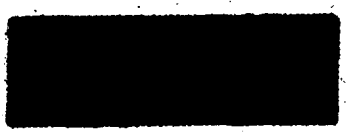
AGH
IFJ

Raport INT 135|I

A PROGRAM FOR ACTIVATION ANALYSIS DATA PROCESSING

JERZY JANCZYSZYN
LUDWIK LOSKA
STEFAN TACZANOWSKI



A PROGRAM FOR ACTIVATION ANALYSIS DATA PROCESSING

PROGRAM DO OPRACOWANIA WYNIKOW POMIAROW W ANALIZIE AKTYWACYJNEJ

ПРОГРАММА ДЛЯ ОБРАБОТКИ РЕЗУЛЬТАТОВ ИЗМЕРЕНИЙ В АКТИВАЦИОННОМ АНАЛИЗЕ

Jerzy Janczyszyn
Ludwik Loska
Stefan Taczanowski

Cracow 1978

Matryce wykonano według dostarczonych oryginałów

Rozprowadza – Распространяет – Available from:
OŚRODEK INFORMACJI O ENERGII JĄDROWEJ
00-901 Warszawa, PKiN, XI p.

Wydaje:
INSTYTUT FIZYKI I TECHNIKI JĄDROWEJ AGH – KRAKÓW
30-059 Kraków, al. Mickiewicza 30

Wydanie 1. Nakład 400+16+100 autorskich egz. Ark. druk. 1,75
Papier powiel. kl. III, A3, 71 g Oddano do produkcji 14. XI. 1978
Zamówienie nr 370/78 GP. II/1757/70 Powielanie ukończono w grudniu 1978
Data złożenia maszynopisu przez autora: sierpień 1978

Wykonano w Powielarni Akademii Górniczo-Hutniczej im. S. Staszica, Kraków, ul. Manifestu Lipcowego 16

Summary

An ALGOL program for activation analysis data handling is presented. The program may be used either for single channel spectrometry data or for multichannel spectrometry. The calculation of instrumental error and of analysis standard deviation is carried out. The outliers are tested, and the regression line diagram with the related observations are plotted by the program. (author)

Streszczenie

Opisano program, w języku ALGOL, służący do obróbki danych pomiarowych uzyskiwanych w analizie aktywacyjnej. Program ten może być stosowany zarówno w spektrometrii jednokanałowej jak i wielokanałowej. W oparciu o krzywą cechowania obliczany jest błąd aparaturowy. Przeprowadzane jest testowanie odstających wyników pomiarów, a także wykreślany jest wykres prostej regresji.

Резюме

Описано программу на языке Алгол предназначенную для обработки результатов измерений полученных в активационном анализе. Программа эта может быть использована, как в одноканальной так и в многоканальной спектрометрии. На основании кривой калибрации вычисляется погрешность аппаратуры. Проводится проверка результатов измерений у которых выступает наибольший разброс, а также вычерчивается график линейной регрессии.

Introduction

Activation analysis data processing requires some time-consuming calculations. Simultaneously, their routine character points out the suitability of the use of electronic calculating techniques.

Thus, a program which handles results of several succeeding series of measurements of analysed samples has been worked out. The essence of the statistical calculations used in the program and the way of its functionning compose the content of this paper.

The program performs analysis of weighted linear regression $Y = bX$, according to the well-known rules, therefore, we confine ourselves only to the explanations of several discussible points. The choice of this particular regression is substantiated below.

Theoretical principles

Definitions of symbols:

- Y - signal, proportional to the concentration of the determinand
- X - amount of the element to be determined
- b - slope of regression line.

The signal is defined as:

$$Y_i = \frac{N_i - B}{m_i \cdot n_{fi}} \cdot k_i \qquad (1)$$

where:

- i - index indicating the i-th observation
- N - total number of counts
- B - total background
- m - sample or standard mass
- n_f - number of counts of the neutron flux monitor
- k - correction factor for decay, selfabsorption and other

sources of systematic errors.

