THE FITTING OF POLYNOMIALS BY THE METHOD OF WEIGHTED GROUPING

By P. G. Guest

University of Sydney, Australia

Summary. A method of fitting polynomials to equally spaced data is developed which is more rapid than the method of least squares. The orthogonal polynomial $T_j(x)$ of the least squares method is replaced by a step function $w_j(x)$, and this greatly reduces the number of multiplications. An efficiency of about 90 per cent is obtained for the estimates of the coefficients and fitted values.

1. Introduction. An appreciable shortening in the time required to fit a curve to a series of n equally spaced observations $y(x)$ is effected by the use of tables of the orthogonal polynomials $T_j(x)$ or $\xi_j'(x)$ [1], [2], [3]. However, the process is still tedious if the number of observations is at all large. A considerable time is spent in the calculation of the orthogonal moments $\Sigma T_j(x)y(x)$, and a mistake in these calculations can easily be made.

In the present paper a method of curve fitting is developed which considerably reduces the time required for the calculation of the moments. The continuous function $T_j(x)$ is replaced by a step function $w_j(x)$. The observations $y(x)$ are summed over each interval of constancy of $w_j(x)$. The groups so formed are multiplied by the weighting factor $w_j(x)$ and added to give $\Sigma w_j(x)y(x)$.

The number of groups required is found to be $\frac{1}{2}(j + 1)$ or $\frac{1}{2}(j + 2)$ according as j is odd or even. Thus for the coefficients of the fourth and fifth degrees the number of multiplications is reduced to three; the number of weights which have to be tabulated is also reduced to three.

The estimates of the polynomial coefficients and fitted values obtained by the method of grouping are all unbiased, and have an efficiency of about 90 per cent. This means firstly that the standard error of the value obtained by the method of grouping is about 5 per cent greater than the standard error obtained by the method of least squares, and secondly that the probability that the difference between the values obtained by these two methods exceeds their standard error is very small. In practically all cases this efficiency will be quite adequate [4].

A pleasing feature of the method of grouping is that the calculation of the coefficients can be carried out easily without a calculating machine, at least for polynomials of lower degree than the fourth. The calculations can of course be done much more rapidly if a machine is used.

2. Estimation of the power series coefficients. To form an unbiased estimate of the coefficient b_{pj} in the polynomial

$$u_p(x) = \sum_{j=0}^{p} b_{pj} x^j \tag{1}$$

which is to fit the observed values $y(x)$, we must (directly or indirectly) multiply each observation by a weight $w_{pj}(x)$ so chosen that

$$\sum_x w_{pj}(x)x^k = 0, \qquad k \leq p, k \neq j, \tag{2}$$

the sum being taken over the n observations. Then the estimate of b_{pj} is

$$b_{pj} = \sum_x w_{pj}(x)y(x)/\sum_x w_{pj}(x)x^j. \tag{3}$$

The fact that $w_{pj}(x)$ depends not only on j but also on p is a disadvantage. A useful system of weights is obtained by selecting weights $w_{pj}(x)$ which are linear functions of weights $w_j(x)$ independent of p, $w_j(x)$ being chosen so that

$$\sum_x w_j(x)x^k = 0, \qquad k < j. \tag{4}$$

We can then write $w_{pj}(x)$ in the form

$$\frac{w_{pj}(x)}{\Sigma w_j x^j} = \frac{w_j(x)}{\Sigma w_j x^j} + \beta_{j+1,j}\frac{w_{j+1}(x)}{\Sigma w_{j+1}x^{j+1}} + \cdots + \beta_{pj}\frac{w_p(x)}{\Sigma w_p x^p}, \tag{5}$$

where w_j represents $w_j(x)$ and where the coefficients β_{kj} are determined from the condition

$$\Sigma w_{pj}(x)x^k = 0 \qquad j < k \leq p, \tag{6}$$

that is,

$$\frac{\Sigma w_j x^k}{\Sigma w_j x^j} + \beta_{j+1,j}\frac{\Sigma w_{j+1}x^k}{\Sigma w_{j+1}x^{j+1}} + \cdots + \beta_{kj}\frac{\Sigma w_k x^k}{\Sigma w_k x^k} = 0. \tag{7}$$

The advantage of using such a system of weights is that we can introduce statistics

$$a_j = \sum_x w_j(x)y(x)/\sum_x w_j(x)x^j \tag{8}$$

which are independent of p and express b_{pj} as a linear function of these statistics. In fact it follows from equations (3) and (5) that

$$b_{pj} = a_j + \beta_{j+1,j}a_{j+1} + \cdots + \beta_{pj}a_p. \tag{9}$$

The method of least squares is a particular case of this method of weighting [5]. In the method of least squares $w_j(x) = T_j(x)$, the orthogonal polynomial of degree j.

