

Distribution: limited

SHC/WS/291
PARIS, 10 June 1973
Translated from the French

UNITED NATIONS EDUCATIONAL,
SCIENTIFIC AND CULTURAL ORGANIZATION

STUDY XXIV

A method of establishing a list of development indicators

by

B. Ivanovic
Director of the Office of Statistics
of the
United Nations Conference on Trade
and Development

Division of Methods and Analysis
Department of Social Sciences

Note: This document may not be quoted
without authorization from Unesco

1. Introduction

The choice of social and economic development indicators is certainly one of the most important problems in the project for development evaluation and forecasting during the Second Development Decade: the more complete the information about national development levels contained in the indicators selected, the better the results.

There are a certain number of indicators which we always think of as providing the most reliable information about national social and economic development and which are the most frequently used in evaluating a country's development level, for example, the per capita GDP, the number of illiterates, the percentage of the population not employed in agriculture or the share of the GDP accounted for by industrial products.

However, the lists of development indicators used in various national or international research institutions are not always identical, and there is constant controversy about the value and importance of one or other of these well-established indicators. In actual fact, it may be wondered whether there are any objective reasons why they should all be considered as social and economic development indicators and whether others, so far overlooked or neglected, might not be more suitable than certain of the ones traditionally used.

We have thus endeavoured to find new development indicators, and in so doing we have defined and examined the indicator of "concentration of births in relation to the age of the mother".⁽¹⁾ This indicator provides information about the family situation in a given country and is obviously also an important development indicator. First of all, a high degree of concentration of births in relation to the mother's age in a particular country signifies that the number of children per family is limited, which ensures a higher standard of living in that country. Next, if the area of concentration is situated between the mother's ages of twenty and thirty, this means that the parents are of an age which makes it possible for the children to have a good family education, prolonged economic protection, favourable physical development and a more complete school education. Lastly, where there is a high degree of concentration, one rarely encounters unfavourable cases such as those of mothers who are too young or too old, children born of a second marriage, etc. In short, the more concentrated the births are in relation to the age of the mother, the more ordered is the family life and the more effective the family's activity from the social and economic points of view.

(1) In a country j , the measurement of this indicator is given by

$$C_j = 1000 \sqrt{\frac{1}{n} \sum_{i=1}^n \left(\frac{x_{ij}}{X_j} - \frac{1}{n} \right)^2},$$

where n represents the number of age groups, x_{ij} the number of births in which the mother's age is in the i -th group, and X_j the total number of births in the country j during the year in question.

Quantitative analysis shows that this indicator gives more information about a country's level of development than certain traditional indicators such as per capita exports of manufactured goods, average expectation of life, number of literates, the contribution of the manufacturing industries to the GDP, the number of school enrolments, etc.

The conclusion could be that to arrive at an exhaustive list of development indicators, those which have previously been neglected or overlooked should be added to those appearing in all the lists used hitherto.

This would make a very long list, which we shall call the maximum list of social and economic development indicators. The statistics established for these indicators would give a very complete picture of all aspects of development in the countries concerned.

However, if we examine all the statistical publications of the United Nations Statistical Office, regional economic commissions, and various Specialized Agencies of the United Nations, as well as national publications and other official or semi-official documents, we see that the choice of development indicators is greatly limited by the absence of statistical data about many developing countries.

The absence or poor quality of statistical data hinders the introduction of a large number of development indicators such as those providing information about productivity in industry or agriculture, standard of living, infrastructure, industrial and educational cadres etc.

The introduction of even a single indicator into a country's statistical system frequently involves setting up a new department with suitably qualified staff and adequate equipment for the compilation, processing and publication of the data. These data should be of a high quality and comparable with data from other countries.

The introduction of new items into our list of development indicators consequently requires time, new institutions and great financial efforts. It goes without saying that this problem is even more serious in developing countries.

If we are to keep track of the development of countries during the Second Decade with anything like sufficient accuracy to reach our objectives, some new items ought unquestionably to be included in the list of development indicators. The question is, however, which should be chosen from the many already suggested.

The first problem to be solved in carrying out the project for the evaluation and forecasting of development during the Second Decade will thus be to establish, on a scientific basis, an optimum list of a limited number of development indicators.

