Six ways to look at linear interpolation*

ANWAR H. JOARDER

Department of Mathematical Sciences, King Fahd University of Petroleum and Minerals,
Dhahran 31261, Saudi Arabia
Email: anwarj@kfupm.edu.sa

Linear interpolation has been explained from different perspectives that are likely to be easily understood by students. Some methods seem to be more interesting in some situations. Apart from its suitability in classrooms, it has also a mnemonic value often expected by many readers.

1. Introduction

If a line is used to estimate a functional value between y-values for which the x-values are known, the process is called linear interpolation. In other words it consists of putting a line through two points over small regions on a curve, and then using the line to approximate the curve. Consider three points $A(x_1,\ y_1)$, $C(x,\ y)$, $B(x_2,\ y_2)$ on a line where y is unknown and $y_1 < y < y_2$. Let us represent them by

x_1	y_1
x	y
x_2	y_2

The value of y is obtained by equating slopes as follows:

$$\frac{y-y_1}{x-x_1}=\frac{y_2-y_1}{x_2-x_1} \tag{1}$$

which is popularly known as linear interpolation.

In this note linear interpolation has been viewed in several ways. They have been labeled as the ratio method, the distance method, the weighing method, the determinant method, the least squares method and the expected value method. They are also illustrated with examples. An application is shown to derive the formula for median in the context of a frequency distribution. Finally linear interpolation has been characterized as the expected value of a random variable based on a uniform probability distribution.

* Published in *International Journal of Mathematical Education in Science and Technology*, 32(6), 932-937 [London, UK]

2. Some Representations of the Linear Interpolation

2.1 The Ratio Method

Since A, C and B are assumed to be on a line, it is easy to check that C divides AB at the ratio

$$\frac{AC}{BC} = \frac{x - x_1}{x_2 - x} = r \tag{2}$$

and that

$$\frac{y - y_1}{y_2 - y} = r \,.$$

Since all $x's$ are known, the value of r can be calculated by (2). It then follows from the above equation that

$$y = \frac{y_1 + r\, y_2}{1 + r} \quad . \tag{3}$$

Example 2.1 Let A, C and B be given by

0.7486	0.67
0.75	y
0.7518	0.68

By (2) we have, $r = \dfrac{0.75 - 0.7486}{0.7518 - 0.75} = \dfrac{7}{9}$ so that it follows from (3) that

$$y = \frac{0.67 + r\, 0.68}{1 + r} = 0.674375 \,.$$

Readers acquainted with elementary statistics may recall that it is the 75^{th} percentile of the standard normal distribution. See e.g. [3, p 537].

2.2 The Distance Method

It follows from equation (1) that

$$y = y_1 + \frac{x - x_1}{x_2 - x_1}\,(y_2 - y_1) \,. \tag{4}$$

Since y_1, y and y_2 $(y_1 < y < y_2)$ are points on the real line, one can make the following equivalent statements:

(i) y is $\dfrac{x - x_1}{x_2 - x_1}$ unit away from y_1 towards y_2.

(ii) y is $\dfrac{x - x_1}{x_2 - x_1}$ unit of the way between y_1 and y_2.

(iii) y is $\dfrac{x - x_1}{x_2 - x_1}$ of the way from y_1 to y_2 (cf. [4, p 45] or, [1, pp.10-11]).

Example 2.2 Let A, C and B be given by

0.9495	1.64
0.95	y
0.9505	1.65

Here $\dfrac{x - x_1}{x_2 - x_1} = \dfrac{0.0005}{0.0010} = 0.5$ so that y is 0.5 unit away from y_1 towards y_2. That is $y = 1.64 + 0.5(1.65 - 1.64) = 1.645$. Readers acquainted with elementary statistics may recall that 1.645 is the 95$^{\text{th}}$ percentile of the standard normal distribution. See e.g. [3, p 537].