The slope is obtained from the well-known formula [1] :

$$b = \frac{\sum_{i=1}^{n} W_i \cdot X_i \cdot Y_i}{\sum_{i=1}^{n} W_i \cdot X_i^2} \qquad (2)$$

with:

n - number of measured standards

W_i - weight of the signal Y_i expressed as the reciprocal of signal variance which consists of Poisson and instrumental components:

$$W_i = \frac{1}{S_{ti}^2 + (\delta_{ap} \cdot b \cdot X_i)^2} \qquad (3)$$

where:

S_{ti} - estimate of the standard deviation resulting from the Poisson distribution of N_i, B, n_{fi} ,

δ_{ap} - relative instrumental error,

$b \cdot X_i$ - estimate of the value of the i-th signal .

The formula for S_{ti} from eq.(1) is:

$$S_{ti} = Y_i \cdot \sqrt{\frac{N_i + V_B}{(N_i - B)^2} + \frac{1}{n_{fi}}} \qquad (4)$$

where:

V_B - variance of the background estimate.

The dispersion of observations around the regression line results from two independent factors. The first of these is the random nature of radioactive desintegration, and the second one is the imperfection of the instruments and standards used. Strictly, it is expressed by the relation:

$$\frac{\sum_{i=1}^{n} W_i \cdot (Y_i - bX_i)^2}{(n-1) \cdot \sum_{i=1}^{n} W_i} = \frac{\sum_{i=1}^{n} W_i \cdot \left[S_{ti}^2 + (\delta_{ap} \cdot bX_i)^2\right]}{n \cdot \sum_{i=1}^{n} W_i} \qquad (5)$$

Where the left side is the estimate of the standard deviation of regression, and the right side is the total mean error containing both components of the dispersion.

Hence:

$$\sigma_{ap} = \frac{1}{b} \cdot \sqrt{\frac{1}{\sum_{i=1}^{n} W_i \cdot X_i^2} \cdot \left[\frac{n \cdot \sum_{i=1}^{n} W_i \cdot (Y_i - bX_i)^2}{n-1} - \sum_{i=1}^{n} W_i \cdot s_{ai}^2 \right]} \qquad (6)$$

Here, in agreement with intuition, some assumptions have been made. It has been assumed that the "apparatus" component of the total standard deviation is proportional to the estimate of expected value of the signal bX_i /attempts to put in directly the measured value of the signal Y_i, brought about erroneous results in practice/.

Obviously, since all these quantities are given in indirect form, all calculations have to be carried out with a successive iteration method. At the first iteration only the Poisson component is assigned to the weights, the slope is calculated, and subsequently the instrumental error. Then, the calculations are repeated with weights now including the apparatus component. Usually, a twofold iteration is quite sufficient.

In the case of more than one series of measurements the whole procedure is performed for each of the series. On the basis of the several results obtained for each sample the mean concentrations and the standard deviations are computed.

It is assumed that a functional relationship exists between the relative standard deviation of the results of analysis and the content of determinand. The theoreticaly derived form of the function is [3]:

$$S = \sqrt{A_1 + \frac{A_2}{x} + \frac{A_3}{x^2}} \qquad (7)$$

where:

A_1, A_2, A_3 - constant values,

x - content of determinand [g].

For calculations a more convenient form of the function was adopted:

$$S = A \cdot x^C \qquad (8)$$

This function is fitted to the experimentally obtained values of S and x, and the constants A and C are obtained. Then the function is employed for testing the results of analysis. The value of standard deviation, respective to particular sample, read out from the function is regarded as a well known value as determined on the basis of tenths of observations. This value is used for testing outlying results using the Taczanowski test [2].