2. Criterion for selecting social and economic development indicators

Various statistical indicators may be used to estimate and keep track of the social and economic development of a country. Each of these indicators provides information about one aspect of the level of development, and information

about this aspect is not, generally speaking, identical with other information. The indicator "percentage of the population not employed in agriculture", for example, gives us information about a country's level of development which is not identical with the information given by the indicator "per capita energy consumption". A choice must thus be made and a whole set of indicators has to be prepared in order to give a more complete idea of the level of development of the countries concerned.

First conclusion: Increasing the number of indicators also increases the total amount of information about the country's level of development.

The various development indicators do not contain the same volume of information about the development level of a country, i.e. they are not all of equal importance in relation to development. The "per capita GDP" indicator thus provides more information about the social and economic development level of a country than, for example, the "average life expectancy" indicator. If the quantity of information contained in development indicators is measurable, it should be possible to classify them in order of importance.

Second conclusion: Sets of indicators of the same size (same number of indicators) do not, in general, contain the same quantity of information about the country's development level.

Generally speaking, development indicators are not independent of each other, so that information provided by one indicator is partially contained in the information already provided by others. Taking the whole set of indicators selected, therefore, it could be that certain quantities of information are repeated two or more times. It is obvious that these duplicated items would have to be subtracted from the total volume of information so as to leave only the basic information contained in this set of indicators.

Third conclusion: The total amount of information (without duplication) given by a set of indicators is generally less than the sum of the quantities of information contained individually in each indicator of that set.

Let us take the example of two completely interdependent indicators. If we arrange the countries in question according to the values corresponding to either indicator, we shall obtain two identical lists. This means that both indicators contain the same information about the countries' development level and that consequently only one of the indicators need be considered, the other being superfluous.

Fourth conclusion: Despite the fact that two given indicators may be very important from the point of view of the information they provide, separately, about a country's development level, the contribution of one of them is insignificant if there is a high degree of correlation between the two.

Fifth conclusion: Indicators producing information which is completely contained in the overall information provided by the indicators already taken into account should be discarded.

It follows that a relatively small number of carefully chosen indicators could contain a greater sum total of information than a large number of badly chosen indicators.

Consequently, to obtain as complete an idea as possible of the development level of the countries under observation, it is not enough to increase the number of indicators. To increase the total amount of information, the importance of each indicator, expressed in the form of a quantity of information, and duplications of information, expressed in the form of interdependence between indicators, must be taken into account.

The optimum selection of a limited number of development indicators should give a maximum total amount of information (without duplication), whilst the sum total of duplications should be reduced to a minimum.

The more information an indicator provides which is not already contained in the sum of the information provided by more important indicators, the greater that indicator's contribution to the evaluation of the country's development level will be. Nevertheless, even an indicator which makes quite a considerable contribution will not be very useful if its value is almost constant in all the countries observed. The indicator "percentage of illiterates over ten years of age in the population", for example, is of no significance in the rich countries where it has become stabilized at between 1 and 3 per cent. On the other hand, in the poor countries this value varies between 5 and 99 per cent, so that this same indicator provides a high degree of discrimination in the evaluation of the development level of developing countries. The opposite would be true if we took "contribution of industrial production to the GDP" as the development indicator.

Sixth conclusion: A good development indicator should be discriminatory in all the countries under observation.

We may therefore complete the criterion for selecting a development indicator for a given set of countries as follows:

The greater an indicator's discriminatory capacity in all the countries observed and the greater the quantity of information it provides which is not contained in the sum of the information given by more important indicators already taken into consideration, the greater its contribution to an evaluation of the countries' development level.

3. Objective evaluation of the degree of importance of a development indicator.

The degree of importance or significance of a development indicator may be thought of as the quantity of information about a country's development level it contains. This quantity of information cannot, of course, be expressed in absolute terms. What we are able to determine is more a relative quantity, dependent on which countries are being observed and which indicators are selected.

If a given indicator is completely dependent on the development level of a country, we shall have the same classification of countries if we list them

according to their value in terms of this indicator, or according to their development level. In this case, we may say that all the information about the development level of countries is contained in this single indicator, since one only has to list these countries by their values according to this indicator to obtain at the same time the list by levels of development.

If dependence is partial, the information contained in the given indicator is only part of the total information about the development level of the countries.

Consequently, we may say that the degree of importance of a development indicator corresponds to the degree of dependence between that indicator and the development level.