2.3 The Weighing Method

The equation in (1) can also be written as

$$y = y_1 + \frac{x - x_1}{x_2 - x_1} y_2 - \frac{x - x_1}{x_2 - x_1} y_1 = (1 - w)\, y_1 + w\, y_2 \tag{5}$$

where $w = \dfrac{x - x_1}{x_2 - x_1}$. If $x_1 < x < x_2$, then by subtracting x_1 from both sides of $x < x_2$, it follows that $0 < w < 1$. Similarly it can be proved that $0 < w < 1$ if $x_1 > x > x_2$. Let us apply this method to Example 2.1 so that

$$w = \frac{x - x_1}{x_2 - x_1} = \frac{0.75 - 0.7486}{0.7518 - 0.7486} = 0.4375$$

which, by the Distance Method, means that y is 0.4375 unit away from y_1 towards y_2. Clearly y is closer to y_1 than it is to y_2. By the Weighing Method we then have

$$\begin{aligned} y &= (1 - w)\, y_1 + w\, y_2 = (1 - 0.4375) y_1 + 0.4375\, y_2 \\ &= (0.5625)\, 0.67 + (0.4375)\, 0.68 = 0.674375 \end{aligned}$$

It may be remarked here that in many situations $x_2 - x_1 = 1$ so that $w = x - x_1$. In those situations the weighing method is easily grasped by students. It is explained below by an example.

Example 2.3 Consider a sample [3, p 46] with n =10 observations with the second largest 5.4 and the third largest 5.7. The rank of the lower quartile (Q_1) is $(n+1)/4 = (10+1)/4 = 2+0.75$ so that the lower quartile is an observation between the 2nd and the 3rd as depicted in the following table:

Rank	Observation
2	5.4
2.75	Q_1
3	5.7

Here $w = x - x_1 = 2.75 - 2 = 0.75$ so that it follows from (5) that

$$\begin{aligned} Q_1 &= (1-0.75)\,(\text{ 2nd observation}) + 0.75\,(\text{ 3rd observation}) \\ &= (1-0.75)(5.4) + 0.75(5.7) = 5.625. \end{aligned}$$

It is interesting to note that the weights are intuitively appealing here. The rank 2.75 of the lower quartile implies that it is closer to the 3rd observation than it is to the 2nd . So the weights 0.75 and 0.25 must be attached to the 3rd and the 2nd observations respectively. This type of problems are frequently encountered in statistics and it seems this method is the best, especially in classrooms, for calculating quartiles, deciles, percentiles or in general for quantiles.

2.4 The Determinant Method

Since the three points are on a line, the area of the triangle made by the three points must vanish, hence in general we have

$$\begin{vmatrix} x_1 & y_1 & 1 \\ x & y & 1 \\ x_2 & y_2 & 1 \end{vmatrix} = 0\,. \tag{6}$$

Those who are acquainted with determinants can easily evaluate that

$$x_1(y - y_2) - y_1(x - x_2) + (xy_2 - yx_2) = 0$$

Consider interpolating the value of y from Example 2.1. We have

$$\begin{vmatrix} 0.7486 & 0.67 & 1 \\ 0.75 & y & 1 \\ 0.7518 & 0.68 & 1 \end{vmatrix} = 0$$

whence

$0.7486\,(y - 0.68) - 0.67\,(0.75 - 0.7518) + 1[0.75\,(0.68) - 0.7518\,y] = 0$
and consequently, $y = 0.674375$.

