Note on the "Method of the Arithmetic Mean" as Applied to Rates of Increase

Author(s): George H. Bryan
Source: *The Mathematical Gazette*, Vol. 3, No. 48 (Dec., 1904), pp. 97-99
Published by: Mathematical Association
Stable URL: http://www.jstor.org/stable/3605264
Accessed: 20-01-2016 23:52 UTC

THE

MATHEMATICAL GAZETTE.

EDITED BY

W. J. GREENSTREET, M.A.

WITH THE CO-OPERATION OF

F. S. MACAULAY, M.A., D.Sc.; Prof. H. W. LLOYD-TANNER, M.A., D.Sc., F.R.S.;
E. T. WHITTAKER, M.A.; W. E. HARTLEY, B.A.

LONDON:

GEORGE BELL & SONS, PORTUGAL STREET, LINCOLN'S INN,
AND BOMBAY.

Vol. III. DECEMBER, 1904. No. 48.

NOTE ON THE "METHOD OF THE ARITHMETIC MEAN" AS APPLIED TO RATES OF INCREASE.

It is known to all students of physics that when a number of different observations have been made of the same constant quantity, the best result is, in general, obtained by taking the arithmetic mean of the observed values.

If the observed quantity instead of being constant is increasing at a uniform rate, and we require to determine this rate of increase, we should naturally expect that the proper course to follow would be to find the rates of increase of the quantity in successive equal intervals of time and take their arithmetic mean.

Suppose, then, that $y_0, y_1, y_2, \dots y_n$ are the observed values of the quantity, the interval between every two consecutive observations being equal to i, and there being altogether n intervals. The rates of increase in the several intervals are

$$\frac{y_1-y_0}{i}, \frac{y_2-y_1}{i}, \dots \frac{y_n-y_{n-1}}{i}.$$

The arithmetical mean of these is

$$\frac{y_n-y_0}{ni},$$

and thus depends only on the initial and final values and not on the intermediate ones. It is, however, quite contrary to all ideas of common sense that this result should be regarded as giving the most probable value, since its accuracy depends wholly on two measurements and no weight is attached to the intermediate measurements.

If we plot the observed values on squared paper the above result would mean that the rate of increase was determined by the straight line joining the first and last points plotted. It might, however, be found that the intermediate points were very nearly in a straight line, and that, owing to considerable errors in the first and last values, the first and last points were far from being in a line with the rest. It is obvious that in such cases the method would give a bad result.

This difficulty can be overcome by applying the method of Least Squares in the following manner. Let it be assumed that the correct value of the observed quantity y at time t is given by the formula

$$y = a + bt,$$

where a, b are constants, b denoting the required probable rate of increase. Then by the method of Least Squares, if times are measured from the first observation, we have to make

$$(y_0 - a)^2 + (y_1 - a - ib)^2 + (y_2 - a - 2ib)^2 + \dots + (y_n - a - nib)^2$$

a minimum by the variation of a and b. Differentiating with respect to a and b, we have

$$(y_0 - a) + (y_1 - a - ib) + (y_2 - a - 2ib) + \dots + (y_n - a - nib) = 0,$$

$$(y_1 - a - ib) + 2(y_2 - a - 2ib) + \dots + n(y_n - a - nib) = 0.$$

These give

$$(n+1)a + \tfrac{1}{2}n(n+1)ib = y_0 + y_1 + \dots + y_n, \quad \dots\dots\dots\dots(1)$$

$$\tfrac{1}{2}n(n+1)a + \tfrac{1}{6}n(n+1)(2n+1)ib = y_1 + 2y_2 + \dots + ny_n. \quad \dots\dots\dots\dots(2)$$

Eliminating a, we have

$$\tfrac{1}{6}n(n+1)(n+2)ib = ny_n + (n-2)y_{n-1} + (n-4)y_{n-2} + \dots - (n-2)y_1 - ny_0.$$

This formula gives b, the required probable rate of increase. We may write it in the form:

$$b = \frac{n(y_n - y_0) + (n-2)(y_{n-1} - y_1) + (n-4)(y_{n-2} - y_2) + \dots}{\tfrac{1}{6}n(n+1)(n+2)i}, \quad \dots\dots\dots\dots(3)$$

and it will be seen that, unlike the previous formula, it takes account of intermediate as well as extreme values. But it still remains to find a simple interpretation showing why and wherefore the formula takes this particular form. This interpretation may be given in the following rule:

Take all the different pairs of observed values that can be found by combinations of the $n+1$ observations taken two at a time. Add together the total increases and divide by the sum of the time intervals which these increases represent.