The calculation of the coefficients β_{kj} can be done most conveniently by evaluating the quantities

$$\alpha_{kj} = -\sum_x w_j(x)x^k \Big/ \sum_x w_j(x)x^j. \tag{10}$$

Then equation (7) becomes

$$\beta_{kj} = \alpha_{kj} + \alpha_{k,j+1}\beta_{j+1,j} + \cdots + \alpha_{k,k-1}\beta_{k-1,j}, \tag{11}$$

and the coefficients β can be built up in turn from the coefficients α.

3. The method of weighted grouping. In the method of weighted grouping we replace the continuous function $T_j(x)$ occurring in the least squares solution by a step function $w_j(x)$. In effect we assign the same weight $w_j(x)$ to all observations in a region where $T_j(x)$ is fairly constant.

The criterion used in the choice of groups is that of maximum efficiency for the coefficient a_j defined by (8), with $w_j(x)$ satisfying (4). The minimum number of groups required is thus $j + 1$. It would be possible of course to choose a larger number of groups, but this complicates the method without producing any great increase in efficiency. Adopting the value $j + 1$ for the number of groups, the values of the weights for each method of grouping are uniquely determined (except for an arbitrary multiplying factor) by equation (4).

When the observations are equally spaced, it is most convenient to change to a variable ϵ whose origin is at the centre of the points of observation x, and whose scale is such that the interval between successive observations is unity. An obvious simplification of the method of grouping for equally spaced observations is to make the groups symmetrical about the origin; that is, to take $w_j(-\epsilon) = (-)^j w_j(\epsilon)$. This reduces the number of different weights to $\frac{1}{2}(j + 1)$ or $\frac{1}{2}(j + 2)$ according as j is odd or even. Also $\beta_{jk} = 0$ when $j + k$ is odd. The observations are to be grouped by adding corresponding observations $y(\epsilon)$ of equal $|\epsilon|$ if j is even and subtracting corresponding observations if j is odd.

It does not seem feasible to calculate general formulae for the method of grouping to give maximum efficiency. However, it is relatively simple to calculate the efficiency for a particular value of n and any chosen arrangement of groups, and hence to determine the best method of grouping for each n.

The important question is whether the efficiencies will be high enough to make the method a satisfactory substitute for the method of least squares. The maximum efficiency for large n (greater than about 50) is found to be practically constant for each coefficient. The efficiencies are listed below, and are seen to be all in the region of 90 per cent. For smaller n the efficiencies tend to be somewhat higher than these values.

a_0	100%	b_{20}	93.9%	b_{40}	91.4%
a_1	88.9%	b_{31}	92.5%	b_{51}	91.2%
a_2	89.7%	b_{42}	89.5%		
a_3	90.1%	b_{53}	91.2%		
a_4	90.4%				
a_5	90.6%				

The efficiency of the estimate $u_p(x)$ of the fitted value at a point varies somewhat with the location of the point, but is always close to 90%.

The coefficients b are linear functions of the coefficients a, but the method of grouping which gives the greatest efficiency for the coefficients a does not in general correspond to that method which would give maximum efficiency for a particular b_{pj}, since the coefficients a_j, unlike the corresponding least squares coefficients, are not orthogonal. But the choice of more complicated weights leads to only a slight improvement in the efficiency, and the method using weights w_j independent of p is much more convenient.

4. Tables and illustrative example. The following quantities are tabulated for values of n from 7 to 55 and for polynomials up to the fifth degree:

(a) The best method of grouping for the estimation of a_j, together with the weights $w_j(\epsilon)$.

(b) The divisor $\sum_{\epsilon} w_j(\epsilon)\epsilon^j$.

(c) The coefficients β_{kj}.

The coefficients β_{kj} are not in general integers. β_{20} and β_{31} are tabulated in full (r signifying that the last figure is repeated indefinitely), while β_{42} and β_{53} are given to ten significant figures and β_{40} and β_{51} to nine significant figures.

In the tables the observations are numbered by the value of $|\epsilon|$ if n is odd and by the value of $|\epsilon| + \frac{1}{2}$ if n is even. For example, for 62 observations the numbers are 1 to 31, for 63 observations 0 to 31. For even values of j observations of equal $|\epsilon|$ are added, while for odd values they are subtracted. This is indicated by the suffix $+$ or $-$ under the summation sign. The expression $c(a—b)$means that the observations numbered a to b (inclusive) are to be grouped and multiplied by the weight c.