If we take the I-distance $D^{(1)}$ as the relative measure of a country's level of development and if we express the degree of dependence between indicator X_1 and the development level by the simple correlation coefficient r_1 , the value r_1 will express the degree of importance of X_1 .

Listing the given indicators according to the order of magnitude of their coefficients of normal correlation with the I-distance, we obtain the classification of development indicators in order of importance.

Development indicators could thus be classified objectively according to the correlation between the various indicators actually used and the overall index containing the maximum amount of information. We cannot calculate the correlation coefficients directly, however, since before we can do so we need the series of I-distances, to calculate which we need the indicator classification which we are in the process of establishing.

As we are thus in a vicious circle, we have to see whether it is at least possible to resolve the problem of classifying development indicators by an approximate method.

It should be observed, first of all, that there are two possible extreme cases of our problem. In the first, all indicators are independent and the problem of the same quantities of information being duplicated no longer occurs. In the second case, all the indicators selected are completely dependent on one of them, i.e. the quantities of information of all the indicators are contained in the information of the dominant indicator. In both cases, it is possible to establish the series of I-distances without any difficulty and subsequently to order the indicators by their degree of importance.

In practice, however, one is usually in a position midway between these two extremes. One may say, nevertheless, that one is always either in a position in which one indicator contains most of the sum total of information about the countries' development level, or in a position in which it is impossible to say that one indicator is obviously dominant in relation to the rest.

(1) See Research Memorandum No. 41 of 5 November 1970 (UNCTAD).

The first position is closer to the extreme case of the total dependence of all indicators on one dominant indicator than to the extreme case of the total independence of all indicators. This is why, as a first approximation, we shall take the dominant indicator as containing the sum total of all the information in all the indicators being studied. In other words, we shall assume for the time being that the development level corresponds to the value of the dominant indicator for the country in question.

If we calculate the correlation matrix and list the given indicators in accordance with the order of magnitude of their coefficients of correlation with the dominant indicator, we shall obtain a provisional list of development indicators by degree of importance. This we shall call the first-stage classification.

Using for the time being, the indicators arranged in this order, we are able to calculate the I-distance values for all the countries under consideration. As this I-distance based on all the indicators under consideration contains more information about the development level of the countries than the dominant indicator, we shall, in the second stage, determine the order of the indicators according to the magnitude of their coefficients of correlation with the I-distance of the first stage. This will be the second-stage classification.

If we calculate the series of I-distances according to the second-stage classification and if we calculate the coefficients of correlation between the indicators under consideration and the I-distance thus obtained, we arrive at the third-stage classification of indicators.

This iterative process is continued until such time as the classification of indicators is identical with that of the previous stage. The classification is then definitive since it remains invariable in all succeeding stages.

In the second case (without a dominant indicator) one is closer to the extreme case of the complete independence of all the indicators than to the extreme case of the complete dependence of all the indicators on the dominant one. We thus assume, as a first approximation, that all the indicators are independent and that there is consequently no duplication of information.

The I-distance is reduced in this case to Fréchet's distance F , the calculation of which requires no order of indicators since they enter into the F formula on an equal footing.

If we calculate the values of the F distance for all countries under consideration and if we subsequently determine the coefficients of correlation between the indicators observed and the F distance, this gives us the first-stage classification of development indicators by their degree of importance.

The same iterative method as in the first case is subsequently used to arrive at the definitive list of indicators.

It should be noted that the selection of social and economic development indicators is an instance of the first case, since the indicator "per capita GDP" is quite clearly dominant.

If the development indicators observed $\{X_1, X_2, \dots, X_n\}$ are arranged according to the definitive classification and if the I-distance D for each country is calculated according to this classification, the quantitative measurement of the importance of the indicator X_i will be expressed by the formula:

$$I_i = r(X_i; D).$$

Their order of magnitude corresponds to the definitive classification, namely:

$$r(X_1; D) \geq r(X_2; D) \geq \dots \geq r(X_n; D).$$

Taking the developing countries and restricting the development indicators to 12, the order of importance of these indicators in 1968 was as follows:

<u>Order</u>	<u>Indicator</u>	<u>Degree of importance</u>
1	Per capita GDP	0.97
2	Per capita energy consumption	0.95
3	GDP of the population employed in agriculture	0.81
4	Number of doctors per 100,000 inhabitants	0.81
5	Number of newspapers printed per 1,000 inhabitants	0.80
6	Percentage of the population not employed in agriculture	0.79
7	Concentration of births in relation to the age of the mother	0.77
8	Exports of manufactured goods per head of population	0.76
9	Average expectation of life	0.71
10	Literacy rate	0.69
11	Share of manufacturing industries in the GDP	0.66
12	Rate of school enrolments	0.64

4. Quantitative assessment of the contribution of an indicator to the evaluation of a country's development level

4.1 Statistical parameters

Let $X = \{X_1, \dots, X_n\}$ represent the set of n indicators being considered for possible use in evaluating the development level of all the countries under observation, represented by $P = \{P_1, \dots, P_m\}$.

