V. BIȚĂ

Automatic Control and Correction — a Premise for the Efficient Utilization of Computer in Data Processing

"The computer is having an implosive effect upon the way we deal with a variety of problems."

C. CSUN, JO. SHAFFER

The organization of an informational system based on the use of the electronic computer, usually, aims at the attainment of several objectives. One of the most important objectives is the ensurance of the quality of data, especially of input data. To this purpose, the respective data are systematically and rigorously verified in all the phases of the informational system.

The manual verification of data is, as a rule, an extremely big and expensive operation, without all eventual errors be discovered in this way. Under the conditions of the use of the data processing electronic computer, the automation of their control is becoming a possible operation and, as experience shows, with extremely efficient results.

SOURCES AND TYPES OF ERRORS

The data submitted to electronic processing may, generally, have mistakes for several reasons, among which we mention:

a) the wrong filling in of various primary documents and data collecting printed forms, such as for example: wrongly calculated indicators, incomplete columns and positions, the use of several units of measure (thousands of lei and lei, tons and kilogrammes, etc.) the non-observance of methodological instructions;

b) the wrong codification of various characteristics; especially in the conditions of decentralized codification;

c) the wrong perforation of data on cards, such as for example: double cards, missing cards, diferring perforations, inversion of figures, figures wrongly interpreted by operators.

Although the perforation control operation is made before the processing on the computer, many of the erroneous data continue to be undiscovered, especially because of the wrong filling in of the printed forms and because of the codification.

Several automatic control methods are used for finding out all or almost all errors with the aid of the electronic computer, methods from which we present those with a larger applicability.

DATA AUTOMATIC CONTROL METHODS

By its nature control may be arithmetical or logical.

Arithmetical control is used when there are certain arithmetical relations among the various data contained in the registering form and it is based on one of the following methods:

— Control by zero. Example:

$$i_1 \pm i_2 \pm i_3 \pm \dots \pm i_n = \Sigma i \tag{1}$$

$$i_1 + i_2 + i_3 + \dots \pm i_n - \Sigma i = 0 \tag{2}$$

where i is the size of an indicator.

In this case, the automatic control programme will contain instructions for the verification of relation (2). If the left side of relation (2) is equal to zero, the data are accepted, if not — the data will be refused or

$$R = i_1 \pm i_2 \pm i_3 \pm \dots \pm i_n \leqslant N \tag{3}$$

$$R = i_1 \pm i_2 \pm i_3 \pm \dots \pm i_n - N \leqslant 0. \tag{4}$$

The automatic control programme will also contain instructions for the verification of relation (4):

— if $R \leqslant 0$, the data are accepted

— if $R > 0$, the data are refused.

Control by relation. It is based on placing the value of a relation within certain previously established limits.

Two quantitative-value indicators M and N are verified by this method, the value of the relations between them being used as a control size. For the acceptance of the two indicators, the control size must have a value comprised between a lower and an upper limit.

Consequently, the acceptance of the relation $R = \frac{M}{N}$ depends on the satisfaction of condition

$$l_i \leqslant R \leqslant l_s.$$

In this case, the scheme of the control programme is presented in figure 1.

An example of the application of this method may be the verification of the value and quantity by calculating the price which may vary within generally known limits. A particular case of this method may be the control, instead of the relation, of any other arithmetic expression.

Logical control is based on the correlation which usually exists between the various characteristics of the registered phenomenon. As, for processing, the characteristics are codified, it means that the correlation between the respective codes will always be verified.

Several methods are used in case of logical and arithmetical control.

Verification of the code by itself. This implies that a certain code be comprised between zero and a maximum limit, for example: the code of the localities within a county, the code of the enterprises within a ministry.

In this case, a limit control is made, refusing the cases in which the codes surpass the admitted maximum limit.

Another control method of a code by itself lies in the verification, through the intermediate of a control figure, calculated beforehand, on the basis of a certain algorithm and added to the respective code, of the correctness of the registering of the code on various documents or of its perforation. Let us suppose, for example, that for the codification of the materials in an enterprise, we use a decimal code of 8 figures, the first figure of which represents the synthetic book-keeping account of affiliation, the next two figures the group of materials and the last three figures the assortment proper.

