AND DEFINITIONS

Interpolation helps us in determining the missing links in the data if any, where as extrapolation is used for forecasting.

According to **D.N. Ethanie,** ***"Interpolation may be defined as the technique of obtaining the most likely estimate of certain quantity under certain assumptions."***

According to **Hirsch,** ***"Interpolation is the estimations of a most likely estimate in given conditions. The technique of estimation a past figure is termed as interpolation. While that of estimating a probable figure for the future is called extrapolation."***

According to **W.M. Harper,** ***"Interpolation consists in reading a value which lies between two extreme points. Extrapolation means reading a value that lies outside the two extreme points".***

There is no difference between interpolation and extrapolation so for as the methods are concerned but for distinguishing past from the future we give them two different names. Interpolation relates to the past whereas extrapolation gives us the forecast for the future.

ASSUMPTIONS

Inspite of the fact that we have a number of dependable methods of interpolation and extrapolation to find out most probable values of the past and future but even then there are some basic conditions which must be satisfied to get reliable results. Some of these conditions are as under :

(*i*) A sufficient number of observations on the basis of which, the dependent variables are to be determined, should be available.

(*ii*) As far as possible the given values of X *i.e.* the independent variables should be at equal intervals.

(*iii*) There should not be continuously missing terms in the data.

(*iv*) The values of X should be in a proper sequence. There should not be sudden jumps in values. The data must not contain any period which corresponds to war, famines or floods etc. It means data should relate to normal period.

(*v*) The variables under consideration must be related to each other.

(*vi*) The rate of change of values of independent variable should be uniform.

(*vii*) The data under consideration should not be affected by external forces.

USES OR IMPORTANCE OF INTERPOLATION AND EXTRAPOLATION

The tool of interpolation and extrapolation has vital importance in the following situations :

(*i*) It estimates intermediate values : Interpolation helps us in determining the intermediate values from the given data, which if otherwise collected may need much of time, money and resources. There are a number of figures or data which are collected at some fixed intervals. There is no provision to collect the same data for any other period. In that case only interpolation techniques come to our rescue. For example, population statistics are collected after an interval of ten years. In case we need the population figure for any intermediate period we shall have to use some suitable formula of interpolation. In simple words, we can say that the technique of interpolation helps in filling the gaps in data.

(*ii*) Missing data : Some times it happens that the collected data is misplaced, destroyed or lost for some reasons like non-seriousness, fire, war or flood etc. It is not always feasible to collect it again. The missing figures can be calculated with some suitable technique of interpolation.

(*iii*) Non-availability of data : Some times it becomes practically impossible to collect the data at some particular time or at some particular place. Then the data for the particular place or time can be estimated on the basis of the other part of the given data. For example, it may become impossible to reach some particular region to contact the people with regard to population census because of some natural calamity. In such cases we can estimate the population figure of that region on the basis of the population figures of that region for the other years.

(*iv*) It can bring uniformity in the data : The primary object of the collection of data is to make comparisons. Suppose we have been given the number of students of a class 'A' who have got marks less than 5, 15, 25, 35............etc. and the students of a class 'B' who have got marks less than

10, 20, 30...........etc. Now to compare them properly both the data must be according to the similar class intervals. To overcome this type of difficulty we use the methods of interpolation.

(v) **For Forecasting :** Extrapolation is the best possible technique for forecasting. All economic planning, policy formulations etc. are nothing but future estimations. The methods provide us scientifically based formulae for predictions and forecasting.

(vi) **Calculation of Median and Mode :** The formulae used for median, mode and other positional averages are nothing but interpolation. The formulae of median and mode for locating values in continuous series are based on interpolation. These formulae are :

$$\text{Median} = L + \frac{\frac{N}{2} - c.f.}{f} \times i$$

$$\text{Mode} = L + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times i$$

METHODS OF INTERPOLATION

The following are significant and widely used methods of interpolation or extrapolation :

1. **Newton's Method Finite or Advancing Differences**
2. **Lagrange's Method**
3. **Binomial Expansion Method**

1. NEWTON'S METHOD OF FINITE OR ADVANCING DIFFERENCES

This method is applied when independent variable x advances by equal intervals and all the values of dependent variables corresponding to all x are also given. The value of x for which y is to be interpolated falls with in the given range of x variables. *This formula gives best suited results if the value to be interpolated lies in the beginning or somewhere in the first half of data.*

Conditions : The following conditions must be satisfied for application of Newton's method.

