A. DIMITRIU
A. ANDĂR
P. ANDĂR

A Symmetry Indicator of Statistical Distributions*

A new symmetry indicator of statistical distributions is proposed which, just as V. V. Mogarovski's uniformity degree ρ, proves to be more accurate than the asymmetry degree a, the interquartile asymmetry coefficient α and the variation degree v. Pearson's interquartile asymmetry coefficient, especially, presents erroneous values and its substitution is recommended by the proposed indicator.

Investigations conducted in recent years under the supervision of Professor V. Ianovici, on the nature of the statistical distribution of the chemical components in different rocks and ores — of major importance in the grounding of numerous working hypotheses in geological activity — have spotlighted that the asymmetry indicators used currently [1] in such research works (the asymmetry degree a and the interquartile asymmetry coefficient α) often show important errors.

The first observations in this line have been quite surprising, bearing in mind that these indicators, elaborated by K. Pearson, had been introduced into the practice of statistical researches for several decades now. Still, the number of instances when errors were discovered, in discordance with the nature of distributions, have gradually increased, so that we have considered it useful to establish the amplitude of the errors occurred, to investigate the source of errors of indicators and, as far as possible, to elaborate an indicator boasting of higher accuracy.

This survey presents the results achieved so far. They refer to both the critical examining of different statistical indicators already known [1], [2], (the arithmetic mean $\bar{x}$, the quadratic mean deviation $\sigma_{\bar{x}}$, the modulus Mo, the median Me, the asymmetry degree a, the interquartile asymmetry

* The computations were worked with the assistance of O. Vicenz and T. Vasile.

coefficient α, and the variation degree v, the uniformity degree ρ), and to a new symmetry indicator which we have marked β.

1. THEORETICAL FOUNDATION OF INDICATORS α, a AND β (PROPOSED)

An analysis of the way in which indicators α and a are defined shows the following:

Both indicators characterize the distribution symmetry introducing in the computation only some elements that do not inform satisfactorily of the essential properties of the distribution; the interquartile asymmetry coefficient α uses in the computation three quantities (q_{I}, q_{II} and q_{III}), and the variation degree a, in fact, two quantities, which stand in the central part of the distribution ($\bar{x}$ and Mo).

It follows that these indicators try to characterize the symmetry of the distribution by taking into account either limited portions of the distribution extension (the case of the interquartile asymmetry coefficient α), or some grouping indicators, neglecting the way in which the allotment of cases in the domain of distribution occurs (the case of the asymmetry degree a).

Taking into account these deficiencies in the drawing up of indicators α and a, we have planned to have in view in the elaboration of a new symmetry indicator the use in computation of both a larger distribution area and the allotment of cases along the distribution.

According to the properties of normal distributions, the positions lying above or below the arithmetic mean value are characterized by the same extension: $\lambda_A = \lambda_B$, where $\lambda_A + \lambda_B = \lambda(\lambda =$ the variance range) and the same number of instances: $n_A = n_B$, where $n_A + n_B = N(N =$ volume of selection).

It has been considered that these properties assume, by a satisfactory approximation, the qualities of a criterion over the symmetry of distributions. Consequently, the quantity of the indicator has been defined as a function of λ_A, λ_B, n_A and n_B, having the expression:

$$\beta = \frac{\dfrac{n_A}{\lambda_A}}{\dfrac{n_B}{\lambda_B}}.$$

It has been noted that $\beta = 1$ in case of distributions with perfect symmetry. In order to set the strict variation limits, it has been agreed that the term $\dfrac{n_i}{\lambda_i}$, of a smaller value, should be always included in the expression numerator. When this term corresponds to the right side branch of the distribution the value of the coefficient is provided with the positive sign, in keeping with the convention adopted in mathematical statistics when referring to left side asymmetry.

2. DISCRIMINATION POWER OF THE NATURE OF DISTRIBUTIONS OF INDICATORS α, a, v, ρ AND β.

In order to spotlight the degree to which diverse indicators (α, a, v and ρ), just as the proposed indicator β, have a regular occurrance in the mode of variation, in keeping with the nature of distributions, the values of these indicators have been computed and their diagrams drawn for 108 distributions; the law of allotment of these distributions has been checked up, generally, through the chi-square method or the method of moments (see annexes 1—5). As far as possible, the distributions under survey comprise the whole range of types of distributions occurring frequently: from the normal distributions to the *j*-shaped distributions.

A first remark, of a general nature, is the high variance of the values of indicators, suggesting a certain confusion in their mode of variation in relation with the character of distributions. Obviously, to discuss the individual values of each indicator in part would be difficult and would not succeed in spotlighting the essential aspects of the mode of variation. To this end, the regression lines have been drawn by the method of conditioned means.

A careful examination of the diagrams (see annexes 1—5) reveal that:

— the angular value of the regression lines of indicators α, a and v is small (in contrast with that of indicators ρ and β), expressing by this lower sensibility in the way in which these indicators reflect the alteration of the type of distribution;

— the known indicators (α, a, v and ρ) seldom present values characteristic of the type of distribution and, this happens only in connection with the normal and *j*-shaped distributions, thus not permitting the differentiation of the lognormal distributions from the other distributions. In the event of the proposed indicator β, the positions taken into consideration with a high degree of probability as characteristic of some distributions, show that this can help the discrimination of several types of distributions (Table I).

In conclusion, the comparison of different indicators in relation to the nature of distributions demonstrates that the proposed indicator β presents rather a high sensibility and accuracy.

3. STABILITY OF DIFFERENT INDICATORS

3.1. *Influence of the interval*

It is known that to choose the value of the interval is still an open question. F. Chayes [3] just as D. M. Shaw [4] advocate that the choosing of value d of the interval should proceed in keeping with the standard deviation ($d \leqslant 0.5\ \sigma_{\bar{x}}$), I. P. Sharapov [5] recommends to do this choosing in compliance with the variance range λ and the selection volume N, namely

$$d = \frac{\lambda}{1 + 3.322 \log N}$$

Table I

Limits of variation of diverse statistical indicators in relation to the types of distribution

Indicator \ Type of distribution	Normal	Intermediate	Lognormal	*j*-shaped
a	0 — 0.60			
	0.20		0.60	
v	0 — 0.60			
	< 0.20	0.60	0.80	> 0.80
α	0 — 1.00			
ρ	0 — 0.60			
	> 0.90			< 0.60
β	0.50 — 1.00			
	< 0.50			
	> 0.70		0.15 — 0.45	< 0.15

while G. Mihoc and M. Urseanu [1] consider the value of the interval as an arbitrary question depending on the aim in view.

The question arises under these conditions: how could the extension of the interval influence the value of different indicators?

A brief analysis shows that indicators φ, ρ, just as β, are not influenced by the interval, owing precisely to the way in which they had been elaborated. With regard to the other indicators (the interquartile asymmetry coefficient α and the asymmetry degree a), it is possible for them to be influenced and, consequently, their investigation in this light appears necessary. At the same time, we think it is useful in this respect to examine the modulus — bearing in mind that this indicator affects the influence of the interval upon the asymmetry degree a — and, hence, to explain the way in which this effect is propagated and, especially, to estimate the stability, under this aspect, of the modulus — one of the indicators frequently used in geological investigations.

To this end, indicators α, a and Mo have been computed for a set of distributions in case of two different values of the interval, determined with the help of the formula advised by F. Chayes [3] and D. M. Shaw [4] (interval l_I) and, respectively by applying the formula according to I. P. Sharapov [5] (interval l_{II}). A large number of distributions have been used for the research, this enabling to make up a broad range of cases on the amplification degree K of the interval $\left(K = \frac{l_{II}}{l_I}\right)$, namely, from $K = 1.0$

to $K = 2.0$. The results have been represented by means of diagrams (see annexes 6, 7 and 8).

