THE ESTABLISHING OF THE COEFFICIENTS OF CORRELATION BETWEEN THE COMPENSATED VARIABLES OR BETWEEN THE FUNCTIONS OF THE COMPENSATED VARIABLES BY RIGOROUS METHODS

Gheorghe NISTOR, Professor Ph. D. Eng., "Gheorghe Asachi" Technical University of Iaşi, Romania, e-mail: ghnistor@yahoo.fr

***Abstract**: The correlated or dependent measurements are these ones for which the set of conditions in which they are accomplished, influence totally or partially the result of other measurement. The real correlation between the originally measurements are expressed by the aid of a coefficient of correlation. The concept of correlation is important in statistical studies and it concerns the connection between the observed variables. Among the measured variables which enter a compensation, the correlation can be mathematical or physical one. The modality of calculation for the coefficients of correlation between the compensated variables and / or between the functions of compensated variables are presented by rigorous method.*

***Keywords**: variable, coefficient, correlation, rigorous.*

1. Introduction

The concept of correlation is a very important notion used in the study of the statistical repartitions and concerns the connection between the observed variables and between the represent phenomena.

Two phenomena one can situated in a strict connection, rigid one, expressed by formula $y = f(x)$ where, for a determined value of independent variable x, the dependent variable y is also a discreet determined value. The opposed situation is that of the complete independence of two phenomena. Between these two extremes is stretched the wide field of the supple univocal connections of statistical or stochastical type, when to every value x correspond not a single value y, but a repartition connected to values y. The statistical connection is generated by multiple factors with influences in various senses and intensities. It is the expression of the co-action of the necessity and random, the necessity being realized by means of random factors.

The correlation includes two fundamental problems: the first one consists in describing the law of mean variation of a variable by respect to another variable, and the second one in characterization of the intensity of the connection by means of a numerical coefficient independent of the unit of measure of the variables.

In this work are presented the modality of the calculation for the coefficients of correlation, as long between the variables compensated by rigorous methods, as between the functions of these variable, for the case of geodetic measurements.

2. Physical correlations and mathematical correlations

Among the variables which enter a compensation, the correlations can be: physical and mathematical.

A. The physical correlations are produced by the influence of the multiply exterior conditions about whole process of measuring. In this case, the correlation of the variable measured is expressed with the aid of a empirical coefficient of correlation determined experimentally by using the statistical processing of a great number of measurements. In practice, the rigorous physical correlations are aproximative enough because of great and complex difficulties of their study. For this reason, the present paper doesn't make any reference about practical establishing of physical correlation.

B. The mathematical correlations are created especially by using of a mathematical model, incomplete or affected by errors of conception, named by F.R. Helmert, errors of the theory.

Taking into consideration the sequence of measurements M_1^0, $M_2^0, \ldots,$ M_n^0 resulted from a complex process of measuring, in a geodetic network and eliminating the influence of the systematic errors, the vector M^0 of the measurements will be considered as an aleatory variable. The real or true modulus of the vector will be expressed by the relation $\overline{M} = F(M_0)$ where each component of the vector will results as a mean of a greatest number of measurements. The real errors represent the differences $\varepsilon_i = M_i^0 - M_i$, $i = \overline{1,n}$ and the stochastical of the real errors ε_i are determined by the covariance (correlation) matrix:

$$C_M = F\left(\varepsilon\varepsilon^T\right) = \begin{bmatrix} \sigma_1^2 & r_{12}\sigma_1\sigma_2 & \dots & r_{1n}\sigma_1\sigma_n \\ r_{21}\sigma_2\sigma_1 & \sigma_2^2 & \dots & r_{2n}\sigma_2\sigma_n \\ \dots & \dots & \dots & \dots \\ r_{n1}\sigma_n\sigma_1 & r_{n2}\sigma_n\sigma_2 & \dots & \sigma_n^2 \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \dots & \sigma_{1n} \\ \sigma_{21} & \sigma_2^2 & \dots & \sigma_{2n} \\ \dots & \dots & \dots & \dots \\ \sigma_{n1} & \sigma_{n2} & \dots & \sigma_n^2 \end{bmatrix}, \tag{1}$$

where we denote by:

σ_i – (theoretical) variance of measurement M_i^0, $\left(i = \overline{1,n}\right)$ expressed as a function of real errors

$$\sigma_i^2 = F\left(\varepsilon_i^2\right), \tag{2}$$

σ_{ij} – (theoretical) covariance of the measurements M_i^0 and M_j^0

$$\sigma_{ij} = F\left(\varepsilon_i, \varepsilon_j\right);\ i, j = \overline{1,n};\ i \neq j, \tag{3}$$

r_{ij} – *the correlation coefficient* between the variables M_i^0 and M_j^0, expressed by formula

$$r_{ij} = \frac{\sigma_{ij}}{\sigma_i \sigma_j}. \tag{4}$$

In the mathematical statistics, σ_i is named *standard deviation*. These values are situated along the principal diagonal of matrix (1) and represent the spreadings of one-dimensional aleatory variable. The other elements σ_{ij} , ($i \neq j$), characterize the intensity of the correlation between the components of the considered vector. The equality $\sigma_{ij} = \sigma_{ji}$ is named *moment of correlation* of the two aleatory variables. The intensity of the correlation between elements M_i^0 and M_j^0 are expressed by the size of the correlation coefficient r_{ij} and has the properties as follows:

a) Characterizes the connection between two variables taken simultaneous, being situated in [–1, 1] interval *i.e.*

$$-1 \leq r_{ij} \leq +1. \tag{5}$$

b) If $r = \pm 1$, between the two values there are a linear connection; if $r > 0$, the both values vary along the same direction, and in the contrar direction for $r < 0$.

c) When the coefficient of correlation takes values close by –1 or +1, then the rectilinial correlation is more intense.

d) If the two variables are independent, the correlation coefficient vanishes. Reciprocal property is not true always; it is not correct to consider that the two variables are independent, whenever the correlation coefficient vanishes. Also, although the coefficient of correlation can be small, between the two studied variables there exists a significant correlation.

When the n components of the studied vector are independent, then $\sigma_{ij} = 0$, and the *covariant matrix* has the form

$$\sigma_{ij} = \begin{bmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \sigma_n^2 \end{bmatrix}. \tag{6}$$

To emphasize better the correlation intensity one use the *normed covariant matrix*

$$\left[r_{ij}\right] = \begin{bmatrix} 1 & r_{12} & \dots & r_{1n} \\ & 1 & \dots & r_{2n} \\ & & \dots & \dots \\ & & & 1 \end{bmatrix}. \tag{7}$$

In the geodesical measurement practical, the real errors are replaced by the probable errors, and the *standard deviation* σ_0 are replaced by the square mean errors s_0, in the case of measurements of the same precision, or by the errors of the weight unit μ in the case of weighted measurements. Theoretically, one admite that $s_0 \to \sigma_0$, when $n \to \infty$.

There are known the connections

$$C_M = s_0^2 Q_M, \text{ or } \overline{C}_M = \mu^2 \tilde{Q}_M, \tag{8}$$

where s_0^2 and μ^2 are some constants representing the variance of the errors, and Q_M or $\tilde{Q}_M$ are the *matrices of the cofactors* of the measurements represented by

$$Q_M = \frac{C_M}{s_0^2} = \begin{bmatrix} Q_{11} & Q_{12} & \dots & Q_{1n} \\ Q_{21} & Q_{22} & \dots & Q_{2n} \\ \dots & \dots & \dots & \dots \\ Q_{n1} & Q_{n2} & \dots & Q_{nn} \end{bmatrix}; \quad \tilde{Q}_M = \frac{\overline{C}_M}{\mu^2} = \begin{bmatrix} \overline{Q}_{11} & \overline{Q}_{12} & \dots & \overline{Q}_{1n} \\ \overline{Q}_{21} & \overline{Q}_{22} & \dots & \overline{Q}_{2n} \\ \dots & \dots & \dots & \dots \\ \overline{Q}_{n1} & \overline{Q}_{n2} & \dots & \overline{Q}_{nn} \end{bmatrix}. \tag{9}$$