It is convenient to illustrate the use of the tables by a specific example. We shall use the example of Birge and Shea [6], the measurements being the frequencies of the first 25 lines of the P branch of a CuH band. The frequencies vary from 22,330.52 to 23,295.47. After subtracting the constant amount 22,300 from each observation, the values are written down as in Model Form 1, starting from the bottom of the left-hand column and working up this column down the right-hand column.

The groups, weights, and β-coefficients are then entered in Model Form 2. Lines are drawn in Model Form 1 to indicate the sums required. For example, $\sum_{+} (6—10)$ occurs in a_4, so a line is drawn to the right between 5 and 6 and another line between 10 and 11.

The sums of the corresponding terms in the columns are added starting from the top, the progressive totals being entered at the right wherever a line is drawn. The differences between the corresponding terms are next added, the progressive totals being entered at the left. The required sums are obtained as differences of the progressive totals; for example,

$$\sum_{+} (6—10) = 12873.55 - 7017.03.$$

As a check, the right-hand and left-hand columns are added. If the sums are R and L, the final $\sum_{-}$ total should be $R - L$, the final $\sum_{+}$ total $R + L(+ y(0))$.

The calculations indicated in Model Form 2 are then carried through. It is not necessary to record the actual sums. In working out a_2 on a calculating machine, the steps are

$$11(14968.38) - 11(11765.71) - 6(7017.03) \div 7370.$$

TABLE OF WEIGHTS

j = 0 | *j* = 1

$a_0 = \Sigma \underset{+}{y}/n$ | $a_1 = \Sigma \underset{-}{w_1} y / \Sigma \underset{-}{w_1} \epsilon$

$b_{p0} = a_0 + \beta_{20} a_2 + \beta_{40} a_4$ | $b_{p1} = a_1 + \beta_{31} a_3 + \beta_{51} a_5$

n	β_{20}	β_{40}	w_1^*	$\Sigma \underset{-}{w_1} \epsilon$	β_{31}	β_{51}
7	−4	10.56	1(2—3)	10	−7	26.6666667
8	−5.25	21.8352273	1(2—4)	15	−8.25	46.5625
9	−6.6*r*	32.3478261	1(2—4)	18	−11	86.6666667
10	−8.25	56.0782895	1(3—5)	21	−14.25	152.0625
11	−10	56.4324324	1(2—5)	28	−16	171.578947
12	−11.916*r*	92.4430970	1(3—6)	32	−19.75	277.198864
13	−14	118.484210	1(3—6)	36	−24	400.307692
14	−16.25	178.9125	1(3—7)	45	−26.25	555.757415
15	−18.6*r*	259.2	1(3—7)	50	−31	761.176471
16	−21.25	363.085227	1(4—8)	55	−36.25	1054.98355
17	−24	440.228571	1(3—8)	66	−39	1200.25
18	−26.916*r*	591.596399	1(4—9)	72	−44.75	1550.67361
19	−30	596.689655	1(4—9)	78	−51	2053.08475
20	−33.25	700.141810	1(4—10)	91	−54.25	2471.49762
21	−36.6*r*	915.2	1(4—10)	98	−61	2677.97849
22	−40.25	1174.97540	1(5—11)	105	−68.25	3303.74133
23	−44	1485	1(4—11)	120	−72	4017.72973
24	−47.916*r*	1699.51705	1(5—12)	128	−79.75	4867.71991
25	−52	2103.90448	1(5—12)	136	−88	5462
26	−56.25	2575.15754	1(5—13)	153	−92.25	6518.06250
27	−60.6*r*	2593.69038	1(5—13)	162	−101	7717.36937
28	−65.25	2902.9725	1(6—14)	171	−110.25	9321.53708
29	−70	3505.46342	1(5—14)	190	−115	10591.5486
30	−74.916*r*	4194.79821	1(6—15)	200	−124.75	12598.6870
31	−80	4653.87610	1(6—15)	210	−135	13542.5807
32	−85.25	5500.39123	1(6—16)	231	−140.25	15553.9937
33	−90.6*r*	6455.21590	1(6—16)	242	−151	17882.4828
34	−96.25	7105.31250	1(7—17)	253	−162.25	20818.9764
35	−102	7099.2	1(6—17)	276	−168	23577
36	−107.916*r*	8260.38603	1(7—18)	288	−179.75	24189.5306
37	−114	9554.95384	1(7—18)	300	−192	27885.1777
38	−120.25	10396.6582	1(7—19)	325	−198.25	28898.9048
39	−126.6*r*	11927.1611	1(7—19)	338	−211	33115.0058
40	−133.25	13617.3606	1(8—20)	351	−224.25	37136.3294
41	−140	14739.4967	1(7—20)	378	−231	41372.4706
42	−146.916*r*	14734.2725	1(8—21)	392	−244.75	46135.2764
43	−154	16722.8852	1(8—21)	406	−259	52035.7419
44	−161.25	18901.3609	1(8—22)	435	−266.25	57467.5970
45	−168.6*r*	20294.6356	1(8—22)	450	−281	60364.8750
46	−176.25	22803.7383	1(9—23)	465	−296.25	67631.1107
47	−184	25533.8680	1(8—23)	496	−304	73189.0720
48	−191.916*r*	28497.1290	1(9—24)	512	−319.75	81530.2345
49	−200	30428.0980	1(9—24)	528	−336	89405.9520
50	−208.25	30449.5231	1(9—25)	561	−344.25	93132.9528
51	−216.6*r*	33843.3333	1(9—25)	578	−361	96082.0856
52	−225.25	37506.4392	1(10—26)	595	−378.25	104845.616
53	−234	39823.0675	1(9—26)	630	−387	114079.560
54	−242.916*r*	43950.4347	1(10—27)	648	−404.75	124075.662
55	−252	48384	1(10—27)	666	−423	136438.061