The representation of the values proceeded in such a way as to best spotlight the ampleness with which the alteration of the indicators value from one distribution to another occurs within the same amplification degree of the interval. To this end (the case of indicators α and a) the values computed for l_{I}, respectively l_{II}, have been connected separately within the diagram, thus the variation area proportional to the influence exerted by the alteration of the interval being defined. In the case of modulus, the influence due to degree K has been given by the relative deviation

$$\frac{Mo_{l_{\mathrm{II}}} - Mo_{l_{\mathrm{I}}}}{Mo_{l_{\mathrm{I}}}} \times 100.$$

Such representations, made for diverse values K, have been connected in the same diagram for each indicator in part, thus being possible to specify the measure in which indicators α, a and Mo vary in step with the increase of K, which is precisely what we had in mind.

It has been noted in the case of each indicator that distributions may appear for which the indicators are influenced to the same extent irrespectively of the amplification degree K. On the other hand, to the same degree K, the alteration of indicators may be utterly different from one distribution to another. Moreover, sometimes the alterations are important, in some instances even with the alteration of the sign of the indicator's value, providing thus different information on the nature of distribution.

If we take account of the fact that at present it is difficult to state which of the two expressions of the interval quantities (after D. Chayes and I. P. Sharapov, respectively) is the correct one, it follows that indicators α and a, just as the modulus, may be given arbitrary values. In conclusion, it appears as an obvious thing, under this aspect too, that less reliability must be granted to indicators α and a.

3.2. *Influence of the selection volume*

Another investigated aspect is the stability of the indicators, according to the selection volume.

To this end, five distributions of chemical elements of diverse rocks (A — normal distribution of ferrous oxide in the manganese ore at Dadu; B — normal distribution of manganous oxide in the manganese ore at Moneasa; C — lognormal distribution of silicon dioxide in the manganese ore at Dadu; D — lognormal distribution of iron trioxide in the dolomite of Băiţa-Troiţa-Crăciuneşti) have been surveyed. For each distribution in part several subgroups have been formed, through the repeated subtraction of the values at chance. So, a wide range of the reduction degree of the selection volume θ (from 0.15 to 1.00) has been created. The com puted values (either as absolute values in the case of indicators a, α, φ and β, or as relative percentage values,

such as in the case of indicators $\bar{x}$, σ_x, *Mo*, *Me* and ρ) have been included in the variational diagrams (see annexes 9 and 10).

For most of indicators, excepting v, $\bar{x}$, *Me* and $\sigma_{\bar{x}}$ it has been noticed a low stability, according to the selection volume. It follows that the non-observance of the representative conditions of the sampling entails important errors in the estimation of indicators; this has to be held as an extra argument for the necessity of improving, based on statistical data, the sampling work in geology.

An interesting observation appears conspicuous with regard to the stability proved by indicators $\bar{x}$, *Me*, $\sigma_{\bar{x}}$ and ρ. It is known that these indicators provide rather accurate information on the chemical composition of rocks, the quality of a useful mineral substance. It follows that a sampling limited to a small number of samples would lead to a sufficiently accurate estimation of the representative chemical composition of the investigated geological formation. Obviously, this observation is valid for the investigated rocks (limestones, dolomites, manganese and iron ores).

As for indicator β, the influence of the reduction degree of the selection volume θ, correlated with the extension of domain $\lambda_{A(B)}$, has been investigated. To this end, the value of indicator β has been computed for diverse domains $\lambda_{A(B)}$, whose extension on one side or another of the arithmetic mean has been taken in compliance with the quantities $\sigma(a)$, $2\sigma(b)$ and $3\sigma(c)$ as well as with the quartiles of 25 per cent and 75 per cent (d).

	λ_A	λ_B
(a)	(x_F per cent = 15.87, $\bar{x}$)	($\bar{x}$, x_F per cent = 84.13)
(b)	(x_F per cent = 2.28, $\bar{x}$)	($\bar{x}$, x_F per cent = 97.72)
(c)	(x_F per cent = 0.14, $\bar{x}$)	($\bar{x}$, x_F per cent = 99.86)
(d)	(x_F per cent = 25.00, $\bar{x}$)	($\bar{x}$, x_F per cent = 75.00)

Then, for each domain $\lambda_{A(B)}$ in part, indicator β has been computed at different reduction degrees of the selection volume. The investigations have been carried out on a number of distributions checked up as normal or lognormal (see annex 11).

In one single instance (the distribution of manganous oxide in the dolomites of Hunedoara), the variation of indicator β corresponding to diverse areas of variance for three selections of identical volume has also been investigated, important differences being observed only in case of high reductions ($\theta = 0.26$).

A survey of the results showed that the best relative stability, especially for small selection volumes, occurs only in the case of indicator β, the domain of variation having the limits according to the magnitude 2σ, that is (x_F per cent = $= 2.28\,\bar{x}$); ($\bar{x}$, x_F per cent = 97.72). It is generally advisable that this domain be adopted when determining indicator β for any distribution under survey.

4. CONCLUSIONS

The investigations conducted show that the proposed indicator and the uniformity degree [3] prove a much higher accuracy than the asymmetry degree a, the variation degree v and, especially, the interquartile variation coefficient α. It is advisable that the estimation of the distributions asymmetry be made on the basis of indicator β or of the uniformity degree ρ, which, as a matter of fact, may also serve for the approximation of the statistical nature of distributions.

REFERENCES

1. MIHOC, G. and URSEANU, V., *Matematici aplicate în statistică*, Editura Academiei, Bucureşti, 1962.
2. МОГАРОВСКИЙ, В. В., *О количественной характеристике степени равномерности распределения химических элементов в природных объектах*, « Геохимия » No. 2, 1962, pp. 181—184.
3. CHAYES, F., *The Lognormal Distributions of the Elements: A Discussion*, "Geochimica et Cosmochimica Acta", vol. 6, No. 2, 1954, pp. 119—120.
4. SHAW, D. M., *Interprétation géochimique des éléments en traces dans les roches cristallines*, Ed. Masson et Cie, Paris, 1964.
5. ШАРАПОВ, И, П., *Применение математической статистики в геологии*, Издательство Недра, Москва, 1965.

UN INDICE DE SYMETRIE DES DISTRIBUTIONS STATISTIQUES

L'examen de l'évaluation de différents indices (degré d'asymétrie a, coefficient interquartile d'asymétrie α, degré de variation v et degré d'uniformité ρ) en rapport avec la nature de la distribution a mis en évidence — sur 108 cas de distribution des composants chimiques de différentes roches — que ces indices sont généralement d'une précision non-satisfaisante, surtout pour ce qui est du coefficient α de K. Pearson.

Afin de permettre d'apprécier de façon plus précise la symétrie des distributions, les auteurs ont élaboré sur la base des propriétés des distributions normales, un nouvel indice, β, dont l'expression est:

$$\beta = \frac{\frac{n_A}{\lambda_A}}{\frac{n_B}{\lambda_B}} \qquad \left(\frac{n_A}{\lambda_A} < \frac{n_B}{\lambda_B}\right)$$

$$(0 < \beta < 1)$$

où: λ_A, λ_B = l'amplitude de variance d'un côté et de l'autre de la moyenne arithmétique ($\lambda_A = \bar{x} - x_{F\%} = 2{,}28$ şi $\lambda_B = x_{F\%} = 97{,}72 - x$) n_A, n_B = le nombre de valeurs correspondant aux domaines partiels de variation λ_A, respectivement λ_B.

