Board 1981
B.D.6

UNITED NATIONS RESEARCH INSTITUTE FOR SOCIAL DEVELOPMENT

MEASUREMENT AND ANALYSIS OF SOCIO-ECONOMIC DEVELOPMENT

Chapter 2: Problems of International Data Quality and Usage

Chapter 3: Selection of Development Indicators and Screening of Data

Chapter 4: Analytic Approaches to International Development Data

Geneva, June 1981
UNRISD/81/C.16

GE.81-01883

CHAPTER 2

PROBLEMS OF INTERNATIONAL DATA QUALITY AND USAGE

Each year numerous compilations of economic and social statistics, covering the world as a whole or major regions of it, are published by international, regional and national agencies. 1/ There is much copying and duplication - and at the same time there are striking inconsistencies among the published statistics. Numerous analytic studies are carried out on these statistics by organizations and by individual scholars and issued in official or scholarly publications. Many of the published statistics are unreliable or for other reasons not comparable, and much of the comparative analysis of them is a useless exercise. Even when the statistics used are sufficiently comparable to justify their use in analysis, the methods of analysis employed are very often unsuitable for the kind of data involved, involving assumptions that are not valid. The present chapter will be concerned with the question of data quality; Chapter 4 will be concerned with methods of analysis.

Publications like the United Nations Statistical and Demographic Yearbooks while sometimes giving estimates (e.g., demographic estimates prepared by the United Nations Population Division), primarily record governmental statistics that have been received in response to a query to national statistical offices or that have been taken from official governmental publications. These statistics are mainly based on national censuses, household surveys, registration systems, and various kinds of administrative records (school enrolments, hospital records, tax records, licensing records, etc.). The national figures are often seriously incomplete (and unrepresentative) in their coverage of the population, especially in the poorer countries; or they may be nominally complete but unreliable because of untrained data collectors, ignorance, apathy or hostility on the part of respondents and registrants, excessive use of guesswork or "estimates" by statisticians to fill in information gaps or to

---

1/ For example, within the United Nations system alone, statistics on birth rates in developing countries are issued annually or at regular intervals in the United Nations Demographic Yearbooks, the United Nations Statistical Yearbooks, the United Nations Population and Vital Statistics Reports, the United Nations Monthly Statistical Bulletins, the United Nations Compendiums of Social Statistics, the Statistical Yearbooks of the Regional Commissions of the United Nations (for Africa, Latin America, Asia and the Pacific, and South West Asia), the World Health Organization World Health Statistics Annuals, the WHO Summaries of Vital and Health Statistics for different regions, the World Bank World Development Reports and the World Bank World Tables.

revise ("adjust") figures that have been collected. Apart from such defects of individual figures, the "indicators" that are used to measure development may be defective because they reflect conditions other than or additional to what they are intended to measure or because they do not meaningfully apply to non-industrialized societies, or are used with different operational definitions in different countries.

International statisticians (and national statisticians dealing with international data) are not unaware of the defects in many of the data. Seriously defective figures are nonetheless issued by international organizations, partly because this is felt necessary to encourage statistical activity in countries where it is being built up, and it is difficult for an international secretariat to refuse to publish figures submitted by governments in response to request. However, the international publications originally issuing such statistics usually provide a commentary on the scope, meaning and comparability of the figures. They often evaluate the individual statistics in the tables as "complete" or "incomplete" in coverage of the population and provide footnotes as appropriate - e.g., on figures known to be based on "estimates", or to be collected by methods other than the standard method,or to refer to years different from the years of reference for other countries, or to use definitions different from the standard definition.

Analysts who employ the statistics, however, often overlook these commentaries and notes, or read them and forget them. In many instances, the analyst relies on statistics from secondary sources that do not reproduce the commentaries and footnotes of the original compilations, or reproduce them in inconspicuous ways. In any case it is an unfortunate fact that statistics from international compilations very often are mechanically poured into computer programmes today to get trends, correlations and regressions or more complex forms of multi-variate analysis, with little heed to their quality (or even to the elementary problem that they may refer to widely different years). Results are published with at most a perfunctory warning although they are quite often subjected scrupulously to tests of statistical "significance" which state on probability grounds the degree of confidence one can have that the relationships found between the variables are valid - but which take no account of the degree of confidence (often approaching zero) one can have in the data, or in the suitability to the data of the methods of analysis employed.

It is only the occasional scholar who undertakes a serious evaluation of the international data he or she uses. Most seem to feel it is not their responsibility or within their competence: if a figure has been published, it is usable (even if there is a footnote casting doubt on it). For many authors (and editors, book reviewers and Ph.D committees) the important thing seems to be to locate statistics for the largest possible number of countries and subject them to the latest most sophisticated methods of analysis. Such studies may demonstrate technical proficiency and confer status but the results cannot be taken seriously. Thus, a recent publication based on development data for a very large number of countries and using a complex combination of factor analysis, principal components analysis and path analysis comes out with the finding that education is not an important component of socioeconomic development. From examination of the kinds of data used it is evident that this finding is simply the result of use of bad educational data. It would be unfortunate if development policy-makers of the Third World revised their plans and policies on the basis of such studies.

In justification of bad data it is sometimes argued that the figures used are the best available and scientific progress cannot be made without data. In practice, better indicators and better statistics can often be found; or existing data can be improved by screening and adjusment to make them comparable (as in time reference). It is true that blanket exclusion of suspect data can, under some circumstances, produce misleading results, particularly when in effect it means exclusion of developing countries. For these countries the relationships between variables may be quite different from those found for developed countries. The answer to this problem is to choose indicators for which relatively reliable and comparable data for a substantial number of developing as well as developed countries are available or can be derived. Otherwise an analysis proposing to cover developing countries or the generality of countries should not be attempted. It is better to omit than to mislead.

ILLUSTRATIONS OF DEFECTIVE INDICATORS IN CURRENT USAGE

Some economic and social statistical series, however important theoretically or useful within countries, are obviously deficient in international coverage, or the available data are so obviously lacking in comparability, that very few even of the more uninhibited researchers venture to use them in large-scale comparative studies of development - although in some cases they may be quite properly used in the limited context of comparative study of a selected small set of countries with comparable data. These statistics include such important items as land distribution, social security, human freedoms, over-all environmental pollution. Other statistical measures, more widely used, have been severely criticized in print. 1/ The following pages contain notes on still other items that are quite widely used today in relatively serious studies and often put forward

---

1/ Notably by Gunnar Myrdal. The reader is referred to his volume entitled Against the Stream, The MacMillan Press Ltd., London, 1974.

The following critique of the use of unemployment in comparative measurement is cited from Myrdal's Asian Drama, Vol. II, Pantheon, New York, 1968, pages 1019-1021.

"Most of the official statistical inquiries dealing with the labor force have posed the wrong questions. This is the basic - though not the only - weakness in data collection to which we wish to draw attention. On labor force matters, the main statistical effort in these countries has been devoted to the compilation of data on 'unemployment' and its appendage 'underemployment'....

"For the reasons spelled out in this chapter, this approach fails to come to grips with the essential realities. On the contrary, it tends to obscure them. Conditions in South Asia are so different from those in the Western countries where the modern approach originated that it is unreasonable to expect the same questions to be relevant in both contexts.
....

"The basic criticism, therefore, is that when the collection of empirical data is guided by the modern concept of unemployment, the questions raised are unreal, and consequently irrelevant. They do not illuminate the essential realities of the economies in question, which is, after all, the purpose of empirical inquiries. Instead, they have the unintended effect of camouflaging underutilization of labor.......

"...In Western countries, estimates of the volume of unemployment are generally prepared from two primary sources of information - unemployment compensation rolls and sample surveys of the population of working age. In a Western context, these techniques yield reasonably valid measures of unemployment, though, of course, identical definitions are not used in all countries. Similar procedures are far from being even approximately reliable when transferred to South Asia."

in lists of key indicators of development, but that are fundamentally defective and unacceptable for comparative analysis. They are chosen to illustrate particular kinds of difficulties.

Crude death rate, which is widely published, has long been known to demographers and health experts to be a very poor international indicator of health, although, remarkably, it is still being used. Its defect is characteristic of a number of other unacceptable indicators of development: it does not simply measure what it is intended to measure (in this case, extent of bad health in a country) but also reflects to an important extent something quite different and unintended (age structure and fertility level). Countries with high health levels usually have low birth rates and a large percentage of elderly people, whose mortality level is inevitably high, raising the crude mortality rate. Thus the country with the highest life expectation and lowest infant mortality rate of the world in 1970, namely Sweden, ranked only 34th in crude mortality rate out of 78 countries (behind a dozen or more developing countries having high birth rates in the recent past, a youthful population and fair health levels).

Primary school enrolment, as a percentage of the official primary school age-group in a population, is an inferior indicator for comparative measurement and analysis which is unfortunately used even by education experts and endorsed by international bodies. As an international indicator it has two basic weaknesses:

a) Countries differ significantly in length of primary schooling, the range being from three to ten years, depending on the school system (availability and duration of secondary education, etc.). The most common official primary school duration in 1970 was six years but nearly forty per cent of developing countries had longer or shorter durations than that. A short duration of primary schooling tends to increase the primary enrolment index, a long duration tends to decrease it (presumably in part because of more drop-outs among the older children), undermining comparability. This effect can be clearly seen when individual countries shorten or lengthen their primary school duration (about a third of the developing countries made one or more such changes during the period of 1960-1974 - see UNESCO Statistical Yearbook 1976). Enrolment ratios tend to go down (or to be held down) when durations are lengthened and to go up when they are shortened. 1/ Table 2.1 gives illustrations.

Table 2.1. Examples of Change of Primary School Enrolment Ratio with Change of Primary School Duration

Country	Year	Official age range of children	Duration in years	Enrolment ratio
Malawi	1960	7-11	5	63
	1965	7-14	8	44
Brazil	1960	7-10	4	95
	1965	6-11	6	72
Thailand	1960	7-10	4	136
	1965	7-13	7	78
Jamaica	1965	6-10	5	97
	1970	6-11	6	85
	1973	6-10	5	109
Morocco	1960	6-10	5	47
	1965	6-13	8	39
	1970	7-11	5	52
Iran	1965	7-14	8	49
	1970	6-10	5	83
	1973	6-11	6	69
	1974	6-10	5	90

Source: UNESCO Statistical Yearbook 1976

b) The primary school enrolment ratio, as available, is a gross ratio which relates the number of children of whatever age enrolled in primary school to the size of the official primary school age-group. Many enrolled children are older or younger than the official age range because they started late (or early in some cases), are repeaters, are staying on to study for examinations for primary school certificates, are older students trying to achieve literacy, etc., depending on conditions that vary from country to country. As secondary

1/ The general trend in primary enrolment rates is upwards in developing countries, but half of the cases where duration is lengthened show a significant downwards change in the ratio as a consequence, with most of the remainder held relatively stable (inhibition of increase). When duration is shortened, the ratios increase almost without exception. The primary enrolment ratio is obviously unsatisfactory for measurement of progress within countries that have changed durations, as well as cross-nationally.

schools become generally available and educational systems develop, gross enrolment ratios decline and differences generally decrease between gross and net enrolment figures (net enrolment is the percentage of children actually of primary school age who are enrolled - a small minority of countries have both gross and net enrolment figures). The highest gross enrolment figures (over 120% up to 200% or more) are found today (1975 or later) almost exclusively in developing countries, not in developed countries where gross enrolment ratios have been mostly declining in recent years. In developing countries the gross ratio may be twice the net ratio or almost equal to it or anything in between, and it may increase when the net ratio decreases over time or decrease when it increases. Conclusions about the relation of education to other factors of development dependent on use of this indicator can be quite misleading. 1/

Per cent of population in urban areas lacks international comparability because there is no internationally standardized definition of "urban". Each country has its own operational definition, and the differences between definitions are large. 2/ Thus in Europe the minimum size of localities classified as "urban" ranges from built-up places with 200 inhabitants (Sweden) to communes of 10,000 (Switzerland). It is not surprising that Sweden is recorded in the United Nations Demographic Yearbooks in 1970 as 81.1 per cent urban by this indicator while Switzerland is only 54.6 per cent urban. In 1960, by the 20,000 criterion, Sweden was 40.4 per cent urban. Austria was nearly the same (38.2 per cent urban). While Sweden rose to 81.1 per cent in 1970 with the new definition, Austria, however, rose only to 51.5 per cent in 1970, using a 1970 definition of urban in terms of "communes of more than 5,000 inhabitants".

Per cent of dwellings with piped water is in principle one of the better social development indicators, but the statistics that have been used in analytic work covering the year 1970 - taken directly or indirectly from the United Nations 1974 Statistical Yearbook - are wholly unsuitable for comparative

---

1/ The problem is not solved but made worse by relating available gross primary school enrolment figures not to the size of the official age group but to that of a fixed age-group, such as the 6-11 or 5-14 group, for every country. Combined enrolment in primary and secondary school as a percentage of children aged 5-19 is a much better indicator and gives higher correlations with other development indicators in developing countries.

2/ In the past, international publications have used a criterion of 20,000 population or more but this was criticized for various reasons and the governments' own definitions were then used. (See UNRISD, Research Data Bank, Vol. IV, pp. 57-66). Attempts are being made by the United Nations Statistical Office to improve the present situation.

analysis. As that Yearbook indicates it in its footnotes, while most of the figures refer only to the per cent of dwellings with water inside the house, one-third of the relevant figures for developing countries refer to the per cent of dwellings with water either inside or outside; also, while most refer to the entire country, a number of them refer to urban areas only. It is not possible to compare, for example, a figure of 36 per cent of dwellings without piped water inside or outside in selected urban areas in Sudan (which had a total urban population equal to only 12 per cent of the national population) with 77 per cent without piped water inside the dwelling for the total population of Algeria. Yet correlations using such non-comparable figures have been published by distinguished scholars. Through a certain amount of research on sources, it was possible for UNRISD to screen and adjust figures for this indicator so as to obtain a relatively comparable set, but the number of countries, especially developing countries, for which comparable figures could be obtained was unfortunately too small to justify general use of the indicator for analytic purposes. 1/

Income distribution statistics have been used increasingly in recent years in comparative studies of development. A review of these statistics has led to the conclusion that "there does not now exist an adequate conceptual and technical basis for common international measurement of income distribution and cross-national comparisons of it". 2/ There are wide differences between countries in what is counted as personal or household income in the data sources used (primarily income tax records, household budget enquiries or censuses) with varying treatment, for example, of public and private transfers and of imputation of monetary values to goods and services received as income in kind. There is no agreed-upon standard definition of the income-receiving unit (individual earner, tax paying unit, household as a unit, household per capita, age-adjusted household per capita, etc.). Depending on a country's

---

1/ Making use of fuller information received from the 1970 census round, the United Nations Statistical Office has recently issued a more complete source of information on this indicator, in the Compendium of Housing Statistics, published in 1980; but the number of countries with comparable data for 1970 is still too small for many analytic purposes.

2/ UNRISD report no. 79.6, International Comparability of Statistics on Income Distribution, Geneva, 1979.

structural characteristics (size and composition of households, rural-urban structure, structure of economic production) and on its distributional and welfare system, conspicuous and even radical changes of the position of a country in relation to other countries may be seen on a scale of equality-inequality of income when different operational definitions are used for income and for the income-receiving unit. The present situation does not justify the issuance of international tables giving comparative data for countries in terms of over-all indices of equality/inequality, nor the use of income distribution indices in analytic work other than in very limited and carefully defined comparisons.

Transportation is an example of a component of development where there are specific indicators but no common measure, and use of one or other specific measure in comparative analytic work would be misleading. Different countries rely to quite different degrees on different modes of transportation (trains, automobiles, buses, boats, airplanes), depending partly on choice, partly on necessity. There is no satisfactory way, given present data, of comparing measurements of these different modes and converting them into a common standard.1/ The situation is complicated by the fact that different types of country may require both different amounts of transportation per capita and different kinds of transportation. Geographically large countries require large amounts of domestic transport, small countries require much less per capita. Countries consisting of islands or archipelagos and countries with extensive shorelines and waterways need relatively more boats, countries with very high mountains and jungles to be crossed need relatively more airplanes, etc.

The situation regarding tansportation is quite different from that in communications where it is possible to use different indicators like relative number of telephones or of television sets, because they are not to any significant degree competitive alternatives but tend to go together; absence of data on one or the other of these indicators is not likely to give a distorted

---

1/ For example, Japan in 1970 reported 2762 railway passenger-kilometres per capita, Canada only 172. But Canada had 310 passenger cars per 1000 population, Japan only 84. Assuming no other transport, which had the higher overall level of development of transportation by what amount?

impression of the overall level of communications. 1/ It is theoretically not impossible to work out an internationally applicable "adequacy scale" for transportation but this has not yet been done.

Energy consumption, like transport, appears in alternative and somewhat competitive forms in different countries, one country using primarily oil, another coal, a third relying heavily on hydro electricity, etc. Unlike transport, however, equivalences between different forms have been established in energy in terms of the amount of each category that is required to create a given amount of heat in water under optimal laboratory conditions. 2/ Thus 1000 kilowatt hours of electricity is equated to .123 metric tons of coal (until recently .125 metric tons was used), each producing the same amount of heat of water.

By this means the various forms of energy consumed in a country are converted to a common standard of coal equivalency and combined into a national aggregate, leading to the indicator: consumption of energy per capita in kilograms of coal equivalent. This is widely used as an international indicator, based mostly on the regularly issued United Nations statistical reports on World Energy Supplies, which include a table giving aggregate and per capita "commercial" energy consumption data for over 100 countries with population of one million or more. 3/ "Consumption" - or, more precisely, "apparent inland consumption" - is determined in this computation by measuring first stage domestic production of the primary commercial energy sources (coal, lignite and brown coal, peat in some countries, crude petroleum and natural gas liquids, natural gas, and hydro, nuclear and geothermal electricity), then adding imports, subtracting exports and bunkers and adjusting for changes in stock.

The resulting indicator, while valuable for some purposes, unfortunately has certain weaknesses from the point of view of comparative measurement and analysis of socioeconomic development when the interest is in the amount of energy consumed in a country for final use (heat, light, electronics, mechanical movement). There are problems arising from the limitation of coverage (exclusion of fuelwood, charcoal and animal and crop residues, which are widely used as household fuels in developing areas but on which data are not regularly collected) and from the fact that the energy values of different sources in

1/ The correlation between telephones and television receivers is 0.89 (UNRISD data); between railway passenger kilometres per capita and number of passenger cars per 1000 population it is 0.43.

2/ See, for example, World Energy Supplies, 1950-1974, United Nations Publications no. E.76.XVII.5, New York, 1976, page xviii.

3/ Beginning with the 1979 issue of World Energy Supplies 1973-78, energy consumption data are also given in terms of megajoules per capita and oil equivalency.

producing heat under ideal laboratory conditions may be rather remote from their relative utility and efficiency in practical matters (as in moving trains, automobiles, small boats). Most important, the per capita energy consumption measure obtained in the manner described above has the defect that it does not, for the most part, represent final consumption, or even a comparable stage of consumption for all sources. Some energy items, such as hydro or nuclear electricity, or imported electricity, gasolene and kerosene, are much closer to final consumption than others, such as coal that is used to create thermal electricity (at the inefficient rate of about .6 metric tons of coal for 1000 kilowatt hours of electricity production). The energy value of coal at first stage production (as produced from the mines) is counted in the aggregate national energy figure, not the electrical energy generated from coal, while the energy of the electricity generated by hydro or nuclear power is counted in the aggregate, not theenergy of hydro or nuclear power consumed in creating this electricity. The national per capita energy consumption figures can thus be misleading, particularly when countries that rely heavily on hydro electric power, such as Norway, Sweden, Switzerland, New Zealand and Ghana, are compared with the countries that have relatively little hydro power and rely mainly on domestic coal to produce their electricity, with loss of three-fourths to two-thirds of the energy in the process. The hydro power countries tend to appear unduly low in energy consumption for their general level of development, compared with other countries.

Since available data do not permit aggregation of the different forms of national energy use in terms of final consumption, it has been judged that, from the point of view of the kinds of comparative measurement and analysis of concern to UNRISD and to other research units of similar interests, the best way under the circumstances to correct the distortions that arise, for example, from attributing (in effect) a much higher national energy figure to thermal electricity (from coal or oil) than to the same amount of electricity generated from hydro or nuclear power, is to return to the older method of international measurement of energy (prevailing prior to 1955). That means converting hydro, nuclear and geo-thermal electricity into coal equivalent by computing the amount of coal that would be required to produce the hydro, nuclear or geothermal electricity (that is using the .6 conversion rate rather than the .123 conversion rate). When this is done, countries like Norway no longer appear significantly out of line (sub-normal) in their energy consumption, and the correlations of energy consumption per capita with GDP per capita and with other development indicators are increased (in particular, for developed countries). 1/

A number of fairly widely used "development indicators" concern expenditure on a given item as per cent of the GDP or related national accounts figure or of the government budget or household budget: for example, expenditure on food as per cent of total private consumption expenditure; expenditure on education or health as per cent of total government expenditure or as per cent of GDP. Data on such items, when available, generally lack international comparability and are suitable for use only in very limited circumstances.