* $c(a—b)$ means that observations numbered a to b (inclusive) are to be grouped and multiplied by the weight c.

TABLE OF WEIGHTS—*Continued*

$j = 2$

$a_2 = \underset{+}{\Sigma} w_2 y / \underset{+}{\Sigma} w_2 \epsilon^2$ $\qquad b_{p2} = a_2 + \beta_{42} a_4$

n	w_2^*	$\underset{+}{\Sigma} w_2 \epsilon^2$	β_{42}
7	3(3) − 2(0—1)	50	−9.64
8	2(4) − 1(1—2)	44	−13.40909091
9	3(4) − 2(0—1)	92	−16.65217391
10	2(5) − 1(1—2)	76	−21.44736842
11	5(4—5) − 4(0—2)	370	−23.44324324
12	3(5—6) − 2(1—3)	268	−29.00746269
13	5(5—6) − 4(0—2)	570	−33.46315789
14	3(6—7) − 2(1—3)	400	−40.06
15	7(6—7) − 4(0—3)	1,078	−47.28571429
16	2(7—8) − 1(1—4)	352	−55.13636364
17	7(7—8) − 4(0—3)	1,470	−61.34285714
18	2(8—9) − 1(1—4)	472	−70.22881356
19	3(7—9) − 2(0—4)	1,044	−73.68965517
20	4(8—10) − 3(1—4)	1,624	−80.70689655
21	3(8—10) − 2(0—4)	1,350	−90.76
22	5(9—11) − 3(1—5)	2,480	−101.4419355
23	11(9—11) − 6(0—5)	5,984	−112.75
24	5(10—12) − 3(1—5)	3,080	−121.5181818
25	11(10—12) − 6(0—5)	7,370	−133.8597015
26	2(11—13) − 1(1—6)	1,452	−146.8305785
27	13(10—13) − 8(0—6)	12,428	−151.7531381
28	3(11—14) − 2(1—6)	3,200	−161.74
29	13(11—14) − 8(0—6)	14,924	−175.8780488
30	7(12—15) − 4(1—7)	8,624	−190.6428571
31	13(12—15) − 8(0—6)	17,628	−201.9734513
32	7(13—16) − 4(1—7)	10,136	−217.7707182
33	15(13—16) − 8(0—7)	23,140	−234.1972342
34	7(14—17) − 4(1—7)	11,760	−246.8714286
35	3(13—17) − 2(0—7)	6,250	−253
36	8(14—18) − 5(1—8)	17,680	−270.5941176
37	17(14—18) − 10(0—8)	39,780	−288.8153846
38	8(15—19) − 5(1—8)	20,240	−302.7086957
39	17(15—19) − 10(0—8)	45,390	−321.9617978
40	9(16—20) − 5(1—9)	25,320	−341.8440758
41	17(16—20) − 10(0—8)	51,340	−357.0821192
42	3(16—21) − 2(1—9)	10,800	−364.54
43	19(16—21) − 12(0—9)	71,858	−385.5901639
44	5(17—22) − 3(1—10)	19,840	−407.2677419
45	19(17—22) − 12(0—9)	80,522	−423.7239264
46	5(18—23) − 3(1—10)	22,180	−446.4329125
47	7(18—23) − 4(0—10)	32,466	−469.7710220
48	11(19—24) − 6(1—11)	53,284	−493.7369942
49	7(19—24) − 4(0—10)	35,994	−511.9404901
50	11(19—25) − 7(1—11)	65,604	−520.8661972
51	23(19—25) − 14(0—11)	142,968	−546
52	12(20—26) − 7(1—12)	77,672	−571.7602740
53	23(20—26) − 14(0—11)	157,458	−591.1840491
54	12(21—27) − 7(1—12)	85,400	−617.9780328
55	25(21—27) − 14(0—12)	184,800	−645.4

* $c(a—b)$ means that observations numbered a to b (inclusive) are to be grouped and multiplied by the weight c.