Conventionnellement le signe de l'indice est positif pour l'asymétrie de gauche et négatif pour l'asymétrie de droite.

Les auteurs démontrent que, contrairement aux autres indices, et tout comme l'indice β, l'indice ρ ne dépend pas de la valeur de l'intervalle, dont la détermination est encore controversée.

Ils présentent également certains résultats concernant l'influence du volume de sélection sur les indices les plus importants ($\bar{x}$, *Mo*, *Me*, $\sigma_{\bar{x}}$, v, a, α, ρ, β) et les conséquences qui en dérivent pour les travaux d'épreuve en géologie.

ПОКАЗАТЕЛЬ СИММЕТРИИ СТАТИСТИЧЕСКИХ РАСПРЕДЕЛЕНИЙ

Исследование развития разных показателей (степень ассимметрии a; интерквартильный коэффициент ассимметрии α; степень вариации v и степень равномерности ρ) по отношению к природе распределения показало — в случае 108 распределений химических компонентов природы — что, вообще, у них неудовлетворительная точность, особенно у коэффинциента α Пирсона.

С целью более точной оценки симметрии распределений, авторы разработали на основе свойств нормальных распределений новый показатель, β, имеюший следующий вид:

$$\beta - \frac{\frac{n_A}{\lambda_A}}{\frac{n_B}{\lambda_B}} \quad \left(\frac{n_A}{\lambda_A} < \frac{n_B}{\lambda_B}\right)$$

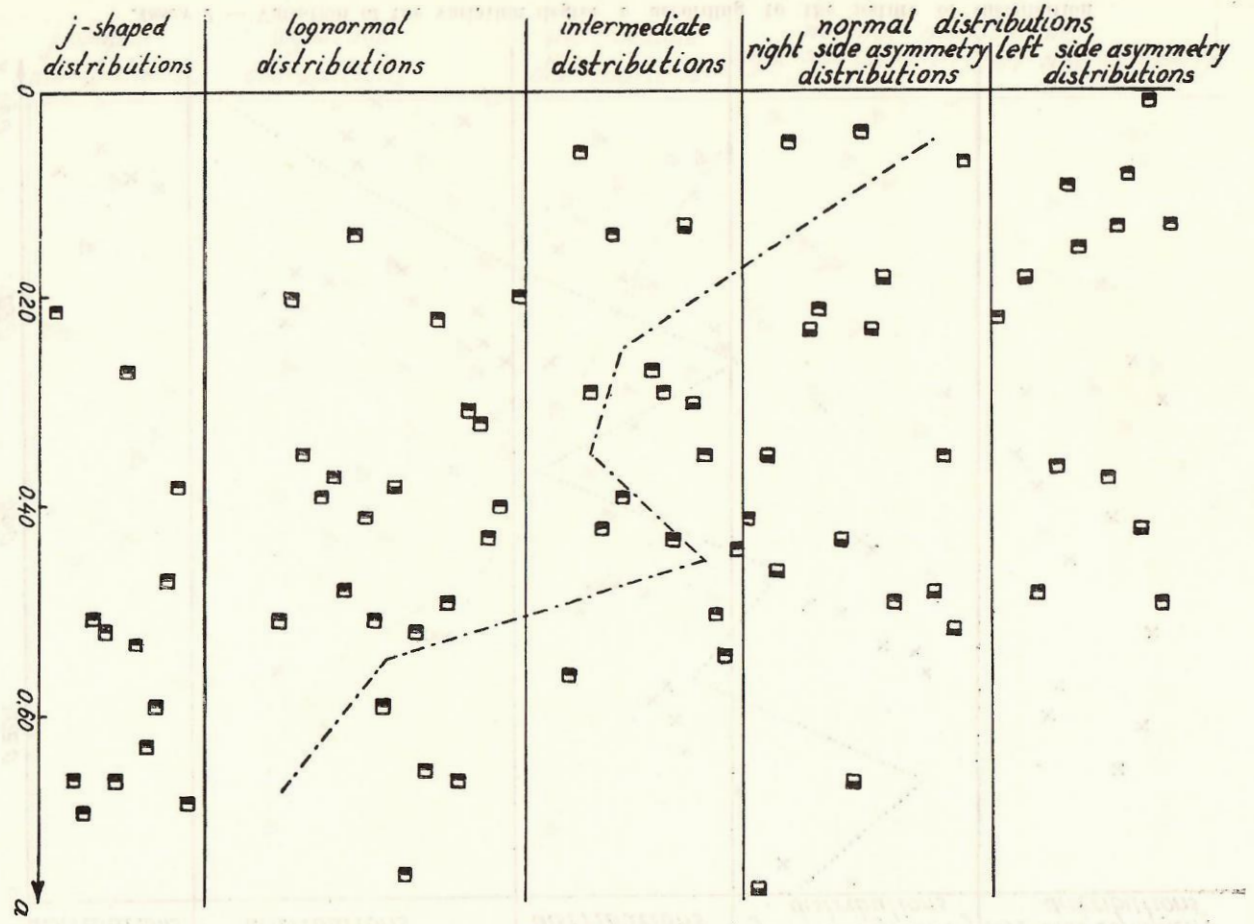
$$(0 < \beta < 1)$$

где: λ_A, λ_B = амплитуда рассеивания в одну и другую строну средней арифметической ($\lambda_A = \bar{x} - x_{F\%} = 2{,}28$ и $\lambda_B = x_{F\%} = 97{,}72 - \bar{x}$ n_A, n_B — число величин, соответствующих частичным областям вариации λ_A соответственно λ_B.