The control figure C might be calculated by using the following algorithm:

Code a_1, a_2, a_3, a_4, a_5, a_6, a_7, a_8 — figures of the code

Weights 2 1 5 7 8 4 2 1

The control figure C will be equal to the rest obtained by effectuating the division:

$$\frac{\sum_{k=1}^{n} A_k W_k}{M}$$

where:

A_k represents the figures of the code

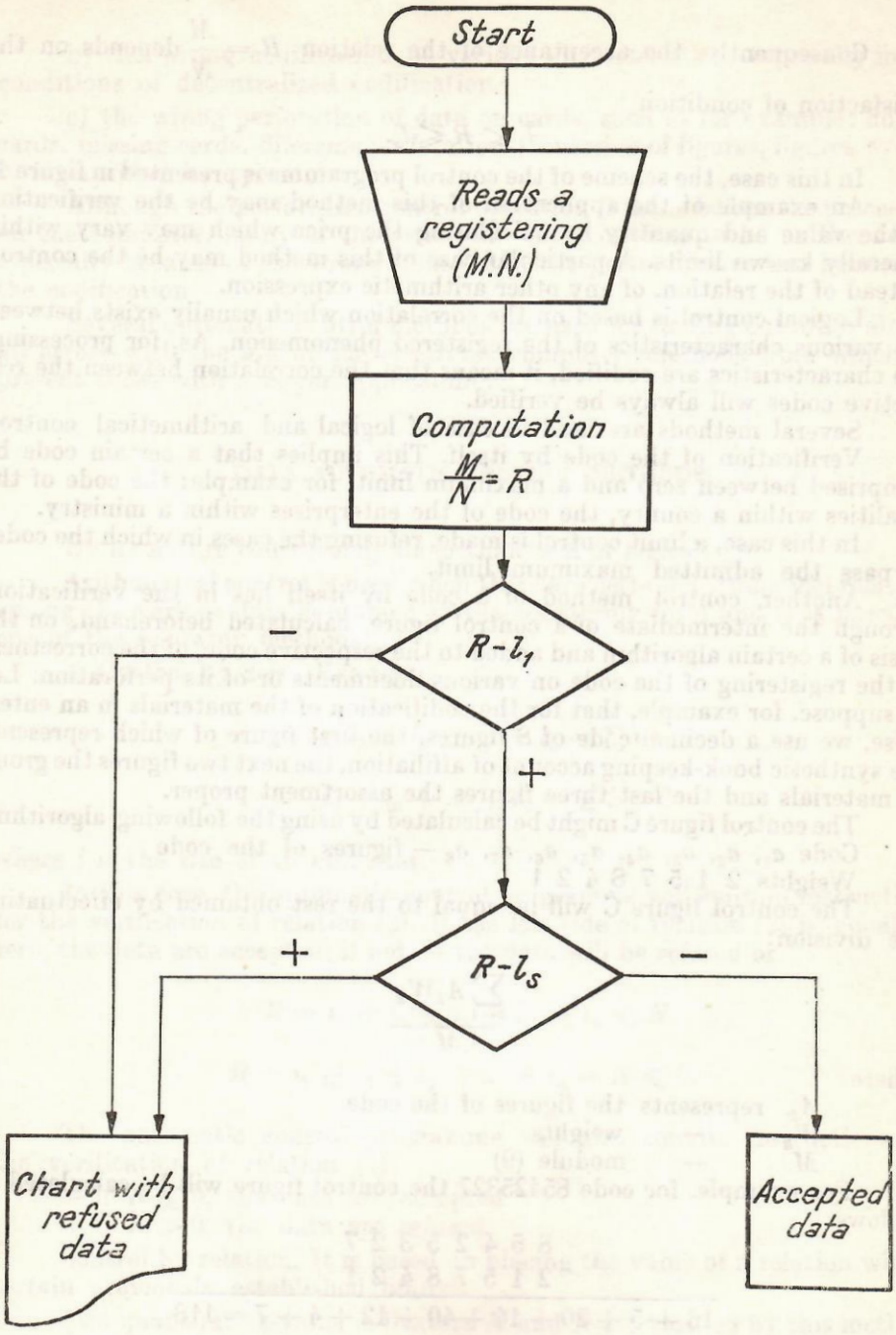
W_k — weights

M — module (9)

For example, for code 85425327 the control figure will be calculated as follows:

$$\begin{array}{r} 8\ 5\ 4\ 2\ 5\ 3\ 2\ 7 \\ 2\ 1\ 5\ 7\ 8\ 4\ 2\ 1 \\ \hline 16 + 5 + 20 + 14 + 40 + 12 + 4 + 7 = 118 \end{array}$$

$$\frac{118}{9} = 13 \text{ rest } 1.$$



Scheme of control program by method of relation

Fig. 1

If $C = 1$ is added to the initial code, it becomes 854253271. The verification of the code is made by effectuating the same calculation and by comparing the control figure obtained to that inscribed, perforated or registered. If there is a discordance between these control figures, the respective code will be automatically refused for being corrected.