Assuming that statistics are available for each indicator of X and for each country of P , if the value of the indicator X_i for country P_j is expressed by the formula x_{ij} , these statistics may be presented in the form of the matrix $M = [x_{ij}]$, $i = 1, \dots, n$, $j = 1, \dots, m$.

Calculation of the statistical parameters of the indicators X_i requires the values x_{ij} to be weighted. Designating the matrix of the weighting co-efficients by the formula $N = [n_{ij}]$, the arithmetic mean will be defined by the formula:

$$\bar{x}_i = \frac{1}{\sum_j n_{ij}} \sum_{j=1}^m n_{ij} x_{ij}, \quad i = 1, \dots, n$$

and the standard deviations by:

$$\sigma_i = \left\{ \frac{1}{\sum_j n_{ij}} \sum_{j=1}^m n_{ij} (x_{ij} - \bar{x}_i)^2 \right\}^{1/2}, \quad i = 1, \dots, n.$$

Using the covariance w_{ic} of the indicators X_i and X_c , it is possible to calculate normal correlation coefficients:

$$r_{ic} = \frac{w_{ic}}{\sigma_i \sigma_c}, \quad i, c = 1, \dots, n,$$

partial correlation coefficients:

$$r_{ic.s} = \frac{r_{ic} - r_{is} r_{cs}}{\sqrt{(1-r_{is}^2)(1-r_{cs}^2)}},$$

for $i < c$, $i, c = 1, \dots, n$ and $s \neq i, c$,

and multiple correlation coefficients:

$$r_{1.2 \dots n} = \sqrt{1 - \frac{R}{R_{11}}},$$

where R represents the determinant of the correlation matrix and R_{11} the minor of R corresponding to the element r_{11} .

4.2 The discriminatory capacity of an indicator

As a measurement of the discriminatory capacity of the indicator X_i in the countries P as a whole, we suggest the coefficient of discrimination which can be expressed algebraically in the following form:

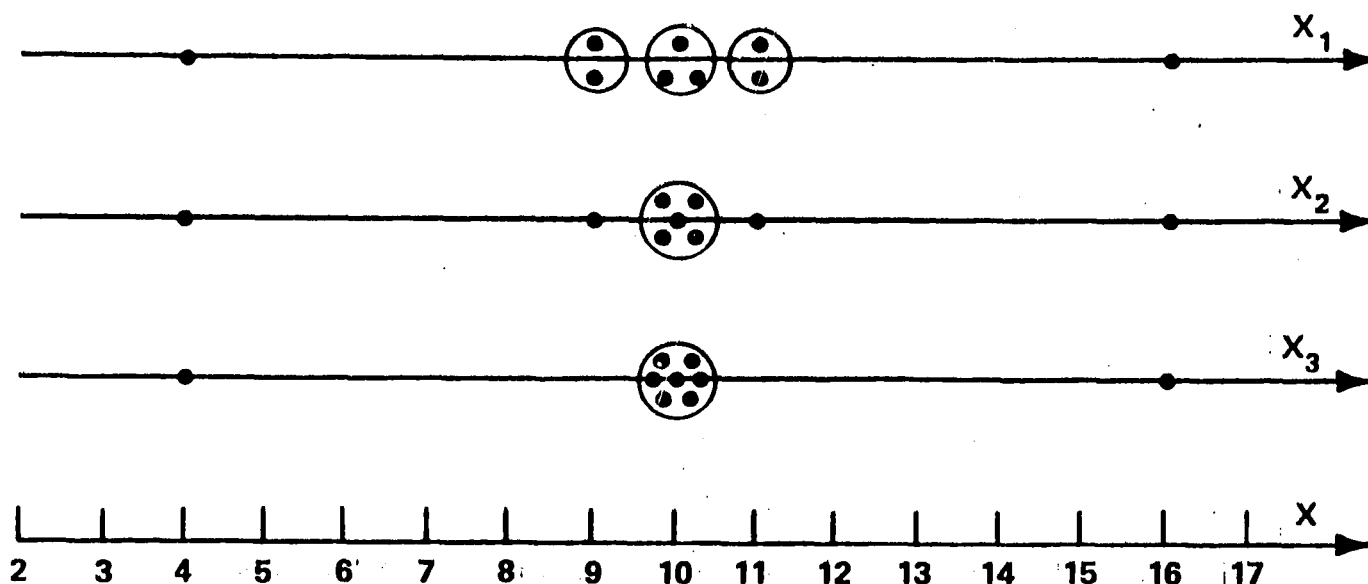
$$CD_i(P) = \frac{\bar{u}_i}{\bar{x}_i} = \frac{1}{\binom{n}{2} \bar{x}_i} \sum_{j, l > j}^m n_{ij} n_{il} |x_{ij} - x_{il}|,$$