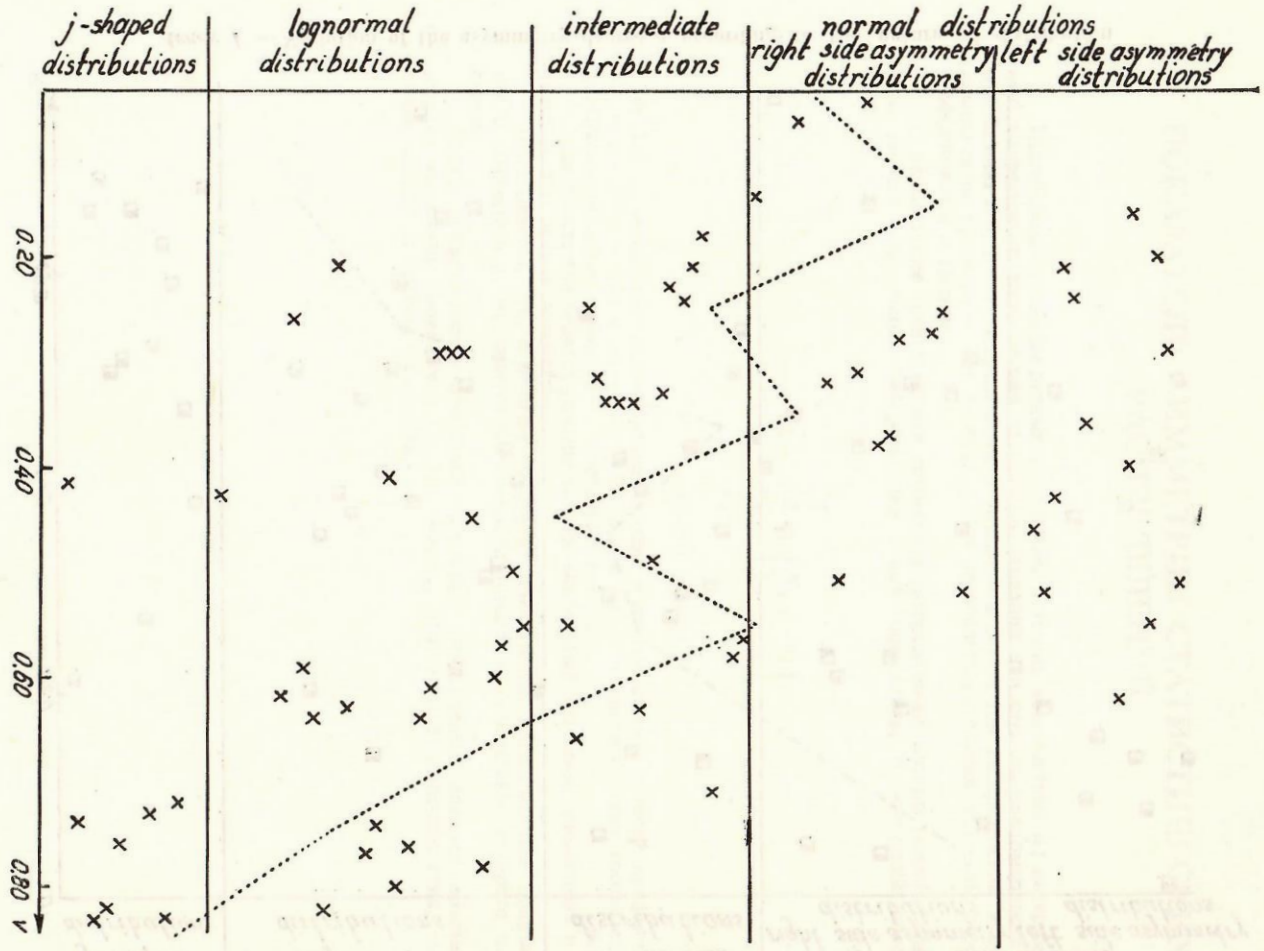
Знак показателя берется условно как положительный для левой асимметрии и отрицательный для правой ассиметрии.

Доказывается, что, в отличие от других показателей, предложенный показатель β, как и показатель ρ, не зависит от величины интервала, определение которого пока еще спорно.

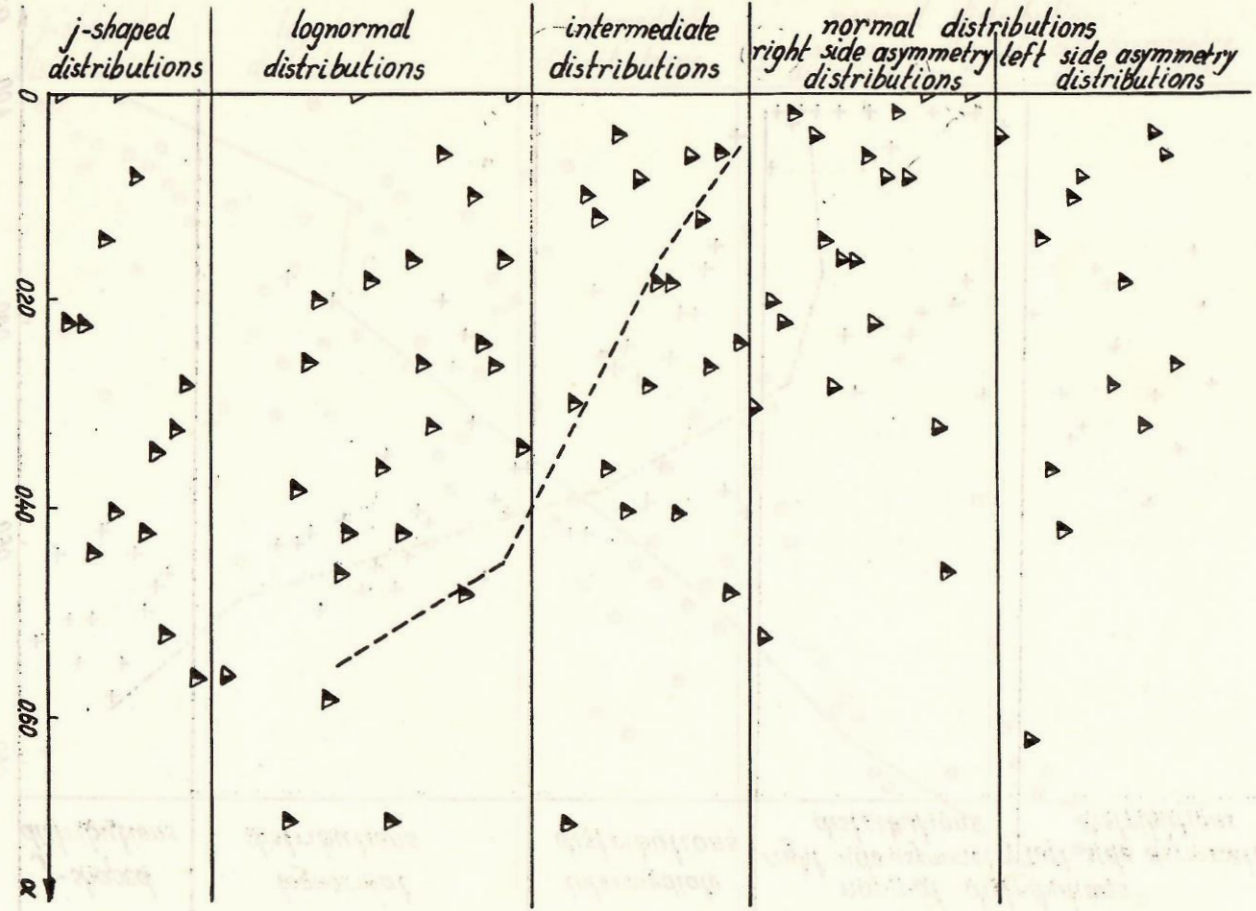
Также, представляются некоторые результаты относительно влияния объема селекции на важнейшие показатели ($\bar{x}$, *Mo*, *Me*, $\sigma_{\bar{x}}$, v, a, α, ρ, β) и последствия, затрудняющие работы по пробам в геологии.



