(3) ELLIOTT AND YOST: J. Am. Chem. Soc. **56**, 1057 (1934).
(4) GARNER, GREEN, AND YOST: J. Am. Chem. Soc. **57**, 2055 (1935).
(5) LARSEN AND HUNT: J. Phys. Chem. **38**, 801 (1934).
(6) PLESKOV AND MONOSGON: J. Phys. Chem. (U. S. S. R.) **6**, 513 (1935).
(7) RITCHEY AND HUNT: J. Phys. Chem. **43**, 407 (1939).
(8) SCOTT: *Standard Methods of Chemical Analysis*, p. 372. D. Van Nostrand Company, Inc., New York (1939).
(9) WILLARD AND FURMAN: *Elementary Quantitative Analysis*, p. 295. D. Van Nostrand Company, Inc., New York (1935).

# AN ABBREVIATION OF THE METHOD OF LEAST SQUARES

GERALD J. COX AND MARGARET C. MATUSCHAK

*The Nutrition Fellowship of The Buhl Foundation, Mellon Institute, Pittsburgh, Pennsylvania*

*Received October 1, 1938*

Many physical and chemical data can be summarized by equations of the type

$$y = a + bx + cx^2 + gx^3$$

but the labor of accurate reduction by the method of least squares is too frequently discouraging or even prohibitive.

In the special case in which the successive values of $x$ are in arithmetic progression, the calculation of the parameters of the power series can be much abbreviated. If an arbitrary origin is set in the middle of the range of observations, then for every positive value of $x$ there will be a corresponding negative value of equal absolute magnitude and the sums of the odd powers of $x$ will all be zero. The equation given above would then be altered to

$$y = A + BX + CX^2 + GX^3$$

in which the capital letters indicate that the equation has been referred to the arbitrary origin and $X = x - h$.

Baily (1) has shown that further abbreviation of the calculations can be made by extracting certain constant values which recur in the solution of the normal equations, such as those shown below for a third-order power series:

$$A = \frac{\Sigma y\, \Sigma X^4 - \Sigma X^2 y\, \Sigma X^2}{\Sigma X^4\, \Sigma X^0 - (\Sigma X^2)^2}$$

$$B = \frac{\Sigma Xy\, \Sigma X^6 - \Sigma X^3 y\, \Sigma X^4}{\Sigma X^6\, \Sigma X^2 - (\Sigma X^4)^2}$$

$$C = \frac{\Sigma X^2 y \, \Sigma X^0 - \Sigma y \, \Sigma X^2}{\Sigma X^4 \, \Sigma X^0 - (\Sigma X^2)^2}$$

$$G = \frac{\Sigma X^3 y \, \Sigma X^2 - \Sigma X y \, \Sigma X^4}{\Sigma X^6 \, \Sigma X^2 - (\Sigma X^4)^2}$$

In these formulae the factors in $y$ must be derived from the observations, but those involving $X$ alone, including the denominators, are independent of the data. Baily (1) has tabulated the functions in $X$.

Use of the tables of Baily requires the determination of a quotient of the difference of two products. We have found that the calculation can be reduced to the difference of two products by the following consideration:

$$\begin{aligned} A &= \frac{\Sigma y \, \Sigma X^4 - \Sigma X^2 y \, \Sigma X^2}{\Sigma X^4 \, \Sigma X^0 - (\Sigma X^2)^2} \\ &= \frac{\Sigma y \, \Sigma X^4}{\Sigma X^4 \, \Sigma X^0 - (\Sigma X^2)^2} - \frac{\Sigma X^2 y \, \Sigma X^2}{\Sigma X^4 \, \Sigma X^0 - (\Sigma X^2)^2} \\ &= k_3 \Sigma y - k_4 \Sigma X^2 y \end{aligned}$$

in which

$$k_3 = \frac{\Sigma X^4}{\Sigma X^4 \, \Sigma X^0 - (\Sigma X^2)^2}$$

and

$$k_4 = \frac{\Sigma X^2}{\Sigma X^4 \, \Sigma X^0 - (\Sigma X^2)^2}$$

Similarly, the values of such functions in $X$ necessary for the derivation of equations of order lower or higher than the third can be obtained. We have calculated the values of $k_n$ necessary for the reduction of data to first-, second-, or third-degree equations through those required for fifty-one observations, as given in tables 1 and 2.