Expenditure on food as a per cent of total private consumption expenditure has long been assumed to be a valid measure of the poverty of a nation and has strong theoretical arguments behind it, but in practice there are very few data for developing countries and the figures that do exist must be regarded with considerable suspicion - primarily because of the problem of collecting reliable and meaningful figures on food expenditure as a per cent of total expenditure on a nationwide basis in developing countries, especially figures for rural areas where much or most of the food consumed (along with other goods and services) is self-produced and "imputations" of monetary expenditure (a highly uncertain procedure) have to be made. In practice, food is often the only non-purchased item of household consumption in subsistence or semi-subsistence societies for which monetary imputations are made - household or family provisioning in housing, fuel, clothing, furniture and other homecraft products, as well as administrative, recreational, harvest labour, health, child-care and other services provided by community or kin group, along with barters and exchanges, all tend to be overlooked, with the result that food expenditure necessarily absorbs nearly all of the imputed household or personal expenditure in mainly self-provisioning societies. Furthermore, estimates of the monetary value of the food consumption (and of other goods and services) are supposed to be made on an annual basis, but people have difficulty in

1/ The published volumes of the UNRISD 1970 Data Bank do not include the adjustment under discussion but it has been incorporated in the Institute's computerized data bank and used in the analytic work presented in this report. Sometimes the efficiency of a country's use of energy is computed by relating aggregate energy consumption to GDP. This picture is considerably influenced by the choice of aggregate energy indicator when certain countries are compared.

censuses and household surveys/in recalling how much they consumed of their own produce a few days ago, to say nothing of consumption over a whole year; the last few days cannot be considered representative of the year, so that collection of adequate household consumption data for a country as a whole for the period of a year can become extremely difficult and expensive.

There is no reason to believe that estimates of personal consumption of food obtained from national accounts estimates of the value of total food production minus the value of marketed food will give any better results for this indicator than household consumption surveys.

Statistics on expenditure on education or health as a percentage of the national budget or of the GDP (along with per capita expenditures on education and health) tend to lack comparability because of wide national differences in defining and counting educational expenditure and health expenditure. Items classified under "education" or "health" by one country are not so classified by another country. Interlacing national, regional and local expenditures , along with private expenditures, confuse the picture, as do the different forms of expenditure, such as capital expenditure (school construction and hospital construction are often recorded as expenditure of the department of public works, not as education or health expenditures), current expenditure (which also tends to be classified by ministry or agency rather than purpose or function), tax exemption (usually overlooked), and private expenditure and voluntary contributions (also overlooked and very difficult to determine). An examination of this subject some 20 years/ago concluded that "social expenditure statistics are relatively very inadequate", 1/ the picture does not seem to have changed greatly since that time.

Per cent of budgetary of GDP expenditure on education has little or no correlation with educational level as measured by level of literacy or schooling. This points to a conceptual flaw in the indicator: one should not in fact expect a poor country that spends a large proportion of its budget or GDP on education to have as high a level of education as a much richer country that spends a smaller proportion but much more in absolute terms. Percentage of

1/ See United Nations, Report on the World Social Situation, 1961, Chapter IV, Expenditures for Social Purposes, page 81. Evidence is cited of about 50 per cent understatement of social expenditure in a developing country because of non-functional classification of government expenditures.

expenditure on education, health or other social service, can be expected to be reflected in the corresponding educational level or health level, etc. only if the countries compared have about the same budgetary or national income level per capita. Otherwise it reflects effort rather than level.

Numerous other examples could be cited of defective and unusable development indicators. 1/ Underlying reasons for indicator inadequacy and unacceptability appear to be mainly the following: transfer from industrialized to non-industrialized countries of concepts that are not suitable to the conditions of the latter countries (e.g. unemployment); influence upon and reflection by the indicator of factors or conditions other than what it is intended to measure (e.g. crude death rate, gross primary enrolment rate) lack of clear and common operational definition of what is being measured (e.g. per cent urbanization, income distribution); impossibility with present data of conversion of alternative modes or manifestations into a common standard (e.g. transportation), or difficulties in conversion (e.g. energy); conceptual and practical weaknesses of "imputation" procedures used to assign monetary values to non-marketed goods and services in subsistence or semi-subsistence societies (e.g. expenditure on food as per cent of total private expenditure); attempts to measure development level in a particular field by indicators suitable only to measure relative effort or input (e.g. expenditure on education or health as a per cent of government budget or GDP). These reasons overlap and a particular indicator may be deficient in several respects.

In some cases development indicators that are inadequate internationally can be used nationally with benefit, or in comparisons of small groups of carefully selected countries. But often the difficulties that render an indicator internationally inadequate will also make it inadequate for time comparisons of a single country, or for comparing small groups of countries or regions within a country.

As noted above, there are no perfect development indicators. Those that

---

1/ The UNRISD volume Notes on Indicators (UNRISD Report no. 77.2, 1977) discusses the weaknesses of a number of indicators, including some of those discussed above (as well as weaknesses in indicators accepted in the present study and listed below).

are used in the present study (the "core" indicators) have weaknesses that can be readily noted. But there are degrees of imperfection and a line must be drawn between the usable and the non-usable. No doubt, those who are seriously concerned with questions of indicator quality may differ on the issue of where exactly this line should be drawn, but they would no doubt agree that it should exclude a great many of the indicators that are in current use, including many of the ones discussed above.

INDIVIDUAL DATA DEFICIENCIES

An indicator that may be accepted as usable on conceptual and technical grounds may not cover a sufficient number of countries to be usable for analytic purposes. It is necessary for such purposes to have a proper sample not only of countries in general but also of certain major categories of countries, in particular developing countries separately considered since they are of special interest. Fairly often, there are enough countries having data on a conceptually and technically acceptable indicator but some of the data are quite unreliable or otherwise non-comparable. Some kind of screening process is required, also a process of adjusting data to make them comparable where that is feasible, especially in relation to time reference. Unfortunately, the screening process often eliminates data from so many countries that it makes the indicator unusable.

The need for examination and screening of individual figures in the case of even simple and well-known statistical series is suggested by the existence of large discrepancies between figures from different sources for the same countries.

Birth rate data

Table 2.2 gives birth rates for selected countries of Asia taken from a published table compiled from a number of different sources, including United Nations sources, World Bank sources, and United States A.I.D. (Agency for International Development) sources. It is obvious that some of these figures must be wrong.

Table 2.2: Estimated Birth Rates in Selected Countries of Asia with over 10 Million Inhabitants, c. 1975

Country	United Nations	World Bank	U.S.A.I.D.
Afghanistan	49	51	43
Burma	39	34	40
China	26	26	14
Malaysia	38	31	35
Philippines	43	36	35
Thailand	43	34	31

Source: Dudley Kirk, "World population and birth rates: agreements and disagreements", Population and Development Review, 5, no. 3 (September 1979), pages 387-403. The data, while coming from the organizations identified, are not necessarily official data.

Infant mortality rate data

Differences in estimates of infant mortality rate are equally striking. Of 18 Latin American (mainland) countries with population of 1 million or more, only five are listed in the United Nations Demographic Yearbooks (1973 and 1974 issues) as having relatively complete (and presumably reliable) statistics on infant mortality for 1970. A publication entitled Indicators on the situation of children in Latin America and the Caribbean, issued in 1979 jointly by UNICEF and the United Nations Economic Commission for Latin America (ECLA), however, provides data on 14 of these countries. Table 2.3 shows the 1970 infant mortality rates of the five countries having acceptable statistics according to the Demographic Yearbooks and the figures for the same five countries as given by the UNICEF/ECLA volume for the same year.

Table 2.3. Infant mortality rate, 1970
(per 1,000 live births)

Country	United Nations Demographic Yearbook	UNICEF/ECLA volume
Chile	79	71
Costa Rica	62	66
El Salvador	67	116
Guatemala	87	120
Uruguay	43	54

The striking differences may be due to reliance on different primary sources or to the fact that one agency has accepted figures arising from registration (or other source) while the other agency has made "adjustments" or corrections on the basis of a model relating infant mortality to other demographic variables. (Such adjustments are common. An infant mortality rate in an African country was found to be 80 per 1,000 births on the basis of a sample survey finding, but this was adjusted and issued as 200.) In India, serious differences have been found between the infant mortality rates given by the National Sample Survey and those given by the Sample Registration System. The discrepancies are attributed, in a recent technical report, to underestimates by the National Sample Survey, caused by "underreporting of the events by the respondents owing to memory lapses and failure to place correctly the occurrence of events in relation to the moving reference period of one year".1/

Statistics on infant mortality - which are issued in a very large number of international, regional and national publications - can be questionable in specific cases not only because of failure of respondents in censuses and

1/ N.S. Sastry, "Household surveys in India: Quality of data collected and their usefulness for planning and policy purposes", Development Centre, Organization for Economic Co-operation and Development (OECD), Working Document No. 16, for study session 14-18 November 1977 on Multipurpose Household Surveys in Developing Countries.

surveys to give correct answers (or to be asked the right questions) and failure of people to report infant deaths in civil registration systems, but also because infant death is, statistically, a "rare item", with a high degree of variance, which calls for a very large sample in household surveys (and a fair-sized population in censuses and registration systems).1/ Household sample surveys, which are frequently used in developing areas without registration systems to get infant mortality rates face the problem that the households in which an infant birth occurs during a given year cannot ordinarily be pre-selected for sampling. At the same time, what is needed to determine an infant mortality rate through a national household sample survey is, in the end, a representative sample of the infants born in the year, not a representative sample of households, especially where, as commonly the case, there are marked differences in frequency of births as between rural and urban households. This calls for special sampling arrangements.

Table 2.4 illustrates the surprisingly large amount of random fluctuation of infant mortality rates that can be found in smallish countries and populations where the data on births and deaths are judged complete and reliable. For these same populations, birth rates remain quite constant or show a constant trend. Even Luxembourg, with a population of 360,000 (equal to about 116,000 households) shows considerable fluctuation in infant mortality rate, although nothing compared to the fluctuation of the population of 10,000 Asiatics in Southern Rhodesia (now Zimbabwe) in the early 1970s. This shows the need for caution in the use in analytic work of infant mortality statistics taken for a particular year from smallish countries or subregions of countries, as well

---

1/ An infant mortality rate of 100 deaths per 1000 live births in a community of 1000 population with a crude birth rate of 40 per thousand population would mean, in theory, four infant deaths per year in the community, but in practice the actual number in a given year could fluctuate from 0 to 8 or more deaths, giving wildly different annual rates; the "true rate" will not be revealed for a given year by averaging these different annual rates over a few years. According to theoretical probability estimates, a sample as large as 100,000 households (which becomes extremely expensive) may be needed in order to get a proper infant mortality rate, depending on various conditions and assumptions including levels of fertility and child mortality (high levels reduce the required sample size). However, the random variance in infant mortality may be greater than is theoretically assumed in such estimates, suggesting the need for still larger samples.

as from smallish national samples.

Table 2.4 Fluctuation of infant mortality rates in small countries and populations

Country	Estimated population (1975)	1973	1974	1975	1976	1977
Cape Verde Islands	300,000	111(29)	79(29)	105(28)	--	--
Southern Rhodesia (now Zimbabwe) Asiatic population	10,000	25(21)	14(21)	63(21)	38(21)	--
Antigua	70,000	12(18)	31(18)	38(19)	--	24(20)
Bahamas	200,000	31(23)	26(22)	35(20)	25(25)	--
Faroe Islands	40,000	14(21)	6(20)	13(19)	10(18)	--
Greenland	50,000	39(19)	32(18)	39(18)	52(17)	--
Luxembourg	360,000	14(11)	12(11)	14(11)	18(11)	11(11)

Birth rates are given in parenthesis after the infant mortality rates.

Source: 1977 and 1978 United Nations Demographic Yearbooks. In some cases where rates are not given directly, they have been derived from figures on live births and infant deaths reported in those volumes.

Life expectation data

Since life expectation - another very popular social indicator - is constructed in principle out of age-specific mortality data, including infant mortality, similar questions arise about the reliability of some of the published statistics for this indicator, which is conceptually perhaps the best of the health indicators. There is less contradiction between rates published by different sources in this case, however, than in the case of infant mortality rates. The reason is that many of the published figures on life expectation - which is a complexly constructed variable for which the required data are often missing - are essentially "estimates" derived by the United Nations Population Division from "models". The models make various assumptions about the relation of life expectation to other variables (for which data may be available) and about changes over time. These estimates in some cases are constructed on the basis of limited available information, and while highly useful for giving a best guess about demographic conditions and trends in the areas concerned, are not primarily designed for use in comparative analytic work and can yield misleading results. 1/ The observational data may not be only thin but also remote. For example, there are a number of countries in Africa for which the latest observational source of information (as available in 1978-79) on national demographic conditions is a census or sample survey conducted prior to 1965: Benin (sample survey, 1961), Chad (sample survey, 1964), Equatorial Guinea (census, 1960), Gabon (census, 1961), Guinea (sample survey, 1955), Niger (sample survey, 1960), Nigeria (census, 1963), Zaire (sample survey, 1958). In all of these cases, the United Nations Population Division has made rough life expectation estimates for the period 1965-70 and the period 1970-75 (along with estimates for the same two periods for 36 other African countries, most of them without reliable age-specific mortality rates). 2/ It is the average of the estimates for these two periods that tends to be picked up and published by other organizations. 3/

Literacy data

Literacy would seem a fairly straightforward and simple indicator but figures for particular countries are often unreliable or non-comparable for one or other reason. Statistics for most countries cover literacy of the population 15 years or more of age, but 5+, 6+, 10+, 13+ or 14+ years of age may be used in the national statistics of some countries, also ranges such as 9-49 and 9-50 years as in U.S.S.R. and Mongolia. All of these other uses tend to inflate the

---

1/ Thus if the model assumes a high (negative) correlation between fertility rate and life expectation in making estimates of life expectation, then the data derived from it will spuriously demonstrate a high (negative) statistical correlation between fertility rate and life expectation.

2/ See United Nations, World Population Trends and Prospects by Country, 1950-2000: Summary report of the 1978 assessment, ST/ESA/SER.R/33, United Nations, New York, 1979.

3/ The UNRISD 1970 Data Bank also made use of the Population Division estimates by averaging 1965-1970 and 1970-1975 estimates (after tests showed little difference from known 1970 figures) but limited this usage to countries where there had been a presumably reliable nation-wide census or sample survey within no more than three or four years of 1970.

amount of literacy compared with the 15+ definition, except where literacy is practically complete by any criterion. Ability to read and write is the most common test of literacy but sometimes reading only or writing only is used. Countries with several languages may define literacy in terms of one of these languages only or any one of them. Most of the more highly developed countries no longer record literacy but simply assume it to be about 99 per cent. Literacy is usually derived from censuses (or sometimes sample surveys) which are taken at long intervals, commonly ten years, sometimes twenty or more years, and the results may not be available for as much as four or five years after. Statisticians compiling international literacy data face a difficult problem: they must either publish statistics that are not comparable to each other in years of reference, many being out of date 1/, or else try to make estimates for a common year such as 1970. This is particularly a problem where, as in many African countries, censuses are usually taken not at the beginning of each decade (1960, 1970, etc.) but some year in between, such as 1962 or 1963, or 1965. 2/

Unesco has carried out a study, reported in a 1978 publication entitled Estimates and Projections of Literacy, which gives 1970 literacy estimates for practically all countries. 3/ The difficulties that are faced in such computations (not counting difficulties arising from the use of regression, as discussed in Chapter 4) are demonstrated by the fact that the estimates of literacy that are produced have to be based in good part on figures that are themselves estimates, and these in turn are often based on figures which are also estimates, etc. In fact, there may be no previous census or survey that

1/ This is the policy generally followed by the Unesco Statistical Yearbooks, although some estimates are made.

2/ The United Nations 1977 Compendium of Social Statistics (published in 1980) contains a table (III.30), giving illiteracy statistics with dates of reference which indicates that for only five out of 37 African countries with a population of 1 million or more were data available from censuses or surveys covering years as late as 1968 16 countries had data referring to 1962 and the rest were scattered, some going back as far as 1950.

3/ The estimation procedure, relying on regression analysis, uses a complex formula involving: the percentage of literate aged 15-19 in the latest census or survey year and in 1970 (the 1970 figures are estimates derived from school enrolment 6-11 as of nine years previous) the population 15-19 in 1970 (often an estimated figure) divided by the population aged 55+ in the last census year (or previous estimate) and the time-lag between 1970 and the year of the last census or survey or previous estimate.

can be used in making the projection, just a previous estimate.

In the same year (1978) the World Bank issued a volume World Development 1978 which gives literacy figures for a large number of countries but for 1974, not 1970, based on World Bank and Unesco estimates. The United Nations 1977 Compendium of Social Statistics, in addition to the table cited above, also contains a table (Table II.1) giving literacy (or more precisely illiteracy) figures for around 1970 for a smaller number of countries.

The figures from these sources are generally consistent when a census or survey result for around 1970 covering the population aged 15+ is available. But when estimates are made on the basis of older censuses or surveys (or previous estimates) or when the census or survey data are given for an age-group other than 15+ population, then inconsistencies often appear. Figures are shown in Table 2.5 for a selected group of ten countries. 1/

Per capita GDP data

The per capita GDP (or GNP) has been criticized severely in recent years particularly when used as a general measure of development, presumed to cover social as well as economic aspects. 2/ There are problems in general of currency conversion rates, inflationary trends, treatment of "statistical discrepancies", and imputation procedures. The national accounts indicators face a number of special difficulties when comparisons are attempted between countries of different socioeconomic structures (developed market economies, socialist economies, developing semi-subsistence economies). 3/ Conversions of the "material product" calculated by Eastern European countries into national accounts figures comparable to those of market economies is precarious and controversial. Adjustments for comparability are necessary to deal with the fact that some countries reported their 1970 national income figures according to the old system of national accounts, others according to the new revised system recommended internationally some years prior to 1970. The practice of deriving private consumption expenditure in developing countries as a residual after subtracting other expenditures from the total GDP is questionable in view of discrepancies with other data (household surveys of private consumption) and is related to the problems of imputing income to subsistence goods and services.

1/ Since the Bank volume cited gives 1974 estimates, some differences are to be expected, but some of the differences shown in Table 2.4 are difficult to understand, especially when the literacy figures for 1974 are given as substantially lower than those for 1970.

2/ See discussion in UNRISD Research Data Bank of Development Indicators, Volume IV, Notes on Indicators, Geneva, 1977, pages 26-29.

3/ See, for example, the criticisms in Gunnar Myrdal, Against the Stream, MacMillan, London, 1974.

Table 2.5: Estimates of Literacy of Population 15 years or more for Selected Countries in 1970, 1974 or thereabout

Country, and year of last information (a)	Unesco source estimate for 1970 or thereabout	World Bank source estimate for 1974 or thereabout	United Nations source estimate for 1970 or thereabout
Burundi (b) 1962	18	10	...
Cameroon (b) 1962	34	12	...
Ghana (c) 1970	30	25	43
Rwanda (d) 1962	34	23	...
Somalia (e) 1962	4	50	...
Tanzania (f) 1967	37	63	28
Tunisia (g) 1975 (1966)	32	55	24
Zaire (b) 1962	43	15	...
Iran (h) 1971	30	50	37
Yemen (P.D.R.) (i) 1973	27	10	27

Unesco source: Estimates and Projections of Illiteracy, 1978, table 3.

World Bank source: World Development Report 1978, table 18.

United Nations source: 1977 Compendium of Social Statistics, published in 1980, table II.1.

Note: For these particular countries there are no explanatory or qualifying footnotes in the sources cited.

(a) The year of last information (census, survey, previous estimate) is taken from the Unesco Statistical Yearbooks 1977, and 1978-79, table 1.3.

(b) Burundi, Cameroon and Zaire did not have a national census or survey in 1962, but a literacy estimate was made for that year, the basis for which is not evident. The Bank's World Development Report 1980 gives a literacy figure of 25 per cent for Burundi in 1975. The Bank's World Tables 1980 gives the figure of 12 per cent literacy for Cameroon in 1970 as well as in 1974.

(c) The 1970 census in Ghana gave a figure of 43 per cent literate for the population aged 6+. The Unesco and World Bank figures are presumably adjustments to cover the population aged 15+.

(d) The World Bank figure may possibly derive from a 1970 demographic survey which reported 23.9 per cent adult literacy in Rwanda.

(e) The relatively high World Bank figure of 50 is explained in a subsequent World Bank publication, World Tables 1980, as due to a 1974 literacy campaign. Presumably the figure emerged from the campaign, since there was no reported census or survey of literacy in 1974.

(f) Accordingto the 1967 census, literacy in that year in Tanzania was 28.1. The relatively high World Bank figure of 63 may reflect the results of a literacy campaign (or possibly a clerical error in using the Unesco 1970 estimates of 63 per cent illiterate without converting it into a literacy figure).

(g) The 1966 census of Tunisia gave a figure of 24 per cent literate, the 1975 census gave 38 per cent. The Unesco figure of 32 is presumably an interpolation between 24 and 38. The United Nations Compendium of Social Statistics lists the 1966 census but not that of 1975. The World Bank World Tables 1980 gives the figure of 24 per cent literate in Tunisia in 1970.

(h) The 1971 estimate for Iran gave a figure of 37 per cent literate for the population 6+ in years.

(i) The 1973 census in Democratic Yemen gave a figure of 27 per cent literate for the population aged 10+.

In view of these difficulties it is not surprising that GDP or GNP estimates for some countries, as made by different international, regional and national authorities, can differ by as much as 100 per cent.

Perhaps the major distortion in current national accounts statistics is the exaggeration of the difference between developing and developed countries, a result of several of the problems mentioned above. It is illustrated by the fact that it would be obviously impossible for a person to survive on $0.15 to $0.20 a day in a developed country today, although that is the reported per capita GDP in several countries (converted to a per diem figure). This exaggeration of differences, however, which is a scaling problem, is not too serious for analytic work that makes use of transformations of scale, in relating per capita GDP to other development indicators cross-nationally, as proposed in this study (Chapter 5).