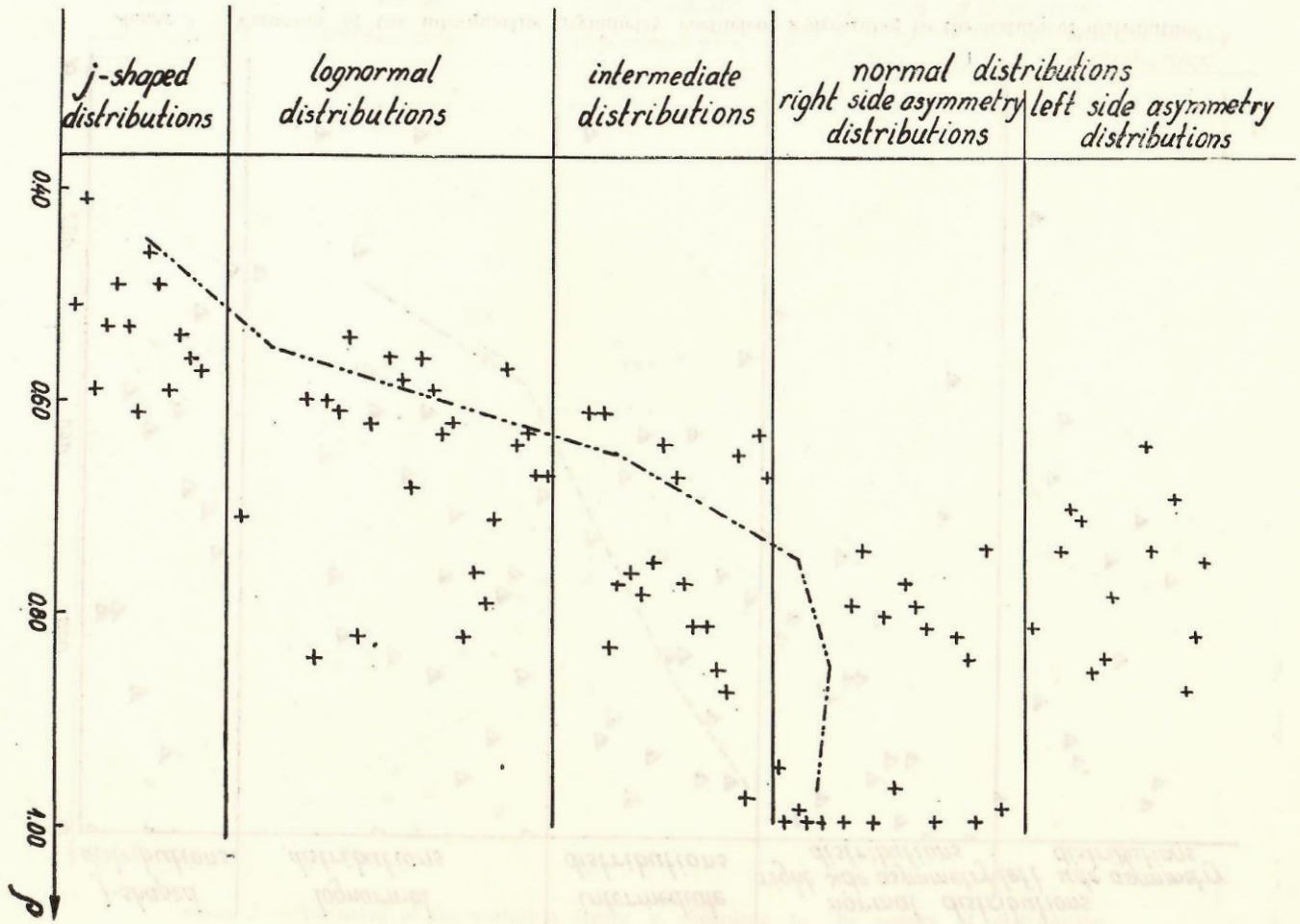
Annex 1 — Variation of the asymmetry degree a according to the nature of distribution



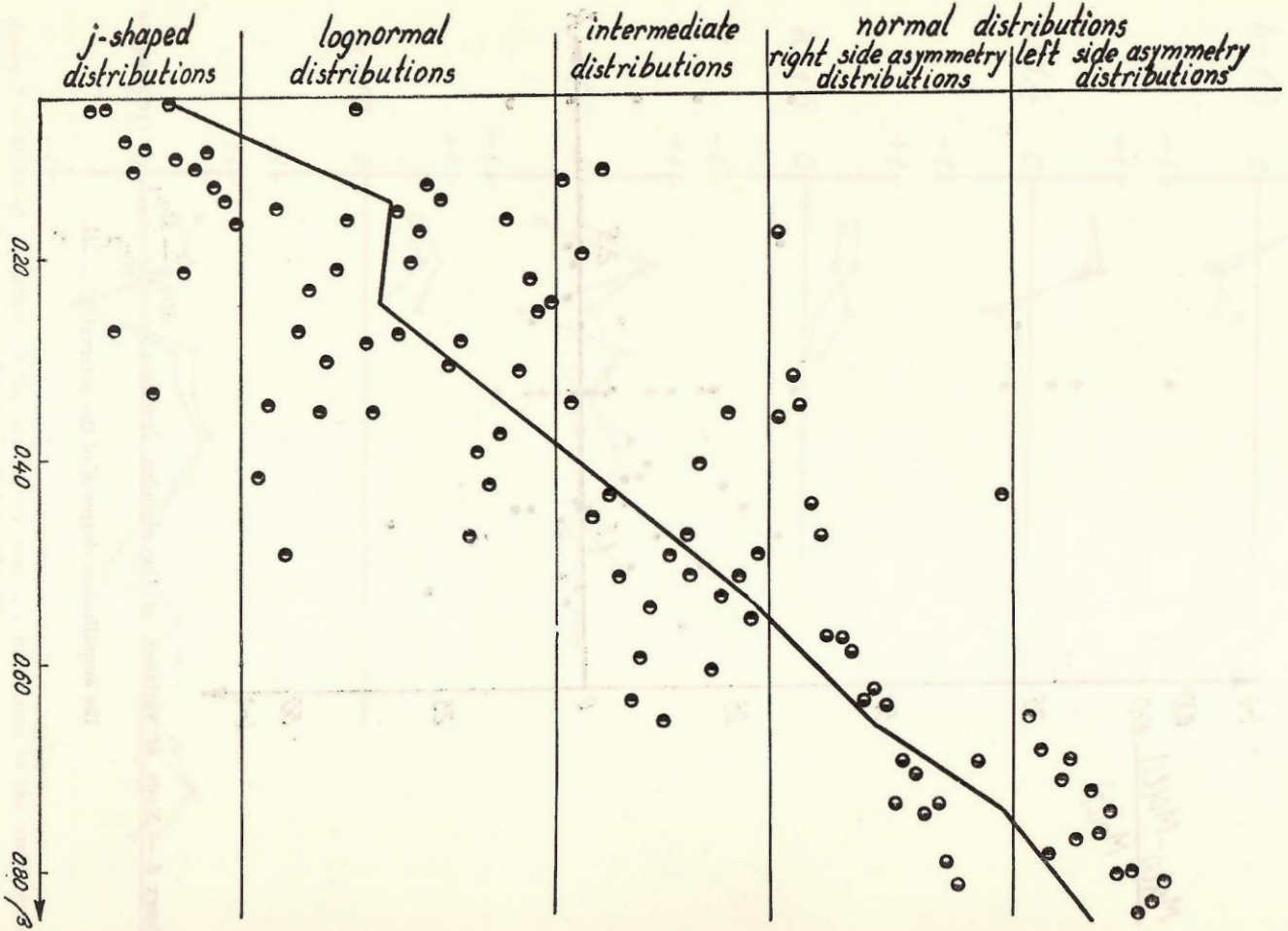
Annex 2 — Variation of the variation degree v according to the nature of distribution