TABLE OF WEIGHTS—*Continued*

$j = 3$

$a_3 = \Sigma \underline{w}_3 y / \Sigma \underline{w}_3 \epsilon^3$ $b_{p3} = a_3 + \beta_{53} a_5$

n	$\underline{w}_3^*$	$\Sigma \underline{w}_3 \epsilon^3$	β_{53}
7	1(3) − 1(1—2)	36	−11.66666667
8	9(4) − 7(1—3)	504	−15.83333333
9	3(4) − 2(1—3)	240	−21
10	5(5) − 3(2—4)	540	−27.16666667
11	6(5) − 5(1—3)	1140	−30.47368421
12	15(6) − 11(2—4)	3630	−37.77272727
13	5(6) − 3(1—4)	1560	−44.84615385
14	24(7) − 13(2—5)	9204	−53.51694915
15	15(7) − 7(1—5)	7140	−61.94117647
16	7(8) − 3(2—6)	3990	−71.97368421
17	7(8) − 4(2—5)	5376	−78.75
18	35(9) − 17(2—6)	32130	−88.72222222
19	20(9) − 9(2—6)	21240	−100.8644068
20	48(10) − 19(2—7)	59736	−112.1946565
21	27(9—10) − 19(2—7)	63612	−117.6881720
22	63(10—11) − 40(2—8)	172620	−129.7846715
23	5(10—11) − 3(2—8)	15540	−144.1351351
24	20(11—12) − 11(2—9)	71280	−157.5740741
25	35(11—12) − 23(2—8)	154560	−167.25
26	77(12—13) − 48(3—9)	378840	−183.7195122
27	44(12—13) − 25(2—9)	244200	−198.7297297
28	24(13—14) − 13(3—10)	147264	−216.5677966
29	2(13—14) − 1(2—10)	13716	−232.9265092
30	117(14—15) − 56(3—11)	881244	−252.1282528
31	54(14—15) − 29(2—10)	485460	−262.2043011
32	39(15—16) − 20(3—11)	382590	−282.5244648
33	65(15—16) − 31(2—11)	701220	−301.4137931
34	35(16—17) − 16(3—12)	409920	−323.1065574
35	25(16—17) − 11(3—12)	316800	−345.7916667
36	5(16—18) − 3(3—13)	93060	−352.4432624
37	88(16—18) − 51(3—13)	1,768272	−375.9644670
38	11(17—19) − 7(3—13)	256410	−388.0855856
39	44(17—19) − 27(3—13)	1,102464	−412.7298851
40	64(18—20) − 37(3—14)	1,736928	−435.3016360
41	34(18—20) − 19(3—14)	988380	−461.3137255
42	17(19—21) − 9(3—15)	533052	−485.2363184
43	39(19—21) − 20(3—15)	1,305720	−512.6129032
44	247(20—22) − 123(4—16)	8,810490	−540.9827586
45	13(20—22) − 7(3—15)	524160	−555.5416667
46	247(21—23) − 129(4—16)	10,578516	−585.0301205
47	133(21—23) − 66(3—16)	6,091932	−611.8587896
48	56(22—24) − 27(4—17)	2,717820	−642.7178952
49	50(22—24) − 23(3—17)	2,587500	−670.8986667
50	280(23—25) − 141(4—17)	16,009140	−690.6405672
51	147(22—25) − 94(4—17)	10,971492	−703.9420655
52	105(23—26) − 64(4—18)	8,336160	−733.0989520
53	165(23—26) − 98(4—18)	13,809180	−766.9110070
54	44(24—27) − 25(4—19)	3,907200	−797.4189189
55	92(24—27) − 51(4—19)	8,595744	−832.5982533

* $c(a—b)$ means that observations numbered a to b (inclusive) are to be grouped and multiplied by the weight c.