The elements of the two matrices Q_M and $\tilde{Q}_M$ are named *cofactors* and represent the weight coefficients of the measured variables. By the aid of these cofactors, the correlation coefficients of the compensated values one can calculate:

$$r_{ij} = \frac{Q_{ij}}{\sqrt{QiiQjj}}, \text{ respectively } \overline{r}_{ij} = \frac{\overline{Q}_{ij}}{\sqrt{\overline{Q}_{ii}\overline{Q}_{jj}}}. \tag{10}$$

The analyses of the relations (10) lead us to the conclusion: the necessary and sufficient condition that the measured values M_i^0 to be independent is vanishing of all rectangular coefficients of weight of matrix (9) ($Q_{ij} = 0$, $i \neq j$).

3. The establishing of the correlation coefficients between the compensated values, treated by indirect measurements method

In the compensating operations of the geodesic measurements by rigorous methods, for example by indirect measurements method, one ask to determine the square mean errors of some function of compensated variables of which the principal terms are cofactors.

In the case of a function of compensated variable of the unknown quantity, that is of a function of variables of which values were determined by indirect measurements of general appearance

$$F = f_{1n}^T X_{n1}, \tag{11}$$

where f^T is a line matrix of some constant variables

$$f_{1n}^T = \begin{bmatrix} f_1 & f_2 & \dots & f_n \end{bmatrix}, \tag{12}$$

and X – column matrix of the unknown quantities, expressed by formula

$$X_{n1} = -\tilde{Q}_{nn} B_{nr}^T L_{r1} = -C_{nr} L_{r1}. \tag{13}$$

Replacing the vector of the unknowns, the function one expresses direct by respect to measured values on the ground

$$-F = f_{1n}^T C_{nr} L_{r1}. \tag{14}$$

Differentiating by respect to L, one obtain

$$-\mathrm{d}(F) = f_{1n}^T C_{nr} \mathrm{d}(L). \tag{15}$$

Then applying the expression of the errors of a function of direct measured variables, it results for the square mean error of the function of compensated values of the unknowns, as follows

$$s_F^2 = s_0^2 f_{1n}^T C_{nr} C_{rn}^T f_{n1}. \tag{16}$$

Because of $C_{nr} C_{rn}^T = \tilde{Q}_{nn}$, the square of the error of the function will have the form

$$s_F^2 = s_0^2 f_{1n}^T \tilde{Q}_{nn} f_{n1}, \tag{17}$$

or, in developed appearance

$$\begin{aligned} s_F^2 = s_0^2 \Big\{ & f_1^2 Q_{11} + 2 f_1 f_1 Q_{12} + \ldots + 2 f_1 f_n Q_{1n} + \\ & + f_2^2 Q_{22} + \ldots + 2 f_2 f_n Q_{2n} + \\ & \ldots\ldots\ldots\ldots\ldots \\ & + f_n^2 Q_{nn} \Big\}. \end{aligned} \tag{18}$$

Accomplishing the operations of multiplication and using the factor $\sqrt{Q_{ii} Q_{jj}} \Big/ \sqrt{Q_{ii} Q_{jj}}$, one obtain

$$\left.\begin{aligned} s_F^2 = {} & f_1^2 s_0^2 Q_{11} + f_2^2 s_0^2 Q_{22} + \ldots + f_n^2 s_0^2 Q_{nn} + \ldots\ldots \\ & + 2 f_1 f_2 \frac{Q_{12}}{\sqrt{Q_{11}Q_{22}}} s_0 \sqrt{Q_{11}} s_0 \sqrt{Q_{22}} + \ldots\ldots\ldots\ldots \\ & \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ & + 2 f_{n-1} f_n \frac{Q_{n-1,n}}{\sqrt{Q_{n-1,n-1}Q_{nn}}} s_0 \sqrt{Q_{n-1,n-1}} s_0 \sqrt{Q_{nn}} \end{aligned}\right\}. \tag{19}$$