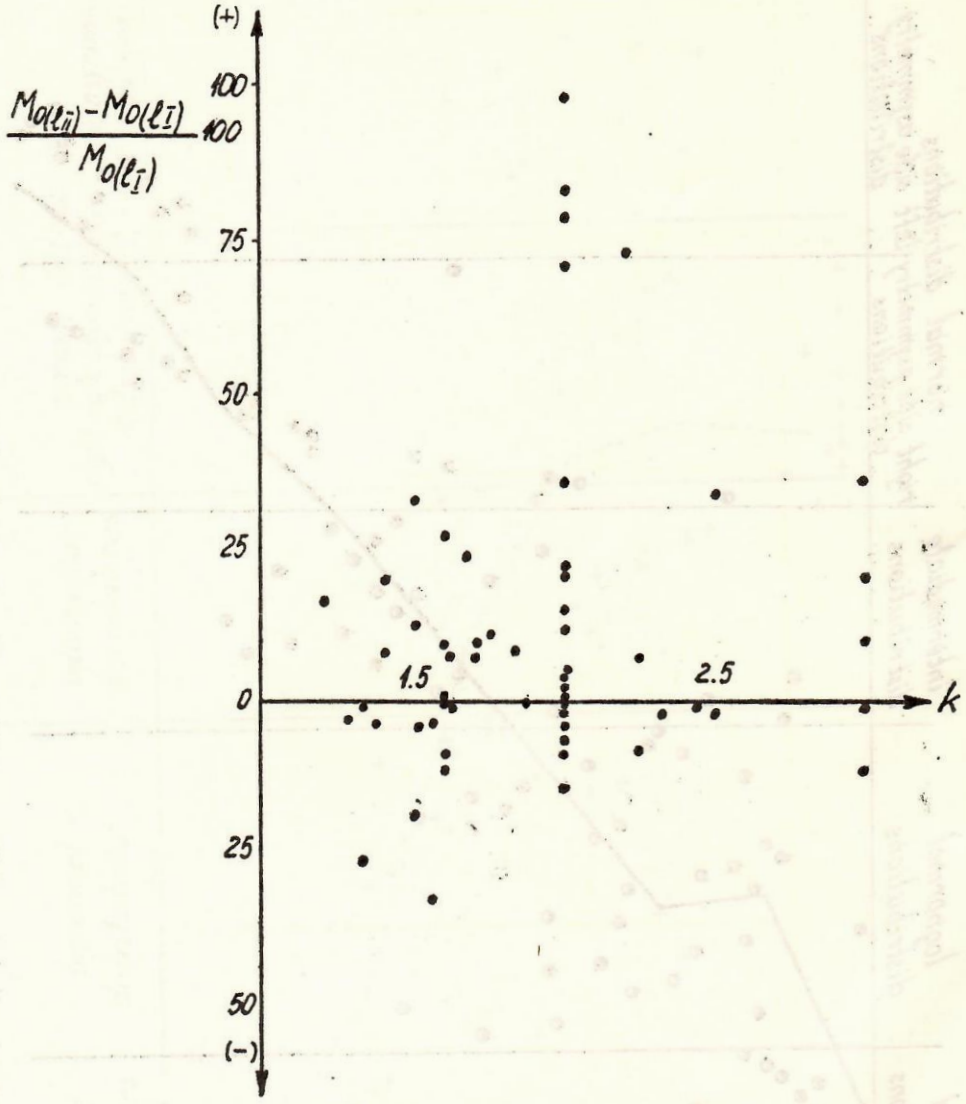
Annex 3 — Variation of the interquartile asymmetry coefficient α according to the nature of distribution



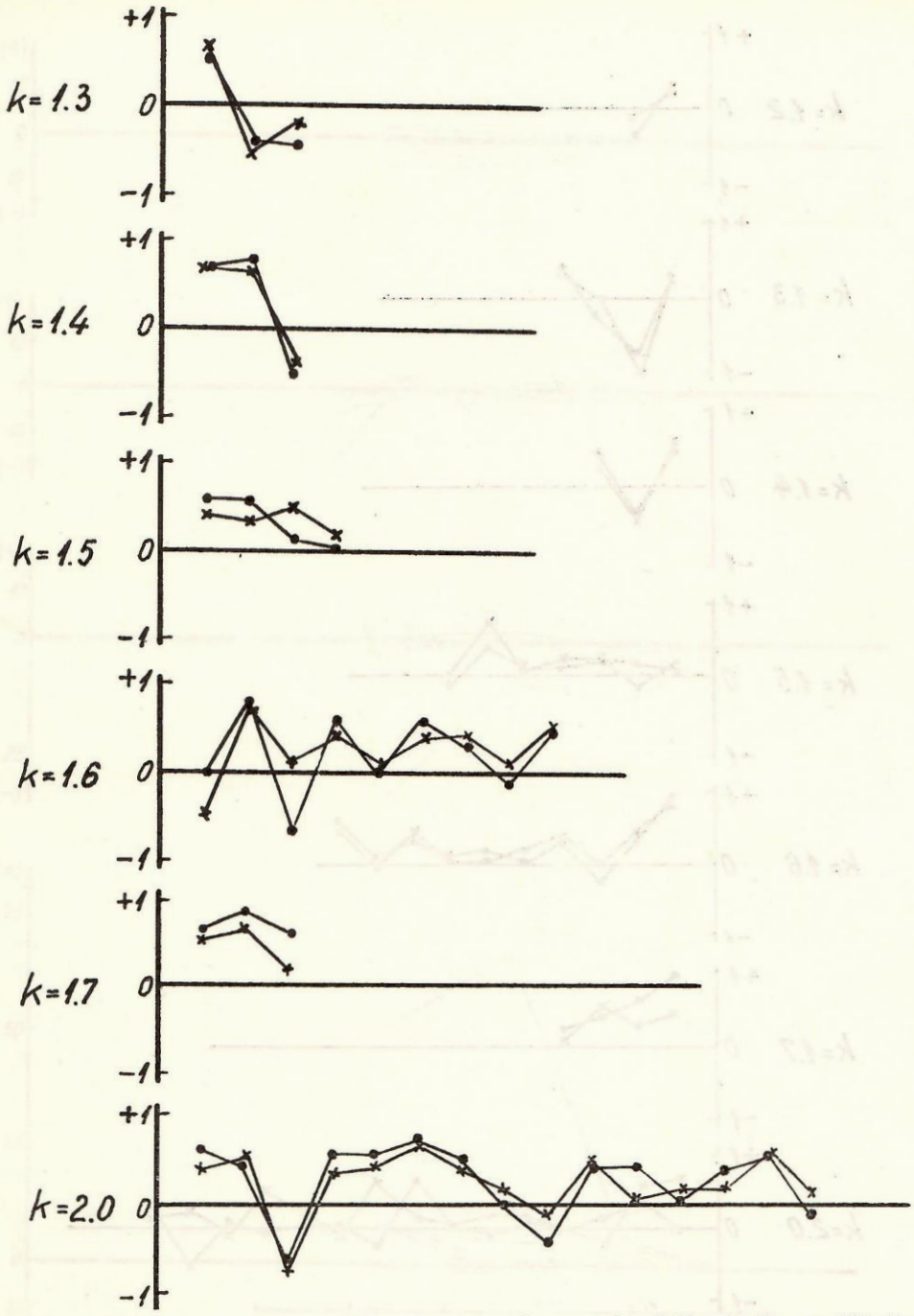
Annex 4 — Variation of the uniformity degree ρ according to the nature of distribution



