A PROBLEM IN LEAST SQUARES

By Jan K. Wiśniewski

§1. We are dealing with two variables, the observed values of which are denoted x and y respectively. The pairs of observations are divided into r groups, numbering $n_1, n_2, \cdots n_r$ pairs. Suppose in each group we determine a regression equation of the following shape:

$$y_i = a_i + b_i x + \cdots m_i x^s \tag{1}$$

where y_i denotes the value of the "dependent" variable obtained from the regression equation, while y without any subscript denotes its observed value. The r regression equations of type (1) are not assumed independent; on the contrary, we postulate that

$$\sum_1^r y_i = a_0 + b_0 x + \cdots m_0 x^s \tag{2}$$

be fulfilled identically in x; $a_0, b_0, \cdots m_0$ being predetermined numbers. This leads to the following conditions:

$$\sum_1^r a_i = a_0 \qquad \sum_1^r b_i = b_0 \cdots \sum_1^r m_i = m_0. \tag{3}$$

The magnitude to be minimized under the theory of least squares is now

$$Z = \sum_1^{r-1} \sum\nolimits_i [y - (a_i + b_i x + \cdots m_i x^s)]^2 + \sum\nolimits_r \left\{ y - \left[\left(a_0 - \sum_1^{r-1} a_i \right) + \left(b_0 - \sum_1^{r-1} b_i \right) x + \cdots \left(m_0 - \sum_1^{r-1} m_i \right) x^s \right] \right\}^2. \tag{4}$$

The normal equations derived from (4) are of the following shape:

$$\begin{aligned} &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ &n_j a_j + n_r \sum_1^{r-1} a_i + b_j \sum\nolimits_j x + \left(\sum_1^{r-1} b_i \right) \left(\sum\nolimits_r x \right) + \cdots m_j \sum\nolimits_j x^s \\ &+ \left(\sum_1^{r-1} m_i \right) \left(\sum\nolimits_r x^s \right) = \sum\nolimits_j y - \sum\nolimits_r y + n_r a_0 + b_0 \sum\nolimits_r x + \cdots m_0 \sum\nolimits_r x^s \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \end{aligned} \tag{5}$$

$$a_j \sum_j x + \left(\sum_1^{r-1} a_i\right)(\textstyle\sum_r x) + b_j \sum_j x^2 + \left(\sum_1^{r-1} b_i\right)(\textstyle\sum_r x^2)$$
$$+ \cdots m_j \sum_j x^{s+1} + \left(\sum_1^{r-1} m_i\right)(\textstyle\sum_r x^{s+1}) = \sum_j xy - \sum_r xy + a_0 \sum_r x$$
$$+ b_0 \sum_r x^2 + \cdots m_0 \sum_r x^{s+1}$$

$$\cdots\cdots\cdots\cdots\cdots\cdots \tag{5}$$

$$a_j \sum_j x^s + \left(\sum_1^{r} a_i\right)(\textstyle\sum_r x^s) + b_j \sum_j x^{s+1} + \left(\sum_1^{r-1} b_i\right)(\textstyle\sum_r x^{s+1})$$
$$+ \cdots m_j \sum_j x^{2s} + \left(\sum_1^{r-1} m_i\right)(\textstyle\sum_r x^{2s}) = \sum_j x^s y - \sum_r x^s y$$
$$+ a_0 \sum_r x^s + b_0 \sum_r x^{s+1} + \cdots m_0 \sum_r x^{2s}$$

$\sum_i$ meaning a summation extended over the i-th group. As (1) is of the s-th degree, we have $(s + 1)$ $(r - 1)$ parameters to determine and as many equations, the problem thus being in theory solved.* As to the numerical solution, Doolittle's method or any other may be applied. We do not enter at present the question, how much labor would the actual solution require.

Examples. Allen and Bowley in their book on "Family Expenditure" (London, 1935) assume the expenditure on some defined item f to be a linear function of the total expenditure e

$$f = ke + c. \tag{6}$$

Evidently $\sum k = 1$, $\sum c = 0$ (cfr. pp. 10–11). Another example I give in a paper on seasonal variation, which appeared in "Economic Studies" III (Kraków). Actual values y of a time series are assumed to be linear functions of certain "normal" values x

$$y = a + bx \tag{7}$$

a and b changing from month to month but constant from year to year. Then $\sum a = 0$, $\sum b = 12$.

§2. **Methods of solution in special cases.** The generally recognized methods of solving normal equations become extremely laborious as the product $(s + 1)$ $(r - 1)$ grows large. As a matter of fact, the amount of computer's work is approximately proportional to the cube of the number of parameters to determine. Therefore short cuts seem to be indispensable. A most elegant one is at our disposal in the special case[1] when the values of x in the several groups

* The remaining $s + 1$ parameters a_r, b_r, $\cdots$ m are, of course, found from (3).

[1] This seems to be realized in Allen and Bowley's work.

are identical, or, at least, the sums n_i, $\sum_i x$, $\sum_i x^2$, $\cdots$ $\sum_i x^{2s}$ are identical in i. Instead of (1) we shall write

$$y_i = A_i + B_i X_1 + \cdots MX_s \tag{8}$$

where X_1, X_2, $\cdots$ X_s are orthogonal polynomials, i.e. such that $\sum X_i X_j = 0$ if and only if $i \neq j$. In general, $X_h = X^h + \alpha_{h-1}^h X^{h-1} + \cdots \alpha_0^h$, the coefficients being rational functions of n, $\sum x$, $\sum x^2$, $\cdots$ $\sum x^{2s-1}$.