In table 3 are shown the products and differences of products used in the derivation of equations from data. The factors $\Sigma y$, $\Sigma Xy$, $\Sigma X^2 y$, and $\Sigma X^3 y$, as needed for the degree of the equation chosen, are calculated from the data arranged with the midpoint as the origin. The factors $k_1$ to $k_8$, inclusive, are taken from table 1 if the number of observations is odd; from table 2, if even.

If the number of observations is even, the midpoint of the data is taken halfway between the two middle items, and half an interval is used as the unit.

Equations derived by use of tables 3 and 1 or 2 generally should be transformed with respect to the independent variable to either the point

TABLE 1

*Value of $k_n$. Odd number of observations*

| $N$ | $k_1$ | $k_2$ | $k_3$ | $k_4$ | $k_5$ | $k_6$ | $k_7$ | $k_8$ |
|---|---|---|---|---|---|---|---|---|
| 3 | 3333 3333 (8) | 5000 0000 (8) | 1000 0000 (7) | 1000 0000 (7) | 1500 0000 (7) | | | |
| 5 | 2000 0000 | 1000 0000 | 4857 1429 (8) | 1428 5714 (8) | 7142 8571 (9) | 9027 7778 (8) | 2361 1111 (8) | 6944 4444 (9) |
| 7 | 1428 5714 | 3571 4286 (9) | 3333 3333 | 4761 9048 (9) | 1190 4762 | 2625 6614 | 3240 7407 (9) | 4629 6296 (10) |
| 9 | 1111 1111 | 1666 6667 | 2554 1126 | 2164 5022 | 3246 7532 (10) | 1143 3782 | 8277 2166 (10) | 7014 5903 (11) |
| 11 | 9090 9091 (9) | 9090 9091 (10) | 2074 5921 | 1165 5012 | 1165 5012 | 6037 9435 (9) | 2881 3779 | 1618 7516 |
| 13 | 7692 3077 | 5494 5055 | 1748 2517 | 6993 0070 (10) | 4995 0050 (11) | 3584 6098 | 1214 0637 | 4856 2549 (12) |
| 15 | 6666 6667 | 3571 4286 | 1511 3122 | 4524 8869 | 2424 0465 | 2304 5899 | 5830 6799 (11) | 1745 7125 |
| 17 | 5882 3529 | 2450 9803 | 1331 2693 | 3095 9752 | 1289 9897 | 1570 2041 | 3081 6420 | 7166 6093 (13) |
| 19 | 5263 1579 | 1754 3860 | 1189 7391 | 2211 4109 | 7371 3696 (12) | 1118 3168 | 1752 5617 | 3257 5497 |
| 21 | 4761 9048 | 1298 7013 | 1075 5149 | 1634 5211 | 4457 7848 | 8248 5070 (10) | 1056 2015 | 1605 1694 |
| 23 | 4347 8261 | 9881 4229 (11) | 9813 6646 (9) | 1242 2360 | 2823 2637 | 6259 0791 | 6672 0719 (12) | 8445 6606 (14) |
| 25 | 4000 0000 | 7692 3077 | 9024 1546 | 9661 8357 (11) | 1858 0453 | 4862 3545 | 4382 3595 | 4692 0337 |
| 27 | 3703 7037 | 6105 0061 | 8352 4904 | 7662 8352 | 1263 1047 | 3852 7423 | 2974 5336 | 2728 9299 |
| 29 | 3448 2759 | 4926 1084 | 7774 0700 | 6179 7058 | 8828 1512 (13) | 3104 7316 | 2076 4076 | 1650 5625 |
| 31 | 3225 8065 | 4032 2581 | 7270 7048 | 5056 1230 | 6320 1537 | 2538 6983 | 1485 0296 | 1032 7049 |
| 33 | 3030 3030 | 3342 2460 | 6828 6552 | 4189 3590 | 4620 6166 | 2102 4471 | 1084 7991 | 6655 2091 (15) |
| 35 | 2857 1429 | 2801 1204 | 6437 3464 | 3510 0035 | 3441 1799 | 1760 7811 | 8073 4407 (13) | 4402 0942 |
| 37 | 2702 7027 | 2370 7918 | 6088 5061 | 2970 0030 | 2605 2658 | 1489 3734 | 6108 7522 | 2979 8791 |
| 39 | 2564 1026 | 2024 2915 | 5775 5692 | 2535 3684 | 2001 6066 | 1271 0408 | 4691 0081 | 2059 2661 |
| 41 | 2439 0244 | 1742 1603 | 5493 2589 | 2181 5961 | 1558 2829 | 1093 4097 | 3650 4910 | 1449 7581 |
| 43 | 2325 5814 | 1510 1178 | 5237 2849 | 1890 7166 | 1227 7380 | 9474 1490 (11) | 2875 1015 | 1037 9428 |
| 45 | 2222 2222 | 1317 5231 | 5004 1234 | 1649 3485 | 9778 7461 (14) | 8263 1159 | 2289 2527 | 7545 3288 (16) |
| 47 | 2127 6596 | 1156 3367 | 4790 8525 | 1447 3875 | 7866 2362 | 7250 1033 | 1841 0171 | 5561 9852 |
| 49 | 2040 8163 | 1020 4082 | 4595 0295 | 1277 1066 | 6385 5329 | 6396 2170 | 1494 1103 | 4152 6134 |
| 51 | 1960 7843 | 9049 7738 (12) | 4414 5960 | 1132 5285 | 5227 0545 | 5671 3855 | 1222 7830 | 3136 9497 |