1. The independent variable X should advance be equal intervals.
2. The value of variable X for which Y is to be interpolated should [redacted] lie in the class limits of X series.

Newton's formula for interpolation is given as :

$$y_x = y_0 + x\,\Delta_0^1 + \frac{x(x-1)}{|\underline{2}}\Delta_0^2 + \frac{x(x-1)(n-2)}{|\underline{3}}\Delta_0^3 + \frac{x(x-1)(x-2)(x-3)}{|\underline{4}}\Delta_0^4 + \ldots\ldots$$

where y_x represents value of y to be interpolated.

$$x = \frac{\text{Value of X for which Y is to be interpolated} - \text{inital value of X}}{\text{Equal gap between two adjoining values of X}}$$

Δ's are finite or advancing differences.

Following table elaborates the process of taking differences.

X	Y	Finite or Advancing Differences			
		1st	2nd	3rd	4th
x_0	y_0				
		$y_1 - y_0 = \Delta_0^1$			
x_1	y_1		$\Delta_1^1 - \Delta_0^1 = \Delta_0^2$		
		$y_2 - y_1 = \Delta_1^1$		$\Delta_1^2 - \Delta_0^2 = \Delta_0^3$	
x_2	y_2		$\Delta_2^1 - \Delta_1^1 = \Delta_1^2$		$\Delta_1^3 - \Delta_0^3 = \Delta_0^4$
		$y_3 - y_2 = \Delta_2^1$		$\Delta_2^2 - \Delta_1^2 = \Delta_1^3$	
x_3	y_3		$\Delta_3^1 - \Delta_2^1 = \Delta_2^2$		
		$y_4 - y_3 = \Delta_3^1$			
x_4	y_4				

Finite differences of y series taking them in pairs are calculated in stages till there remains only one figure. While taking differences proper signs positive or negative are taken into consideration at the time of subtracting previous value from the immediate next value.

BINOMIAL EXPANSION METHOD

This method is based on Binomial Theorem. This method is easy to understand and apply. The basic requirement of this method is that the intervals of dependent variables must be equal throughout the data and missing value or values of the independent variables must correspond to these intervals.

Conditions : This method is applicable, if the following conditions are satisfied :

1. The independent variable X should advance by equal intervals.
2. The value of X for which Y is to be interpolated should lie in one of the class limits of X series.

For example this method is applicable when the data is of the form

Marks Obtained :	10	20	30	40	50
No. of Students :	5	23	?	45	75

Followings steps should be followed for interpolation by this method

(*i*) Mark the values of x variable including those for which y is to be interpolated as x_0, x_1, x_2,......etc. and the corresponding values of y variable as y_0, , y_1 , y_2 ....etc. including those which are missing.

(*ii*) Depending upon the given values of y and unknown values of y the leading differences are put equal to zero. If the number of known variables are n then the leading difference when taken as zero will give

$$\Delta_0^n = y_n - n\, y_{n-1} + \frac{n(n-1)}{\lfloor 2} y_{n-2} - \frac{n(n-1)(n-2)}{\lfloor 3} y_{n-3} + \ldots\ldots + (-1)^n\, y_0 = 0$$

In other words we put $\Delta_0^n = 0$ or $(y-1)^n = 0$.

The binomial expansion is done according to the known values and taking leading difference equal to zero.

Following formulae will be of much help in solving the problems :

Number of Known Values	Binomial Expansion
2	$\Delta_0^2 = y_2 - 2\,y_1 + y_0 = 0$
3	$\Delta_0^3 = y_3 - 3\,y_2 + 3\,y_1 - y_0 = 0$
4	$\Delta_0^4 = y_4 - 4\,y_3 + 6\,y_2 - 4\,y_1 + y_0 = 0$
5	$\Delta_0^5 = y_5 - 5\,y_4 + 10\,y_3 - 10\,y_2 + 5\,y_1 - y_0 = 0$
6	$\Delta_0^6 = y_6 - 6\,y_5 + 15\,y_4 - 20\,y_3 + 15\,y_2 - 6\,y_1 + y_0 = 0$
7	$\Delta_0^7 = y_7 - 7\,y_6 + 21\,y_5 - 35\,y_4 + 35\,y_3 - 21\,y_2 + 7\,y_1 - y_0 = 0$
8	$\Delta_0^8 = y_8 - 8\,y_7 + 28\,y_6 - 56\,y_5 + 70\,y_4 - 56\,y_3 + 28\,y_2 - 8\,y_1 + y_0 = 0$

LAGRANGE'S METHOD

This method was developed by French Mathematician, Lagrange. This formula is most suitable when x series does not advance by equal intervals. This is a condition free method and can be used when class intervals are equal or unequal.

