ON THE SYSTEMATIC FITTING OF STRAIGHT LINE TRENDS BY STENCIL AND CALCULATING MACHINE

By

HERBERT A. TOOPS,
Ohio State University

Whenever there is only one plotting point corresponding to N successive abscissal values equally spaced, it is possible greatly to simplify the fitting of straight lines to the empirical observations. Let the N several absicissal values (ordinarily time) be

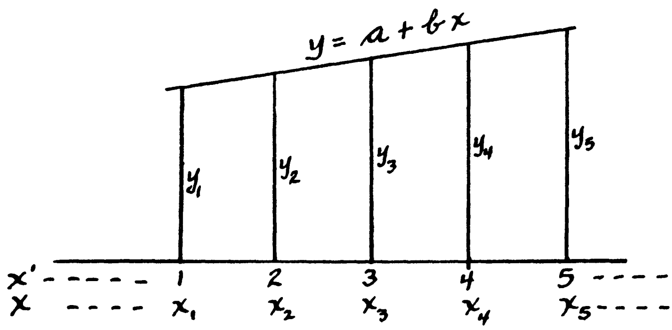
$$x_1, x_2, x_3, \cdots, x_N.$$

Let these several x values be replaced by a series of transmuted steps, $x'_i \quad (i = 1, 2, 3, \cdots, N)$.

Let the several corresponding ordinates be $y_1, y_2, y_3, \cdots, y_N$.

The situation is represented in Figure 1.

FIG. 1. Illustrating the Notation Employed.



Letting the equation of the fitted straight line be

(1) $\qquad y = a + bx'.$

it is well known that the solutions, by least squares, for the two constants are,

(2) $$a = \frac{\Sigma y \cdot \Sigma (x')^2 - \Sigma x' \, \Sigma x'y}{N \, \Sigma (x')^2 - (\Sigma x')^2},$$

and

(3) $$b = \frac{N \, \Sigma x'y - \Sigma x' \, \Sigma y}{N \, \Sigma (x')^2 - (\Sigma x')^2}.$$

Also that, inasmuch as the x' coordinates are an arithmetical series, we may substitute in the above for $\Sigma x'$ and $\Sigma(x')^2$ as follows:

$$(4) \qquad \Sigma x' = \frac{1}{2} N(N+1),$$

$$(5) \qquad \Sigma(x')^2 = \frac{1}{6} N(N+1)(2n+1),$$

thus yielding,

$$(6) \qquad a = \frac{(4N+2)\Sigma y - 6\,\Sigma x'y}{N(N-1)},$$

$$(7) \qquad b = \frac{12\,\Sigma x'y - 6(N+1)\Sigma y}{N(N^2-1)}.$$

which equations, if of infrequent usage, are highly serviceable. It is possible, however, to proceed to the derivation of formulae still more useful for systematic fitting of straight line trends. Thus, there being only one ordinate to each abscissal value as assumed,

$$(8) \qquad \Sigma y = y_1 + y_2 + y_3 + \cdots + y_N,$$

$$(9) \qquad \Sigma x'y = 1y_1 + 2y_2 + 3y_3 + \cdots + Ny_N.$$

It will be observed further that the denominator $N(N-1)$ of (6) is invariably an even number, and therefore exactly divisible by 2.[1] Substituting (8) and (9) in (6), and then multiplying both numerator and denominator by ½, we obtain

$$(10) \qquad a = \frac{(2N+1)(y_1+y_2+\cdots+y_N) - 3(1y_1+2y_2+\cdots+Ny_N)}{\frac{N(N-1)}{2}},$$

an equation which is a function only of the several ordinates and of N. Furthermore, this equation when solved for specific values of N leads to a system of equations remarkably simple; and moreover one easily extended indefinitely.

Thus, when

$$(11) \qquad N = 2, \qquad a_2 = \frac{1}{1}(2y_1 - y_2)$$

[1] The desiderata are: 1. To obtain a formula which shall obtain as small multipliers of the several Y's as possible consistent with

2. Integral multipliers, and

3. An integral numerator and denominator for the A and B (i.e. if and when the Y's are integral).

(12) When $N = 3,\quad a_3 = \frac{1}{3}(4y_1 + 1y_2 - 2y_3)$

(13) When $N = 4,\quad a_4 = \frac{1}{6}(6y_1 + 3y_2 + 0y_3 - 3y_4)$

etc., etc.

The symmetry of arrangement is more readily grasped if the several coefficients of the y's, and the denominators, D_a, be collected into an orderly table thus (Figure 2):—

FIG. 2. Systematic Solution of Equation (10), for Specific Values of for Finding a.

$$y = a + bx'.$$

RULE: Extend and cumulate the successive y's by the stencil multipliers of the row of the table appropriate to the problem (determined by N) in question. Divide the accumulated sum by the denominator, D_a, of the same row. The resulting quotient is a.

