Remark on the residuals obtained by least squares

Wesselink, A.J.

Remark on the residuals obtained by least squares, by *A. J. Wesselink.*

The residuals of a least-squares solution, to be expected a priori, differ from one equation of condition to the other. A formula is derived, which predicts the dispersion of a residual for each equation.

The present note deals with a problem in least squares as it occurred to the writer in the course of some computations. The resulting formula was thought unknown; it may possibly be useful to others and is therefore communicated here.

Before we treat the general case we will illustrate the problem by a simple example. Consider the determination by least squares of a straight line through a number of points the abscissae of which are exact but whose ordinates are subject to accidental errors.

Suppose the distribution of the abscissae such that the great majority of the abscissae are clustered together whereas there is a single point of which the abscissa differs considerably from those of the rest. It is then obvious and well known that the most probable straight line determined by least squares is likely to pass almost exactly through this extreme point, even though the weight may be the same as that of the other points.

Even if the situation is not so extreme but if the distribution over the abscissae is rather even, the probable value of a residual differs from point to point. Our problem in this particular case consists in the determination of the dispersion of each residual as a function of the abscissae of all points and that of the corresponding point in particular.

In a more abstract but at the same time more general way we may state our problem as follows. Consider an arbitrary number of equations of condition, which for the sake of simplicity are taken of equal weight. Let the number of unknowns be two; the extension to more unknowns and to conditional equations with unequal weights is indicated at the end of this note.

We have: $a_r x + b_r y = O_r$, n equations with the same weight. Let the equations be solved by means of least squares and let residuals have been formed. We express $(O-C)_r$ as a linear function of the independent quantities O_r, each with a mean error ε.

The dispersion of the r^{th} residual $(O-C)_r$ is then found by means of the well known formula for the mean error of a linear function of independent quantities.

The unknowns may be written:

$x = [\alpha O]$ The α's and β's are functions of the
$y = [\beta O]$ coefficients a, b.

We have

$$a_r[\alpha O] + b_r[\beta O] = C_r$$
$$\sum_s (a_r\alpha_s + b_r\beta_s)\, O_s = C_r \qquad s = 1, 2 \ldots . n$$

Hence

$$(1 - a_r\alpha_r - b_r\beta_r)\, O_r - \sum_{\substack{s \\ r \neq s}} (a_r\alpha_s + b_r\beta_s)\, O_s = (O-C)_r.$$

The square of the dispersion of the r^{th} residual is therefore:

$$\{a_r^2\,[\alpha\alpha] + b_r^2\,[\beta\beta] + 2a_rb_r\,[\alpha\beta] + 1 - 2\,(a_r\alpha_r + b_r\beta_r)\}\,\varepsilon^2.$$

We introduce quantities p_{ij}. They express the unknowns in the right-hand members X, Y, of the normal equations:

$$x = p_{11}X + p_{12}Y$$
$$y = p_{21}X + p_{22}Y$$
$$X = [paO],\ Y = [pbO].$$

These quantities are usually obtained during the course of the solution with the purpose of deriving the weights of the unknowns.

We have

$$p_{11} = [\alpha\alpha] \quad p_{22} = [\beta\beta] \quad p_{12} = p_{21} = [\alpha\beta]$$

further

$$\alpha_r = p_{11}a_r + p_{12}b_r$$
$$a_r\alpha_r = p_{11}a_r^2 + p_{12}a_rb_r$$
$$b_r\beta_r = p_{21}a_rb_r + p_{22}b_r^2$$

The required result is therefore: dispersion^2 of the r^{th} residual equals $(1 - p_{11}a_r^2 - p_{22}b_r^2 - 2p_{12}a_rb_r)\,\varepsilon^2$.

As a check on the result we may form the sum of this expression over r. The result is $(n-2)\varepsilon^2$, as it should be in the case of two unknowns.

When we have a problem with more than two unknowns the extension of the formula is obvious.

When the equations have different weights p_r, and ε is the mean error corresponding to unit weight, the disp^2 of the r^{th} residual is:

$$(p_r^{-1} - p_{11}a_r^2 - p_{22}b_r^2 - 2\,p_{12}a_rb_r)\,\varepsilon^2.$$

Examples.

a) Consider the determination of a straight line through three points, the abscissae of which are exact and equidistant. The ordinates are of equal weight. The disp^2 of each of the end-ordinates is $\frac{1}{6}\varepsilon^2$. The disp^2 of the central ordinate is $\frac{2}{3}\varepsilon^2$.

When there are four equidistant points we find for each of the end-ordinates a disp^2 of $\cdot 3\varepsilon^2$. Each of the central ordinates has a disp^2 of $\cdot 7\varepsilon^2$.

b) When a double-star orbit is determined from a set of observations which do not cover a complete period, it often happens that there is a single solitary position (e.g. an old position by HERSCHEL), which has a considerable effect on the orbit and which shows a small residual. The formulae given above allow to compute how closely calculated and observed positions should agree. This is of particular interest for a solitary position.

c) Our formula applied to the case of one unknown reduces to the well known formula for the determination of the mean error from differences from the arithmetic mean value: $\varepsilon^2 = \dfrac{\Sigma\Delta^2}{n-1}$.