For each sample the interval is computed using standard deviation and a selected confidence level: $x \pm k \cdot s$. Table I shows the values of the coefficient k [2] for the most commonly used parameters. All results that do not belong to the interval are rejected and new values of constants A and C are computed. Then the testing operation is repeated and the loop works until none of the results is rejected. The final results for each sample consist of:

- the mean content: x
- the concentration: $X = x / m$
- the number of accepted results of analyses: n
- the standard deviation of the concentration:

$$S_X = \frac{S}{m \cdot \sqrt{n}}$$

- the mass of the sample: m

Table I

Values of the coefficient k for testing outlying results

n \ 1 - α	0.70	0.80	0.90	0.95	0.98
3	1.94	2.21	2.59	2.92	3.32
4	1.99	2.22	2.57	2.88	3.24
5	2.03	2.26	2.58	2.87	3.22
6	2.08	2.29	2.60	2.88	3.21
7	2.12	2.32	2.63	2.90	3.22
8	2.16	2.36	2.65	2.92	3.22
α - significance level					

Structure of the program

For all mathematical operations described in the preceeding part an ALGOL 1900 program has been written flow diagram of which is presented on Fig.1.

Input data

The form of input data is shown in the Fig.2. An introductory text is followed by auxiliary data. Their exemplary values are given below:

```
2   0.95   1.96   0.8   60
- - - - -
10.83  4.75  3.76  3.39  3.20  3.10  3.03  2.99  2.96  2.94
 2.92  2.91  2.91  2.90  2.90  2.90  2.90  2.90  2.91  2.91
- - - - -
12.71  4.30  3.18  2.78  2.57  2.45  2.37  2.31  2.26  2.23
 2.20  2.18  2.16  2.14  2.13  2.12  2.11  2.10  2.09  2.09
- - - - -
 0.89  0.59  0.49  0.43  0.40  0.37  0.35  0.34  0.32  0.31
```