N	D_a	Multiplier of the ordinate:—											
		y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9	y_{10}	y_{11}	y_{12}
2	1	2	-1										
3	3	4	1	-2									
4	6	6	3	0	3								
5	10	8	5	2	-1	-4							
6	15	10	7	4	1	-2	-5						
7	21	12	9	6	3	0	-3	-6					
8	28	14	11	8	5	2	-1	-4	-7				
9	36	16	13	10	7	4	1	-2	-5	-8			
10	45	18	15	12	9	6	3	0	-3	-6	-9		
11	55	20	17	14	11	8	5	2	-1	-4	-7	-10	
12	66	22	19	16	13	10	7	4	1	-2	-5	-8	-11

Having such a table at hand it is obvious that a may be determined quickly by

1. Simply choosing the appropriate row of multipliers for the number, N, of successive plotting points available[2]; and

2. Extending the several y's by the appropriate multipliers, most conveniently done by calculating machine;

3. Dividing the sum so obtained by the appropriate divisor, D_a.

[2] Obviously if any plotting point intermediate between y_1 and y_N is missing it must be supplied (by interpolation) before employing this method.

The multipliers may be extended indefinitely for larger and larger values of N by simply noting that the diagonal marginal row increases by the successive addition of –1, while columns increase by the successive addition of 2; and rows decrease by the successive addition of –3. The denominators have a constant second order difference, $\Delta^2 = 1$, and consequently may be prolonged readily.

Let us now return to b, equation (7). The denominator $N(N^2-1)$ is always divisible by 6. Hence, substituting (8) and (9) in (7) and dividing both numerator and denominator by 6, we obtain,

$$(14)\quad b = \frac{2(1y_1+2y_2+3y_3+\cdots+Ny_N)-(N+1)(y_1+y_2+y_3+\cdots+y_N)}{\frac{N(N^2-1)}{6}}.$$

In like manner, this equation when solved for specific values of N leads to a systematic series of equations:

$$(15)\quad b_2 = \frac{1}{1}(-1y_1+y_2) \qquad \text{(where } N=2\text{)}$$

$$(16)\quad b_3 = \frac{1}{4}(-2y_1+0y_2+2y_3) \qquad \text{(where } N=3\text{)}$$

$$(17)\quad b_4 = \frac{1}{10}(-3y_1-1y_2+1y_3+3y_4) \qquad \text{(where } N=4\text{)}.$$

The corresponding table yields Figure 3.

FIG. 3. Systematic Solution of Equation (14), for Specific Values of N, for Finding b

$$y = a + bx'.$$

RULE: Extend and cumulate the successive y's by the stencil multipliers of the row of the table appropriate to the problem (determined by N) in question. Divide the accumulated sum by the denominator, D_b, of the same row. The resulting quotient is b

N	D_b	Multiplier of the ordinate:—											
		y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9	y_{10}	y_{11}	y_{12}
2	1	-1	1										
3	4	-2	0	2									
4	10	-3	-1	1	3								
5	20	-4	-2	0	2	4							
6	35	-5	-3	-1	1	3	5						
7	56	-6	-4	-2	0	2	4	6					
8	84	-7	-5	-3	-1	1	3	5	7				
9	120	-8	-6	-4	-2	0	2	4	6	8			
10	165	-9	-7	-5	-3	-1	1	3	5	7	9		
11	220	-10	-8	-6	-4	-2	0	2	4	6	8	10	
12	286	-11	-9	-7	-5	-3	-1	1	3	5	7	9	11

The extension of this table is readily made by observing that the diagonal marginal row increases by adding 1; the columns, by adding −1, and the rows, by adding 2; while the denominator has a constant third order difference, $\Delta^3 = 1$.

For hand computations these two tables, Figures 2 and 3, are undoubtedly simplest because the multipliers are smallest. If, however, a calculating machine is available, the magnitude of the multipliers is of relatively small moment if anything is to be gained by using different multipliers. It is obvious, for example, that the several multipliers of a row may be divided by the appropriate denominator, the resulting decimal multipliers, to replace the present integral multipliers, being presented in tables of N columns or sections.

An even more useful set of tables for general purposes may be derived by reducing equations (10) and (14) to a common (integral) denominator, so that the same denominator may be employed for calculating both a and b.

We may obtain the least common denominator, $\frac{N(N^2-1)}{2}$ by multiplying equation (10) by $\frac{N+1}{N+1}$; and, equation (14), by multiplying by $\frac{3}{3}$, thus:

$$a = \frac{(N+1)\left[(2N+1)(y_1+y_2+\cdots+y_N)-3(1y_1+2y_2+\cdots+Ny_N)\right]}{\frac{N(N^2-1)}{2}}. \tag{18}$$

$$b = \frac{6(1y_1+2y_2+\cdots+Ny_N)-3(N+1)(y_1+y_2+\cdots+y_N)}{\frac{N(N^2-1)}{2}}. \tag{19}$$

Accordingly, it follows that if the three following changes be effected, we shall have an integral system:

1. The previous table values of a to be multiplied by $(N+1)$ of the row in question throughout.

2. The previous table values of b to be multiplied by 3 throughout.

3. The common denominator, $\frac{N(N^2-1)}{2}$, of any row in question to be made to be 3 times the previous denominator of b of the row.