TABLE 2

*Value of $k_n$. Even number of observations*

| $N$ | $k_1$ | $k_2$ | $k_3$ | $k_4$ | $k_5$ | $k_6$ | $k_7$ | $k_8$ |
|---|---|---|---|---|---|---|---|---|
| 4 | 2500 0000(8) | 5000 0000 (9) | 6406 2500(8) | 7812 5000 (9) | 1562 5000 (9) | 6336 8056 (8) | 7118 0556 (9) | 8680 5556(10) |
| 6 | 1666 6667 | 1428 5714 | 3945 3125 | 1953 1250 | 1674 1071(10) | 1126 7499 | 4870 7562(10) | 2411 2654(11) |
| 8 | 1250 0000 | 5952 3810(10) | 2890 6250 | 7812 5000(10) | 3720 2381(11) | 4196 3534 (9) | 9732 7441(11) | 2630 4714(12) |
| 10 | 1000 0000 | 3030 3030 | 2289 0625 | 3906 2500 | 1183 7121 | 2040 1329 | 2964 3389 | 5058 5988(13) |
| 12 | 8333 3333(9) | 1748 2517 | 1897 3214 | 2232 1429 | 4682 8172(12) | 1149 4485 | 1146 6157 | 1348 9597 |
| 14 | 7142 8571 | 1098 9011 | 1621 0938 | 1395 0893 | 2146 2912 | 7125 6741(10) | 5186 5517(12) | 4463 4695(14) |
| 16 | 6250 0000 | 7352 9412(11) | 1415 5506 | 9300 5952(11) | 1094 1877 | 4725 9999 | 2622 0143 | 1722 7426 |
| 18 | 5555 5556 | 5159 9587 | 1256 5104 | 6510 4167 | 6046 8266(13) | 3296 7149 | 1440 7871 | 7465 2181(15) |
| 20 | 5000 0000 | 3759 3985 | 1129 7349 | 4734 8485 | 3560 0365 | 2391 7243 | 8448 3844(13) | 3540 8149 |
| 22 | 4545 4545 | 2823 2637 | 1026 2784 | 3551 1364 | 2205 6748 | 1790 5616 | 5218 8071 | 1805 8156 |
| 24 | 4166 6667 | 2173 9130 | 9402 3164(9) | 2731 6434 | 1425 2052 | 1375 4794 | 3364 5791 | 9775 0702(16) |
| 26 | 3846 1538 | 1709 4017 | 8675 3091 | 2146 2912 | 9539 0720(14) | 1079 5940 | 2248 0302 | 5561 6779 |
| 28 | 3571 4286 | 1368 3634 | 8052 8846 | 1717 0330 | 6578 6704 | 8629 5508(11) | 1548 2276 | 3301 1249 |
| 30 | 3333 3333 | 1112 3471 | 7513 9509 | 1395 0893 | 4655 4704 | 7006 8080 | 1094 4042 | 2031 9424 |
| 32 | 3125 0000 | 9164 2229(12) | 7012 7390 | 1148 8971 | 3369 1996 | 5767 1532 | 7913 1009(14) | 1290 8811 |
| 34 | 2941 1765 | 7639 4194 | 6627 2213 | 9574 1423(12) | 2486 7902 | 4803 7846 | 5836 2361 | 8431 4304(17) |
| 36 | 2777 7778 | 6435 0064 | 6258 0624 | 8062 4358 | 1867 7458 | 4043 7597 | 4380 6481 | 5643 7105 |
| 38 | 2631 5789 | 5471 0581 | 5927 9058 | 6853 0703 | 1424 7547 | 3436 0952 | 3339 8722 | 3861 1239 |
| 40 | 2500 0000 | 4690 4315 | 5630 8741 | 5874 0602 | 1102 0751 | 2944 4203 | 2582 2837 | 2693 8074 |
| 42 | 2380 9524 | 4051 5355 | 5362 2160 | 5073 0520 | 8632 5332(15) | 2542 3116 | 2021 9092 | 1912 8753 |
| 44 | 2272 7273 | 3523 6081 | 5118 0477 | 4411 3495 | 6839 3016 | 2210 2564 | 1601 3580 | 1380 2431 |
| 46 | 2173 9130 | 3083 5646 | 4895 1643 | 3859 9309 | 5475 0792 | 1933 6316 | 1281 5606 | 1010 5351 |
| 48 | 2083 3333 | 2713 8515 | 4690 8968 | 3396 7392 | 4424 7580 | 1701 3314 | 1035 4426 | 7497 7742(18) |
| 50 | 2000 0000 | 2400 9604 | 4503 0048 | 3004 8077 | 3607 2121 | 1504 8177 | 8439 3512(15) | 5631 4922 |