As usual the values of independent variable are expressed as x_0 , x_1 , x_2 , x_3 ....x_n and corresponding values of dependent variable by y_0 , y_1 , y_2 ....y_n. If x is the value corresponding to which y is to be interpolated then

$$y_x = \frac{(x-x_1)(x-x_2)(x-x_3)......(x-x_n)}{(x_0-x_1)(x_0-x_2)(x_0-x_3)......(x_0-x_n)} y_0 + \frac{(x-x_0)(x-x_2)(x-x_3)......(x-x_n)}{(x_1-x_0)(x_1-x_2)(x_1-x_3)......(x_1-x_n)} y_1 +$$

$$\frac{(x-x_0)(x-x_1)(x-x_3)......(x-x_n)}{(x_2-x_0)(x_2-x_1)(x_2-x_3)......(x_2-x_n)} y_2 + \frac{(x-x_0)(x-x_1)(x-x_2)......(x-x_n)}{(x_3-x_0)(x_3-x_1)(x_3-x_2)......(x_3-x_n)} y_3 +$$

$$.......... + ... + ...$$

$$\frac{(x-x_0)(x-x_1)(x-x_2)......(x-x_{n-1})}{(x_n-x_0)(x_n-x_1)(x_n-x_2)......(x_n-x_{n-1})} \cdot y_n$$

Illustration 10. From the following data, applying Lagrange's formula, find the value of Y, when X = 5 for the data :

x :	2	3	4	6	7
y :	1	5	13	61	125

Solution :

X	Y
$2 = x_0$	$1 = y_0$
$3 = x_1$	$5 = y_1$
$4 = x_2$	$13 = y_2$
$6 = x_3$	$61 = y_3$
$7 = x_4$	$125 = y_4$

By Langrange's formula, we have

$$y_x = \frac{(x-x_1)(x-x_2)(x-x_3)(x-x_4)}{(x_0-x_1)(x_0-x_2)(x_0-x_3)(x_0-x_4)} y_0 + \frac{(x-x_0)(x-x_2)(x-x_3)(x-x_4)}{(x_1-x_0)(x_1-x_2)(x_1-x_3)(x_1-x_4)} y_1$$

$$+ \frac{(x-x_0)(x-x_1)(x-x_3)(x-x_4)}{(x_2-x_0)(x_2-x_1)(x_2-x_3)(x_2-x_4)} y_2$$

$$+ \frac{(x-x_0)(x-x_1)(x-x_2)(x-x_3)}{(x_4-x_0)(x_4-x_1)(x_4-x_2)(x_4-x_4)} y_4$$

Substituting the values

$$y_5 = \frac{(5-3)(5-4)(5-6)(5-7)}{(2-3)(2-4)(2-6)(2-7)}(1) + \frac{(5-2)(5-4)(5-6)(5-7)}{(3-2)(3-4)(3-6)(3-7)}(5)$$

$$+ \frac{(5-2)(5-3)(5-6)(5-7)}{(4-2)(4-3)(4-6)(4-7)}(13) + \frac{(5-2)(5-3)(5-4)(5-7)}{(6-2)(6-3)(6-4)(6-7)}(6$$

$$+ \frac{(5-2)(5-3)(5-4)(5-6)}{(7-2)(7-3)(7-4)(7-6)}(12$$

$$= \frac{(2)(1)(-1)(-2)}{(-1)(-2)(-4)(-5)}(1) + \frac{(3)(1)(-1)(-2)}{(1)(-1)(-3)(-4)}(5) + \frac{(3)(2)(-1)(-2)}{(2)(1)(-2)(-3)}(13)$$

$$+ \frac{(3)(2)(1)(-2)}{(4)(3)(2)(-1)}(61) + \frac{(3)(2)(1)(-1)}{(5)(4)(3)(1)}(12$$

$$= \frac{4}{40}(1) - \frac{6}{12}(5) + \frac{12}{12}(13) + \frac{12}{24}(61) - \frac{6}{60}(125)$$

$= 0{\cdot}1 - 2{\cdot}5 + 13 + 30{\cdot}5 - 12{\cdot}5$

$= 43{\cdot}6 - 15 = 28{\cdot}6$ **Ans.**