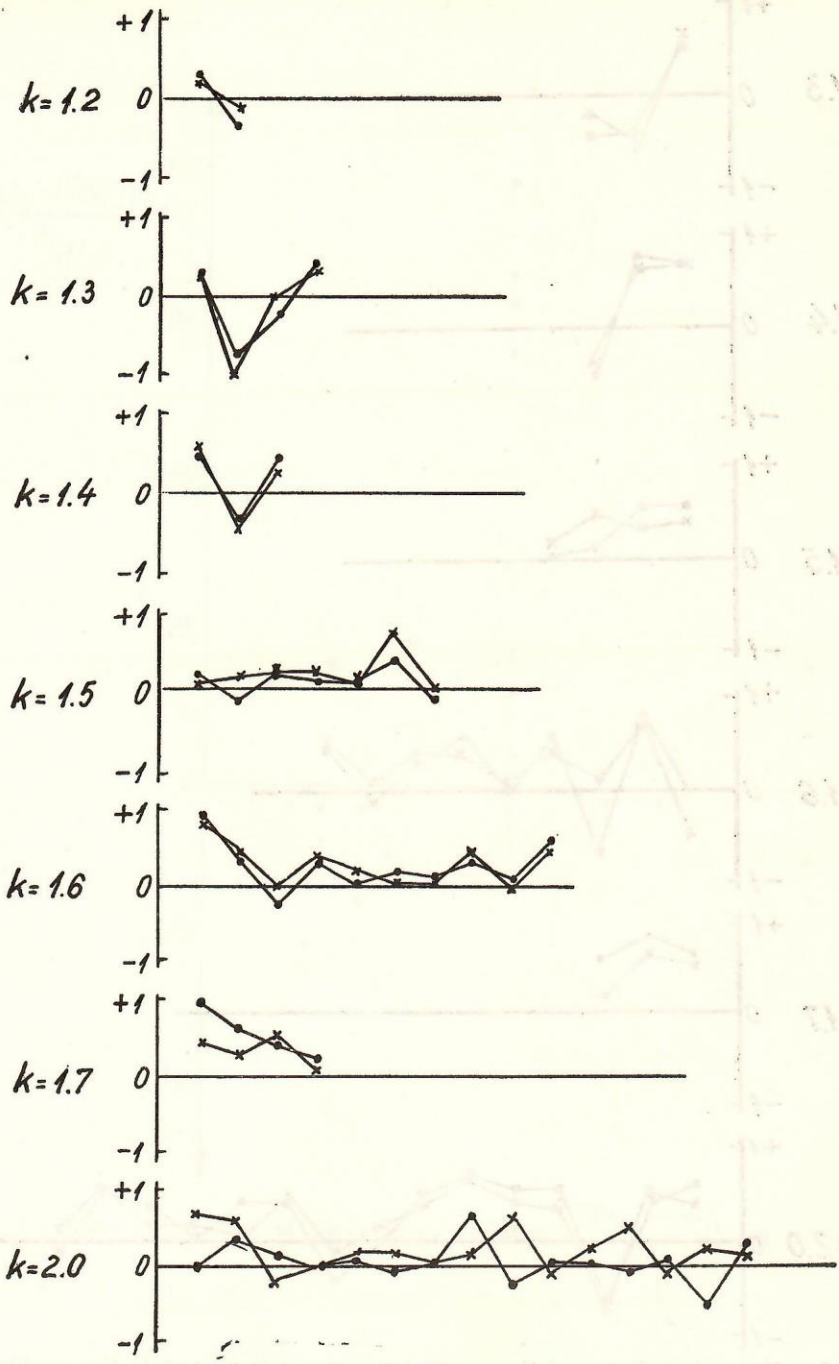
Annex 5 — Variation of the symmetry coefficient β according to the nature of distribution



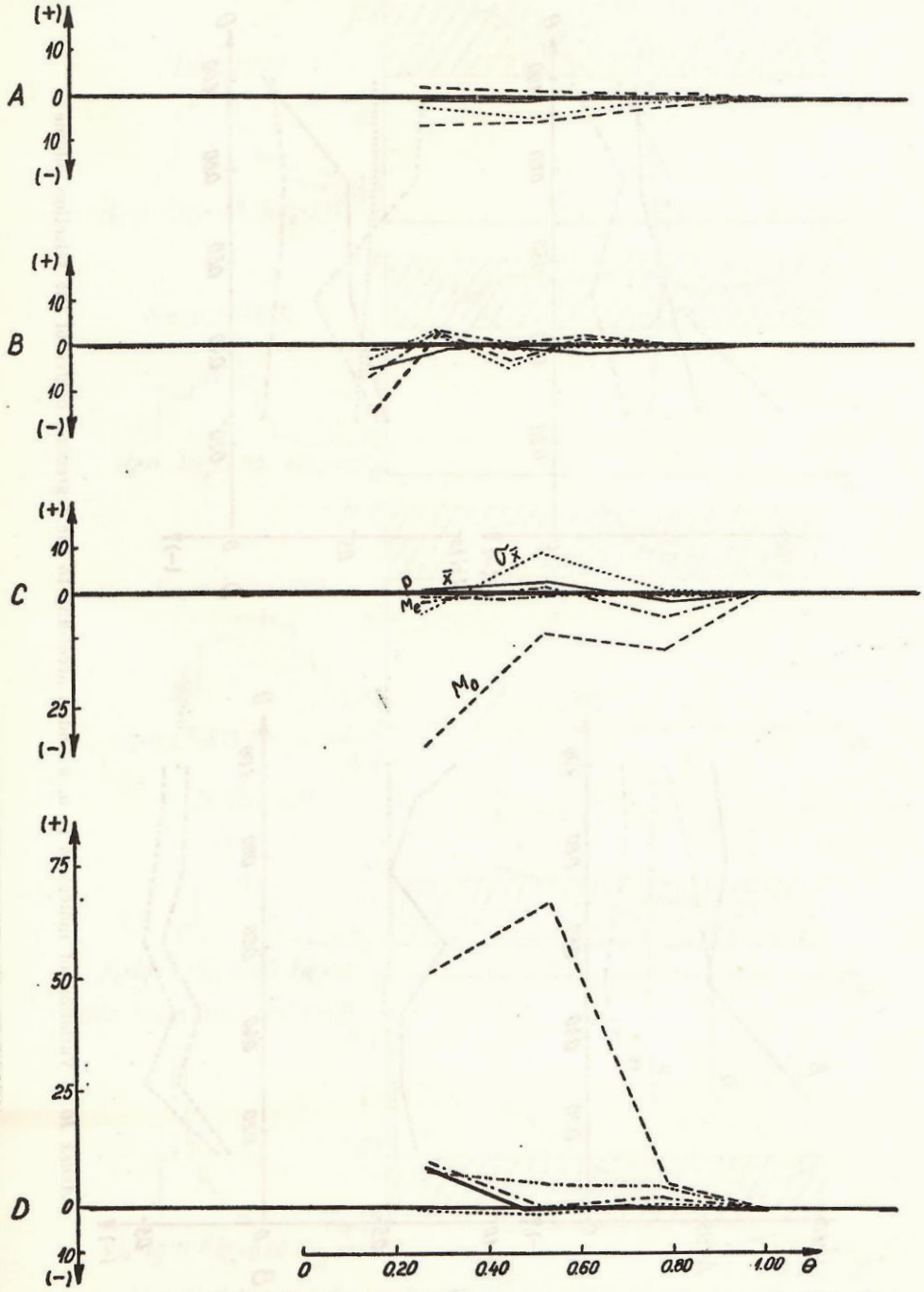
Annex 6 — Mode of variation of the modulus deviation $\frac{Mol_{II} - Mol_{I}}{Mol_{I}}$ in relation to the amplification degree K of the interval $K = \frac{l_{II}}{l_{I}}$