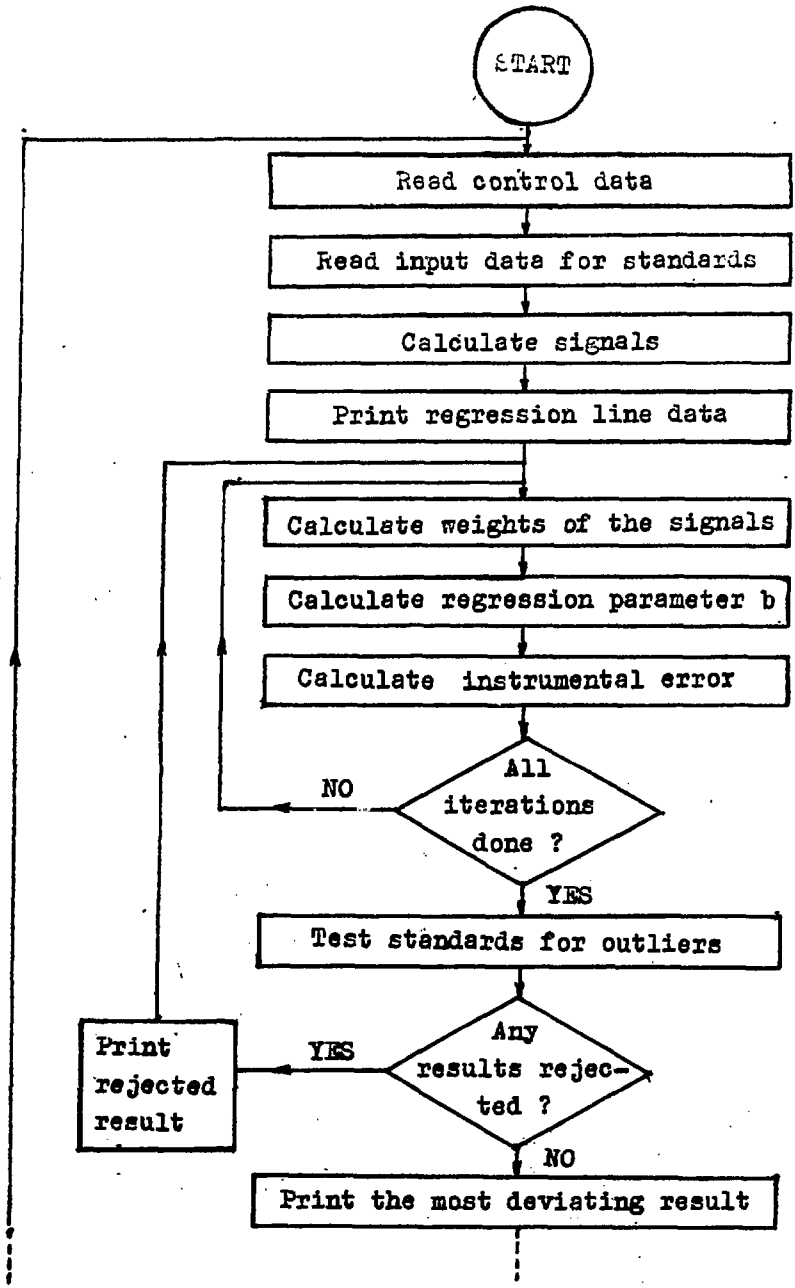


Fig.1 Flow diagram of the program

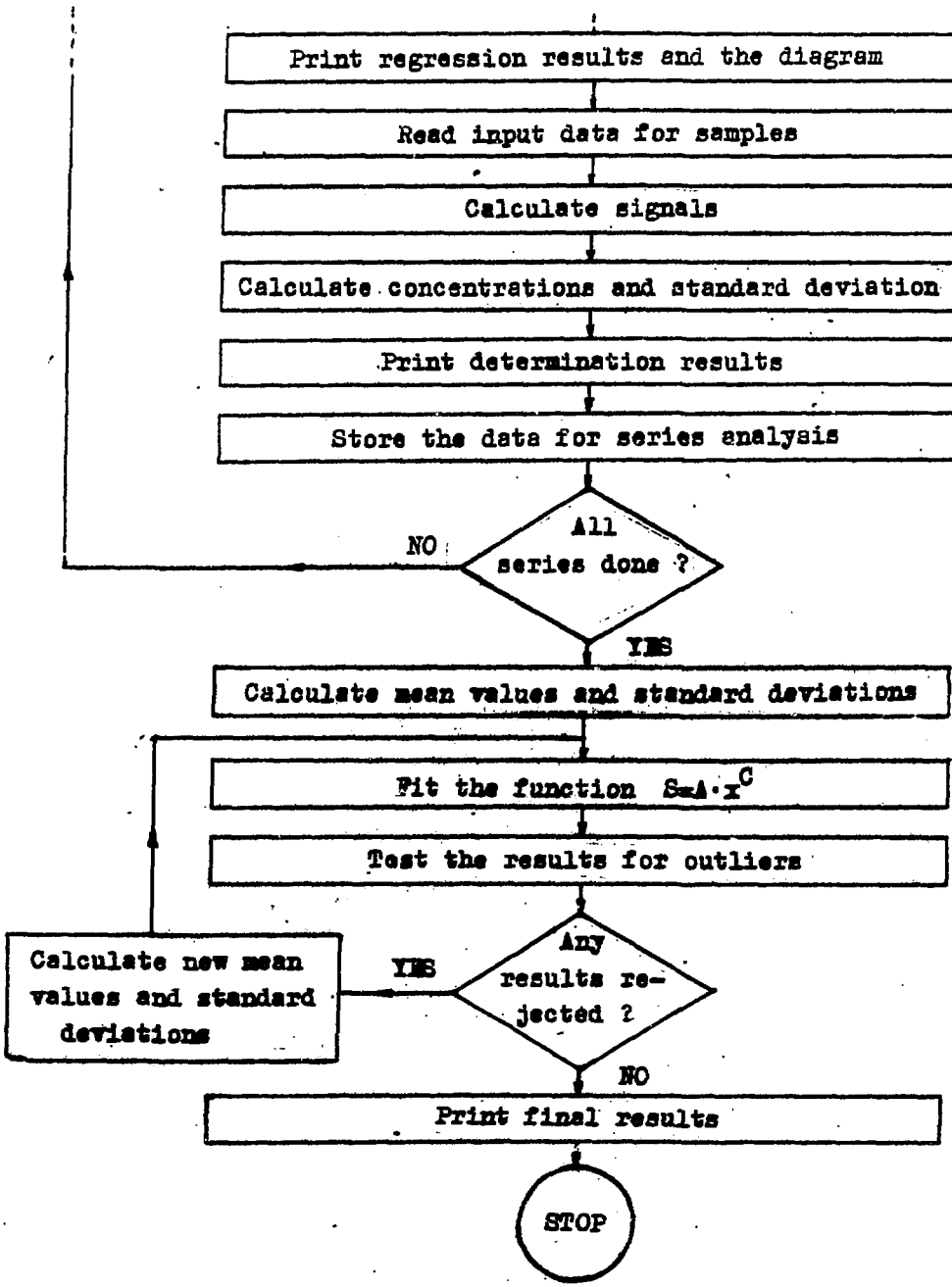