One denotes by

$$r_{ij} = \frac{Q_{ij}}{\sqrt{Q_{ii}Q_{jj}}};\quad i, j = \overline{1,n};\ i \neq j, \tag{20}$$

the coefficient of correlation between the compensated values of the unknowns (of the parameters indirect determined) and by

$$s_i^2 = s_0^2 Q_{ii},\quad s_i = \pm s_0 \sqrt{Q_{ii}};\quad i = \overline{1,n} \tag{21}$$

the square mean errors of the unknowns.

Finally, the mean error square of a function of the compensated values of the unknowns indirectly determined will be

$$\left.\begin{aligned} s_F^2 = {} & f_1^2 s_1^2 + 2 f_1 f_2 s_1 s_2 r_{12} + \ldots + 2 f_1 f_n s_1 s_n r_{1n} + \\ & + f_2^2 s_2^2 + \ldots\ldots\ldots\ldots\ldots + 2 f_2 f_n s_2 s_n r_{2n} + \\ & \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ & + f_{n-1}^2 s_{n-1}^2 + 2 f_{n-1} f_n s_{n-1} s_n r_{n-1,n} + \\ & + f_n^2 s_n^2 \end{aligned}\right\}. \tag{22}$$

One observe that in the case when the general function (11) has a linear form, the constant coefficients f_{1n}^T are represented by values of the partial derivatives of the function by respect to provisional values X_{n1}^0 of the parameters indirectly determined $\left(f_j = \partial F / \partial X_j^0 \right)$.

In the automatic calculus case, the values of the cofactors are obtained by inversation of the coefficients matrix of the normal equations $\left(\tilde{Q}_{nn} = N_{nn}^{-1} \right)$.

4. The establishing of the coefficients of correlation between functions, case of a composed function of adjusted values rigorously treated by indirect measurements methods

In the case of a composed function of the form

$$\Phi = m'\varphi + m''\psi, \tag{23}$$

the values m' and m'' are some constant, and φ *and* ψ are linear functions of adjusted values of the unknowns, in fact, of parameters determined by indirect measurements

$$\varphi = f_{1n}^{'T} X_{n1};\quad \psi = f_{1n}^{''T} X_{n1}. \tag{24}$$

The quantities f' and f'' appearing in (24) and having the form

$$f_{1n}^{'T} = \begin{bmatrix} f_1^{'} & f_2^{'} & \dots & f_n^{'} \end{bmatrix}; \; f_{1n}^{"T} = \begin{bmatrix} f_1^{"} & f_2^{"} & \dots & f_n^{"} \end{bmatrix}, \tag{25}$$

are constant values and X is the vector of unknowns

$$X_{n1} = X_{1n}^T = \begin{bmatrix} x_1 & x_2 & \dots & x_n \end{bmatrix}. \tag{26}$$

Replacing φ and ψ into (23), the composed function will have the form

$$\Phi = m' \, f_{1n}^{'T} X_{n1} + m" \, f_{1n}^{"T} X_{n1}. \tag{27}$$

The vector of unknowns being

$$X_{n1} = -\tilde{Q}_{nn} B_{nr}^T L_{r1} = -C_{nr} L_{r1}, \tag{28}$$

the composed function (function of function) appears as a simple function expressed linearly by the vector of free terms, which represents the direct measurements accomplished on the ground

$$-\Phi = \left(m' \, f_{1n}^{'T} C_{nr} + m" \, f_{1n}^{"T} C_{nr} \right) L_{r1}. \tag{29}$$