All in all, our judgment has been that the per capita GDP can be retained for analytic work in comparative studies - provided the individual country data are examined carefully and provided this indicator is not given an importance it should not have and a role that it cannot fulfill as a measure of total progress.

Time comparability

Many tables containing development data, and analyses using such data, are misleading in that they compare values for one country with values for another country at a different time, or values for one indicator with values for another indicator at a different time. Sometimes there are differences of as much as ten or twenty years between figures for different countries or for different variables for the same country, which may or may not be pointed out in the statistical compilation, and if pointed out, may or may not be taken seriously by the analyst. The problem of time discrepancy is more serious for some indicators than others, since figures for some indicators change relatively little in a few years, others change rapidly. For example, demographic or economic structural variables, like per cent of children under 15 years of age in the population or per cent of adult male labour in agriculture, tend to change rather slowly while production and trade indicators and indicators of facilities, such as per cent of dwellings with piped water or telephones per 100,000 population, may change rapidly in a few years. (The per cent of dwellings in Poland with piped water changed from 39% to 55% between 1966 and 1970, telephones per 1000 population in Korea (Rep.) increased from 12 in 1966 to 21 in 1970.) Mortality figures also have been changing rapidly in some countries.

Data for comparative cross-national analysis should relate to the same year. The time reference of each published country value needs to be scrutinized in order to screen out or adjust non-comparable figures and avoid the distortions otherwise arising. The reason that published quantitative studies of development so often use data that are not comparable temporally is simply that the data often come that way - they are published as the "latest available" - and are put into the computer without adjustment. In a great many cases, however, it is possible to make simple interpolations or extrapolations to get reasonable figures for a desired common date (e.g., 1970). In other cases, it is best to drop the statistics with dates that are significantly out of line.

Adjustments for age structure

Many development indicators are of a per capita type in which the total amount produced, consumed or available in a given year of a given item (e.g. amount of gross domestic production in dollars, of energy consumption in kilograms of coal equivalent, number of hospital beds, number of births) is adjusted for population size by dividing it by the national population for that year. Countries differ considerably in the age structure of their population, however, and these demographic differences may affect the significance of measurement by the indicators in question. Children under 15 years of age should not be expected to produce or consume as much of various goods and services as mature adults, while those 65 years or more of age will also have different production capacities and consumption requirements.

This suggests the need for adjustments to per capita type indicators in order to take account of age structure along with total population size in measuring development. The adjustment or weighting process for this purpose is necessarily somewhat arbitrary. 1/ Use of age-adjusted data is important for comparative measurement, particularly for trend analysis. Without adjustment, an exaggerated impression of economic growth as measured by GDP per capita can be given by a decline in fertility rate or an increase in infant mortality rate. For most development indicators, use of age-adjusted data does not make a great deal of difference as far as correlations are concerned. It generally reduces correlations slightly (in part because of the effet of reducing differences between more developed and less developed countries - see below on the "elongation effect").

Age structure adjustment and fertility measurement

The greatest effect of adjustment for age structure is found in relation to fertility measurement in developing countries. 2/ The following examples show

---

1/ In work at UNRISD, three age-groups are used: 0-14, 15-64 and 65 and over; the relative weighting is changed according to the nature of the indicator. Thus for birth rate the respective weightings for the three groups are 0, 1, 0; for telephones 1/3, 1, 1; for food, energy consumption and GDP per capita they are 1/2, 1, 2/3. Where a particular indicator could be weighted in terms of either production capacity or consumption requirements of the different age groups, the use of consumption requirements was favoured. Indicator values adjusted for age-structure are given in the UNRISD Research Data Bank of Development Indicators, Volume II, Compilation of Indicators for 1970 with adjustments for age structure.

2/ A better adjustment than relating births to the size of the 15-64 year old population would no doubt be to relate the number of births to the size of the 15-49 year-old population, but tests showed little difference from use of the 15-64 range and data on the latter were more accessible. Adjustment can be made by use of more elaborate demographic fertility measures, such as the "female gross reproduction rate", which is related to females aged 15-44 - or sometimes 10-49 - and eliminates males from the picture. This latter measure has generally more effect in reducing correlations than does the simpler adjusted birth rate. The adjusted birth rate covers more countries, however, and is much less complex than the female gross reproduction rate (and various other elaborate fertility measures), which may be constructed in different ways in different countries and in practice involve a considerable amount of estimation. Thus the gross reproduction rate purports to describe the fertility experience of a generation of women but the rates presented in international tables are actually based on the fertility reported or estimated for a limited reference period, usually a single year or a five-year period. (See United Nations Demographic Yearbook 1975, pages 33-36.) Low correlations found with use of the gross reproduction rate thus may be due in part to lack of comparability in the data arising from the use of estimates and of different methods of construction.

the effect of adjustment for age structure on the correlation of birth rate with selected development indicators in the case of developing countries with available data. Correlations are clearly reduced.

	Crude birth rate	Adjusted birth rate	Female gross reproduction rate
Infant mortality rate	-.66	-.56	-.54
Life expectation	-.78	-.68	-.64
Combined school enrolment	-.70	-.62	-.58
Telephones per 1000 population	-.58	-.48	-.38
Foreign trade per capita	-.41	-.32	-.30
GDP per capita	-.58	-.46	-.41

Use of the crude birth rate can thus give a misleading impression of the extent to which fertility is associated with the level of GDP per capita and of other development indicators in developing countries. The basic problem is that the crude birth rate reflects mortality conditions as well as fertility, just as the crude death rate reflects fertility conditions as well as mortality. Because of high mortality, some of the least developed countries as in Africa south of the Sahara have a much smaller percentage of children and old people in their populations than do more advanced developing countries that have had better health and lower mortality of children and old people but high fertility, as in North Africa, Latin America, and elsewhere. The following comparative sets of dependency ratios (children under 15 plus persons 65 and over as per cent of population 15-64) will illustrate (1970 data):

Less advanced in health		More advanced in health	
Angola	79.5	Algeria	110.3
Benin	89.8	Iran	96.4
Burundi	83.0	Jamaica	110.5
Cameroon	79.6	Mexico	100.0
Cent. Afr. Rep.	82.4	Venezuela	98.9

The crude birth rate tends to exaggerate the fertility in least developed countries with relatively low dependency ratios, and to conceal the fertility in more advanced developing countries that have a very high percentage of non-procreative children and old people in their populations. With adjustments for age structure, most of the more advanced developing countries still have lower fertility levels after adjustment than do the least developed countries but the differences are reduced. In cases where the crude birth rates are initially about the same for individual countries in these two groups, the adjusted rates will tend to be greater for the more advanced developing countries. Data for two of the very poorest countries, Burundi and Ethiopia, may be compared in this respect with data for two more advanced developing countries, Algeria and Iran, which had about the same 1970 crude birth rates as Burundi and Ethiopia.

Country	Crude birth rate	Dependency ratio	Adjusted birth rate	Female gross reproduction rate
Burundi	48.0	83.0	87.8	3.1
Ethiopia	49.7	84.0	91.6	3.3
Algeria	49.1	110.3	103.3	3.5
Iran	47.9	96.4	94.1	3.4

The crude birth rate is an exception to the rule that technically inferior indicators tend to have lower correlations than technically superior indicators.

Other structural adjustments

There are other structural adjustments that in theory would be desirable but in practice are not feasible for lack of proper data. This applies, for example, to indicators involving measurement of income or expenditure, such as GDP per capita, personal consumption expenditure, or household income per adult equivalent household member. In practically all countries, recorded urban income per person is substantially higher than recorded rural farm income - usually up to two or three times as high in developing countries, to judge from the sparse data. Levels of living as measured by "real" indicators are also usually higher in urban than in rural areas. But even if average farm and non-farm levels of living in a country are the same, average farm income will tend to be recorded as lower, because of problems of imputation of money values to "income in kind" and other problems having to do with the way farm income is treated in income tax returns or household inquiries used for getting income data. 1/ The result is a bias towards underestimation of national or personal income in countries with relatively large rural farm populations. There is no way at present, however, of making appropriate adjustments of income to deal with this problem.

Countries that are relatively small in population and geographic size show significant differences from large countries with respect to several indicators. For example, at whatever level of development, they tend to have significantly more foreign trade per capita - large and variegated countries rely more on domestic trade between regions. In the present report, this matter is taken up under the discussion of "typological factors" in Chapter 5 where methods of adjustment are discussed.

1/ In one developed country (New Zealand) it has been estimated that a self-employed farmer on a pre-tax income of $2,465 has the same standard of living as an urban dweller on $4,065 per annum.

CHAPTER 3

SELECTION OF DEVELOPMENT INDICATORS AND SCREENING OF DATA

The present chapter outlines methods of selecting socioeconomic development indicators and screening data, as employed in the UNRISD study. It discusses in some detail the role of correlation in the selection of indicators.

DEFINITION OF DEVELOPMENT

The first question that arises in attempts to measure development through a set of indicators is what the term "development" should embrace. Definitions range from narrow conceptions of economic growth to broad conceptions that cover practically all good things - all economic, social, ethical, political and other goals and values that a society might pursue. The latter approach, which is in part an over-reaction to the first, can easily become offensive to countries that are prepared to be described as "developing" but not as lacking in all societal virtues. There is a tendency in many definitions of development, to project into the concept the professional and the idealogical interests and predilictions of those who do the defining. At the same time, there are schools of thought that declare, in the spirit of "cultural relativity", that the people of developing countries have quite different interests and values from those of the people of developed countries, their development goals are not and should not be made to be the same, and their progress should not be measured by the same indicators. Representatives of developing countries generally do not share this point of view, any more than they share the view that development is all good things. Some have remarked that it represents an imperialistic approach, serving to keep the people of developing countries in an inferior position.

Although particular means to development may vary from place to place, there is a universal final goal which few would dispute: improvement of the human condition; and there are universally recognized components of this condition (health, nutrition, education, housing, communications, security). There are also certain specific means to some of these conditions that are universal today: good water as a means to health, schooling as a means to education and learning, household income as a means to consumption. The first step in selecting international indicators is to identify the universally accepted goals and means. This is, in fact, not too difficult a task, except for some borderline cases.

In the present study, development has been understood, in the first instance, to mean essentially what the international community, as represented by the United Nations and specialized agencies, have agreed upon as the universal components of development. These are indicated in international declarations, programmes and reports, in the organizational structures set up and the kinds of work carried out under the heading of "development activities" or "development cooperation". So defined, development covers progress in a number of well-known fields, both economic and social: agriculture, industry, trade, transportation, energy, income and investment, health, nutrition, education, housing, social security, communications, science and technology. 1/

For the purpose of this study, in addition to indicators of the items listed above, a certain number of structural variables have been examined as possible empirical correlates (and in that sense indicators) of development: e.g., percentage of adult labour in agriculture, percentage of salaried and wage-earners in the economically active population, degree of urbanization, percentage of GDP derived from manufacturing, various indicators of demographic structure (especially age-structure and dependency ratios). Many of these were eliminated from the final selection but some that were highly correlated with value indicators and meaningfully related to development in a conceptual way were maintained. Thus, percentage of the adult male labour force engaged in agriculture was retained as a structural indicator because it is very highly correlated

---

1/ The social components of development overlap considerably with the recognized components of "levels of living". See United Nations, Report on International Definition and Measurement of Standards and Levels of Living, 1954 (Sales No.: 1954.IV.5) and International Definition and Measurement of Levels of Living, an Interim Guide, 1961 (Sales No.: 61.IV.7). The more recent concept of "basic needs" in practice covers much the same set of items.

(negatively) with indicators representing social or economic development values, and there are substantive reasons why the relative size of the agricultural labour force should decline with development.

It can be debated whether the concept of development should embrace such matters as (progress against) inflation, environmental pollution and crime, which are important for national welfare but do not necessarily distinguish developed from developing countries. (The international definition of development is not clear on these items.) But whatever the theoretical position may be, the practical situation is that these items are today not subject to comparative international measurement - comparable data are simply not available - so that, like many other items, they cannot appear in an international quantitative analysis of development. This is without prejudice to the study of such items in individual countries or in small groups of countries with available and comparable data.

There are various background items for which data are available and which are important for the study of development but which are obviously not in themselves components or correlates of it. Population size and density and geographical area are examples. They may condition certain aspects of development but are not part of development.

The position taken in this report is that social factors and social development must be considered jointly with economic factors and economic development. This has long been affirmed to be an essential approach to development by the United Nations Economic and Social Council and General Assembly. It creates difficulties in measurement, however, and many development economists would be happier not to have to bother about the social aspects (although others are constantly discovering them and giving them a central role, under such headings as "basic needs approach"). Good social development indicators for international usage are basically distributional - they tell what percentage of the total population or relevant population reaches or does not reach specified conditions or specified levels of nutrition, health, education, housing, etc. They show the extent of development in a country with respect to various items by indicating how much of the population is participating in the development.

Good development statistics should also provide breakdowns by regions within a country and by social and economic groupings and should identify those who are not participating in development. Such statistics, disaggregative by region or socioeconomic category, are relatively sparse, however, and tend to lack international comparability.

Economic development statistics used in international work primarily consist of aggregative indicators expressed in a per capita form for comparative purposes: production or consumption is divided by the relevant population, as in per capita GDP, per capita trade, per capita investment, manufacturing production per person active in manufacturing, per capita steel consumption or energy consumption. These are quite different in their nature and behaviour from the percentage type social distributional indicators, and there are difficulties in bringing the two types together. For lack of adequate data, however, many of the so-called "social" indicators are in fact economic indicators in disguise - for example, per capita calorie consumption, which is derived from production and trade data but does not tell what percentage of the people are actually undernourished. These are inferior as social indicators. The relatively small number of social distributional indicators with available comparable data include: percentage of adult population that is literate or has specified educational qualifications; percentage of school-age children in school; number of infant deaths (or survivals) per 1,000 births and other age-specific mortality rates; percentage of population with access to pure water; percentage of dwellings with electricity or piped water. These are relatively crude distributional measures but better than nothing. A major need of development indicators today is to build up more and better social distributional statistics. Studies such as the present one are seriously handicapped by the existing data situation.

SELECTION OF INDICATORS

Selection of indicators in the UNRISD study was carried out in a series of steps involving progressive application of stricter criteria.

1. An initial set of 100 variables was compiled, after review of published statistical series for 1970. 1/ In a number of cases, new indicators were constructed out of existing data - e.g., per cent of the adult male labour force in agriculture was constructed from basic data because the published indicator on per cent of labour force in agriculture covered both male and female labour and was judged to have very poor comparability (owing to marked national differences in the extent to which females in agricultural households are counted in the labour force). For national accounts indicators, adjustments were made to deal with the fact that some countries in 1970 used the new system of national accounts (SNA), while others used the old system, and the corresponding figures were not comparable.

2. This list of 100 was reduced to 73 by elimination of indicators that had insufficient data or conspicuous defects (e.g. crude mortality rate, discussed above), or that obviously did not distinguish between developed and developing countries (e.g. terms of trade). A number of these indicators, excluded as development indicators, were kept as "background indicators". 2/

3. The list of 73 was reduced to 60 by reduction of duplication.

4. A reduction was then made to a "reservoir" of 40 indicators and finally to a "core" list of 20 indicators to be used for major analytic purposes. These indicators are shown in Table 3.1.

The criteria used in selection of the reservoir of 40 indicators and then, by stricter application, the core list of 20, are as follows:

1/ Account was taken in this compilation of experiences in compiling the UNRISD 1960 Data Bank.

2/ Data for the 73 indicators and also for 20 background indicators, are given in UNRISD, Research Data Bank of Development Indicators, op.cit., Vol.I.

Table 3.1

RESERVOIR AND CORE INDICATORS IN 1970 UNRISD DATA BANK, WITH DATA AVAILABILITY BY INDICATORS (FOR 120 COUNTRIES)

Core indicators shown by asterisk

	Number of countries having data
HEALTH AND DEMOGRAPHY	
* Infant mortality rate per 1,000 live births	65
* Expectation of life at birth - both sexes	76
Proportional mortality ratio	50
Crude birth rate per 1,000 population	81
NUTRITION	
Calories, apparent consumption, per capita, per day	98
Protein, apparent consumption, per capita, per day	98
* Apparent consumption of protein of animal origin, per capita, per day	98
Apparent consumption of calories derived from cereals and starchy roots as per cent of total calorie consumption	97
EDUCATION	
* Literate as per cent of total population 15 and over	65
* Combined primary and secondary education enrolment as per cent of population 5-19	99
Per cent of population 12-17 enrolled at school	86
Vocational education enrolment as per cent of population 15-19	83
Higher education enrolment per 1,000 population 20-29	97
HOUSING AND URBANIZATION	
Average number of persons per room	44
Dwelling with piped water as per cent of total dwellings	39
Dwelling with electricity as per cent of total dwellings	39
Population in urban areas as per cent of total population	116
Population in localities of 100,000 and over as per cent of total population	95
COMMUNICATIONS	
* Newspapers ('Daily general interest') circulation per 1,000 population	95
* Telephones per 100,000 population	111
* Television receivers per 1,000 population	90
Domestic mail traffic per 1,000 population	73

	Number of countries having data
TRANSPORT AND SERVICES	
Motor vehicles (passenger cars plus commercial vehicles) per 1,000 population	105
* Per cent economically active population in electricity, gas, water, transport, storage and communications (isic div. 4 and 7)	56
AGRICULTURE	
* Agricultural production per male agricultural worker (isic div.1) at current prices (in purchasers' values), in U.S. Dollars	70
* Adult male labour in agriculture (isic div.1) as per cent of total adult male labour	77
GDP (in purchasers' values) derived from agriculture as per cent of total GDP	100
Fertilizer, apparent consumption, kilograms per adult male agricultural worker (isic div.1)	76
INDUSTRY	
* Electricity, apparent consumption per capita (in kilowatt hours)	116
* Steel, apparent consumption per capita (in kilograms)	108
* Energy, apparent consumption per capita (in kilograms of coal equivalent) revised	116
Economically active population in manufacturing industries (isic div.3) as per cent of total economically active population	58
GDP (in purchasers' values) derived from manufacturing industries as per cent of total GDP	87
* Manufacturing production per person active in manufacturing industries (isic div.3) in U.S. Dollars	45
FOREIGN TRADE	
* Foreign trade (exports plus imports) per capita in U.S. Dollars	111
GENERAL ECONOMIC	
* Investment per economically active person (annual average 1960-1970) in 1970 U.S. Dollars	69
* Salaried and wage earners as per cent of total economically active population	56
TECHNOLOGY	
* Professional, technical and related workers (isco Div.0-1) as per cent of total economically active population	53

For details of indicators, see UNRISD Research Data Bank of Development Indicators, Volumes I and IV.

Conceptual significance: whether the indicator represents a conceptually important aspect or element of development for countries in general, and particularly for developing countries.

Balance and avoidance of duplication: whether the indicator contributes to sectoral balance or to imbalance and duplication in the set of indicators. The attempt was made to cover as many sectors as possible, without an undue number of indicators for any particular sector. Some overlap was maintained, as in infant mortality rate and life expectation, where it was felt necessary in view of the importance of the component and the absence of other usable indicators.

Validity: whether an indicator measures what it is intended to measure and does not reflect other conditions. (See example of crude death rate in preceding chapter.)

Discriminative power and correlations: whether the indicator effectively distinguishes between different levels of development, particularly with regard to developing countries 1/, and is well correlated with other development indicators, both within and without the sector.

Comparability: whether the indicator is operationally defined in the same way and used to count or measure the same things in different countries. The indicator radio receivers per 1000 population, which apparently counts small transistor sets in some countries but not in others, was eventually dropped partly for this reason. Published statistics on income distribution and on per cent of dwellings with piped water are other examples of non-comparability of data.

1/ Some indicators that have little or no relevance today for measuring progress in developed countries, such as literacy and water supply (most developed countries have reached the highest level in these indicators), were, however, included since the primary purpose of the UNRISD Data Bank was to promote measurement and analysis in relation to developing countries.

Availability of data: whether a sufficient number of countries, developing as well as developed, have data on the indicator for the year 1970, or data from which a value can be reasonably computed for 1970. In general, indicators for which there were less than around 45 or 50 country values for 1970 were excluded.

Data quality and basis: whether a sufficient number of country values for the indicator can be assumed to be reliable and to be based on observations. Indicators involving an excessive use of estimates and assumptions were excluded, such as "calorie consumption per capita as a percentage of requirements". 1/

The indicators remaining after application of these criteria are by no means without defects but they are considered usable as a group.

THE SCREENING OF DATA

Screening of individual country values for defective figures was carried out for all indicators accepted in the "reservoir", and consequently also for all those in the core.

The following are the main methods of screening that were applied, some more rigorously than others:

Careful reading of notes and footnotes: As mentioned in Chapter 2, the organization compiling and issuing international statistics very often provides notes and footnotes indicating irregularities and weaknesses in specific figures. These are highly important for data screening but surprisingly neglected.

1/ This indicator not only requires estimates (often involving large amounts of guesswork) of total domestic food production, imports and exports, stockpiling and reduction of stocks, losses and wastage (in storage, transport, retailing, and household usage), of total calories in the net amounts of food remaining, and of overall population size and age structure; it also requires guesses at average body size or body weight of the population, levels of physical activity, and other such items affecting calorie requirements for which there are generally no factual data at the national level.