or

$$CD_i(P) = \frac{2}{n(n-1)} \sum_{j, l > j}^m n_{ij} n_{il} \left| \frac{x_{il}}{\bar{x}_i} - \frac{x_{jl}}{\bar{x}_i} \right|, \quad i = 1, \dots, n.$$

It should be noted that this coefficient does not have the same significance as the coefficient of variation which measures the relative mean dispersion of X_1 in P.

Thus in our example, which illustrates three distributions in which the discriminatory capacity obviously varies (with respective coefficients of discrimination of 3.22, 3.00 and 2.67), the dispersion is practically constant (2.88, 2.86 and 2.83).



4.3 Total information of a group of development indicators

To determine the total quantity of information contained in the set of k indicators selected X_1, X_2, \dots, X_k , we must follow the criterion concerning the choice of development indicators formulated in section 2, according to which we must take into account each indicator's degree of importance, the duplications of information contained in the various indicators, and each indicator's discriminatory capacity.

Given that the indicators X_1, X_2, \dots, X_k are already arranged in order of their importance, the overall contribution of this set of indicators to the evaluation of the development level of the countries under observation is expressed by the formula:

$$(4.3.1) \quad C(1,2,\dots,k) = \sum_{i=1}^k CD_i(P) \prod_{j=1}^{i-1} (1 - r_{j1.12 \dots j-1}) ,$$

where $CD_i(P)$ represents the coefficient of discrimination of indicator X_i in all the countries P and $r_{j1.12 \dots j-1}$ the coefficient of partial correlation between indicators X_1 and X_j .

This formula is increasing and homogeneous in relation to the discriminatory capacities of the indicators selected. Each capacity is weighted so that the higher the indicator's rank and the smaller the partial correlations between it and the indicators preceding it, the greater the importance of its contribution in the sum total of information.

If all indicators are mutually independent, the total amount of information becomes the sum of the discriminatory capacities of all the indicators selected, namely:

$$(4.3.2) \quad C(1,2,\dots,k) = \sum_{i=1}^k CD_i(P) ,$$

and if a functional-linear dependence exists between them, the total amount of information is reduced to the discriminatory capacity of the first indicator, namely to:

$$(4.3.3) \quad C(1,2,\dots,k) = CD_1(P).$$

If the set of indicators selected can be divided into two mutually independent groups, the total amount of information will be the sum of the total information corresponding to the two groups.

It should be emphasized again that in accordance with the analytical formula for the total amount of information (4.3.1), the discriminatory capacities of all the indicators represent additive quantities and that the quantity of information from indicator X_1 already contained in the information from the indicators preceding it in order of importance may be subtracted by using the weighting factor

$$(1-r_{11})(1-r_{21.1})(1-r_{31.12})\dots (1-r_{i-1,1.12 \dots i-2}) .$$

4.4 Optimum list of a limited number of development indicators

The fundamental question involved in the choice of a list of development indicators is: if we have a maximum list of n indicators, how can it be reduced to a list of k indicators ($k < n$) such that this set of k indicators shall contain a total quantity of information about the development level of the countries under observation greater than any other set of k indicators selected from the n indicators of the maximum list?