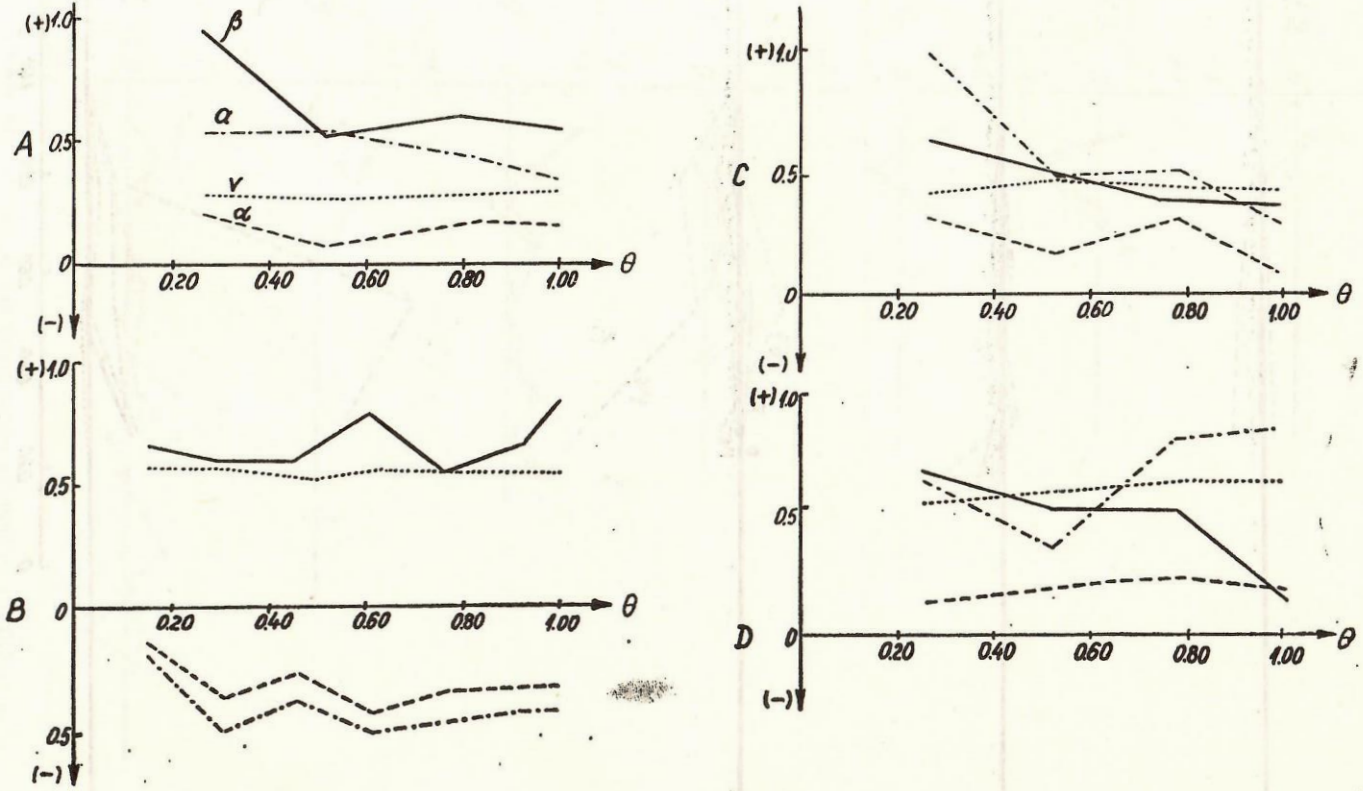
Annex 7 — Mode of variation of the asymmetry degree α in relation to the amplification degree K of the interval



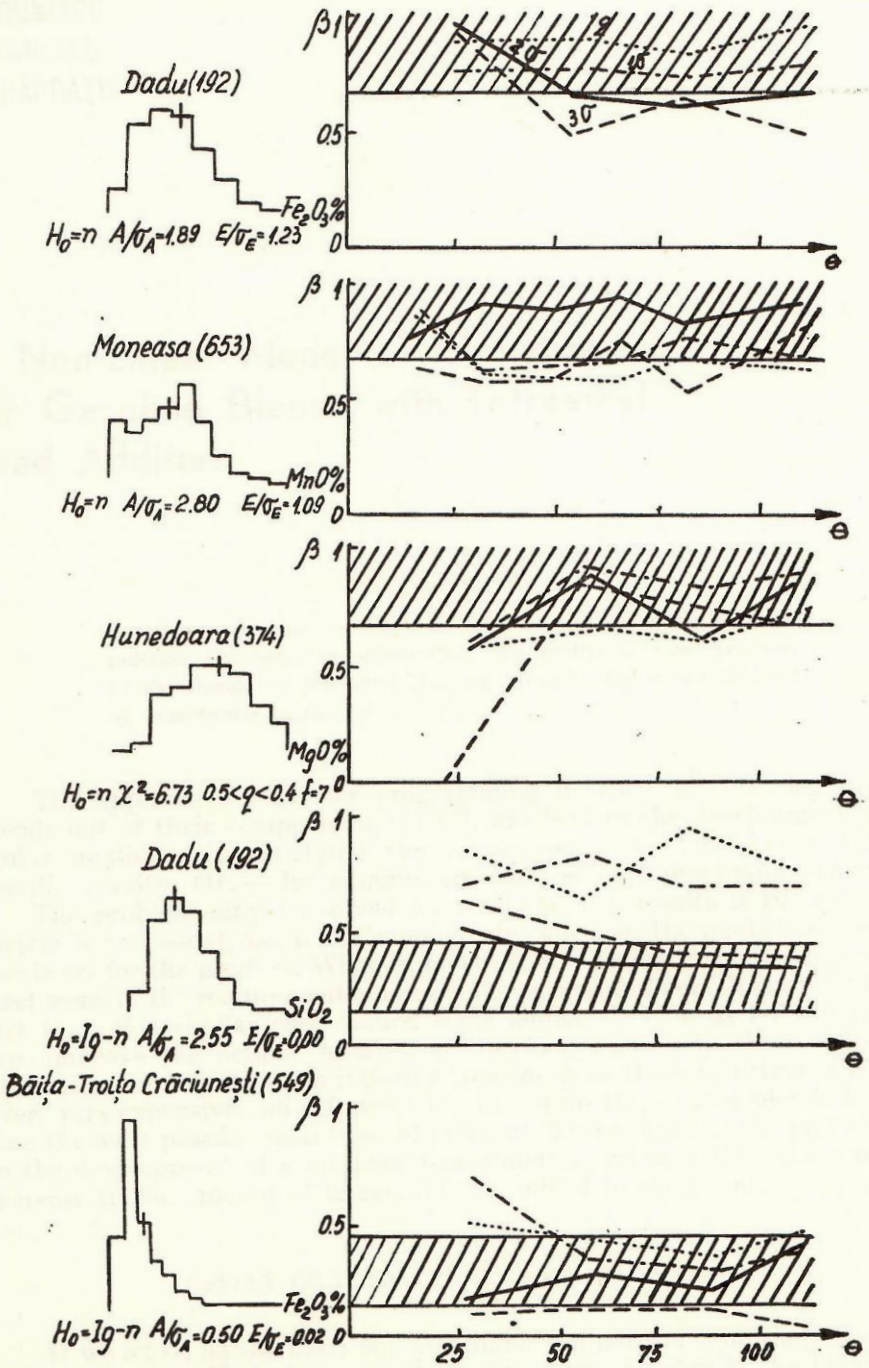
Annex 8 — Mode of variation of the interquartile asymmetry coefficient α in relation to the amplification degree K of the interval



Annex 9 — Variation of indicators $\bar{x}$, $\sigma_{\bar{x}}$, ρ, *Me* and *Mo* according to reduction degree θ of the selection volume



Annex 10 — Variation of indicators α, a, v and β according to the regression degree of the selection volume



Annex 11 — Variation of indicator β with the extension of the domain $\lambda_{A(B)}$ and of the reduction degree θ of the selection volume