Inspection of tables of country values: By simple inspection of a table of values of different countries for a given indicator, it is often possible to detect values that are way out of line and that may, on checking, prove to be mistakes made at one stage or other of data compilation - often elementary mistakes in copying, or in adding, subtracting or dividing. 1/

Inspection of trends: When a value on an indicator for a given date such as 1970 is far out of line with other values for the same country over time and contrary to the historical trend, there is a good possibility it is a mistake.

Comparison of sources: Contradictory figures issued by different authoritative sources for the same indicators in the same countries, as illustrated in Chapter 2, give reason for pause and investigation.

Consistency of values for closely related indicators: Scatter diagrams for indicator pairs that are closely related, such as life expectation and infant mortality, may show marked inconsistencies in relative position of some of the countries, raising questions about the correctness of some of the values and giving reason for checking.

Consultations: Figures and series that seemed questionable (as a result of preceding screening operations) were discussed by correspondence or visit with international and national statistical offices and other authorities.

In spite of the indicator selection process and the screening operations described above, some of the indicators kept in the UNRISD core group used for major analytic purposes are nevertheless conceptually or technically weak - much weaker than indicators that had to be eliminated for lack of data or for reasons of duplication. Most of the individual

---

1/ A more rigorous application of this method would have prevented, for example, publication in the unrevised Volume I of the UNRISD Research Data Bank of the heavily mistaken value of 2.3 metric tons of consumption of cement per person in Hong Kong in 1970, a mistake due to a misplaced decimal point. (The correct value is .23 metric tons or 230 kilograms.)

figures remaining after screening are also undoubtedly incorrect to varying degrees. Even the most statistically advanced countries with relatively ample technical and financial resources can have difficulty in getting a correct statement (through census enumeration) on such an elementary matter as population size. Development statistics give an impression of a precision that does not exist. As noted above, the problem in international statistics is not that of getting exact figures but of getting tolerable figures, and the findings from analytic use of such figures are not truths but approximations.

THE USE OF CORRELATION IN INDICATOR SELECTION

Correlation was used as one of the criteria of indicator selection. More specifically, an indicator showing relatively low average correlation with the mass of other development indicators was usually not selected, although high correlation was not in itself a sufficient basis for selection if other criteria were not met. In view of the complexities of correlations between development indicators, as discussed in the following chapter, both product moment and rank order correlation were considered.

In some forms of measurement, absence of correlation with other items, or even negative correlation, may be a reason for selecting an item, not for rejecting it. Measurement of socioeconomic development, however, is a different matter. In certain respects, it can be likened to the measurement of growth or development of a child by means of different variables - height, weight, strength, agility, vocabulary size, problem solving ability, etc., all of which are necessarily correlated in a population of children of different ages. If not, they would not be measures of child development. However, if children of a specific age are taken, such as those who have just passed their tenth birthday, then there may be very little correlation or even a negative correlation between, say, weight and vocabulary size. The correlation between these two variables in the 3-13 year-old population, on the other hand, will be high because both variables increase regularly as children grow older. Similarly, there may be high correlation between development indicators covering countries at a number of successive levels of development, although no correlation at all for countries at exactly the same level (if sufficient such countries could be found). This issue will be discussed in more detail in the following chapter.

In principle, correlations should be highest between indicators concerned with the same component of development - between different indicators of health or education or energy, etc. But in practice that is often not true because of the variable quality of indicators and the empirically demonstrable fact that technically inferior indicators tend to have low correlations. The higher correlations that should emerge in theory between indicators related to the same subject matter tend to be overwhelmed by problems of indicator and data quality. Thus a good indicator of education will tend to have a higher correlation in available data with a good indicator of health or economic production, etc. than with a bad indicator of education. Table 3.2 gives illustrations. 1/

One result of this situation is that a "factor analysis" of development data is likely to lead to the identification, as a second factor, of a group of technically bad indicators which happen to be highly correlated with each other but not with other development indicators.

The above argument can be illustrated with reference to the measurement of industrialization as a "factor" of development. It might be initially assumed that per cent of GDP derived from industry is a good measure of degree of industrialization. However, this indicator has relatively low correlation with the mass of other development indicators (e.g., correlation with literacy = 0.46, with telephones per 100,000 population = 0.37, with per cent of GDP from manufacturing

---

1/ The "relatively poor" indicators in this Table are discussed in the preceding chapter.

Table 3.2

Correlations of relatively good and relatively poor indicators from different sectors
Adjusted data 117 countries
(Rank order correlations in parentheses below product moment figures)

	Comb. Enrol. A	Gross P. Enr B1	Educ. Expend B2	Life Expect A	Crude Death B1	Health Expend B2	Tele-phones A	Radios B	GDP Manuf. A
Education									
A Combined Enrolment									
B1 Gross Primary Enrol.	0.84 (0.75)								
B2 Gov. Expenditure on Education as % of GDP	0.33 (0.44)	0.32 (0.31)							
Health									
A Life Expectation	0.93 (0.93)	0.78 (0.61)	0.28 (0.38)						
B1 Crude Death Rate	-0.79 (0.64)	-0.85 (0.72)	-0.25 (0.31)	-0.88 (0.73)					
B2 Gov. Expenditure on Health as % of GDP	0.41 (0.30)	0.28 (0.22)	0.56 (0.60)	0.45 (0.34)	-0.38 (0.39)				
Communication									
A Telephones per 100.000 inhabitants	0.68 (0.91)	0.33 (0.66)	0.22 (0.46)	0.64 (0.92)	-0.41 (0.63)	0.38 (0.31)			
B Radios per 100.000 inhabitants	0.60 (0.77)	0.48 (0.54)	0.27 (0.32)	0.54 (0.75)	-0.42 (0.53)	0.24 (0.08)	0.74 (0.82)		
Industry									
A % of GDP from Manuf. Industries	0.79 (0.80)	0.65 (0.67)	0.18 (0.22)	0.80 (0.82)	-0.62 (0.57)	0.12 (0.06)	0.65 (0.80)	0.50 (0.66)	
B % of GDP from Industry	0.52 (0.56)	0.42 (0.49)	0.25 (0.31)	0.51 (0.55)	-0.39 (0.34)	0.16 (0.07)	0.37 (0.67)	0.33 (0.54)	0.43 (0.57)

A = Relatively good indicators
B = Relatively poor indicators (B2 indicators are inferior to B1 indicators)

Average Correlation: relatively good indicators with relatively good indicators (A with A)
r (product moment)= 0.748 r (rank order)= 0.863

Average Correlation: relatively poor indicators with relatively poor indicators (B with B)
r (product moment)= 0.373 r (rank order)= 0.385

Average Correlation: relatively good indicators with relatively poor indicators (A with B)
r (product moment)= 0.495 r (rank order)= 0.546

industries = 0.43, with GDP per capita = 0.52, with energy consumption = 0.42). 1/ On examination this appears to be due to the fact that in the national accounts statistical definition, industry includes mining, and a number of very poor developing countries get a large part of their computed income from mining of ores (copper, etc.) and their export. Zambia with 52 per cent of its GDP derived from industry, Mauritania with 36 per cent and Liberia with 35 per cent are nevertheless not highly industrialized nations, compared, for example, with Norway, Sweden, and the United States (30 per cent, 30 per cent and 29 per cent respectively). Oil, also counted as mining, adds to the confusion. (Libya with 65 per cent of GDP derived from industry and Saudi Arabia with 66 per cent are the most industrialized countries in the world by this indicator of industrialization.)

While per cent of GDP derived from industry is a poor indicator and has generally low correlations with other economic as well as social development indicators, it may nevertheless have quite high correlations with a small group of indicators that are closely related to it such as per cent of labour force in industry (with industry defined in the same way). Such a collection of inferior indicators is what the computer is likely to find as a separate "factor" of development, additional to the central factor defined by the relatively highly intercorrelated indicators of better quality. 2/

The impact of indicator quality on correlations (and on regressions, etc.) is so great that many development studies which purport to show the quantitative impact of different factors upon a given variable, or to say how much of its variation is "explained" by these factors, would yield quite different results with a different selection of indicators. It is easy to forget that in development analysis one is relating indicators of varying adequacy - they are generally not direct and full measures

---

1/ In the case of development indicators, correlations below 0.55 are relatively low. In a correlation matrix of 60 by 60 development variables yielding 1770 specific bivariate correlations, 781 of the correlations have a value of 0.70 or more. (UNRISD Data Bank)

2/ This has been the experience at UNRISD.

of the objective conditions - and the results may tell us more about the quality of the indicator than about the conditions of the components of development they are supposed to measure.

The fact that, empirically, indicators that are technically inferior tend to have relatively low correlations with most other development indicators, inside or outside the sector, is a practical reason for use of average correlation as one of the criteria in indicator selection, taking account, however, of the weaknesses and complications of correlation analysis of development data.

ADJUSTMENTS OF DATA

The main forms of adjustment made on data to promote comparability have been already discussed and need only be summarized here:

1. Adjustments for age structure. (Adjusted figures are given in UNRISD Data Bank Volume II.) These adjustments substantially change the levels of per capita income and other per capita figures. Readers should take this into account when examining tables and charts using "adjusted data". The general relation between the original and the adjusted values for GDP per capita is indicated in figure 3.1.

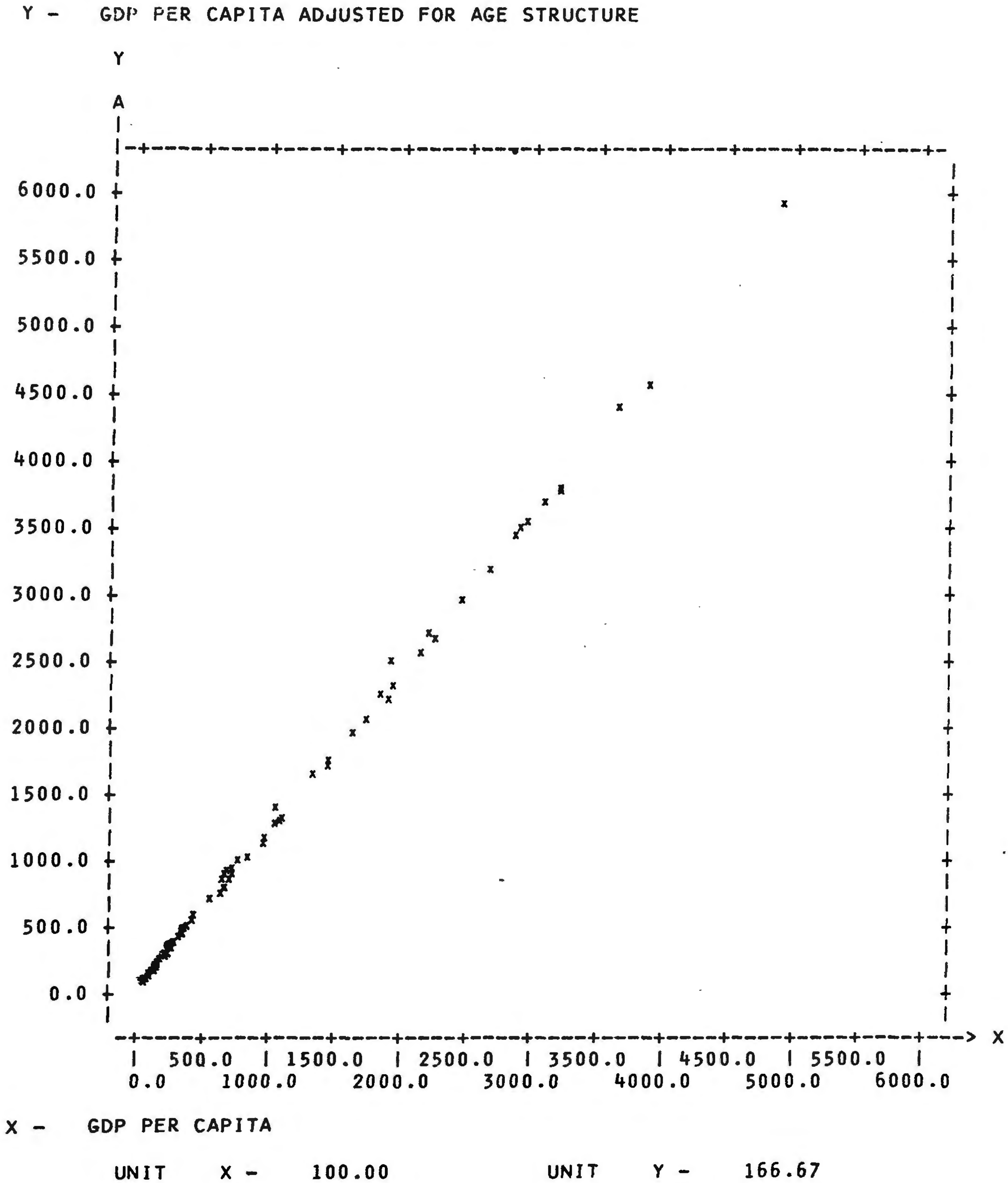
2. Adjustments for time comparability (1970). Adjustments were made only for data within a few years of 1970, by use of extrapolations or interpolations. Data available covering other years were not adjusted or used.

3. Adjustments of national accounts figures to take account of problems of inflation and currency exchange rates affecting international comparability, also of differences in treatment of "statistical discrepancies" and in use of the new vs. the old national accounts system.

Figure 3.1

Relation between GDP per capita and GDP per capita adjusted for age structure

100 Countries



CHAPTER 4

ANALYTIC APPROACHES TO INTERNATIONAL DEVELOPMENT DATA

CHARACTERISTICS OF DEVELOPMENT DATA AFFECTING ANALYSIS

It is conventional wisdom to say that one chooses a method of data analysis suitable to one's purpose. It is less conventional but no less wise to choose a method that is suitable to the nature of the data. Unfortunately, the nature of the data may exclude the purpose. Thus, the nature of development data is such as largely to exclude derivation by statistical methods of the amount of impact of one variable upon another. Much of the quantitative analysis carried out on development data rests on assumptions about the data that are not valid. It is assumed, explicitly or implicitly, that they have certain characteristics required by the method that is used, while in fact they do not have these characteristics, and the analysis is essentially an exercise in make-believe.

Part of the difficulty arises from the fact that methods of statistical analysis that are now conventional were developed primarily in relation to statistics describing individuals (humans, animals, insects, plants) as the primary units of observation and measurement, but not in relation to statistics describing countries as the primary units of observation and measurement. Conditions commonly found in relation to measurements of populations of individuals, such as the "normal frequency distribution" of values (or approximations thereto), are not normal - in fact not found at all - and cannot be assumed in relation to development data on countries. A group of countries used in comparative analysis is a population of populations, an aggregate of aggregates. It is irregularly distributed in terms of frequency, and small in the number of units (individual countries), although at the same time the individual countries and territories listed in statistical compilations have tremendous variation in the size of their populations, from around a hundred persons to nearly a billion, and also a tremendous range of variation in other characteristics. 1/ Furthermore, the kinds of variables that apply to countries tend to be technically different from those that measure individuals within countries, and vice versa. (Individual persons, for example, do not have mortality rates or percentages enrolled in school or various other percentage type characteristics that apply to populations.) Finally, some of the conceptual framework concerning individuals in which statistical methods of analysis originated can not be said to apply to development data concerning countries. Thus Galton, the 19th century eugenicist, conceived of the idea of "regression to the mean" (the origin of statistical regression analysis) in connection with a theory of a biological central tendency in successive populations of individuals; but countries do not have such a biological central tendency in relation to development -- the very concept of development implies something quite different, namely a steady progression upwards - and the idea of regression to the mean makes no sense in development theory. In the paragraphs that follow, some of the elementary characteristics of international development data that affect analytic possibilities are described; implications for three basic kinds of analysis, trend analysis, correlation and regression, are then considered.

1. Limitations of country coverage

Not only is the total number of countries small, but also for any one indicator, one set of countries will have data, for a second indicator a second set, a third indicator a still different set, etc. The result is that for a group of indicators covering various sectors there will be few countries having data on all the indicators, and those that do are likely to constitute a sample badly biased in favour of more developed countries. This makes it difficult to apply methods like conventional factor analysis which require data on all indicators for all countries under consideration. It means that one must go about the process of analysis in a piecemeal fashion, building it up from

---

1/ There are about 200 countries and territories listed in international statistical yearbooks. For analytic purposes, small countries with population under one million, which tend to have different characteristics from larger countries, are frequently omitted (as by UNRISD), leaving about 120 for comparative analysis, although the number of countries having data for a particular analytic purpose is likely to be under 100. The 120 countries constitute the total universe of countries above one million population, not a sample of some much larger universe of countries, and analytic procedures based on probability models assuming a larger universe do not properly apply.

relationships of pairs of indicators where sufficient numbers of observations can be obtained. Estimates can of course be made to a limited extent to fill in missing data, but as noted above, extensive estimation of missing values carries the danger that the analytic findings may turn out to be nothing more than a discovery of the assumptions made in the estimates. The question of the validity of estimates based on limited evidence (however elaborate the model used) has been discussed in Chapter 2.

2. Technically contrasting types of indicators

Two main types of development indicators have been identified above, namely the percentage type indicator, used for social-distributional and structural measurement, and the per capita type indicator, used primarily for economic measurement but also for various kinds of social measurement particularly where percentage type data are not available. 1/ There are two other types of indicators, less commonly found, that need to be identified at this point:

Ratio type indicators which relate two separate categories of persons, goods, services, expenditures, etc. within a country by dividing the number or amount of one category by the number or amount of another category, then sometimes multiplying by 100 or more. 2/ Examples are: "dependency ratio" relating number of children under 15 to number of adults 15-64 (and multiplying by 100); teacher/pupil ratio; the male/female ratio in school enrolment; the ratio of imports to exports, of GDP from manufacturing to GDP from agriculture, of investment to consumption, etc.

All percentage type indicators can be easily converted into ratio indicators. Thus the percentage of literates in a country can be converted into the ratio

---

1/ Some of the more complexly constructed indicators are not easily classifiable. Thus, life expectation, while not giving a percentage figure, is constructed out of percentage type figures (age specific mortality rates), behaves like a social distributional percentage type indicator and can be conveniently so classified. On the other hand, the indicator "calorie consumption as per cent of requirements" is given in percentages and may appear to be a social-distributional indicator but is not - it relates the estimated aggregate consumption of calories to the estimated aggregate requirement but gives no idea of the distribution of consumption in a country in terms of the percentage of individuals consuming an amount of food below or above specified levels or standards.

2/ The word "ratio" is often used broadly to cover all kinds of fractions but in the present context it is not meant to include relation of part to whole (percentage type indicator) or of objects, services, expenditures, etc. to persons (per capita type indicator).

of literates to illiterates, or of illiterates to literates. 1/ Infant mortality rate can be converted into the ratio of infant deaths to infant survivals or infant survivals to infant deaths. Per cent of labour in agriculture converts into the ratio of farmers to non-farmers or non-farmers to farmers. However, as will shortly be demonstrated, the two ratio type indicators can behave quite differently from the percentage type indicator (and from each other) because they involve transformations of scale.

Inverse per capita indicators, which are obtained, not by dividing the amount of a good, service or expenditure by the size of the relevant population (per capita type indicator), but by dividing the population by the amount of the good, service, or expenditure (and perhaps multiplying by 100 or 1000 or more as convenient). The most common indicators of this type are persons per room, inhabitants per doctor, inhabitants per hospital bed, persons per square kilometre. There is no reason why any per capita type indicator cannot be turned into an inverse per capita indicator, so that one could measure, for example, persons per telephone or persons per 100 kilowats of electricity consumed or per $10,000 of GDP, rather than telephones, electricity or GDP per capita. It simply involves reversing the fraction and multiplying by an appropriate number. (Thus $500 per capita national income becomes 2 persons per $1000 or 20 persons per $10,000 of national income.)

Here again, seemingly innocent changes in the way of presenting the same basic data can have large, sometimes extraordinary, effects on correlations and regressions (see below); they can also have important implications for the measurement of progress or rate of growth, as well as for any kind of comparative measurement that involves conversion of different indicators to a common scale (e.g., a 0-100 scale) where a given interval on one indicator is established as equal to a given interval on another.

3. Frequency distribution of indicator values

It is an important fact for the analysis of development data that, as noted above, the country values on the individual development indicators are not distributed according to the normal bell-shaped curve, or to approximations of

---

1/ If per cent literate = a, then the two ratios are simply

$$\frac{a}{100 - a} \text{ and } \frac{100 - a}{a} .$$

it. The distributions are elongated and irregular, and the main concentrations of frequency tend to be at one end or another of the range of values, depending on the technical nature of the indicator (changing the technical nature can shift the apparent concentrations from one end to the other). 1/

In the case of percentage-type indicators, the values tend to become more concentrated as they approach or reach the maximum (100 per cent for literacy, around 75 years for life expectation); or in negative percentage type indicators as they approach the minimum (around 3 or 4 per cent for per cent male labour in agriculture). On the other hand the values for per capita type indicators, in their ordinary scales, tend to be most heavily concentrated around the lowest values. (The 1970 per capita GDP ranged from $55 to $4,880 but nearly a third of the countries were below $200. Telephones per 10,000 population ranged from 5 to nearly 6,000, but well over a third of the countries had values less than 100.)