Let us suppose, for example that on the registering of the code, two figures (a_4 and a_5) are inversed, a rather frequent mistake and the code appears as 854523271, which will obviously have an entirely different significance.

In this case, the control figure will be

$$\begin{array}{r} 8\ 5\ 4\ 5\ 2\ 3\ 2\ 7 \\ 2\ 1\ 5\ 7\ 8\ 4\ 2\ 1 \\ \hline 16 + 5 + 20 + 35 + 16 + 12 + 4 + 7 = 115 \end{array}$$

$$\frac{115}{9} = 12 \text{ rest } 7$$

Consequently, $7 \neq 1$, the respective code will be refused for being corrected manually.

This automatic control method is extremely efficient, a very large range of errors being eliminated through its application. We should also mention the existence of other methods for determining the control figure.

VERIFICATION OF THE CORRELATION BETWEEN TWO OR SEVERAL CODES

Two aspects are distinguished in this case:

— the presence of a code implies the compulsory presence of another code; for example, the code for "place of work" requires the presence of the code for "occupation";

— the presence of a code implies the compulsory absence of another code;

The incompatibilities automatically discovered by the electronic computer may be:

— unquestionable, in which case the correction is, as a rule, automatically carried out, for example: in case of the age under 15, there cannot be a code for the last graduated school;

— questionable, in which case, the correction is carried out automatically, from each case to another, for example: in case of the age 15—19 for men and 15—17 for women, the codes for married, divorced, widow are questionable.

AUTOMATIC CORRECTION METHODS

Through the data automatic control, the erroneous data will be discovered and refused, going to be corrected. As a rule, the correction operation causes interruptions in the processing flow having negative repercussions on the operativity and efficiency of processing.

By using the data processing electronic computer, the automation, to a great extent, of the operation of correction of the erroneous cases is also possible.

The use of the electronic computer also allows the automatic attribution of codes when they are missing, the respective case being also considered as an error.

Of course, the error automatic correction methods have a relatively restricted applicability, as they cannot be used, for example, in financial-book-keeping record works, in the calculation of wages and the administration of materials.

The use of methods of automatic correction of the errors cropping up in processing (for example, the perforation errors) is also improper, they going to be corrected manually in strict conformity with reality.

The application of the automatic correction methods can be extremely efficient in big statistical works, such as for example in case of censuses and statistical inquiries.

In this case, several automatic correction methods can be mainly used in keeping with the probability of appearance of each code which is going to be attributed instead of the erroneous one.

The systematic correction method. By this method, a certain constant value is being each time attributed to the erroneous characteristic.

The systematic correction method is used when the correlation admits a single value for the corrected characteristic or when there is a very high probability for the indicator to acquire a single value.

For example, if the age for men is less than 18 and the code for legal status corresponds to the legal status of widow, the respective code will be corrected for unmarried, respectively code 4 (widow) is replaced by code 1 (unmarried).

Method of correction with alternative values. This method may be used when a code can be corrected with one of several values with a probability of approximately equal appearance.

The method may have two aspects:

— The correction of the erroneous characteristic is being made by attributing to it a value out of the whole series of the values of possible codes, namely the one immediately following that with which the correction of the previous case was made.

For example, the hours of the appearance of an incident in the electric network may be considered equiprobable.

If the hour of a certain incident will be erroneous, for example 26, it will be corrected by attributing the hour 16 in case that the hour of the previous incident had been corrected by 15.

— The correction of the erroneous characteristic is being made by attributing to it a value of a certain group of values of the series of the possible values (determined by the higher probability of appearance).

The correction is made on the same principle as in the case of the first aspects. For example:

Age	Sex	Number of children	Attributed for number of children
20	2	12	1 or 2

Method of aleatory correction. By this method, the correction is made by replacing the erroneous code by the code of the last correct case for the respective characteristic.

For example, in case that the code of the social category is erroneous being outside the admitted limits, it will be replaced by the code of social category from the previous information on condition that it should be corrected. In this way, the corrections will observe the distribution of the studied collectivity.