Appling the formulas of the mean errors for a function of direct measured values, expressed in differential form, it results

$$-\mathrm{d}(\Phi) = \left(m' \, f_{1n}^{'T} C_{nr} + m" \, f_{1n}^{"T} C_{nr} \right) \mathrm{d}(L). \tag{30}$$

Passing from differential to the square mean error and taking into consideration $C_{nr} C_{rn}^T = \tilde{Q}_{nn}$, for the square of the error of composed function, it results the expression

$$s_\Phi^2 = s_0^2 \left(m' \, f_{1n}^{'T} \tilde{Q}_{nn} f_{n1}^{'} m' + m" \, f_{1n}^{"T} \tilde{Q}_{nn} f_{n1}^{"} m" + 2m' m" \, f_{1n}^{'T} \tilde{Q}_{nn} f_{n1}^{"} \right). \tag{31}$$

Because the direct measurements have the same precission, $\left(s_{l_1} = s_{l_2} = ... = s_{l_n} = s_0 \right)$, it means that s_0 is the square mean error of a single direct measurement. The matrix of the cofactors is obtained by inverting the matrix of the normal equations $\left(\tilde{Q}_{nn} = N_{nn}^{-1} \right)$. In formula (31) the first two terms in parenthesis contain the cofactors of the two functions, *i.e.*

$$Q_{\varphi\varphi} = f_{1n}^{'T} \tilde{Q}_{nn} f_{n1}^{'} = \begin{bmatrix} f_1^{'}, f_2^{'}, ..., f_n^{'} \end{bmatrix} = \begin{bmatrix} Q_{11} & Q_{12} & \dots & Q_{1n} \\ Q_{21} & Q_{22} & \dots & Q_{2n} \\ \dots & \dots & \dots & \dots \\ Q_{n1} & Q_{n2} & \dots & Q_{nn} \end{bmatrix} \begin{bmatrix} f_1^{'} \\ f_2^{'} \\ \dots \\ f_n^{'} \end{bmatrix}, \tag{32}$$

$$Q_{\psi\psi} = f_{1n}^{"T} \tilde{Q}_{nn} f_{n1}^{"} = \begin{bmatrix} f_1^{"}, f_2^{"}, ..., f_n^{"} \end{bmatrix} = \begin{bmatrix} Q_{11} & Q_{12} & \dots & Q_{1n} \\ Q_{21} & Q_{22} & \dots & Q_{2n} \\ \dots & \dots & \dots & \dots \\ Q_{n1} & Q_{n2} & \dots & Q_{nn} \end{bmatrix} \begin{bmatrix} f_1^{"} \\ f_2^{"} \\ \dots \\ f_n^{"} \end{bmatrix}, \tag{33}$$

and the third term contains the rectangular cofactor

$$Q_{\varphi\psi} = f_{1n}^{'T} \tilde{Q}_{nn} f_{n1}^{"} = \left[f_1^{'}, f_2^{'}, ..., f_n^{'} \right] = \begin{bmatrix} Q_{11} & Q_{12} & \dots & Q_{1n} \\ Q_{21} & Q_{22} & \dots & Q_{2n} \\ \dots & \dots & \dots & \dots \\ Q_{n1} & Q_{n2} & \dots & Q_{nn} \end{bmatrix} \begin{bmatrix} f_1^{"} \\ f_2^{"} \\ \dots \\ f_n^{"} \end{bmatrix}. \tag{34}$$

Accomplishing adequate substitutions, the expression of the square of the composed function error will be

$$s_\Phi^2 = s_0^2 \left(m^{'2} Q_{\varphi\varphi} + m^{"2} Q_{\psi\psi} + 2m^{'} m^{"} Q_{\varphi\psi} \right). \tag{35}$$

As in the 2 case, one obtains the formula

$$s_\Phi^2 = m^{'2} s_0^2 Q_{\varphi\varphi} + m^{"2} s_0^2 Q_{\psi\psi} + 2m^{'} m^{"} \frac{Q_{\varphi\psi}}{\sqrt{Q_{\varphi\varphi} Q_{\psi\psi}}} s_0 \sqrt{Q_{\varphi\varphi}} s_0 \sqrt{Q_{\psi\psi}}. \tag{36}$$