If we divide the range (between bottom and top value) of each of a group of indicators into five equal intervals according to the scale of each indicator (interval I being the least developed level and interval V the most developed) and count the number of countries falling into each interval (Table 4.1), we find that in only one case (combined school enrolment) is the greatest frequency in the middle group, and even in this case the distribution is by no means "normal". While percentage type indicators will be seen to show greatest frequency at or towards the top level and per capita type indicators at or towards the bottom level, the change of frequency from level to level is not a smooth progression but tends to be irregular and in most cases involves more than one bump (with use of smaller intervals it becomes even more irregular).

Some kind of geometric, curvilinear relationship between percentage and per capita type indicators is inevitable with such distributions. It should also be noted that changing the technical form of an indicator can radically change the frequency distribution. This is illustrated by the marked difference in frequency distribution of the last two items of Table 4.1: inhabitants per physccian and physicians per 100,000 inhabitants (respectively a per capita type indicator and an inverse per capita type indicator). In the case of the former

1/ The frequency distributions are, of course, also affected by availability of data. If very few developing countries have data on an indicator, there will not be a heavy concentration at the lower part of the scale.

Table 4.1: Frequency distribution of country values on selected indicators by equal intervals on indicator scales.

Indicator and Range (lowest to highest development levels)	Interval I	Interval II	Interval III	Interval IV	Inverval V
	——— less developed more developed ———				
Life expectation: Range 38.1 to 74.9 years	2	13	17	18	23
Infant Mortality: Range: 200 to 11 deaths per 1000 live birth (negative indicator)	4	12	6	12	38
Combined School enrolment: Range: 8.6 to 84.7 per cent	13	12	26	21	24
Per cent adult male labour in agriculture: Range: 96 to 3.3 per cent (negative indicator)	7	13	18	14	22
Agricultural Production per capita: Range: $150 to $11,490	44	8	6	9	2
Telephones per 100,000 population: Range: 53 to 58,677	74	10	7	0	5
Steel consumption per capita in kilograms: Range: 1 to 734	68	7	7	11	5
GDP per capita: Range: $55 to $4,880	70	12	9	4	1
Inhabitants per Physician: Range: 92,827 to 406	1	5	4	8	86
Physicians per 100,000 inhabitants: Range: 1.1 to 246.3	63	9	17	12	3

indicator, the greatest frequency concentration is at the highest level (interval V) involving the more developed countries; in the case of the latter indicator, the greatest frequency is at the lowest level (interval I) involving the less developed countries. 1/ The relationship between these two forms of the same indicator is curvilinear. Similar discrepancies in frequency distribution are found in relation to the different technical forms of other indicators.

4. Curvilinear relationships

Figures 1-3, showing the relationship of per capita GDP to life expectation school enrolment and per cent of adult male labour in agriculture, are typical of the exponential or logarithmic kind of relationship found between economic per capita type indicators and social or structural percentage type indicators. 2/ Figures 4-5 relating life expectation to school enrolment and to per cent of adult male labour to agriculture, show the elongated S-shaped curve frequently found in relationships between two percentage type indicators. Figures 6-7, relating per capita GDP to telephones per 100,000 population and to steel consumption, show the mildly curvilinear relationships that tend to be found between two per capita type indicators.

Figure 8 relating per capita GDP and physicians per 100,000 inhabitants (no line fitting attempted) shows a somewhat curved relationship between two per capita type indicators. Figure 9 shows an extremely neat but sharply curvilinear relationship between per capita GDP and inhabitants per physician - although the latter indicator is simply the inversion of the indicator physicians per 100,000 population shown in relation to GDP per capita in Figure 9. The marked contrast between figures 8 and 9 illustrates the extent to which curvilinearity depends upon the particular technical details of indicator construction.

1/ The five intervals for these two indicators are as follows:

	Interval I	Interval II	Interval III	Interval IV	Interval V
Inhabitants per physician	92,827 - 74,343	74,342 - 55,859	55,858 - 37,374	37,373 - 18,890	18,889 - 406
Physicians per 100,000 inhabitants	1.10 - 50.1	50.2 - 99.2	99.3 - 148.2	148.3 - 197.3	197.4 - 246.3

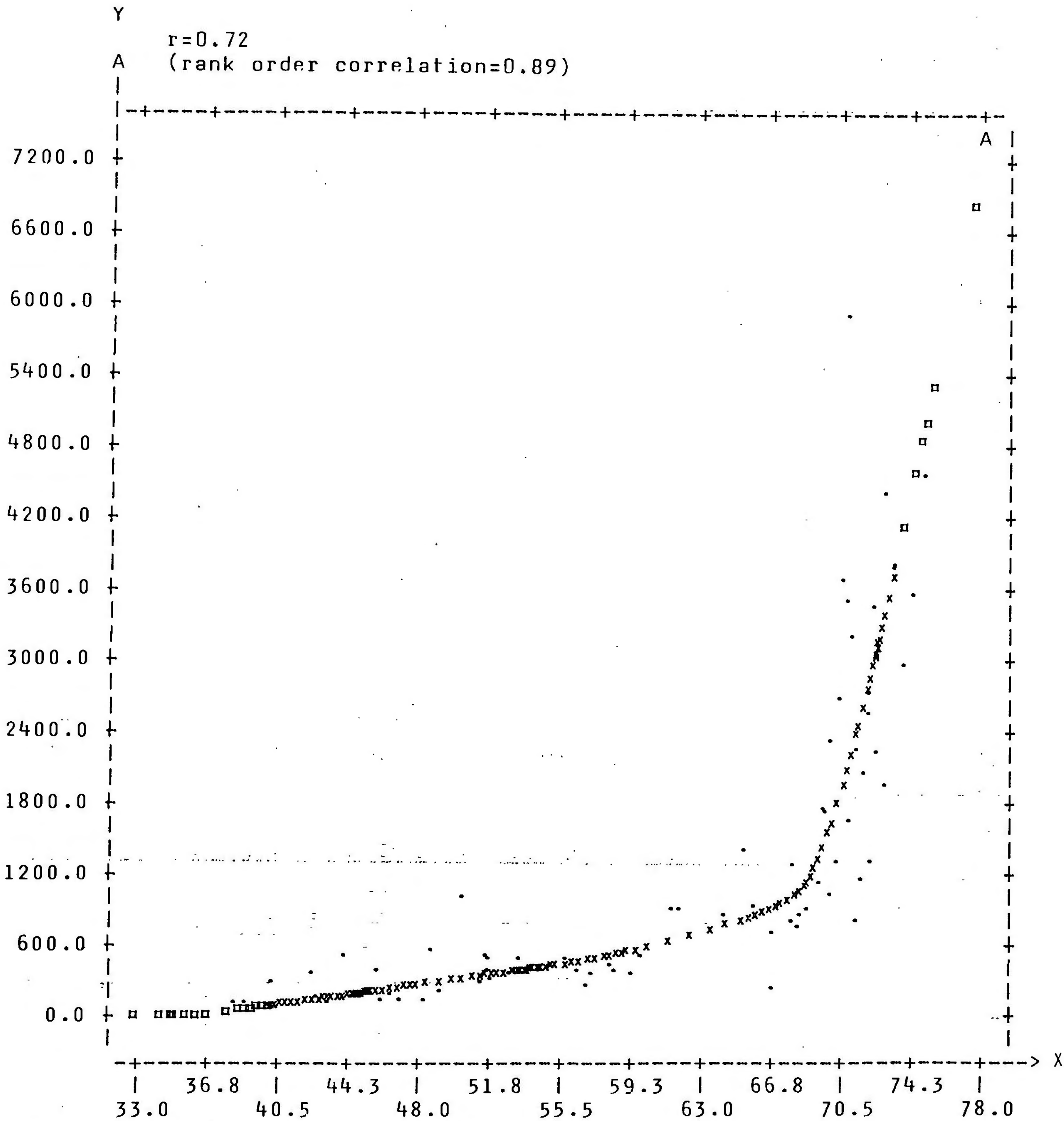
2/ The bivariate distributions under discussion in this paragraph have a line fitted to them by a method described in the following chapter and in the Annex to this report. The curves are by no means simple geometric progressions that can be readily found by inspection or by trial and error "curve-fitting" procedures.

Figure 4.1

G.D.P. per capita and Life Expectation at Birth

Country values adjusted(.)
Best fitting line(xxxx) Extrapolations(¤¤¤¤)

Y - G.D.P. per capita



X - Life Expectation at Birth

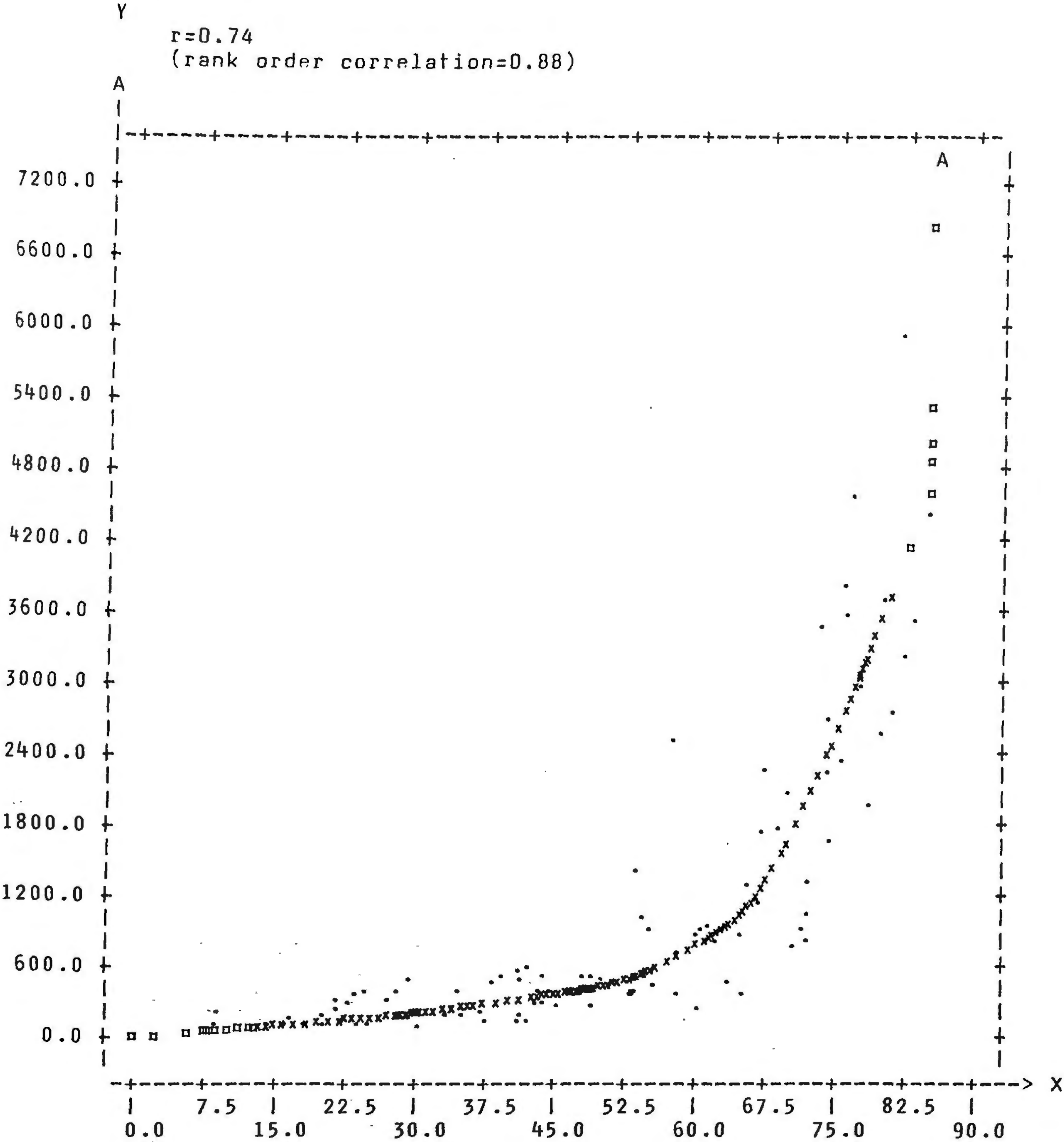
UNIT X - 0.75 UNIT Y - 200.00

Figure 4.2

G.D.P. per capita and Combined Primary and Secondary Enrolment as % of population 5-19

Country values adjusted(.)
Best fitting line(xxxx) Extrapolations(¤¤¤¤)

Y - G.D.P. per capita



X - Combined Primary & Secondary Enrolment as % of Population 5-19

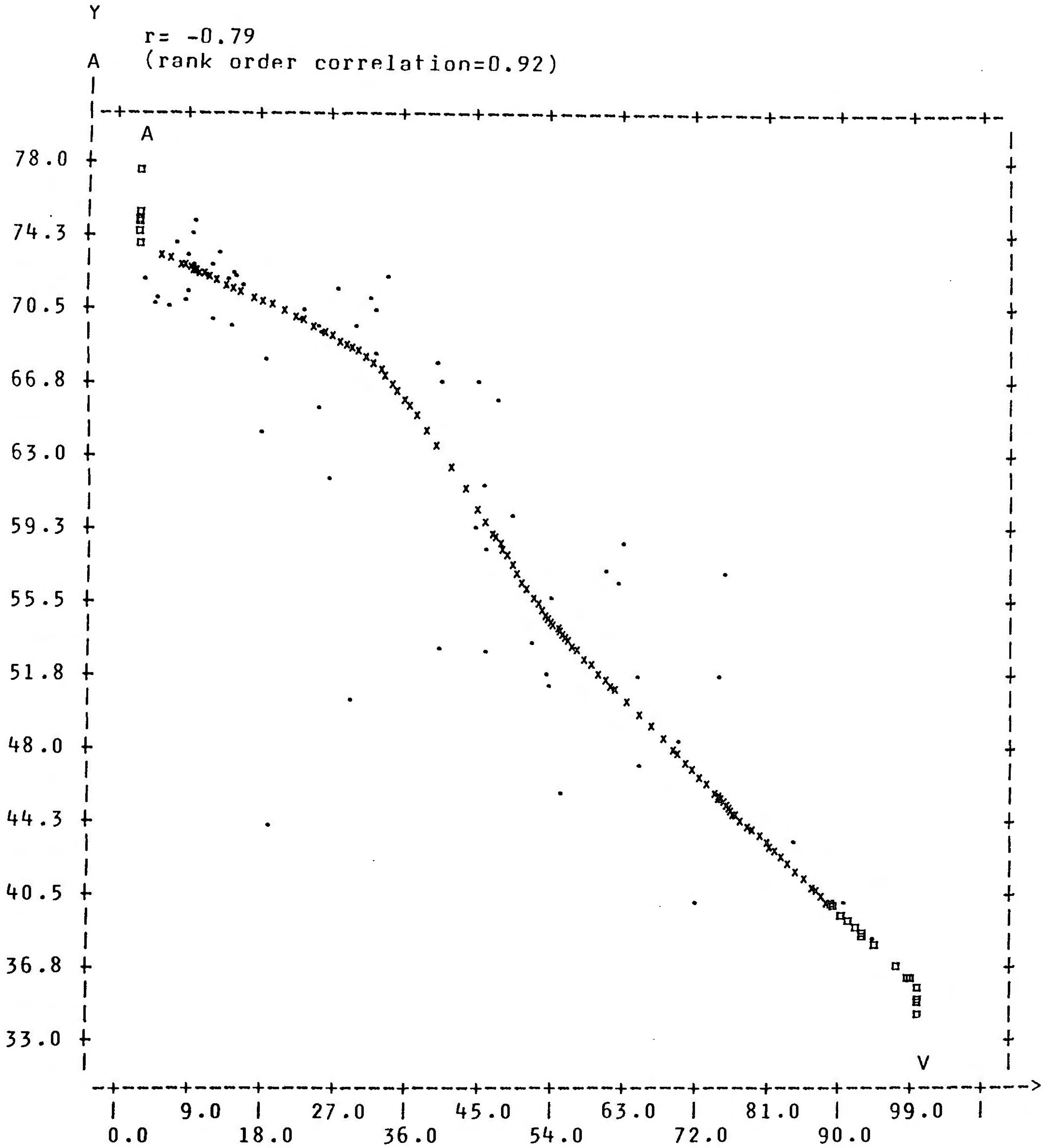
UNIT X - 1.50 UNIT Y - 200.00

Figure 4.3

G.D.P. per capita and % Adult Male Labour in Agriculture

Country values adjusted(.)
Best fitting line(xxxx) Extrapolations(□□□□)

Y - Life Expectation at Birth



X - % Adult Male Labour in Agriculture

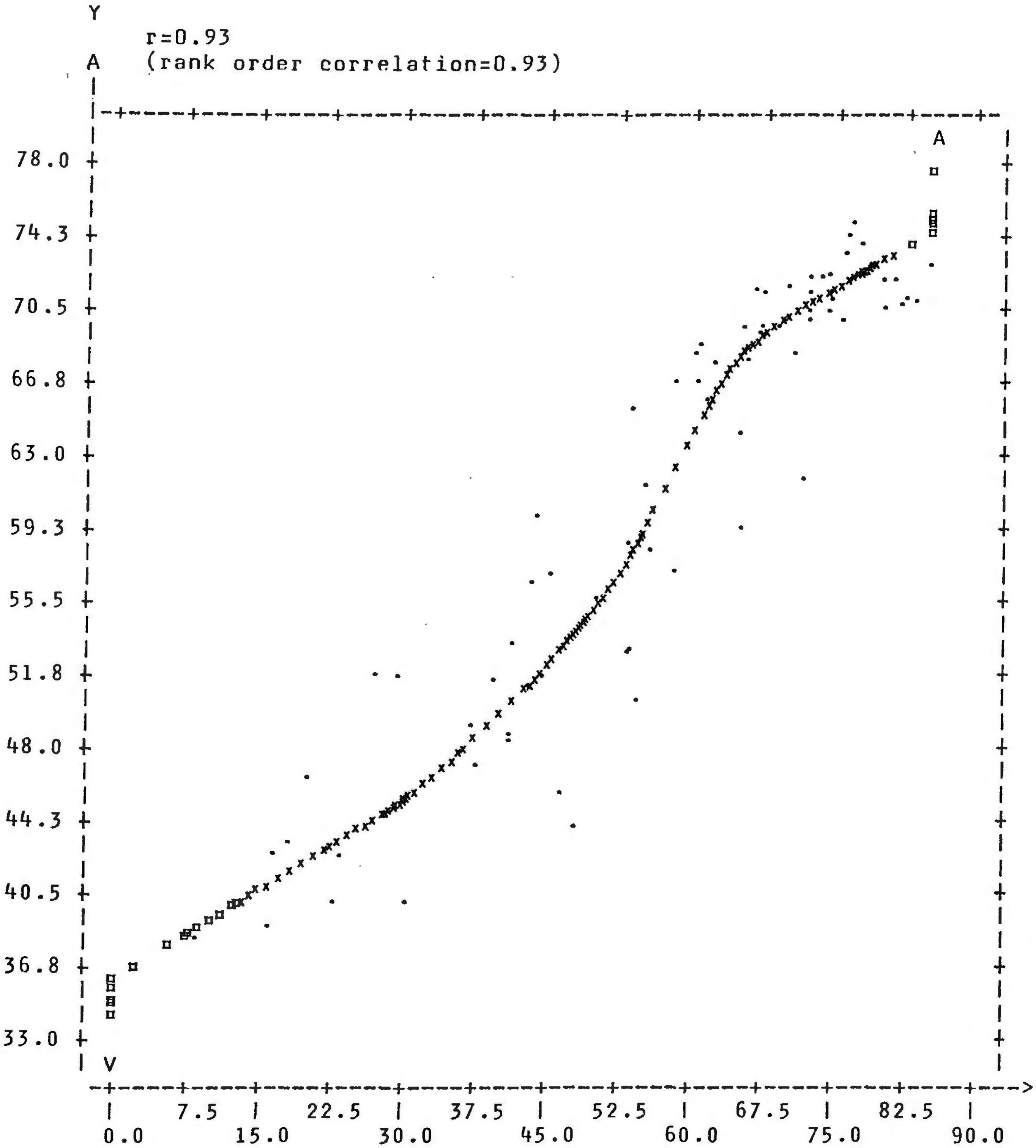
UNIT X - 1.80 UNIT Y - 1.25

Figure 4.4

Life Expectation at Birth and Combined Primary and Secondary Enrolment as % of Population 5-19

Country values adjusted(.)
Best fitting line(xxxx) Extrapolations(¤¤¤¤)