2.5 The Least Squares Method

The equation (1) can be written as

$$y = -\frac{x_1\, y_2 - y_1\, x_2}{x_2 - x_1} + \frac{y_2 - y_1}{x_2 - x_1}\ x\,. \qquad (7)$$

But it is easy to check that the least squares estimate of β_0 and β_1 in the line $y = \beta_0 + \beta_1\, x$ based on the points (x_1, y_1) and (x_2, y_2) are given by

$$\beta_1 = \frac{(x_1 - \bar{x})(y_1 - \bar{y}) + (x_2 - \bar{x})(y_2 - \bar{y})}{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2} = \frac{y_2 - y_1}{x_2 - x_1} \text{ and}$$

$$\beta_0 = \bar{y} - \beta_1\, \bar{x} = \frac{y_1\, x_2 - x_1\, y_2}{x_2 - x_1}$$

respectively, which are the slope and the intercept parameter. Note that the least square estimate of β_0 can simply be written as $\beta_0 = y_1 - \beta_1 x_1$ (or $\beta_0 = y_2 - \beta_1 x_2$).

Consider interpolating the value of y in Example 2.1. It is easily checked that $\beta_1 = 3.125$, $\beta_0 = -1.669375$ and consequently

$$y = -1.669375 + 3.125\, x\,. \qquad (8)$$

If $x = 0.75$, then $y = 0.674375$. It is obvious from (8) that this method is the best if one needs to find several values of y corresponding to several values of x.

2.6 The Expected Value Method

Let us have the following three points

x_1	y_1
x	c
x_2	y_2

The following proposition is obvious from Sections 2.1 and 2.3.

Proposition 1: Linear interpolation to find c may be viewed as the expected value of a discrete random variable Y with the following probability mass function:

$$P(Y = y_1) = \frac{x_2 - x}{x_2 - x_1} = \frac{1}{1+r} = 1 - w$$
$$P(Y = y_2) = \frac{x - x_1}{x_2 - x_1} = \frac{r}{r+1} = w, \qquad 0 \le w \le 1$$

Proposition 2: Linear interpolation to find c may be viewed as the expected value of a discrete random variable Y with the following probability mass function :

$$Y = \begin{cases} y_1 & \text{if } X \ge x \\ y_2 & \text{if } X \le x \end{cases}$$

where X has a continuous uniform probability distribution $U(x_1, x_2)$.

Proof: The expected value of Y is given by

$$\begin{aligned} E(Y) &= y_1\, P(Y = y_1) + y_2\, P(Y = y_2) \\ &= y_1\, P(X \ge x) + y_2\, P(X \le x) \\ &= y_1\, [1 - P(X \le x)] + y_2\, P(X \le x) \\ &= y_1 \left(1 - \frac{x - x_1}{x_2 - x_1}\right) + y_2 \left(\frac{x - x_1}{x_2 - x_1}\right) \end{aligned}$$

which is the formula derived by the Weighing Method.

We conjecture that optimal probability model can be determined to improve upon the usual interpolation methods.

3. An Application to Statistics

Consider a frequency distribution having median class $[y,\ y+h]$ with relative frequency $F_{y+h} - F_y$ where F_{y+h} = cumulative relative frequency up to the median class and F_y = cumulative relative frequency up to the class preceding the median class. Also let $y_{0.50}$ be the median. Then we have the following representation

F_y	y
0.50	$y_{0.50}$
F_{y+h}	$y+h$

Then by (4) we have

$$y_{0.50} = y + \frac{0.50 - F_y}{F_{y+h} - F_h}\ h$$

which is the well-known formula for median of grouped data in a frequency distribution. See e.g. [2, p 72].

Acknowledgement

The author acknowledges the excellent research facilities available at King Fahd University of Petroleum and Minerals, Dhahran, Saudi Arabia.

References

[1] Briddle, D. F., 1982, Analytic Geometry. Wadsworth Publishing Co. California, USA.

[2] Kvanli, Alan H.; Guynes, C.H. and Pavur, Robert J., 1992, *Introduction to Business Statistics: A Computer Integrated Approach*. West Publishing Co. New York, USA.

[3] Lapin, L. L., 1997, *Modern Engineering Statistics*. Wadsworth Publishing Co., New York, USA.

[4] Newbold, P., 1995, *Statistics for Business and Economics*. Prentice Hall. New Jersey, USA.