TABLE OF WEIGHTS—*Continued*

$j = 4$

$b_{p4} = a_4 = \sum\limits_{+} w_4 y / \sum\limits_{+} w_4 \epsilon^4$

n	w_4^*			$\sum\limits_{+} w_4 \epsilon^4$
7	2(3)	− 5(2)	+ 2(0—1)	168
8	1(4)	− 2(3)	+ 1(1)	144
9	25(4)	− 46(3)	+ 14(0—1)	5376
10	9(5)	− 10(3—4)	+ 11(1)	3600
11	71(5)	− 73(3—4)	+ 50(0—1)	39648
12	16(6)	− 15(4—5)	+ 14(1)	12480
13	72(6)	− 53(3—5)	+ 58(0—1)	84768
14	59(7)	− 41(4—6)	+ 32(1—2)	90000
15	45(7)	− 29(4—6)	+ 28(0—1)	89880
16	20(8)	− 11(4—7)	+ 12(1—2)	54960
17	59(8)	− 31(4—7)	+ 26(0—2)	200376
18	144(9)	− 71(5—8)	+ 70(1—2)	613152
19	590(9)	− 395(4—7)	+ 396(0—2)	4,138824
20	206(10)	− 131(5—8)	+ 106(1—3)	1,721280
21	415(10)	− 245(5—8)	+ 226(0—2)	4,182864
22	284(11)	− 161(6—9)	+ 120(1—3)	3,345552
23	235(11)	− 117(5—9)	+ 100(0—3)	3,395784
24	415(12)	− 194(6—10)	+ 185(1—3)	7,072128
25	217(12)	− 98(6—10)	+ 78(0—3)	4,241328
26	288(13)	− 115(6—11)	+ 134(1—3)	6,855936
27	2989(13)	− 1155(6—11)	+ 1126(0—3)	80,879904
28	16(14)	− 6(7—12)	+ 5(1—4)	489312
29	2989(14)	− 1344(6—11)	+ 1450(0—3)	125,116488
30	944(15)	− 410(7—12)	+ 379(1—4)	44,279712
31	535(15)	− 221(7—12)	+ 226(0—3)	28,384776
32	586(16)	− 235(8—13)	+ 206(1—4)	34,551408
33	1533(16)	− 561(7—13)	+ 532(0—4)	103,700520
34	763(16—17)	− 502(8—14)	+ 497(1—4)	97,162464
35	1491(16—17)	− 957(8—14)	+ 826(0—4)	209,629728
36	2240(17—18)	− 1405(9—15)	+ 1071(1—5)	346,514448
37	9420(17—18)	− 5397(8—15)	+ 5408(0—4)	1680,885360
38	141(18—19)	− 79(9—16)	+ 70(1—5)	27,560016
39	11148(18—19)	− 6045(9—16)	+ 5792(0—4)	2408,213808
40	2820(19—20)	− 1765(9—16)	+ 1696(1—5)	763,126848
41	13332(19—20)	− 8151(9—16)	+ 7008(0—5)	3922,631856
42	415(20—21)	− 245(10—17)	+ 226(1—5)	133,851648
43	15620(20—21)	− 9031(10—17)	+ 7456(0—5)	5452,591056
44	4779(21—22)	− 2576(10—18)	+ 2271(1—6)	1855,459872
45	19074(21—22)	− 9955(10—18)	+ 9354(0—5)	8071,065288
46	5535(22—23)	− 2834(11—19)	+ 2406(1—6)	2521,801584
47	21945(22—23)	− 10923(11—19)	+ 9894(0—5)	10844,513064
48	415(23—24)	− 194(11—20)	+ 185(1—6)	226,308096
49	8489(23—24)	− 4667(11—19)	+ 3850(0—6)	5386,084704
50	3785(24—25)	− 1693(12—21)	+ 1560(1—6)	2388,695712
51	30485(24—25)	− 15249(11—20)	+ 14080(0—6)	22841,591136
52	8680(25—26)	− 4263(12—21)	+ 3610(1—7)	6940,591392
53	34645(25—26)	− 16549(12—21)	+ 14800(0—6)	29764,941336
54	1967(26—27)	− 924(13—22)	+ 758(1—7)	1798,408080
55	46255(26—27)	− 20515(12—22)	+ 17754(0—7)	46002,560208

* $c(a—b)$ means that observations numbered a to b (inclusive) are to be grouped and multiplied by the weight c.