Y - Life Expectation at Birth



X - Combined Primary & Secondary Enrolment as % of Population 5-19

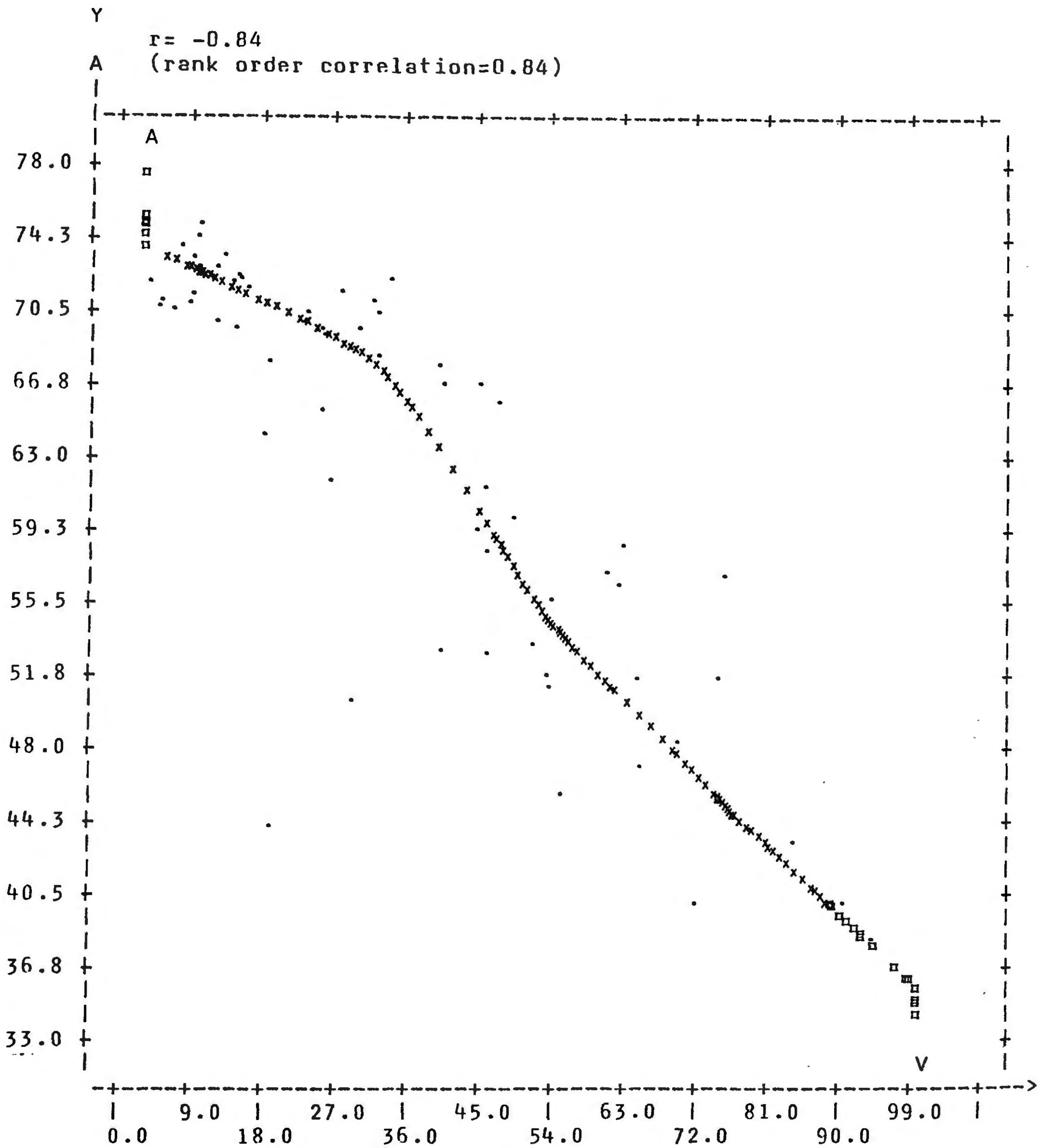
UNIT X - 1.50 UNIT Y - 1.25

Figure 4.5

Life Expectation at Birth and % Adult Male Labour in Agriculture

Country values adjusted(.)
Best fitting line(xxxx) Extrapolations(¤¤¤¤)

Y - Life Expectation at Birth



X - % Adult Male Labour in Agriculture

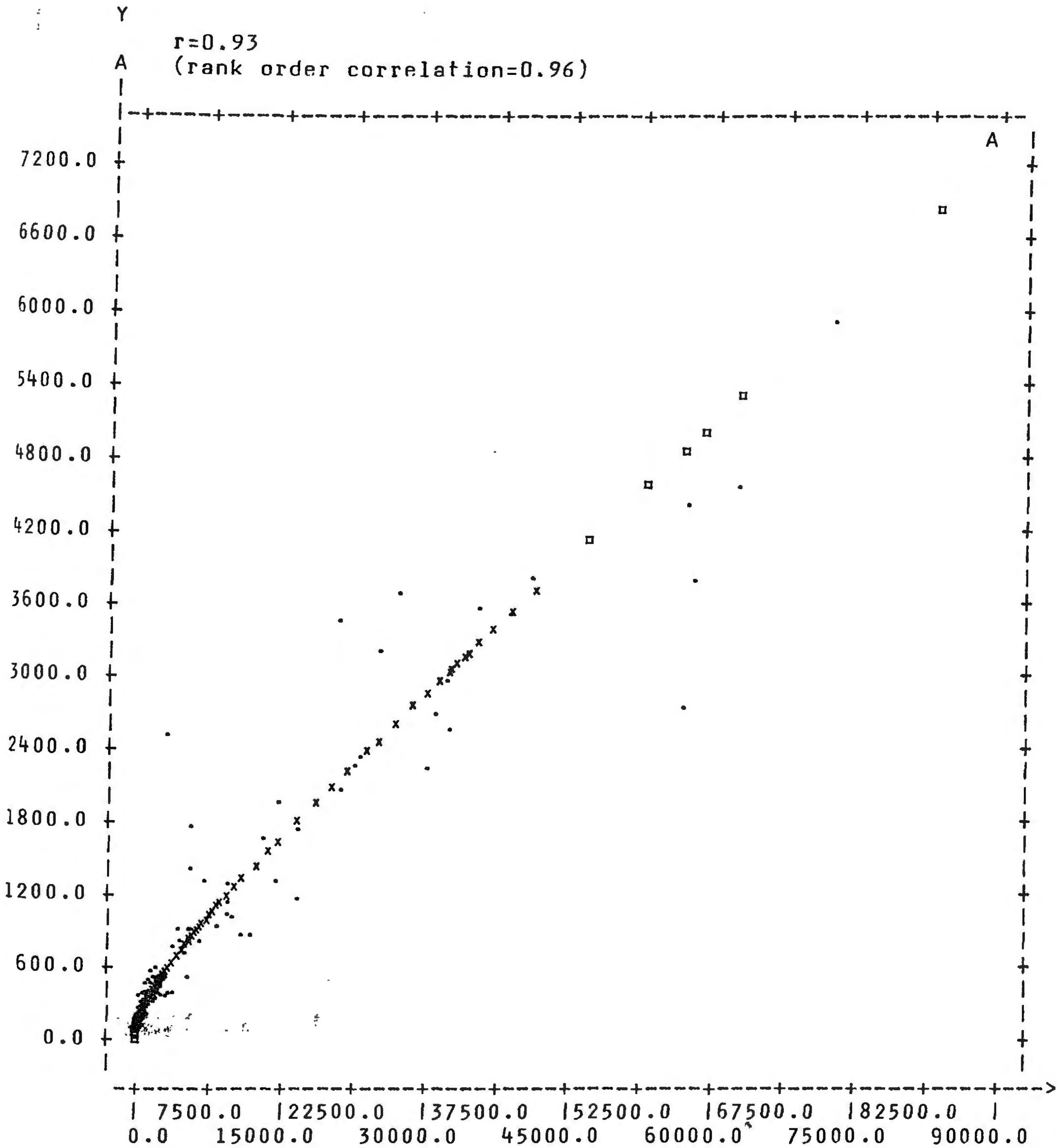
UNIT X - 1.80 UNIT Y - 1.25

Figure 4.6

G.D.P. per capita and Telephones per 100,000 Population

Country values adjusted(.)
Best fitting line(xxxx) Extrapolations(¤¤¤¤)

Y - G.D.P. per capita



X - Telephones per 100,000 Population

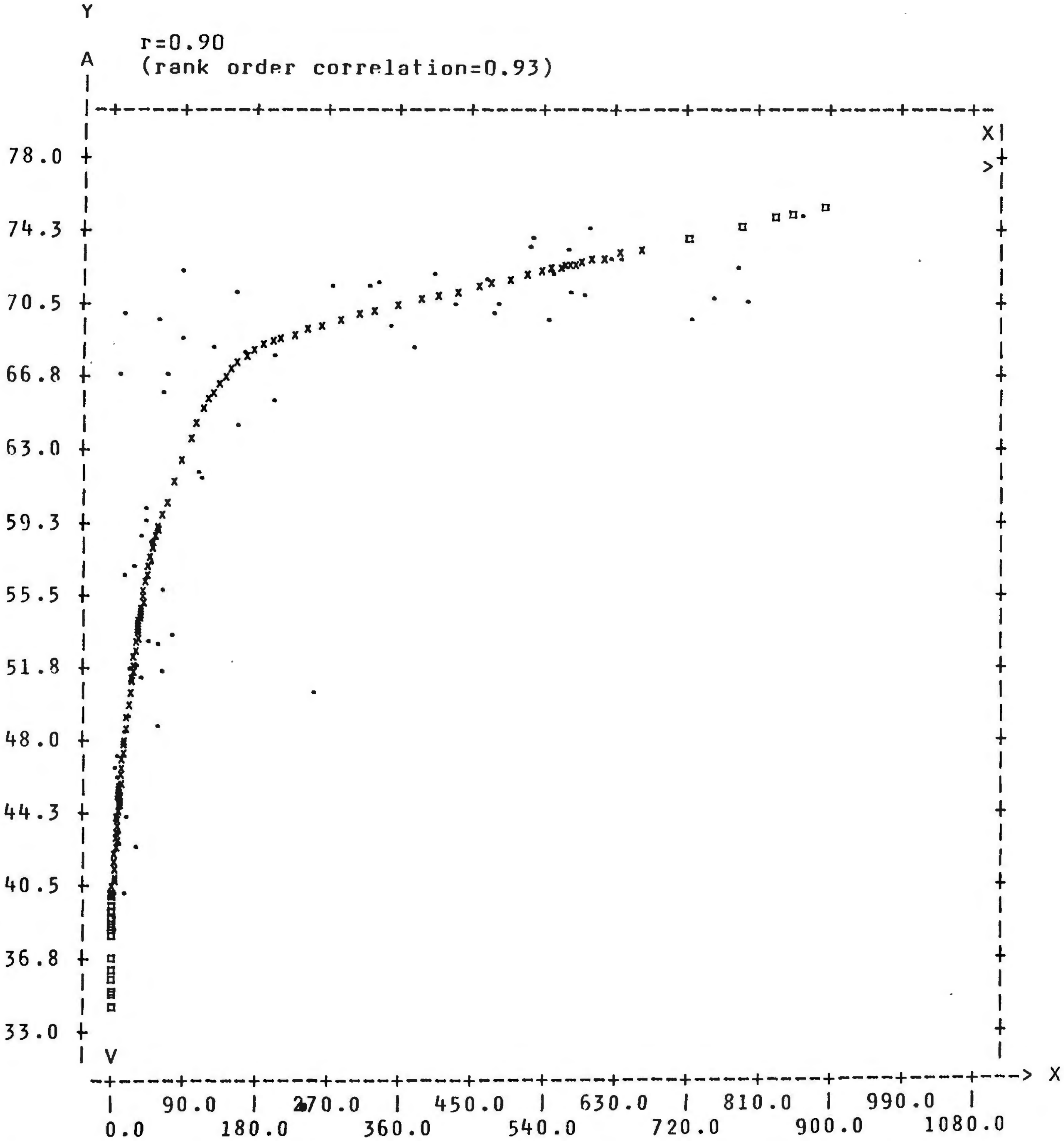
UNIT X - 1500.00 UNIT Y - 200.00

Figure 4.7

G.D.P. per capita and Steel Consumption per capita in Kg.

Country values adjusted(.)
Best fitting line(xxxx) Extrapolations(□□□□)

Y - Life Expectation at Birth



X - Steel Consumption per capita in Kg.

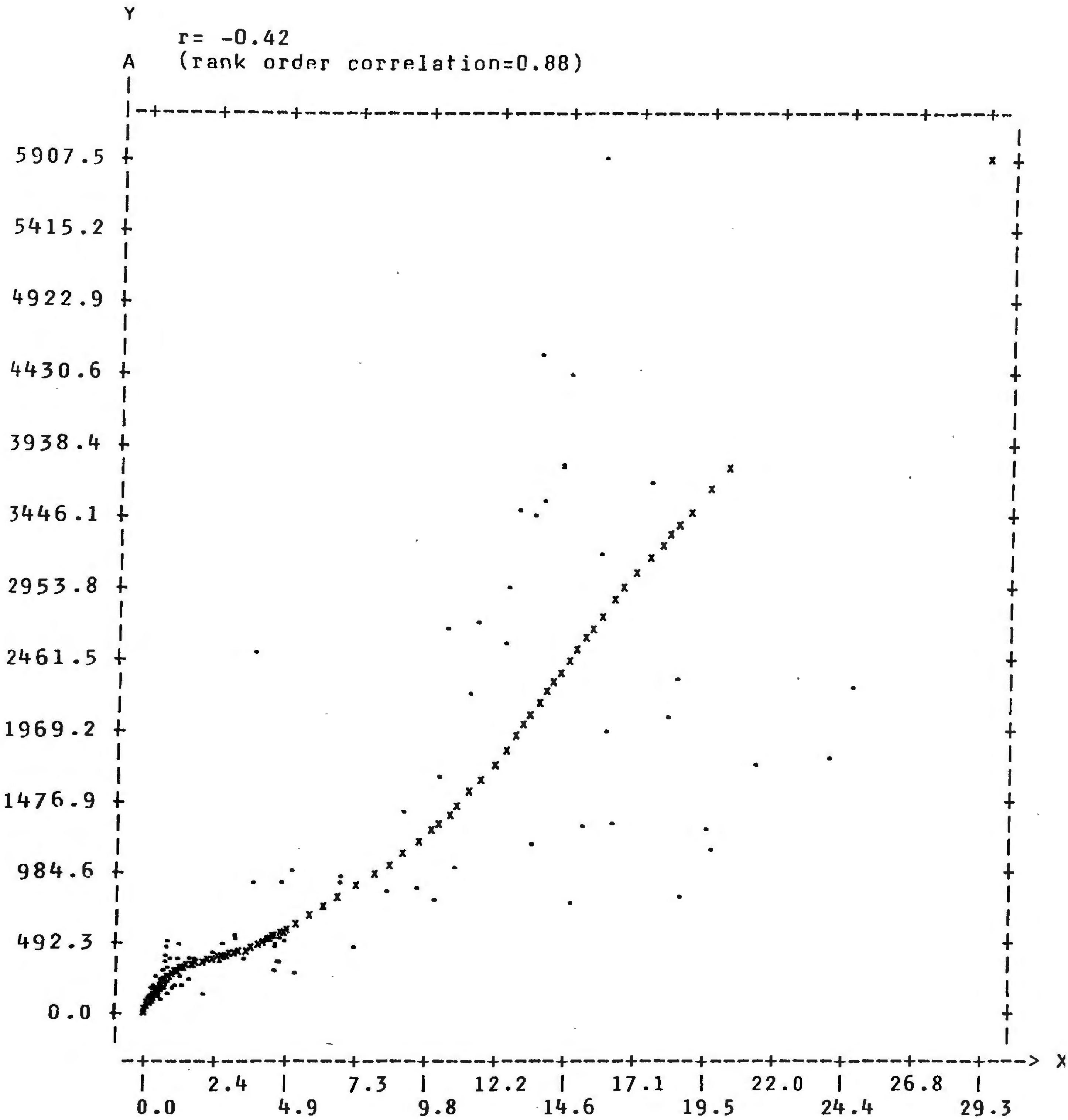
UNIT X - 18.00 UNIT Y - 1.25

Figure 4.8

G.D.P. per capita and Doctors per 10,000 Population

Country values adjusted(.)
Best fitting line(xxxx) Extrapolations(""""")

Y - G.D.P. per capita



X - Doctors per 10,000 Inhabitants

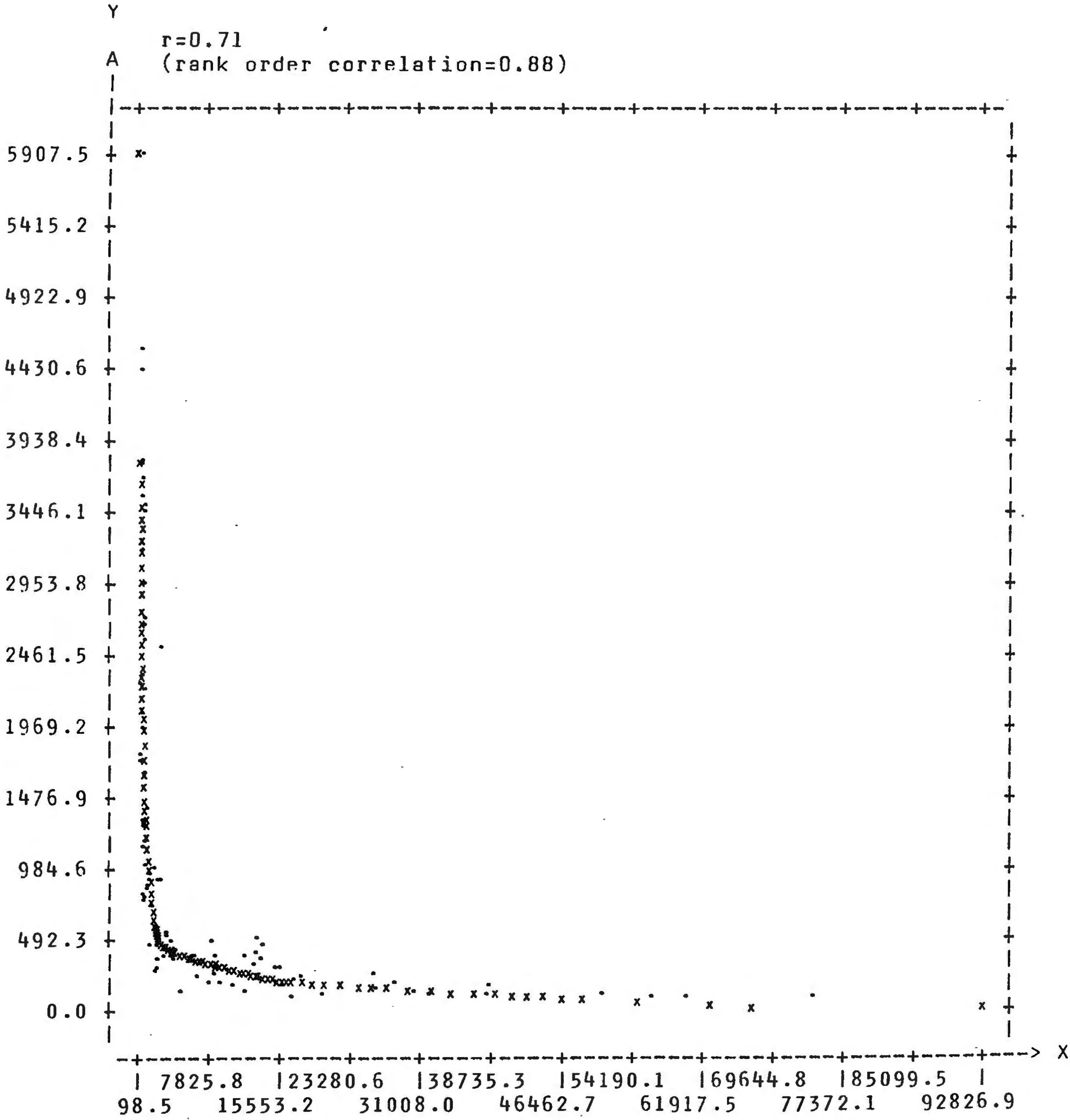
UNIT X - 0.49 UNIT Y - 164.10

Figure 4.9

G.D.P. per capita and Inhabitants per Doctor

Country values adjusted(.)
Best fitting line(xxxx) Extrapolations(¤¤¤¤)

Y - G.D.P. per capita



X - Inhabitants per Doctor

UNIT X - 1545.47 UNIT Y - 164.10

ANALYTIC METHODS APPLIED TO DEVELOPMENT DATA

The following pages discuss the difficulties of applying even the most familiar and elementary kinds of analysis to international development data - difficulties arising from the characteristics of development data discussed above, but also partly from the inherent limitations of the methods. The next chapter attempts to build up an analytic approach more suitable to development data.

The three elementary forms of analysis to be discussed here are: trend analysis, in the simple sense of derivation of rates of growth or change; correlation analysis; and regression analysis. These analytic methods are not only used directly in numerous studies but also are involved in various more complex analytic procedures carried out under other names. They are fundamental to much of the work in estimations (of missing data), projections, identification of "factors" (factor analysis) or "principal components" (principal components analysis), causal analysis and explanations of variance, and in the construction and operation of the various elaborate "development models", which absorb so much of the time and energy of development economists today.

The main themes put forward here are that the customary method of computing rates of economic growth (as in compound interest rates) cannot be applied meaningfully to percentage-type social development or structural variables; that correlation analysis has a certain use in development work, provided it is employed with caution and with full awareness of its limitations; but that regression analysis is of almost no use at all.

Trend analysis

To measure rates of growth, it is customary in development economics to use the percentage increase (or decrease) of the item in question over a base year. By this method a growth of five per cent is the same rate of growth, whether it is from $100 to $105 or $1,000 to $1,050. It is common practice to apply this same method to percentage type social and structural indicators. However, paradoxes quickly arise. Thus, if literacy increases from 10 per cent to 20 per cent and this is called a 100 per cent increase, illiteracy in the same country decreases from 90 per cent to 80 per cent which is only 11 per cent decrease. If country A is a developing country moving from 10 per cent to 20 per cent literacy, while country B is a developed country moving from 90 per cent to 95 per cent literacy, the developing country will be advancing much faster in literacy while the developed country will be moving much faster in the battle against illiteracy (50 per cent drop in illiteracy as against only a 11 per cent drop in the developing country). This problem becomes particularly important when the rate of change of a percentage type indicator like literacy is compared with the rate of change of a per capita type indicator of opposite direction).