We may verify this rule by exhibiting the differences and the time intervals in the following tabular form:

Differences.	*Intervals.*	
$y_n - y_{n-1}$	i	
$y_n - y_{n-2}$	$2i$	
$y_n - y_{n-3}$	$3i$	
....	....	
....	....	
$y_n - y_0$	ni	$\tfrac{1}{2}n(n+1)i$
$y_{n-1} - y_{n-2}$	i	
$y_{n-1} - y_{n-3}$	$2i$	
....	....	
....	....	
$y_{n-1} - y_0$	$(n-1)i$	$\tfrac{1}{2}(n-1)ni$
....	....	
....	....	
$y_1 - y_0$	i	i
$ny_n + (n-2)y_{n-1} + \quad - (n-2)y_1 - ny_0$		$\tfrac{1}{6}n(n+1)(n+2)i$

We may also state the rule as follows: The total increases of value determined from each of the $\frac{1}{2}n(n+1)$ pairs of observed values taken two at a time are of equal weight.

If, instead of taking all possible pairs of observations, we take only the pairs formed by the first and last, the second and last but one, and so on, the total increases in the corresponding intervals are $y_n - y_0$, $y_{n-1} - y_1$ and so on, and the lengths of the intervals are ni, $(n-2)i$, ... With this arrangement the weights of the several intervals are proportional to the lengths of these intervals, as is evident from equation (3).

Equation (1) gives

$$a + \tfrac{1}{2}nib = \frac{y_0 + y_1 + y_2 + \dots + y_n}{n+1}$$

or the most probable value of the quantity y at the middle of the period of observation is the arithmetic mean of the observed values, as it evidently should be.

To take a numerical example let us calculate the average annual rate of increase of the quantity of pig iron produced by Great Britain during the period 1891-1901, the annual quantities, in million tons, being as follows (from Whitaker's *Almanack*):

7·4, 6·7, 7·4, 7·3, 7·6, 8·6, 8·7, 8·5, 9·4, 8·8, 7·8.

The process may be written thus:

	Increase.	No. of years.		Sums.	Corresponding No. of years.
$d_1 = y_{10} - y_0$	0·4	10	d_1	= 0·4	10
$d_2 = y_9 - y_1$	2·1	8	$d_1 + d_2$	= 2·5	18
$d_3 = y_8 - y_2$	2·0	6	$d_1 + d_2 + d_3$	= 4·5	24
$d_4 = y_7 - y_3$	1·2	4	$d_1 + d_2 + d_3 + d_4$	= 5·7	28
$d_5 = y_6 - y_4$	1·1	2	$d_1 + d_2 + d_3 + d_4 + d_5$	= 6·8	30
			$5d_1 + 4d_2 + 3d_3 + 2d_4 + d_5$	= 19·9	110

The mean rate of increase given by the formula is 19·9 ÷ 110 = ·181 million tons per annum; whereas if the increase were calculated by dividing the difference between the first and last values by the number of years, the mean rate of increase would be only ·04 million tons. This is a very good example of the use of the method. There is very little difference between the first and last values of the observed quantity, but the remaining values show that the production was considerably greater as a rule towards the end than towards the beginning of the decade. Here, then, we have a case in which the difference between the initial and final values fails to afford an efficient measure of the general increase.

The tabular working of the above example also shows a useful method of calculating such quantities as $(1a + 2b + 3c + 4d \dots) \div (1 + 2 + 3 + 4 \dots)$. The numbers in the right-hand half are the sums of the corresponding numbers in the left-hand half, added up thus: "0·4 and 2·1 are 2·5 and 2·0 are 4·5 and 1·2 are 5·7 and 1·1 are 6·8." A similar method of adding up may frequently be found of use in obtaining average of marks of candidates in examinations, when the distribution of the candidates over the various percentages is known.

GEORGE H. BRYAN.

UNIVERSITY COLLEGE OF NORTH WALES,
BANGOR, WALES, 11*th October*, 1904.