Generally, the method is used in case that the distribution of the cases is unknown.

The "hot-deck" correction method. By this method, the erroneous code is replaced by the value of the code of the last similar case. This method is thus based on the analysis of several characteristics which determine the resemblance. For example:

Nr. crt.	Social category	Age	Number of children	Corrected code for number of children
1	1	28	2	it is not the case
2	2	37	1	it is not the case
3	1	26	14	2

The number of children for the first two cases was considered correctly while for the third case, it was corrected by value 2.

For this, the automatic correction programme envisages, by this method, first the formation in memory of a chart of the following shape:

Groups of age / Social category	15—19	20—24	25—29	30—34	a.s.o
1	0	1	1	2	
2	0	1	2	3	
3	1	2	2	2	

This chart is initially filled in with probable numbers on the respective phenomenon. In the given example — with number of children.

With the unrolling of the programme, the chart is modified with the values of the respective case and on the appearance of an erroneous case (for example, number of children—14), the corrector code is determined by the value included on the chart for the respective case.

Method of correction by calculated values. By this method, the correction of an indicator, usually a quantitative-value one, is made by calculating the value of the respective indicator on the basis of a previously established calculation formula.

For example, suppose that the information contains the quantity and value for a certain product, sizes which are automatically controlled by the method of relation

$$l_i < \frac{V}{Q} < l_s$$

where

l_i represents the lower limit
l_s — the upper limit
V — value
Q — quantity

If the information is erroneous, that is it does not satisfy the above relation, on the basis of an average price — P_m or of a constant price previously introduced and stocked in the memory of the computer, the calculation $(Q \times P_m)$ is remade, the value obtained replacing the erroneous value.

In this case, the quantity was apriori considered as a corect size.

The results obtained through the application of automatic control and correction methods are both qualitative, all the errors existing in the basic material being found out and quantitative. Due to the execution speed, the automatic control is extended on all keys and possible correlations.

Results of the application of automatic control and correction

Nr. crt.	Denomination of the Work	Nr. of control conditions	Speed	Nr. of mistakes found out	of which: corrected automatically
1	Census of the Population	35	600 cases/min	1,6%	1,5
2	Family budgets	400	60 budgets/min	15%	—
3	Registering of labour force	25	100 forms/min	2%	—

These few examples eloquently prove the importance of this source of efficiency of the electronic computer in case of statistical works.

Apart from the obvious advantages obtained, automatic control is, in case of electronic processing, a compulsory condition from several considerations:

— in case of the utilization of an electronic computer, the charts obtained by processing are usually more detailed and minute than in case of the other processing methods (manual, mechanographic) which makes that the possible errors contained in the basic material be more visible on the charts with the results;

— the charts obtained by electronic processing, due to their multiple possibilities of calculation and imprinting, can be directly multiplied, which makes that the correction of the possible errors on the charts with the results should be almost excluded;

— the electronic processing programmes by which the charts with the results are being obtained most often include instructions which can operate only on correct information; otherwise, these programmes can no longer function and consequently, the chart cannot be obtained.

To carry out the automatic control of data with the aid of an electronic computer, the following are necessary:

— the correct establishing of all relations and arithmetical and logical dependences existing among various characteristics and indicators, including the determination of the limits within which a series of indicators followed by statistical observation may vary;

— the elaboration of control and correction algorithms and the writing of the programmes for the electronic computer by which all data which do not satisfy the control conditions laid down should be found out and refused.

Whereas the second activity is ensured by the analysts and programmers of the computation centre, the first can be efficiently ensured only by specialists in the respective domains who know the works that have been carried out, the basis of the control and correction method consisting of certain general knowledge on the theme solved with the aid of the computer. This general knowledge may be divided into two categories, namely: theoretical and empiric.

Theoretical knowledge appears as an exact knowledge because it is determined by a series of unequivocal preliminary definitions and relations. For example, "the age" should always be equal to the difference between the year of registering and the year of birth. Theoretical knowledge which allows the rejection of the erroneous data is, however, insufficient sometimes for establishing the wrong indicator, for example the age or year of birth.

Empiric knowledge supplies less exact information, being based on a series of hypothetical relations. Generally, empiric knowledge is gained by practical activity, being the result of experience in the respective domain.