When change in percentage type indicators is measured by percentage increase, one can usually predict remarkably well whether developing countries will appear to be progressing faster or slower than developed countries. If the indicator is positive (development represented by increase), developing countries, low on the scale, will change faster while if the indicator is negative (development represented by reduction), then developed countries which are at already lower levels will show greater progress. Thus, expectation of life which is a positive indicator of a percentage type (involving ratios of survivors to age cohorts) shows faster frowth for developing countries while infant mortality which is a negative indicator shows greater progress between 1960 and 1970 for developed countries. (If infant survival rather than infant mortality were used as an indicator, the developing countries would seem to be progressing faster.) Since most percentage type indicators are social (or structural) rather than economic and are positive rather than negative, it is easy to draw false conclusions from routine indicator data about comparative social and economic progress in developing and developed countries - conclusions that are merely reflections of the technical character of the indicators used. One has to ask to what extent conclusions about relative growth of developing and developed countries - or of social and economic aspects of development - are an artifact of the measurement technique. It may be noted that problems of comparing growth rates of developed and developing countries are much the same as problems of comparing growth rates of the same countries at different historical periods.

We can in fact measure change or progress in percentage type indicators in quite another way - by subtracting the initial percentage figure from the later figure and using the resulting difference in percentage points as the measure of change. By this approach, an increase from 10 per cent in literacy is an increase of 10 percentage points, and the decrease in illiteracy will be equally 10 percentage points. When this method is used, there is not such a one-sided and predictable difference in growth rates between developed and developing countries for the percentage type indicators. It is preferred to the method of measurement of change by percentages of percentages. Another approach to the comparative measurement of change will be considered below.

Correlation analysis

It has been mentioned above that the universe of countries is small and highly variegated. Correlations and other relations between variables may be quite different for different types of country, and universal functional relations through all stages of development cannot be assumed. 1/ Probabilistic "representative sampling" is not applied to countries in comparative studies. Data availability tends to be the main criterion of the sample of countries used. There is inherent likelihood of some degree of bias. The best that can be done is to try to ensure a sufficient number of major types (e.g., developing and developed) in any broad analysis, to apply correlations separately to use these major types where possible, and to eliminate some of the more obviously atypical countries from analysis.

Different correlations found in relation to different pairs of variables are not strictly comparable because the samples of countries having data on the two variables being correlated are usually not the same for the different pairs and cannot be made the same for any sizeable set of indicators without seriously reducing the sample size. This points to the need for caution in comparing indicators in terms of average correlation level. 2/ In general,

---

1/ For example, a relatively small expenditure of funds may produce large health effects in developing countries (spraying for malaria, innoculations against contagious diseases), while very large expenditures may be required to make progress against degenerative diseases common in developed countries.

2/ For example, GDP per capita at "purchasing power parity rates" has a higher correlation with most other development indicators than does the regular GDP per capita, and one may be inclined to consider it a better indicator for that reason. But there are only around 23 countries with 1970 data on the purchasing power parity indicator, as against 109 with data on the regular GDP per capita indicator. When the regular GDP indicator is used with the very limited sample of countries available for the purchasing power parity indicator (a sample that, while small in number, nevertheless covers a wide range), the correlations of the regular GDP indicator increase on the average, although still remaining below those of the purchasing power parity indicator in most cases. When rank order correlation is also used, however, the difference between the two indicators in correlation level tends to disappear almost entirely.

however, it appears that the size of the sample (using "sample" in the sense of an incomplete data set) does not affect the correlations as much as the presence or absence in it of certain countries having unusual or atypical features. Thus, the correlations between GNP per capita and life expectation for 81 developing countries may be found to be 0.78 but the addition of a single country, Libya, can cause the correlation to drop to 0.53, illustrating the fragility of cross-national correlations in the face of highly out-of-line data. As will be demonstrated below, the presence or absence of certain countries that are out of line with the rest of the bivariate distribution, not in the relationship of the two variables but in the level of the two variables (being substantially higher or lower in both variables), can also seriously affect correlations. 1/

The impact of contrasting indicator types and of curvilinearity on product moment correlations can be illustrated by reference to figures 8 and 9. The regular (product moment) correlation of physicians per 10,000 inhabitants with per capita GDP is 0.71; the correlation of inhabitants per physicaan with per capita GDP is only -0.42. (The rank order correlation in both cases is 0.88.) If the indicator physicians per 10,000 inhabitants is correlated with inhabitants per physician, the result is a relatively low -0.52 - although these two indicators are supposed to be describing the same objective phenomenon and are constructed out of the same basic figures. 2/ (Their rank order correlation is perfect.) There is a scale transformation involved in changing from one to the other, however, causing much greater curvilinearity in figure 9 and hence much lower correlation, in spite of the more ordered appearance.

1/ It is better to use correlations not involving such countries if the correlations are intended to show the degree of relationship between two variables for the great majority of countries. Thus Libya is not included in much of the correlational and other analytic work at UNRISD on developing countries; nor are Israel, Argentina, Venezuela and South Africa, which are sometimes classified with developing countries but are questionably so classified in view of their levels and characteristics. Hong Kong, Singapore and Puerto Rico are excluded from most over-all analyses (all countries, developed plus developing), because of their special situation as city-states (Hong Kong and Singapore) or unusual status (Puerto Rico). In Hong Kong and Singapore, for example, per capita foreign trade is much greater than the per capita GDP ($1,375 vs. $765 for Hong Kong in 1970, $1,935 vs. $945 for Singapore).

2/ Most internationally published tables use inhabitants per physician, others give physicians per 10,000 or 100,000 inhabitants, still other give both.

Table 4.2 shows the correlations of different technical forms of six different indicators, with each other and with four additional indicators. The results show not only the remarkable effect of change of technical form upon regular correlation but also the extent to which this effect varies with different correlated items. If it is desired to show a high correlation between infant mortality and GDP per capita, one should choose, not the regular infant mortality rate to correlate with GDP, but infant survivals divided by infant deaths in a given year. If a high correlation of school enrolment with GDP is to be demonstrated, one should choose, not the regular enrolment percentage, but number enrolled divided by number not enrolled. Does the relative number of telephones in a country have an association with agricultural productivity? If we measure the availability of telephones by inhabitants per telephone, the correlation is quite low: r=-0.31. If we measure it by telephones per 100,000 inhabitants, the correlation is very high : r = 0.91. On the other hand, inhabitants per telephone has a higher correlation with the regular infant mortality rate (r = 0.76) than does telephones per 100,000 inhabitants (r =-0.65). 1/

Which of the alternative forms of these indicators gives the true correlation? The answer is none, although the higher correlations are nearer the truth. The best way to approximate a true picture of the degree of association between two factors where different technical forms exist is to straighten out the curvilinear distributions and get the correlations from transformed data. The alternative is to use the rank order correlation, in spite of its obvious crudeness and its weakness in handling situations where some of the points are very close together or very far apart. 2/

1/ It should be noted that changing the technical form of an indicator may change its correlations from positive to negative. Inverse per capita indicators are negative indicators having negative correlations, except for correlations with other negative indicators (as in the case of inhabitants per telephone and the regular infant mortality. Of the two ratio type indicators derived from a percentage type indicator, one will be positive, the other negative, depending on the sign of the percentage type indicator.

2/ There is in fact, however, a striking similarity between rank order correlations and transformed data correlations, as shown in Chpater 5, lending support to the views of theoretical statisticians like Kendall who tend to look more favourably than others upon the use of rank order correlations in the case of data with the characteristics of development data.

Table 4.2

Correlations for different technical versions of the same indicator

		INFANT MORTALITY			COMB. ENROLMENT			ADULT MALE IN AGR			TELEPHONE		INH PER DR		ENERGY		EXPLI	LITER	AGRPR	STEEL
		V1	V2	V3	V1	V2	V3	V1	V2	V3	V1	V2	V1	V2	V1	V2	V1	V1	V1	V1
INFMOR	V1																			
	V2	1.00																		
	V3	-0.82	-0.80																	
COMENR	V1	-0.91	-0.91	0.75																
	V2	-0.81	-0.79	0.80	0.90															
	V3	0.72	0.75	-0.45	-0.81	-0.59														
AMLAGR	V1	0.87	0.86	-0.78	-0.90	-0.84	0.68													
	V2	0.66	0.69	-0.38	-0.69	-0.48	0.89	0.68												
	V3	-0.56	-0.55	0.63	0.64	0.79	-0.35	-0.72	-0.31											
TELEPH	V1	-0.65	-0.63	0.83	0.68	0.84	-0.40	-0.71	-0.34	0.74										
	V2	0.76	0.77	-0.52	-0.59	-0.44	0.59	0.59	0.78	-0.27	-0.32									
INH/DR	V1	0.65	0.68	-0.37	-0.67	-0.50	0.72	0.67	0.84	-0.30	-0.35	0.77								
	V2	-0.79	-0.78	0.64	0.79	0.76	-0.53	-0.81	-0.45	0.60	0.60	-0.46	-0.52							
ENERGY	V1	-0.69	-0.67	0.82	0.70	0.85	-0.41	-0.74	-0.36	0.74	0.88	-0.35	-0.39	0.71						
	V2	0.69	0.70	-0.45	-0.58	-0.42	0.60	0.60	0.86	-0.27	-0.29	0.87	0.84	-0.44	-0.33					
EXPLIF	V1	-0.96	-0.96	0.78	0.93	0.82	-0.73	-0.84	-0.62	0.56	0.64	-0.62	-0.66	0.83	0.67	-0.59				
LITER%	V1	-0.92	-0.91	0.74	0.92	0.82	-0.80	-0.80	-0.61	0.55	0.63	-0.52	-0.63	0.84	0.67	-0.53	0.93			
AGRPRO	V1	-0.67	-0.65	0.81	0.72	0.86	-0.38	-0.77	-0.32	0.80	0.92	-0.31	-0.33	0.67	0.85	-0.33	0.68	0.63		
STEELC	V1	-0.75	-0.73	0.86	0.73	0.84	-0.43	-0.79	-0.39	0.75	0.83	-0.39	-0.41	0.80	0.90	-0.37	0.73	0.71	0.82	
GDPPCC	V1	-0.72	-0.70	0.86	0.74	0.88	-0.44	-0.79	-0.39	0.80	0.93	-0.39	-0.42	0.71	0.91	-0.36	0.72	0.69	0.92	0.90

DEFINITION OF INDICATORS:

V1=VERSION 1 .... V2=VERSION 2 .... V3=VERSION 3

INFMOR V1=Infant mortality rate per 1,000 live births .... V2=V1/(1000-V1)... V3=(1000-V1)/V1

COMENR V1=Combined primary & secondary education enrolment as % of population 5-19 .... V2=V1/ (100-V1)... V3= (100-V1)/V1

AMLAGR V1=Adult male labour in agriculture(ISIC DIV.1) as % total adult male labour .... V2=V1/ (100-V1)... V3= (100-V1)/V1

TELEPH V1=Telephones per 100,000 population .... V2= 100000/V1

INH/DR V1=Inhabitants per physician .... V2= 100000/V1

ENERGY V1=Energy, apparent consumption per capita (in Kg of coal equivalent) revised .... V2= 100000/V1

EXPLIF V1=Expectation of life at birth - both sexes

LITER% V1=Literate as % of total population 15 and over

AGRPRO V1=Agricultural production per male agricultural worker (ISIC DIV. 1) at current prices (in purchasers' values), in U.S. dollars

STEELC V1=Steel, apparent consumption per capita (in kilograms)

GDPPCC V1=GDP per capita (in purchasers' values), at current prices in U.S. dollars

Since percentage type indicators tend to be of a social or structural nature and per capita type of an economic nature, there is marked curvilinearity in the relationships between indicators of these two types, with resulting lowered correlations. This gives the impression that economic factors are much more closely associated with each other than with social factors. When rank order correlations (or transformed data correlations) are used, however, the picture is changed. There is now only a limited difference on the average, between the correlations of economic with economic (or social with social) and the correlations of economic with social indicators,and there are various exception. Thus by rank order correlation agricultural productivity correlates 0.94 with combined school enrolment, 0.88 with life expectation, 0.88 with steel production and 0.84 with foreign trade. Energy correlates 0.95 with combined school enrolment, 0.94 with steel, 0.94 with manufacturing productivity, 0.88 with foreign trade. Combined school enrolment correlates 0.86 with literacy and 0.87 with life expectation.

The elongation effect

Even obviously bad indicators, like crude mortality rate, can have fairly high correlations, compared to correlations in other fields. This is because of the stretched out nature of the bivariate (or multivariate) distributions. The analogy was drawn earlier in this report with correlations between variables measuring child development. Two such variables (weight and vocabulary size) may have no correlation at all at a specific age (e.g., 10 years) but as additional age groups are added, the correlation builds up, because older children have both greater weight and greater vocabulary size than younger children. If children grew for 100 years or more, each child increasing his or her weight and vocabulary each year, the correlation would get very high, and would resemble the correlations found in socioeconomic development. 1/

1/ It is customary to speak of different measures of child development as being correlated with a third factor, age or time (apart from any other relations). Age, however, does not exist as a variable applying to countries in comparative socioeconomic development analysis. The only thing that can play an anchor role similar to age and to which one can relate changes in a specific development variable is the mass of other development variables. This problem will be taken up in Chapter 5.

The problem can be considered abstractly by imagining two clusters of, say, 10 or so points on a scatter diagram with coordinates X and Y, with no correlation between X values and Y values in either cluster but a fair distance between the two clusters. For the total set of points (the two clusters taken together), there can nevertheless be a good correlation which gets higher the farther way the clusters are from each other, until it eventually approaches perfect correlation. Even if one or both of the clusters has a negative correlation, the overall correlation can still be strongly positive. If there are three clusters more or less in line in any direction except horizontal or vertical, dropping the middle cluster will increase the product moment correlation. If non-correlated clusters are joined along a line in any direction except horizontal or vertical, dropping the middle points will increase the correlation as will adding more points (uncorrelated among themselves) at either end.

Let us take some illustrations now from development data. 1/ The correlation between GNP per capita and relative size of child population (1970 UNRISD data) is low, or non-existent, for developed and developing countries taken separately (r=-0.25 and +0.11 respectively). But it is quite high (-0.85) for the two groups taken together. If now all countries between $500 and $1,500 per capita GNP are eliminated, the over-all correlation (combining those below $500 with those above $1,500) moves up: from r=-0.85 to r=-0.90. With this elimination, the correlation for countries above $1,500 is: r =-0.10; countries below $500 GNP per capita show a surprising <u>positive</u> correlation of +0.40. 2/ Thus, in a combined set of developed and developing countries there can be very high correlation between two variables even though there is no correlation at all or correlation in the opposite direction when the developed or developing countries are taken separately. Elimination of the middle-income countries in general tends to increase over-all linear correlation between economic per capita type variables and social percentage type variables because their removal lengthens the distance between the centers of gravity of the more developed and less developed groups of countries (and possibly because of their position near the sharpest part of the bend of the curvilinear relationship).

1/ Illustrations that follow are taken from "Development Statistics: A Comment on Hicks and Streeten" by D.V. McGranahan, C. Richard and E. Pizarro, <u>World</u> Development, Vol. 9, No.4, pp. 389-397, 1981.

2/ This correlation is positive presumably because high infant and child mortality rates reduced the child population in the poorest of these countries while better-off countries closer to $500 were healthier but had not in the period prior to 1970 reduced their fertility.

Figures 10 and 11 show visually the effect of eliminating middle-income countries in the case of the relationship between GDP per capita and life expectation (UNRISD data). The over-all correlation (r) is 0.71. The angle between the two regression lines a and b in figure 10 represents (inversely) the level of correlation. In figure 11 countries between $500 and $1,500 have been eliminated. The over-all correlation is now 0.84, represented by a narrower angle between the two regression lines. For the countries below $500 taken separately the correlation is 0.48, for the middle income countries between $500 and $1,500 it is 0.33, for the highest income countries above $1,500 it is 0.20.

It follows from the above that the number of middle-level countries that happen to be present in a distribution relating economic and social development variables can have a considerable impact upon the over-all correlation: The fewer the middle level countries, the higher the correlations. At the same time, because of the "elongation effect", over-all correlation is increased when more countries are added at one or the other extreme of a bivariate distribution, other things remaining equal. 1/ In the case of practically all variables that can be regarded as "development indicators", average correlation with other such variables is substantially higher for the over-all range of countries than for a limited sector of it such as developed or developing countries taken separately. (The technically inferior health indicator "crude death rate" is an exception.) In general, correlations between development indicators have significantly increased in recent years (1960 to 1970), and a possible explanation is that values have increased at the highest levels but have remained relatively constant at the lowest levels (as a result of new countries coming in with statistics), thus elongating the distribution.

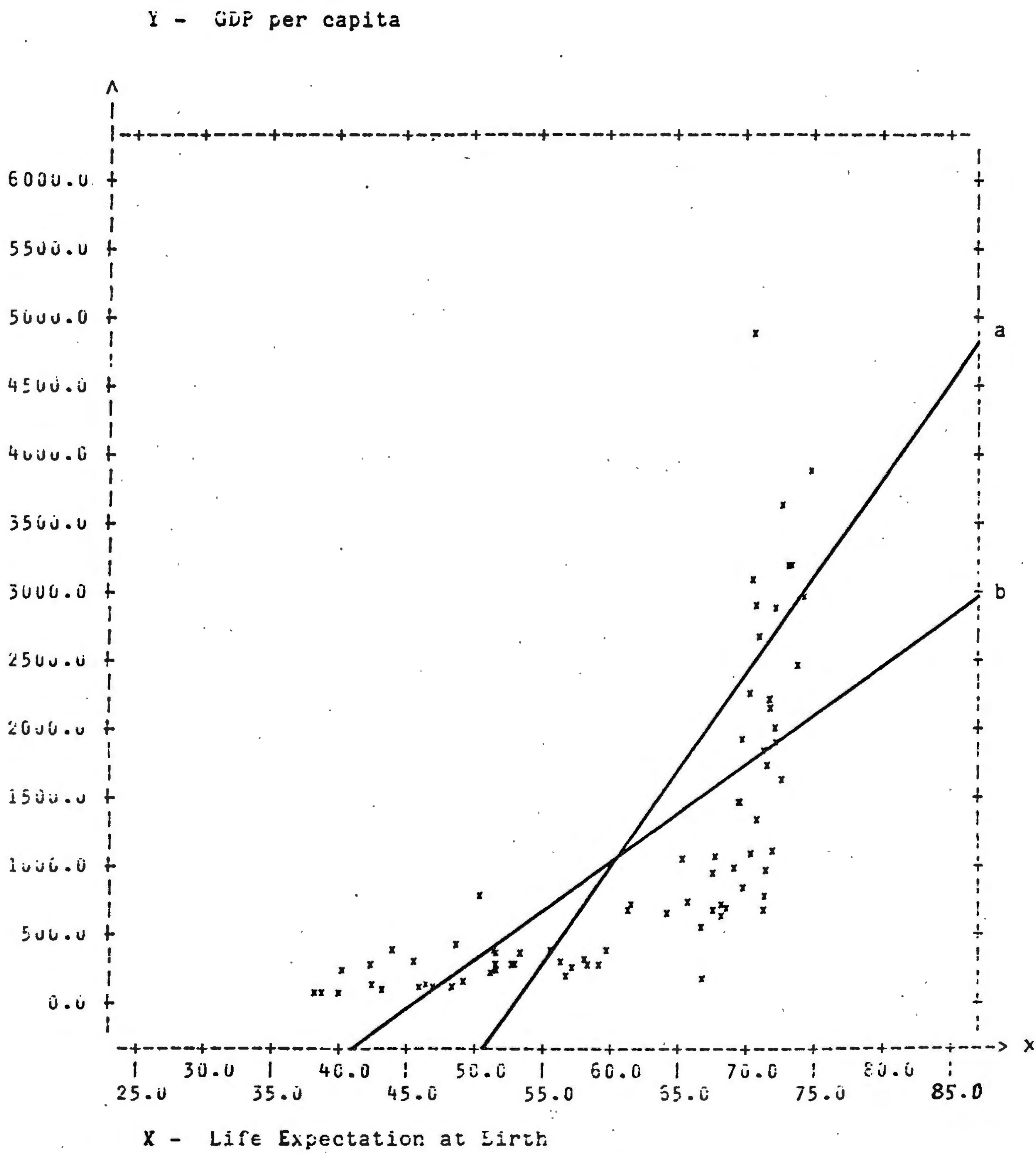
Even the addition of a single country, with relatively high values on two variables, to a limited groups of developing countries (showing little or no correlation between the two variables) can greatly increase the correlation, or create one where non-existent. Thus a group of 20 developing countries may show little correlation (r =-0.23) between GNP per capita and percentage of

---

1/ It is assumed in this paragraph that the points dropped are not significantly more in line with the distribution than the points remaining, and that the points added to the extremes are not significantly more out of line than the existing points (do not cause a funneling out at the extreme greater than the funneling out of the two regression lines).