Fig.1 Continued

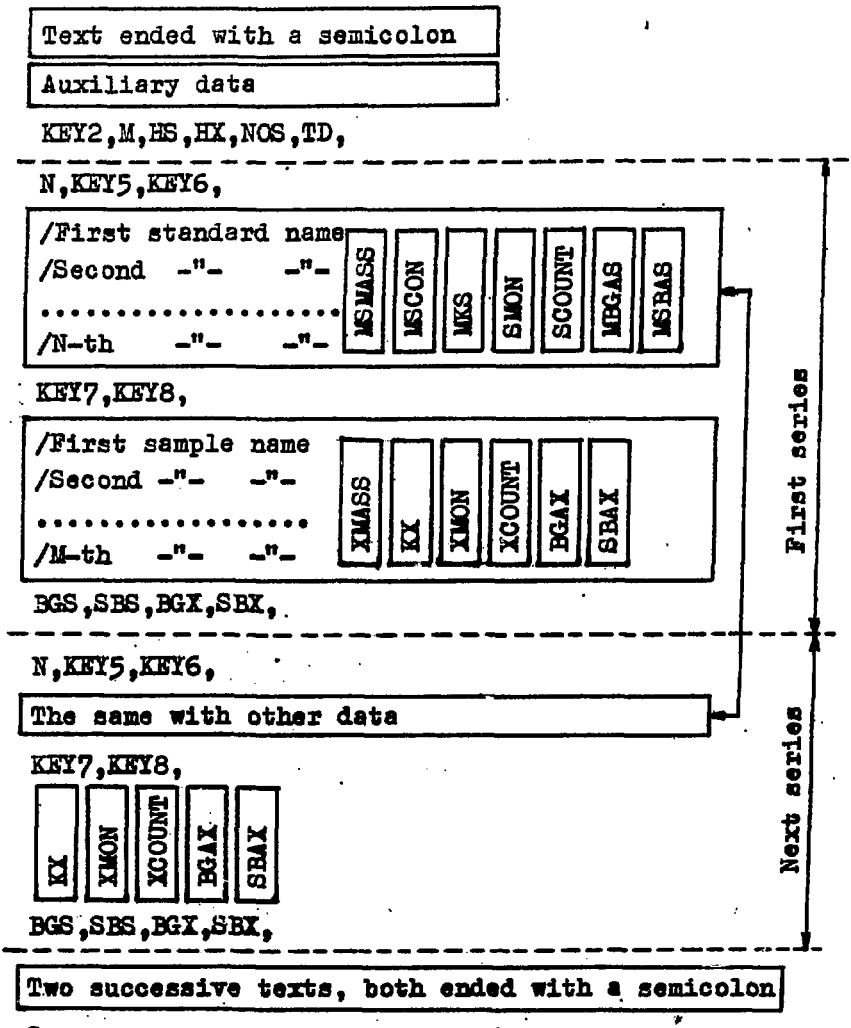


Fig.2 Form of input data

The numbers are respectively:

- number of iterations for the calculation of the slope of regression line,
- confidence level,
- confidence interval corresponding to the above confidence level,
- confidence level of the rejection test,
- dimensions of the linear regression diagram /number of spaces/,
- array of coefficients for rejection test [2],
- array of coefficients for Student test,
- array of coefficients for the estimation of standard deviation from the range.

The remaining data should be introduced according to Fig.2. The meaning of symbols is explained in the list of variables at the end of the paper. A set of data for individual standard or sample begins with a "slash".

The number of standards and their data can vary from one series to another. On the contrary, the number of samples is fixed for all series. In a case when a sample is not analysed in a particular series a big negative number /e.g. $-1.0 \cdot 10^{20}$ / should be put into the array XCOUNT for this sample. There are no particular requirements for the other data of this sample, usually number 1.0 is assigned to each of them.

The form of the arrays SCOUNT, MBGAS, MSBAS, XCOUNT, BGAX and SBAX depends on the counting method and the preliminary data treatment.

Three possibilities are provided, one of which can be selected by the choice of the value of variable TD:

a/ single channel counting - TD < 0.5 - number of counts, background and background variance should be assigned to SCOUNT /XCOUNT/, MBGAS /BGAX/ and MSBAS /SBAX/ respectively,

b/ multichannel counting, Covell method - TD > 0.5 - total counts, number of channels and the sum of side channels should be given,

c/ multichannel counting, other then Covell method - TD > 0.5 - net counts /peak area/, zero and standard deviation of net counts should be given.

At the end of all data two successive texts follow, each of them ended with a semicolon. If the whole program is to be executed for another set of data, each set should be ended with a number greater then 0.5 . The last set should and with a number smaller then 0.5

Results

The results begin with the text read in at the beginning of input data, then follows the information about standards: name, mass, neutron monitor counts, corrected net counts, concentration of the determinand and normalized signal, in succession; each standard in one line. The results of applying the rejection test are then printed out. Next come the data concerning the regression analysis. These are: slope, its absolute and relative standard deviations, correlation coefficient and instrumental error.

Depending on the value of KEY2, the calibration line and the points representing the measurements are plotted. When two or more points fall on the same place in the diagram they are printed as digits indicating the number of measurements of equal values.

Next follows the results of analyses for a given series. Sample name, its mass, concentration of the determinand, relative standard deviation and the confidence interval of the result of analysis are printed in a line.

The set of results described up to this point is repeated for all the series and then the results of analyses, for all samples, obtained in successive series are gathered and printed once again. The parameters of the function approximating the dependance of the standard deviation of a single analysis result on the determinand content are printed below.

The final results of analysis containing: sample name,

mean value of the concentration of the determinand, its relative standard deviation, number of accepted measurements, sample mass, content of the determinand and standard deviation of a single result of analysis /calculated directly from the results of this sample and not from the function/ are printed at the end. They are tripie copied each time ended with the text introduced at the end of input data.

Discussion

The program was designed basing on several assumptions, covering most of the cases in activation analysis. The most important is assumption of the linear relationship between the signal and the concentration of the determinand. It can be true when the spectrometer dead time and the self-absorption and self-shielding effects are independent of this concentration. Also the matrix activity should be negligible, or at least constant.

If these assumptions are not satisfied, the estimation of the influence of all these effects on the results of measurements should be performed separately and introduced in the form of correction factors or as a variable background for each standard and sample. Otherwise all calculations are only fairly correct because of the approximation of the real relationship by the linear regression.