One denotes by

$$r_{\varphi\psi} = \frac{Q_{\varphi\psi}}{\sqrt{Q_{\varphi\varphi} Q_{\psi\psi}}}, \tag{37}$$

the expression of the coefficient of correlation between the two component functions. On the basis of this expression will be established the independence or correlation (dependence) between the two functions. Denoting by

$$s_\varphi = \pm s_0 \sqrt{Q_{\varphi\varphi}}; \;\; s_\psi = \pm s_0 \sqrt{Q_{\psi\psi}}, \tag{38}$$

the sizes of the values of the errors of functions, finally one obtains the expression of the square of error of the composed function

$$s_\Phi^2 = m^{'2} s_\varphi^2 + m^{"2} s_\psi^2 + 2m^{'} m^{"} s_\varphi s_\psi r_{\varphi\psi}. \tag{39}$$

In the particular case when the sizes of the constants of the composed function are equal to unit, $m' = m'' = 1$, one obtains a simplified form

$$s_\Phi^2 = s_\varphi^2 + s_\psi^2 + 2 s_\varphi s_\psi r_{\varphi\psi}, \tag{40}$$

for the formula of the square error of the composed function. One specifies that the cofactors $Q_{\varphi\varphi}$, $Q_{\psi\psi}$ and $Q_{\varphi\psi}$ are calculated on the basis of the coefficients $f_j^{'}$ and $f_j^{"}$, $j = \overline{1,n}$, and of the correlation matrix $\tilde{Q}_{nn}$, also using the relations (32), (33), (34).

When the indirect measurements are weighted, one acts identically.

The solving of the linear system of equations of the weighted corrections will be realized under the condition of minimum $V_{1r}^T P_{rr} V_{r1}$ = min. and the error s_0 will be replaced by weighted unit error denoted by μ.

5. Conclusions

The idea of connection between observed variables in the statistical collectivities, transposed in mathematical language by Galton-Pearson biometric school, one constitutes into statistical correlation theory. The correlation describes the mean variation law of one variable by respect to others variables as well as the characterization of the connection intensity by a numeric coefficient, that is, the coefficient of correlation, which is independent of the units of the variables.

In this paper are presented the modalities of calculation for the correlation coefficients between the rigorously compensated values, by indirect measurements method, formula (20), and between the functions, case of the composed functions of values rigorously compensated, formula (37). On this basis, the independence or dependence of the values or functions are established.

In the compensation of the geodetic measurements one determine also the mean errors of some functions in which the principal terms are cofactors. In this case, the functions for which the conditions $Q_{ij} = 0;\ i \neq j$ are satisfied, are named *orthogonal functions*. These functions can be treated as independent elements in a subsequent processing, having the same character of independence as of the originally measurements.

6. References

1. *Ciumac P., ş.a., Teoria probabilităţilor şi elemente de statistică matematică, Ed. Tehni-că, Chişinău, 2003.*
2. *Ghiţău D., Geodezie – Triangulaţie, Editura Didactică şi Pedagogică, Bucureşti, 1972.*
3. *Moineagu C., ş.a., Statistica, Editura Ştiinţifică şi Enciclopedică, Bucureşti, 1976.*
4. *Nistor Gh., Teoria prelucrării măsurătorilor geodezice, Ed. Univ. Tehn. „Gheorghe Asachi", Iaşi, 1996.*
5. *Nistor Gh., Teoria prelucrării măsurătorilor geodezice – Lucrări practice, Ed. Univ. Tehn. „Gheorghe Asachi", Iaşi, 1998.*
6. *Tiron M., Teoria erorilor de măsurare şi metoda celor mai mici pătrate, Ed. Tehnică, Bucureşti, 1972.*