Whereas theoretical knowledge gives a basis for absolute remarks as regards the quality of observations, empiric knowledge gives probable, approximate remarks. This means that theoretical knowledge will have greater weight in the application of automatic control methods, without neglecting empiric knowledge which may be expressed in a mathematical form.

All this knowledge which through the application of automatic control allows the classification of all registrations in two main zones — acceptance and elimination — should be supplied by specialists-beneficiaries of the electronic computer in a very exact, precise and unambiguous form. On its basis, the automatic control and correction programmes will be written at the computation centre.

As the introduction of modifications in already written programmes is leading to a considerable increase in the programming time, sometimes even doubling it, it is extremely important that the automatic control conditions once announced should suffer no modifications. Experience shows that when such modifications are introduced, the efficiency expected by submitting the work to the electronic computer is seriously diminished. This fact once more shows that the efficient use of the electronic computer, we have in view, first and foremost, the operativity and quality of processing, is conditioned by the way in which the demands for data processing are being formulated, including those on the arithematical and logical conditions which the primary information must satisfy.

As the writing and testing of the programmes for electronic processing and for automatic control inclusively require considerable time which, usually, cannot be reduced by increasing the number of programmers allocated for a theme, the information on the control conditions of data and correction methods which are to be applied is given much before (several months) the starting of processing proper.

As part of the electronic processing technological flow, the control and correction operation precedes the operations for drawing up the charts. This operation will be repeated until the last error in the card-index of initial data is corrected, so as to exclude any possibility of appearance of errors in the charts with final results, because in contradistinction to mecanographic or manual processing, in case of electronic processing, the correction of the possible errors in the final charts is an extremely difficult and expensive operation because of the big number of charts which are obtained on the basis of the same primary information.

In this way, it has become a rule that the technological flow of statistical data electronic processing should contain in the first part, after the operation of converting the information on cards perforated on magnetic tape, the operations of automatic control, automatic and/or manual correction of the possible errors.

Concluding the main aspects of the data automatic control and correction methods, one might as well say that the utilization of the possibilities offered by the electronic computer in this direction, is especially in case of statistical work, one of the main conditions for an efficient use of the electronic computer.

The minimization of the automatic control and correction methods, their giving up or incomplete use because of the apparent difficulties of formulation and definition of arithmetical and logical conditions which should be satisfied by the data to be declared as correct, is leading, in our opinion, to the loss of a considerable part of the efficiency which might be obtained by submitting the work to electronic processing.

LE CONTROLE ET LA CORRECTION AUTOMATIQUE — PREMISSE DE L'UTILISATION EFFICIENTE DU CALCULATEUR ELECTRONIQUE POUR LE TRAITEMENT DES INFORMATIONS

L'organisation d'un système d'informations, fondé sur le traitement des données par calculateur, poursuit, d'habitude, le but d'atteindre plusieurs objectifs, dont un des plus importants est d'assurer la qualité des données et surtout des données d'entrée.

Les données d'entrée peuvent présenter, dans la pratique, de nombreuses erreurs, des données incorrectes, le ramplissage incomplet des documents d'entrée, une codification erronée, des erreurs de perforation etc.

L'auteur présente les méthodes de contrôle automatique des données: la méthode de la correction systématique, la méthode de la correction à valeurs alternatives, la méthode de la correction aléatoire, en démontrant que les résultats obtenus par leur application sont d'ordre qualitatif et quantitatif, ce qui fait que l'efficacité du traitement augmente considérablement.

АВТОМАТИЧЕСКИЙ КОНТРОЛЬ И КОРРЕКТИРОВАНИЕ — ПРЕДПОСЫЛКА ЭФФЕКТИВНОГО ПРИМЕНЕНИЯ ЭЛЕКТРОННЫХ КАЛЬКУЛЯТОРОВ ДЛЯ ПЕРЕРАБОТКИ ДАННЫХ

Организация информационной системы, основанной на электронной переработке данных, направлена, как правило, на достижение нескольких целей, из которых одной из наиболее важных является обеспечение качества данных и, прежде всего, входящих.

Входящие данные могут содержать на практике многочисленные погрешности, неправильные данные, недостаточную комплектацию входящих документов, ошибочное кодирование, ошибки перфорирования и т.д.

Автор приводит методы автоматического контроля данных — метод систематического корректирования, метод корректирования с переменными величинами, метод случайной корректировки и указывает, что полученные результаты их применения имеют качественный и количественный порядок, повышая в значительной мере эффективность переработки.