The two sets of multipliers may now be combined into one systematic stencil (Figure 4) with a common denominator D or common reciprocal, $\frac{1}{D}$. The directions for using this stencil are as follows:—

1. Count the number of plotting points.
2. Find the row of the stencil having the same number of plotting points, (N).
3. Record the y values for the successive plotting points in the little rectangles of the row just located.
4. Using a calculating machine, obtain the summation of the extensions of the several y-values by the multipliers just *immediately above,* employing a fixed decimal point.
5. Divide the sum just found by the divisor, D, at the left hand of the row. The result is b.
6. Similarly obtain the summation of the extensions of the y's by the multipliers of the several respective windows *immediately beneath,* again employing a fixed decimal point.
7. Divide the sum thus obtained by the same divisor, D. The result is a.
8. Substitute values of a and b in

(1) $$y = a + bx'.$$

9. If we summate (1) we obtain the checking equation,

(20) $$\Sigma y = \frac{1}{2}\left[2Na + N(N+1)b\right],$$

since $\Sigma x' = \frac{1}{2}N(N+1)$.

Now, let us employ the revised stencil on a problem (of perfect fit):—

x (Age)	y (Attainment)	x'
3.5	12.72	1
5.5	22.45	2
7.5	32.18	3
9.5	41.91	4
11.5	51.64	5
13.5	61.37	6
15.5	71.10	$7 = N$
$\Sigma x = 66.5$	$293.37 = \Sigma y$	

FIG. 4. Revised Stencil for Solving Formulae (19) and (18) for b and a, respectively, $y = a + bx'$.

$$(19)\qquad b = \frac{6(1y_1 + 2y_2 + \cdots + Ny_N) - 3(N+1)(y_1 + y_2 + \cdots + y_N)}{\frac{N(N^2-1)}{2}}$$

$$(18)\qquad a = \frac{(N+1)[(2N+1)(y_1 + y_2 + \cdots + y_N) - 3(1y_1 + 2y_2 + \cdots + Ny_N)}{\frac{N(N^2-1)}{2}}$$

$$D = \frac{N(N^2-1)}{2}.$$

Multiplier of Ordinate No.:—

N	D		1	2	3	4	5	6	7	8	9	10	11	12
2	3	above	-3	3										
		below	6	-3										
3	12	above	-6	0	6									
		below	16	4	-8									
4	30	above	-9	-3	3	9								
		below	30	15	0	-15								
5	60	above	-12	-6	0	6	12							
		below	48	30	12	-6	-24							

FIG. 4—*Continued*

n	D												
6	105	-15	-9	-3	3	9	15						
		70	49	28	7	-14	-35						
7	168	-18	-12	-6	0	6	12	18					
		96	72	48	24	0	-24	-48					
8	252	-21	-15	-9	-3	3	9	15	21				
		126	99	72	45	18	-9	-36	-63				
9	360	-24	-18	-12	-6	0	6	12	18	24			
		160	130	100	70	40	10	-20	-50	-80			
10	495	-27	-21	-15	-9	-3	3	9	15	21	27		
		198	165	132	99	66	33	0	-33	-66	-99		
11	660	-30	-24	-18	-12	-6	0	6	12	18	24	30	
		240	204	168	132	96	60	24	-12	-48	-84	-120	
12	858	-33	-27	-21	-15	-9	-3	3	9	15	21	27	33
		286	247	208	169	130	91	52	13	-26	-65	-104	-143

Since the X-coordinates are replaced, for computation, by the series, the following transmuting equation prevails:

(21) $x' = .5x - .75$.

The stencil set up, employed for seven y's , is:—

	-18	-12	-6	0	6	12	18
$D = 168$	12.72	22.45	32.18	41.91	51.64	61.37	71.10
	96	72	48	24	0	-24	-48

$$b = \frac{1}{168}\left[-18(12.72) - 12(22.45) - \cdots + 18(71.10)\right] = 9.730$$

$$a = \frac{1}{168}\left[96(12.72) + 72(22.45) + \cdots - 48(71.10)\right] = 2.990$$

whence:

$$y = 2.990 + 9.730\,x'$$

At this stage, application of formula (20) proves the correctness of this equation.

Now substitute for x' its equivalent $(.5x+.75)$ and

$$y = 4.865x - 4.3075,$$

which may be checked by summating,

$$\Sigma y = 4.865\Sigma x - 4.3075N = 4.865(66.5) - 4.3075(7);$$

i.e. $293.37 = 293.37$. The check holds.