The choice of the regression equation /Y = bX/ should be further substantiated. Assuming the regression /Y = bX+a/, the measurement corresponding to the zero content of the determinand becomes most significant because of its great statistical weight. This is especially important for the determination of the intercept. Unfortunately the normalization of its number of counts is impossible since the natural background plus the container activity are present in this measurement. Thus, it is necessary to substract that part of the number of counts of all the measurements, which is not a function of the neutron monitor counts or mass. It is useful

to substract also the matrix activity, if possible. As a result, the obtained intercept has the value approximately equal to zero and the regression /Y = bX/ is acceptable. That choice is equivalent to the assignment of an infinitely great weight to the point /0,0/, obtained after the substraction. Since its weight is in most cases a very great one, then the inexactitude introduced by this approximation is entirely negligible.

Considering the rejection of outliers, we confine ourselves to only a few explanations. The test rejects the outlying observation if its weighted distance from the regression line exceeds the given tolerance limit. These and the related confidence levels are chosen arbitrarily by the experimenter and should be given in the form of constants, which multiplied by the standard deviation, establish the tolerance limits. Since the number of standards allocating the regression line is in most cases rather small, the assumption that only one observation is to be suspected may be accepted. Nevertheless the program operates and the test is still useful and fairly correct, even if this condition is not satisfied. The details of the test can be found in [2].

Summarizing it may be observed that the program presented here was successfully applied in many analyses, as for instance in the determination of oxygen and silicon in metals, oxygen, nitrogen and silicon in ceramic materials, and so on, supplying the results in a suitable, normalized form and saving plenty of the analyst time.

List of variables

M	-	number of samples /in all series the same/
HS	-	numbers of characters used for standards names /not greater than 19/
HX	-	number of characters used for samples names /not greater than 19/
NOS	-	number of series
TD	-	control number
N	-	number of standards
MSMASS	-	array of standard's masses
MSCON	-	array of the determinand contents /per cent/
MKS	-	array of standards self-absorption and self-shielding correction factors
SMON	-	array of neutron flux monitor counts/standards/
SCOUNT	-	array of gamma ray detector counts for standards
MBGAS	-	array of background counts for standards
MSBAS	-	array of background counts variance for standards
XMASS	-	array of sample's mass
KX	-	array of samples self-absorption and self-shielding correction factors
XMON	-	array of neutron flux monitor counts for samples
XCOUNT	-	array of gamma ray detector counts for samples
BGAX	-	array of background counts for samples
SBAX	-	array of background counts variance for samples
BGS	-	common background counts for standards
SBS	-	common background counts variance for standards
BGX	-	common background counts for samples
SBX	-	common background counts variance for samples

If KEY2 > 0.5 then the program plots the regression diagram,

or else it does not,

If KEY5 > 0.5 then the numbers for array MKS must be given in data, or else not,

If KEY6 > 0.5 then MBGAS and MSBAS must be given, BGS and SBS must not, or else BGS and SBS must be given, MBGAS and MSBAS must not,

If KEY7 > 0.5 then KX must be given, or else must not,

IF KEY8 > 0.5 then BGAX and SBAX must be given, BGX and SBX must not or else BGX and SBX must be given, BGAX and SBAX must not,

If S > 0.5 then execute the program again for another set of data, or else end the execution.

REFERENCE

1. W.I.Goldanskii, A.W.Kutsenko, M.I.Podgoretskii: Statistika otschetov pri registratsii yadernykh chastits. Fiz.-Mat.-Giz., Moscow /1959/.
2. S.Taczanowski, Report ITJ Nr 38/I, 1973.
3. J.Janczyszyn, S.Kwieciński, L.Loska, W.Pohorecki, S.Taczanowski, J.Radioanal.Chem., 31 /1976/325.

POWIELANIA
AGH
KRAKÓW