of origin of the data or other origin such as zero. This translation of the axis requires the substitution of $x = X + h$. For ready reference the functions necessary for the transformation of the parameters in terms of $X$ to those in $x$ are given in table 4. The value of $h$ may be obtained from the number of observations, $N$, and the initial observation, $x_i$. Equations for obtaining $h$ are appended to table 4. For odd observations $x_i$ is expressed in the original units, but for even observations in half-units.

In the calculation of equations to fit even numbers of observations the unit is half an interval. Therefore the unit of the original data must be restored by multiplying the parameters $a$, $b$, $c$, and $g$ by 1, 2, 4, and 8,

TABLE 3

*Functions for derivation of parameters for origin at midpoint of observations*

| DEGREE OF EQUATION | PARAMETERS | | | |
|---|---|---|---|---|
| | $A$ | $B$ | $C$ | $G$ |
| 1 | $k_1\Sigma y$ | $k_2\Sigma Xy$ | | |
| 2 | $k_3\Sigma y - k_4\Sigma X^2y$ | $k_2\Sigma Xy$ | $k_5\Sigma X^2y - k_4\Sigma y$ | |
| 3 | $k_3\Sigma y - k_4\Sigma X^2y$ | $k_6\Sigma Xy - k_7\Sigma X^3y$ | $k_5\Sigma X^2y - k_4\Sigma y$ | $k_8\Sigma X^3y - k_7\Sigma Xy$ |

TABLE 4

*Formulae for transformation of equations*

| DEGREE OF EQUATION | PARAMETERS FOR $x = X + h$ | | | |
|---|---|---|---|---|
| | $a$ | $b$ | $c$ | $g$ |
| 1 | $A - Bh$ | $B$ | | |
| 2 | $A - Bh + Ch^2$ | $B - 2Ch$ | $C$ | |
| 3 | $A - Bh + Ch^2 - Gh^3$ | $B - 2Ch + 3Gh^2$ | $C - 3Gh$ | $G$ |

For odd observations $h = \frac{N - 1}{2} + x_i$, with $x_i$ in the original units.

For even observations $h = N - 1 + x_i$, with $x_i$ in half-units.

respectively. If, in addition, it is desired to alter the original measure of the independent variable, $a$, $b$, $c$, and $g$ are to be multiplied by $m^0$, $m^1$, $m^2$, and $m^3$, in which $m$ is the ratio of the desired unit to the original measure.