Let us assume that the indicators $\{X_1, X_2, \dots, X_n\}$ are already arranged according to the definitive classification and that

$$\langle \{X_{i_1}, X_{i_2}, \dots, X_{i_k}\} \rangle$$

is an ordered sub-set of k indicators in which $i_1 \leq i_2 \leq \dots \leq i_k$ and

$$r(X_{i_1}; D) \geq r(X_{i_2}; D) \geq \dots \geq r(X_{i_k}; D) .$$

The total contribution of this set of indicators to the evaluation of the development level of the countries is, according to formula (4.3.1), as follows:

$$(4.4.1) \quad C(i_1, i_2, \dots, i_k) = \sum_{s=1}^k CD_{i_s}(P) \prod_{j=1}^{s-1} (1 - r_{i_j, i_s} \cdot i_1 i_2 \dots i_{j-1}) \quad .$$

The optimum list of k indicators will thus consist of the group $\{X_\alpha, X_\beta, \dots, X_k\}$ which maximizes formula (4.4.1). In other words, the optimum list will consist of the indicators $X_\alpha, X_\beta, \dots, X_k$,

if

$$(4.4.2) \quad C(\alpha, \beta, \dots, k) = \text{Max}_{i_1, i_2, \dots, i_k} \{C(i_1, i_2, \dots, i_k)\} \quad .$$

It will not be easy to determine this optimum list, even using a computer, if the maximum list contains a large number of indicators.

In practice, however, high-level authorities such as the United Nations General Assembly, UNCTAD's Trade and Development Board, ECOSOC, etc., which decide on and adopt a definitive list of development indicators, could insist on a certain number of indicators being included in advance and leave it to the experts to add others to make up an optimum list. This list would perhaps no longer be the best possible, but from the technical point of view it would be much easier to carry out calculations.

Let X_1, X_2, \dots, X_p be the p indicators decided in advance and let

$X_{i_{p+1}}, X_{i_{p+2}}, \dots, X_{i_k}$ be the $(k-p)$ indicators to be chosen from the $(n-p)$

indicators of the maximum list. We can order the set formed by the union of any combination of $(k-p)$ indicators with the p indicators decided in advance, i.e.

$$< \{1, 2, \dots, p\} \cup \{i_{p+1}, i_{p+2}, \dots, i_k\} > \quad .$$

The order of these indices will be established according to the degree of importance of the corresponding indicators.

The optimum list will now be made up of the p indicators decided in advance and the $(k-p)$ indicators $\{X_\nu, X_\mu, \dots, X_k\}$ for which

$$(4.4.3) \quad C \left[< \{1, 2, \dots, p\} \cup \{\nu, \mu, \dots, k\} > \right] = \text{Max}_{i_{p+1}, \dots, i_k} \left\{ C \left[< \{1, 2, \dots, p\} \cup \{i_{p+1}, i_{p+2}, \dots, i_k\} > \right] \right\}$$

Lastly, if the k indicators are already selected, either by establishing them in advance or by identifying them by the formulae (4.4.2) or (4.4.3), it remains to identify the indicator X_{k+1} which would produce the maximum increment in the total amount of information $C(1, 2, \dots, k)$.

This will be the indicator X_{ω} if

$$C\left[\left\langle \left\{1, 2, \dots, k\right\} \cup \left\{\omega\right\} \right\rangle\right] = \max_{i_{k+1}} \left[\left\langle \left\{1, 2, \dots, k\right\} \cup \left\{i_{k+1}\right\} \right\rangle \right]$$

for $i_{k+1} \in \{k+1, k+2, \dots, n\}$.

5. Conclusion

We have tried in this study to outline an objective procedure for establishing an optimum list of development indicators which would provide statistics of use in evaluating and forecasting development in the Second Development Decade.

The first step is to establish the maximum list of development indicators by combining the various lists used hitherto with other indicators which have been neglected or overlooked in the past.

The procedure then consists of making a quantitative evaluation of the discriminatory capacity and importance of each indicator in this maximum list.

The total amount of information of a group of indicators is defined as the sum of the coefficients of the discriminatory capacity of all the indicators in this group whose weighting coefficients correspond to the importance of the indicators and do away with duplications of the same quantities of information. The optimum short list of development indicators will then be formed by the indicators which provide the maximum quantity of total information.

Considering the absence of statistics in certain cases or the poor quality of the statistics provided by the large number of indicators in a maximum list, and the expense and time involved in introducing new indicators into the statistical system of developing countries, this procedure makes it possible to draw up a rational plan for determining a limited number of indicators providing the maximum amount of total information.