The conditions (3) can now be replaced by a set of equivalent ones, viz.

$$\sum_1^r A_i = A_0 \qquad \sum_1^r B_i = B_0 \cdots \sum_1^r M_i = M_0. \tag{9}$$

How the actual values of A_0, B_0, $\cdots$ M_0 are found, will be shown in the next paragraph. The solution becomes now very easy, as the normal equations for the determination of each set of $r - 1$ parameters are independent, i.e. we can calculate the A's separately, then the B's etc., the order of solution being of no importance. Moreover the shape of the normal equations permits of considerable simplification of solution. Suppose we have to determine the values of the coefficients K, corresponding to X_h. The normal equations are now—after certain simplifications—

$$\begin{aligned} 2K_1 + K_2 + K_3 + \cdots K_{r-1} &= \frac{1}{\sum X_h^2} \left(\sum_1 X_h y - \sum_r X_h y\right) + K_0 \\ K_1 + 2K_2 + K_3 + \cdots K_{r-1} &= \frac{1}{\sum X_h^2} \left(\sum_2 X_h y - \sum_r X_h y\right) + K_0 \\ &\cdots\cdots\cdots\cdots\cdots \\ K_1 + K_2 + K_3 + \cdots 2K_{r-1} &= \frac{1}{\sum X_h^2} \left(\sum_{r-1} X_h y - \sum_r X_h y\right) + K_0. \end{aligned} \tag{10}$$

Adding these equations, dividing the sum by r and substracting the quotient from the j-th equation, we get

$$K_j = \frac{\sum_j X_h y}{\sum X_h^2} - \frac{1}{r}\left(\sum_1^r \frac{\sum_i X_h y}{\sum X_h^2} - K_0\right). \tag{11}$$

The first member of the right hand side of (11) should be regarded as the principal term: this is actually the value we would obtain for K_j, were this coefficient independent from the other K's. The second member is a correction term, the necessary amount of correction being distributed equally among the several K's. The simple solution given by (11) is only possible if the sum $\sum X_h^2$ is the same for each group. From the definition of X_h we see that it is equivalent to saying that n_i, $\sum_i x$, $\sum_i x^2$, $\cdots$ $\sum_i x^{2h}$ be identical in i. As h increases to s, we come to the condition given at the beginning of this paragraph.

§3. If this condition is not fulfilled, we can, indeed, replace the power series in x by orthogonal polynomials $X_{h \cdot i}$, the second subscript being appended in order to show that the values of the X polynomials are no more identical for the several groups; these polynomials are now orthogonalized separately within each group. But we are no more able to predetermine the values of $A_0, B_0, \cdots M_0$, as they depend on each other; this will be made clear a little later. Therefore we have to resort to an approximation: the values of the parameters will not be found from simultaneous equations, but successively, step by step, beginning with those corresponding to the highest degree of the independent variable.

The values of $a_0, b_0, \cdots m_0$ are given. It is evident that $m_0 = M_0$. The j-th normal equation is now:

$$M_j \sum_j X_{s \cdot j}^2 - M_0 \sum_r X_{s \cdot r}^2 + \left(\sum_1^{r-1} M_i\right)\left(\sum_r X_{s \cdot r}^2\right) = \sum_j X_{s \cdot j} y - \sum_r X_{s \cdot j} y. \tag{12}$$

We see at once that

$$M_i = \frac{M_j \sum_j X_{s \cdot j}^2 + \sum_i X_{s \cdot i} y - \sum_j X_{s \cdot j} y}{\sum_i X_{s \cdot i}^2}. \tag{13}$$

Inserting this into /12/ we get

$$M_j = \frac{\sum_j X_{s \cdot j} y}{\sum_j X_{s \cdot j}^2} - \frac{1}{\sum_j X_{s \cdot j}^2} \cdot \frac{\sum_1^r \frac{\sum_i X_{s \cdot i} y}{\sum_i X_{s \cdot i}^2} - M_0}{\sum_1^r \frac{1}{\sum_i X_{s \cdot i}^2}}. \tag{14}$$

The second member of the right hand side of /14/ is again a correction term, the necessary amount of correction being distributed in inverse proportion to $\sum_j X_{s \cdot j}^2$. Now we determine the value of L_0, this coefficient corresponding to $s - 1$, the second highest degree of x, and calculate the several L's from equations strictly analogous to (14) thus accomplishing the second step of our work, and so on, down to the A's. L_0 is found from the following equation:

$$L_0 = l_0 - \sum_1^r [\alpha_{s-1}^s(i) \cdot M_i]. \tag{15}$$

To α_{s-1}^s is now appended a bracketed i, this to stress its variation from group to group. We see from (15) that before the several M's are calculated we are not in a position to determine L_0. On the other hand, if α_{s-1}^s is the same for all groups, the second member of the right hand side of (15) simply reduces to $\alpha_{s-1}^s \cdot m_0$ and L_0 can be determined in advance, i.e. before calculating the M's. This is the case treated first (in §2). In any case, if no definite correlation is to be expected between $\alpha_{s-1}^s(i)$ and M_i, the approximative method developed here should give very nearly correct results. The writer applied this method of solution to the simple problem of seasonal variation mentioned in §1 and found the results very satisfactory.