Frequently a series of experiments yields sets of parallel data, namely, with an equal number of observations for which a special set of factors can be derived by combining the values of table 1 or 2 with tables 3 and 4, with the inclusion of $m$ so that the desired equation is obtained directly. Such a combination yields little if any advantage over the procedure of derivation and transformation in the case of third-order equations. But for first- and second-degree expressions such procedures can lead to economies if many sets of data are to be reduced.

### USE OF TABLES 1 AND 2

The values of $k_n$ in tables 1 and 2 as required for the equation are to be multiplied by $10^{-r}$, in which $r$ is appended in parentheses to $k_n$ given to

TABLE 5

*Derivation of an equation for an odd number of observations*

| RAT WEIGHT | | SUM | DIFFERENCE | T |
|---|---|---|---|---|
| *grams* | *grams* | *grams* | *grams* | |
| 58.6 | 11.6 | 70.2 | 47.0 | 21 |
| 57.6 | 12.7 | 70.3 | 44.9 | 20 |
| 57.1 | 13.3 | 70.4 | 43.8 | 19 |
| 54.8 | 14.1 | 68.9 | 40.7 | 18 |
| 54.2 | 15.6 | 69.8 | 38.6 | 17 |
| 53.8 | 16.1 | 69.9 | 37.7 | 16 |
| 51.4 | 17.0 | 68.4 | 34.4 | 15 |
| 50.4 | 18.0 | 68.4 | 32.4 | 14 |
| 50.0 | 18.8 | 68.8 | 31.2 | 13 |
| 47.8 | 20.0 | 67.8 | 27.8 | 12 |
| 46.7 | 21.9 | 68.6 | 24.8 | 11 |
| 46.4 | 22.0 | 68.4 | 24.4 | 10 |
| 45.3 | 22.9 | 68.2 | 22.4 | 9 |
| 42.8 | 24.0 | 66.8 | 18.8 | 8 |
| 42.3 | 24.4 | 66.7 | 17.9 | 7 |
| 40.0 | 25.6 | 65.6 | 14.4 | 6 |
| 39.1 | 27.6 | 66.7 | 11.5 | 5 |
| 39.0 | 28.5 | 67.5 | 10.5 | 4 |
| 37.0 | 29.5 | 66.5 | 7.5 | 3 |
| 35.5 | 32.0 | 67.5 | 3.5 | 2 |
| 35.0 | 32.5 | 67.5 | 2.5 | 1 |
| 984.8 | 448.1 | 1432.9 | 536.7 | |
| 481.1 | 33.0 | 33.0 | 448.1 | |
| 1465.9 | 481.1 | 1465.9 | 984.8 | |

$\Sigma W = 1465.9$; $\Sigma WT = 7630.0$; $\Sigma WT^2 = 229{,}323.0$; $\Sigma WT^3 = 2{,}092{,}861.6$

(1) $W = 33.41 + 1.212T + 0.004388T_1^2 - 0.0002143T_1^3$

(2) $W = 5.913 + 0.2218t + 0.02561t^2 - 0.0002143t^3$

(3) $W = 5.913 + 0.6654d + 0.2305d^2 - 0.005786d^3$

eight places. The exponent $r$ is shown only for the first entry of a series of $k_n$ for which $r$ is the same, subsequent values taking the same $r$.

### EXAMPLES OF CALCULATIONS

In tables 5 and 6 are shown the data of the weight of a suckling rat taken at 8-hr. intervals from the fourth to the eighteenth day. Table 6 contains one more weight (60.9 g. at 18 days, 8 hr.) than table 5; thus the differences in the treatment of almost identical data are illustrated in the cases of odd and even observations.

The weights in the first column of table 5 are, reading from the bottom, the weights of the last twenty-one 8-hr. periods; those of the second