TABLE OF WEIGHTS—*Continued*

$j = 5$

$b_{55} = a_5 = \Sigma w_5 y / \Sigma w_5 \epsilon^5$

n	w_5^*	$\Sigma w_5 \epsilon^5$
7	1(3) − 4(2) + 5(1)	240
8	15(4) − 49(3) + 35(1—2)	6720
9	9(4) − 26(3) + 14(1—2)	6720
10	49(5) − 111(4) + 84(1—2)	65520
11	26(5) − 55(4) + 30(1—2)	51840
12	72(6) − 143(5) + 55(1—3)	208560
13	3(6) − 4(4—5) + 3(1—3)	15120
14	112(7) − 130(5—6) + 143(2—3)	840840
15	275(7) − 301(5—6) + 231(1—3)	2,808960
16	19(8) − 20(6—7) + 13(2—4)	252720
17	481(8) − 464(6—7) + 364(1—3)	8,910720
18	1755(9) − 1360(6—8) + 1547(2—4)	47,895120
19	287(9) − 213(6—8) + 189(1—4)	9,954000
20	819(10) − 589(7—9) + 456(2—5)	35,112000
21	112(10) − 75(7—9) + 68(1—4)	6,283200
22	697(11) − 455(8—10) + 357(2—5)	47,295360
23	1037(11) − 583(7—10) + 561(1—5)	95,729040
24	5586(12) − 3059(8—11) + 2622(2—6)	615,593160
25	510(12) − 430(8—10) + 366(1—5)	92,085120
26	228(13) − 115(9—12) + 100(2—6)	37,044000
27	1824(13) − 1274(8—11) + 1235(2—6)	485,503200
28	2528(14) − 1701(9—12) + 1413(2—7)	788,492880
29	3220(14) − 2030(9—12) + 2009(2—6)	1202,742240
30	35200(15) − 21518(10—13) + 18183(2—7)	15207,265920
31	22149(15) − 13230(10—13) + 10235(2—7)	10916,650080
32	14553(16) − 8463(11—14) + 5735(2—8)	8211,661920
33	29475(16) − 16344(11—14) + 12800(2—7)	19440,662400
34	34125(17) − 16632(11—15) + 15125(2—8)	27440,028000
35	10465(17) − 4998(11—15) + 4199(2—8)	9460,956960
36	3540(18) − 1652(12—16) + 1239(2—9)	3608,463600
37	1925(18) − 861(12—16) + 732(2—8)	2257,995600
38	36736(19) − 14874(12—17) + 14245(2—9)	51234,261120
39	110374(19) − 43890(12—17) + 39121(2—9)	171065,664000
40	12111(20) − 4719(13—18) + 3809(2—10)	20909,168280
41	69223(20) − 25960(13—18) + 23405(2—9)	135787,454880
42	524160(21) − 241736(13—18) + 229395(3—10)	1,379383,716480
43	17949(21) − 8085(13—18) + 6944(2—10)	52229,469600
44	18954(22) − 8385(14—19) + 6794(3—11)	60473,424960
45	6075(22) − 2568(14—19) + 2233(2—10)	21840,477120
46	104091(23) − 43290(15—20) + 35445(3—11)	408520,153080
47	8060(23) − 3289(15—20) + 2461(2—11)	34633,959360
48	128(24) − 47(15—21) + 47(3—11)	639,576000
49	73437(24) − 26468(15—21) + 24192(2—11)	400419,714240
50	34595(25) − 12285(16—22) + 10619(3—12)	204535,553040
51	9709(25) − 3400(16—22) + 2793(3—12)	62103,392400
52	57967(25—26) − 36800(17—23) + 30355(3—12)	696524,337600
53	13650(25—26) − 8075(16—23) + 7514(3—12)	185253,868800
54	289960(26—27) − 168883(17—24) + 144768(3—13)	4,267843,419840
55	411312(26—27) − 236698(17—24) + 193397(3—13)	6,521985,180480

* $c(a—b)$ means that observations numbered a to b (inclusive) are to be grouped and multiplied by the weight c.

MODEL FORM 1

	Σ_-				Σ_+
0			647.29		
1	*81.67*	605.48		687.15	
2		561.83		725.15	
3		516.42		761.27	*4504.59*
4	*816.01*	469.22		795.39	
5		420.29		827.54	*7017.03*
6		369.60		857.71	
7		317.17		885.85	
8	*2928.93*	263.06		911.94	
9		207.40		935.95	*11765.71*
10	*4465.36*	149.98		957.86	*12873.55*
11		91.05		977.79	*13942.39*
12	*6317.05*	30.52		995.47	*14968.38*
		4002.02		*10319.07*	

MODEL FORM 2*

b_0	$\Sigma_+(0\text{—}12)/25$	a_0	*598.735200*
	β_{20} — *52*	$\beta_{20}a_2$	*+48.492011*
	β_{40} + *2103.90448*	$\beta_{40}a_4$	*+0.029212*
		b_0	*647.256423*
b_2	$\Sigma_+\{11(10\text{—}12) - 6(0\text{—}5)\}/7370$	a_2	*−0.9325387*
	β_{42} — *133.8597015*	$\beta_{42}a_4$	*−0.0018586*
		b_2	*−0.9343973*
b_4	$\Sigma_+\{217(12\text{—}\quad) - 98(6\text{—}10) + 78(0\text{—}3)\}/4{,}241328$	$a_4 = b_4$	*0.0000138848*
b_1	$\Sigma_-1(5\text{—}12)/136$	a_1	*40.448824*
	β_{31} — *88*	$\beta_{31}a_3$	*+0.385928*
	β_{51} +	$\beta_{51}a_5$	
		b_1	*40.834752*
b_3	$\Sigma_-\{35(11\text{—}12) - 23(2\text{—}8)\}/154560$	a_3	*−0.004385546*
	β_{53} —	$\beta_{53}a_5$	
		b_3	*−0.004385546*
b_5	$\Sigma_-\{(\quad\text{—}\quad) - (\quad\text{—}\quad) + (\quad\text{—}\quad)\}/$	$a_5 = b_5$	

* $c(a\text{—}b)$ means that observations numbered a to b (inclusive) are to be grouped and multiplied by the weight c.

If a calculating machine is not available, it is best to work out the sums $(a\text{—}b)$ individually. The product $w(a\text{—}b)$ should be multiplied out in full, but seven-figure logarithms may be used for the division by $\Sigma w_j \epsilon^j$ and for the calculation of the terms $\beta_{kj} a_k$.

The values obtained for the polynomial coefficients by the method of least squares and by the method of grouping are shown below, together with the standard errors.

	Least Squares	*Grouping*
$b_0 \times 10^3$	647254.3 ± 12	647256.4
$b_1 \times 10^3$	40834.6 ± 2.2	40834.8
$b_2 \times 10^4$	−9343.5 ± 5	−9344.0
$b_3 \times 10^5$	−437.6 ± 2.2	−438.6
$b_4 \times 10^6$	13.8 ± 3.6	13.9

When the polynomial coefficients have been determined, the fitted values may be worked out. If the polynomial is required in terms of a variable other than ϵ, a Horner shift is performed in the usual way.

If the standard errors are required, the residuals v_p must be calculated. Assuming an efficiency of 90%, the estimated standard error of an observation is given by

$$s_p = [\Sigma v_p^2 / \{n - 0.9(p + 1)\}]^{\frac{1}{2}}.$$

The estimated errors of the polynomial coefficients and fitted values can be found by using the tabulated weight functions [7] for the least squares solution, multiplied by the factor 1.05 to allow for the efficiency of 90%.

It is sometimes necessary to know whether the neglect of higher powers is justified. The quantities $a_j'^2 \Sigma T_j^2$ provide a test for determining the degree of the polynomial to be used, since $a_j'^2 \Sigma T_j^2$ is the amount by which Σv^2 is reduced when the degree is increased from $j - 1$ to j in the least squares method. To a sufficiently good approximation we can use a_j for a_j' and put $\Sigma T_j^2 = n^{2j+1}/\kappa_j$, where

$$\kappa_1 = 12, \quad \kappa_2 = 180, \quad \kappa_3 = 2800, \quad \kappa_4 = 44{,}000, \quad \kappa_5 = 700{,}000.$$

In the example used here we find that

$$a_3^2 \Sigma T_3^2 = 41; \qquad a_4^2 \Sigma T_4^2 = 0.016; \qquad a_5^2 \Sigma T_5^2 = 0.009.$$

Thus a_3 is highly significant, a_4 is of doubtful significance, while terms of higher degree are probably insignificant.

REFERENCES

[1] R. A. Fisher and F. Yates, *Statistical Tables for Biological, Agricultural, and Medical Research*, 3rd ed., Oliver and Boyd, Ltd., Edinburgh, 1948.

[2] R. L. Anderson and E. E. Houseman, "Tables of orthogonal polynomial values extended to $n = 104$," *Research Bulletin 297* (1942), Iowa State College Agricultural Experimental Station.

[3] R. T. Birge, "Least-squares fitting of data by means of polynomials," *Rev. Modern Physics*, Vol. 19 (1947), pp. 298–360.

[4] HAROLD JEFFREYS, *Theory of Probability*, 2nd ed., Clarendon Press, Oxford, 1948, p. 179.

[5] P. G. GUEST, "Orthogonal polynomials in the least-squares fitting of observations," *Philos. Mag.* (7), Vol. 41 (1950), pp. 124–137.

[6] R. T. BIRGE AND J. D. SHEA, "A rapid method for calculating the least squares solution of a polynomial of any degree," *Univ. California Publ. Math.*, Vol. 2 (1927), pp. 67–118.

[7] P. G. GUEST, "Estimation of the error at a point on a least-squares curve," *Australian Jour. Sci. Research. Ser. A*, Vol. 3 (1950), pp. 173–182; "Estimation of the errors of the least-squares polynomial coefficients," *ibid*, pp. 364–375.