TABLE 6

*Derivation of an equation for an even number of observations*

| RAT WEIGHT | | SUM | DIFFERENCE | $T$ |
|---|---|---|---|---|
| *grams* | *grams* | *grams* | *grams* | |
| 60.9 | 11.6 | 72.5 | 49.3 | 43 |
| 58.6 | 12.7 | 71.3 | 45.9 | 41 |
| 57.6 | 13.3 | 70.9 | 44.3 | 39 |
| 57.1 | 14.1 | 71.2 | 43.0 | 37 |
| 54.8 | 15.6 | 70.4 | 39.2 | 35 |
| 54.2 | 16.1 | 70.3 | 38.1 | 33 |
| 53.8 | 17.0 | 70.8 | 36.8 | 31 |
| 51.4 | 18.0 | 69.4 | 33.4 | 29 |
| 50.4 | 18.8 | 69.2 | 31.6 | 27 |
| 50.0 | 20.0 | 70.0 | 30.0 | 25 |
| 47.8 | 21.9 | 69.7 | 25.9 | 23 |
| 46.7 | 22.0 | 68.7 | 24.7 | 21 |
| 46.4 | 22.9 | 69.3 | 23.5 | 19 |
| 45.3 | 24.0 | 69.3 | 21.3 | 17 |
| 42.8 | 24.4 | 67.2 | 18.4 | 15 |
| 42.3 | 25.6 | 67.9 | 16.7 | 13 |
| 40.0 | 27.6 | 67.6 | 12.4 | 11 |
| 39.1 | 28.5 | 67.6 | 10.6 | 9 |
| 39.0 | 29.5 | 68.5 | 9.5 | 7 |
| 37.0 | 32.0 | 69.0 | 5.0 | 5 |
| 35.5 | 32.5 | 68.0 | 3.0 | 3 |
| 35.0 | 33.0 | 68.0 | 2.0 | 1 |
| 1045.7 | 481.1 | 1526.8 | 564.6 | |
| 481.1 | | | 481.1 | |
| 1526.8 | | | 1045.7 | |

$\Sigma W = 1{,}526.8$; $\Sigma WT = 16{,}412.8$; $\Sigma WT^2 = 1{,}000{,}842.0$; $\Sigma WT^3 = 18{,}877{,}307.2$

(1) $W = 33.99 + 0.6047T + 0.001098T^2 - 0.00002275T^3$

(2) $W = 5.247 + 0.1512t_1 + 0.005671t_1^2 - 0.00002275t_1^3$

(2a) $W = 5.247 + 0.3024t + 0.02268t^2 - 0.0001820t^3$

(3) $W = 5.247 + 0.9072d + 0.2041d^2 - 0.004914d^3$

column, reading from the top, are of the first twenty-one periods. The weight at the midperiod was 33.0 g. In table 6, twenty-two weights appear in each column and no weight for a midperiod.

The third and fourth columns show, respectively, the sums and differences of the items of the first and second columns. A check on the accuracy of these sums and differences is shown by the totals of the columns.

$\Sigma W$ is obtained by summing the weights of the forty-three observations of table 5 and of the forty-four observations of table 6.

$\Sigma WT$ is the sum of the products of the items of columns 4 and 5 of each table.

$\Sigma WT^2$ is the sum of the product of the items of column 3 and the square of the items in column 5.

$\Sigma WT^3$ is the sum of the product of the items of column 4 and the cube of the items in column 5.

The values of $k_n$ were taken for $N = 43$ in table 1 and $N = 44$ in table 2 for the data of tables 5 and 6, respectively.

Equation 1 in each table represents the data referred to the midpoint origin in terms of 8-hr. intervals. Equation 2 in each table shows equation 1 translated to origin at birth of the suckling rat through the use of $h = 33$ for table 5 and $h = 67$ for table 6. Equation 2a of table 6 shows the restoration of the original unit of measurement by multiplication of the successive parameters by 1, 2, 4, and 8, respectively. For odd observations the value of $x_i$ was 12, as the first weight was recorded on the 4th day,—namely, twelve 8-hr. periods from zero origin,—but $x_i$ was 24 for even observations, as the 4th day was twenty-four 4-hr. periods removed from the time of birth. Equation 3 in table 5 shows the weight as a function of age in days obtained from equation 2 by multiplying the coefficients of $t$, $t^2$, and $t^3$ by $m$, $m^2$, and $m^3$, respectively, in which $m = d/t = 3$. Equation 3 of table 6 was similarly derived from equation 2a. All calculations and transformations of equations in tables 5 and 6 were carried out to eight significant figures and then rounded to four.

The foregoing equations merely summarize the data and *are not to be regarded as expressions of the law of growth of suckling rats.* The equations are excellent interpolation formulae but fail in extrapolation, as can readily be seen by their difference at zero time.

## REFERENCE

(1) Baily, J. L.: Ann. Math. Statistics **2**, 355 (1931).