FIGURE No. 4.10

75 Countries



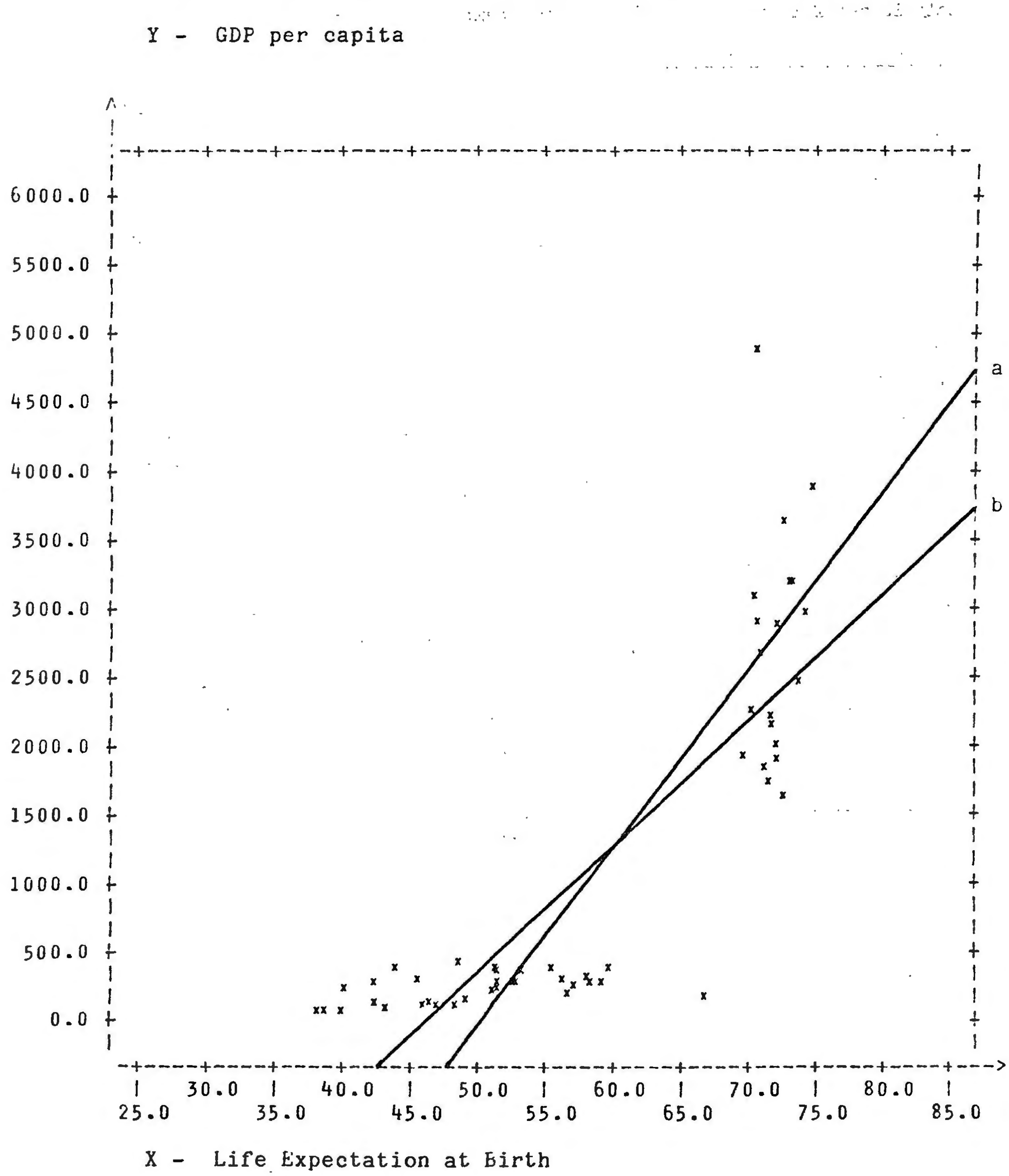
a = Regression line of X on Y r = 0.71

b = Regression line of Y on X Source: UNRISD Data bank

FIGURE No. 4.11

52 COUNTRIES

(Countries between $500 and $1500 GDP per capita excluded)



a = Regression line of X on Y   r = 0.84

b = Regression line of Y on X   Source: UNRISD Data bank dwellings with piped water 1/ but the addition of a single country, Israel (classified by some international sources as developing), can raise the correlation to -0.54 (UNRISD data). This indicates the need for careful consideration of the influence of the definition of developing countries upon analysis and for use of relatively large samples (over 25). 2/

Correlation and determination

The need for caution in drawing conclusions about causal relations from correlations has long been recognized but has not always guided practice in the analysis of development data where the need is particularly compelling. "Coefficients of determination" are frequently used in simple or multiple correlation analysis to indicate the proportion of the variance in one variable that may be explained by another variable or by a group of variables. 3/

The main pitfalls facing attempts to draw causal conclusions from correlations of development variables, including correlations between different types of social and economic indicators, may be summarized briefly as follow:

1. The causal influences between the different variables are but vaguely known on independent grounds - or controversially known. They are complex and multiple and may differ from one level to another. Controls are difficult to establish.

2. At the same time, any one factor like health or agricultural productivity (as measured by life expectation and agricultural production per male agricultural worker) may have high correlations with a large number of other items (in these two cases, at least 15 or 20 correlations of 0.80 or more); and these correlations still do not cover important factors for which adequate indicators or data are lacking and for which correlations cannot therefore be computed.

1/ Using unadjusted and relatively non-comparable data for the piped water variable - see Chapter 2, page

2/ This influence of one or two countries (with relatively high values) which may or may not happen to be included in the definition, along with the impact of just one or two out-of-line countries which may or may not happen to have data on a particular pair of variables, raises questions about the meaningfulness of conventional statistical tests of significance (at the 5 per cent level, etc.) widely applied in the development literature to correlations involving relatively small numbers of countries.

3/ The coefficient of determination is the square of the correlation coefficient. Thus if the correlation between X and Y is 0.71 and Y is expressed as a linear regression on X, then X explains one half the variance of Y.

3. Cross-national socioeconomic development correlations, like child development correlations or other correlations involving time, can be quite high without any direct causal relationship whatsoever between the correlated variables; items with high correlation in a wide sample of countries may not be correlated at all among countries at the same development level.

4. Correlations between indicators are used (hopefully) to show relations between real factors but reflect to an important extent the technical form of the indicator.

5. The (accidental or purposive) presence or absence of just one or two countries in a correlation can greatly influence the level of correlation.

Regression Analysis

Regression is used extensively in the analysis of development data to determine the level of a variable that goes with a given level of another variable, to estimate missing values and make projections from one variable or set of variables to another; it is also used to determine the amount of impact that a given increase of one item will have upon another. The concept of "regression to the mean" or to "mediocrity" was originated in the late 19th century by Galton in his work in eugenics and illustrated by the well-known example of heights of fathers and of sons: the sons of fathers who deviate x amount from the average height of all fathers themselves deviate by less than x amount from the average height of all sons.

This systematic regression to the mean is associated with the existence of normal bell-shaped frequency distribution (of the heights both of fathers and of sons). If there is change of relative height from father to son, the probability is greater that the change at higher or lower levels will be in the direction of the greater frequency. The original regression lines were lines that were drawn through averages of intervals. 1/ Figure 12 illustrates.

1/ Thus the average height of fathers between 74.5 and 75.5 inches, between 73.5 inches and so on is determined, also the average height of the sons of the fathers who are in each interval, yielding a point for each interval. This gives one regression line. The other line is obtained by averaging the heights of sons at successive intervals and of their respective fathers. In the case of development data, however, with stretched out distributions and small numbers of observations, there are not enough cases in successive intervals on a scale to provide a meaningful array, so it is necessary to proceed, not by taking equal intervals on a scale, but by taking equal numbers of ranked individuals (countries): e.g., the ten or twelve highest on X, the next ten or twelve, etc, averaging the X and Y values for each group to get the points for the regression line of Y on X; then the ten or twelve highest on Y, the next ten or twelve etc. averaging the X and Y values for each group to get points for the regression line of X on Y. This somewhat different method of compilation should not change the nature of the lines.

It can be demonstrated mathematically that under the conditions that apply to data like heights of fathers and sons, one can get essentially the same two lines by using a formula that minimizes the sum of the squares of the deviations (or "errors" as they unfortunately tend to be called in statistics) of the points from the line, the one line minimizing the squares of the deviations on Y (taking X as given), the other minimizing the squares of the deviations on X (taking Y as given). This is the modern approach and the use of the "least squares" method to get a line is basically what is meant by "regression analysis".

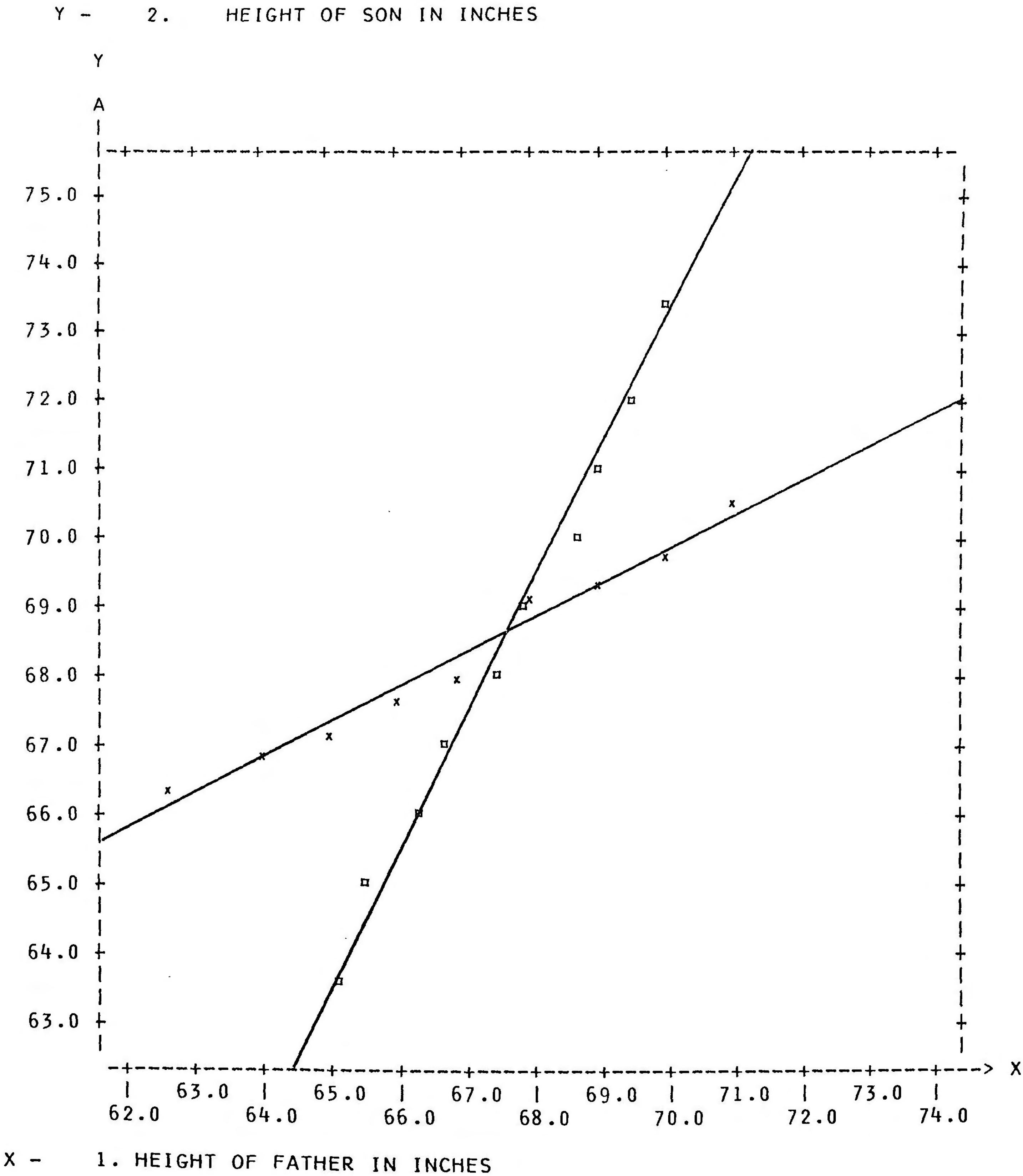
The regression lines obtained by the least squares method are applied today to all kinds of data arrays, not just to those where there is a normal bell-shaped frequency distribution of the values on each of the separate variables. Nevertheless, these lines still imply a regression to the mean in the sense that the individuals (or countries) with the highest value on X will have lower values on Y than those individuals with the highest values on Y; the next highest group on X will tend to average lower Y values than the second highest group on Y, and so on to the group around the mean. In the case of development data, however, such regression does not occur. The least-squares lines have little resemblance to the lines obtained by the averaging method which do not show regression to be mean.

In figures 13, 14, 15, covering three indicator pairs (expectation of life and combined school enrolment, expectation of life and GDP per capita, GDP per capita and steel consumption per capita), countries have been divided into five or more equal groups (quintiles, etc.), according to their rank, first on the X indicator, and then separately by their rank on the Y indicator. For the groups ranked on X, the average X value and the average Y value for each group have been determined and the resulting points joined by a line. The same procedure has been repeated for the groups ranked on Y. These pairs of lines in figures 13, 14, 15 should presumably look like the lines in figure 12, or in figure 10, if regression applied. It is obvious that they do not. The lines criss-cross back and forth. The disparity is not due simply to the curvilinarity of the bivariate distribution. If we straighten out the curves and use transformed data, as shown in figures 16, 17, 18, the averages lines still cross and do not much resemble least squares regression lines, except perhaps at the extremes of the distribution.

Figure 4.12

Regression lines by Interval averages for
Heights of Fathers and Sons

By interval average of fathers' heights (xxx)
By interval average of sons' heights (¤¤¤)



Adapted from Yule and Kendall,
An Introduction to the Theory of Statistics,
14th ed., Griffen and Co. London, 1950, page 218

Figure 4.13

Combined Primary & Secondary Enrolment as % of 5-19 population and Expectation of life at birth

Values by ranked, equal-size groups

Ranked groups on X(▫)
Ranked groups on Y(×)

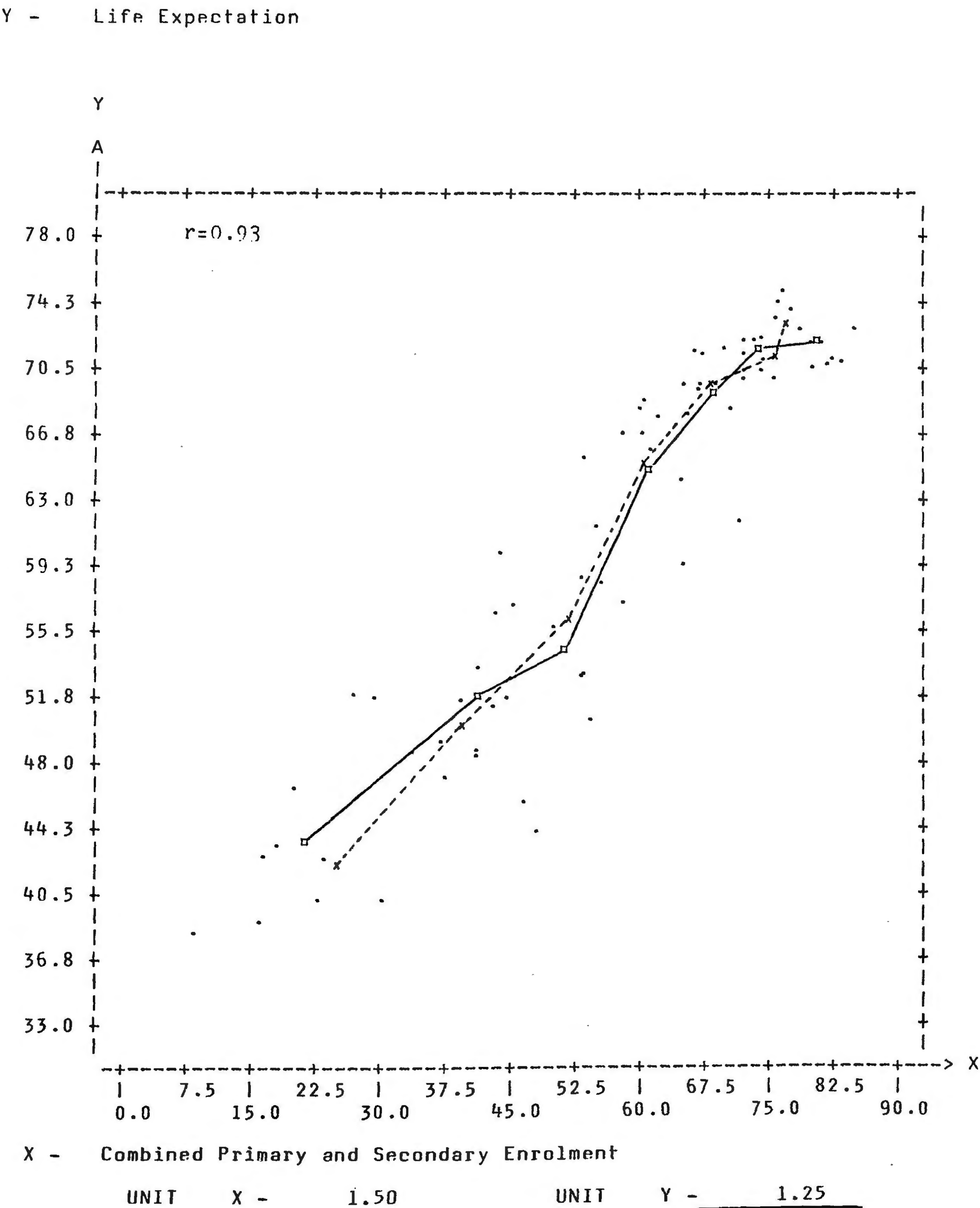


Figure 4.14

Expectation of life at birth and Gross Domestic Product per capita

Values by ranked, equal-size groups

Ranked groups on X(▫)
Ranked groups on Y(x)

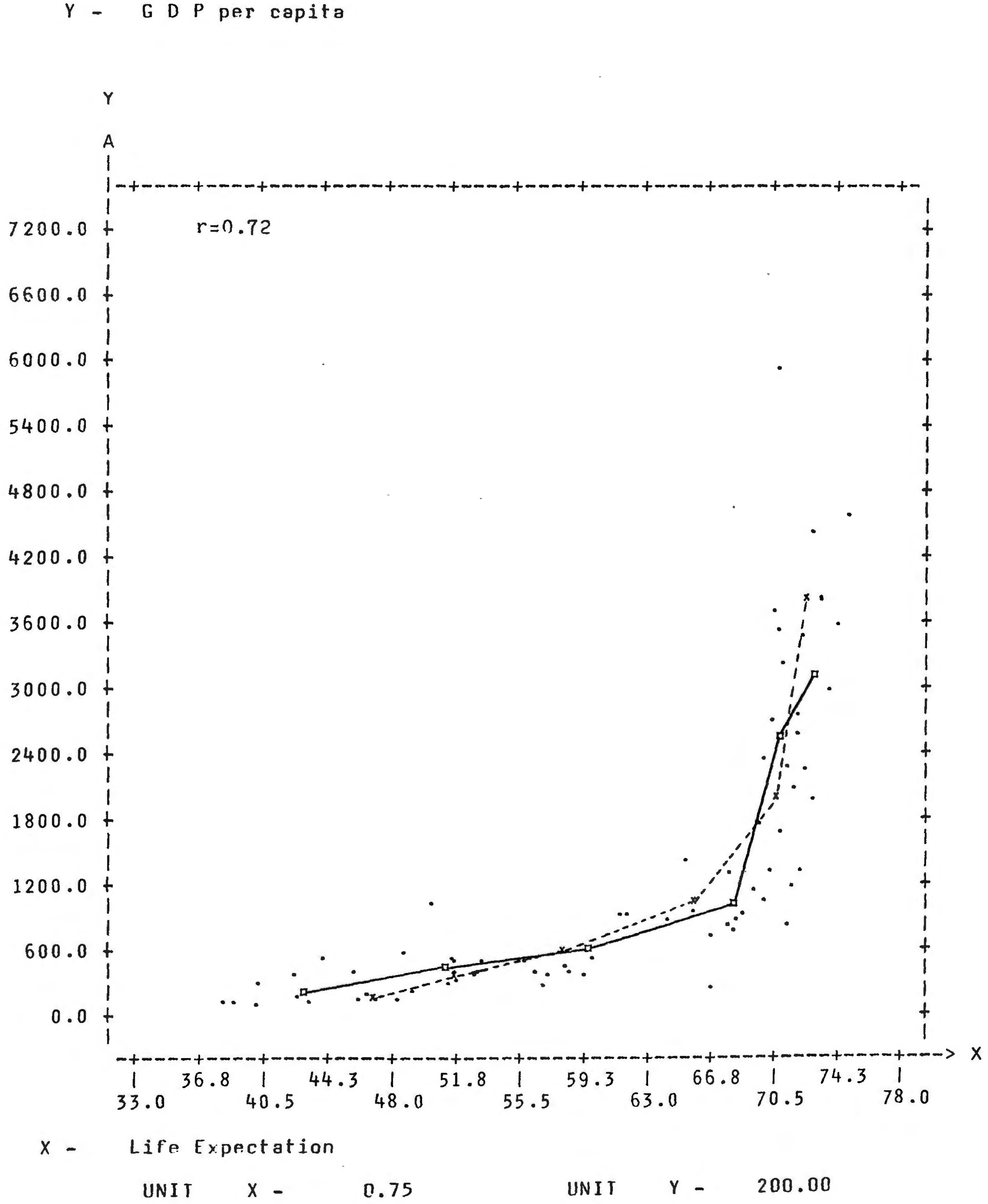


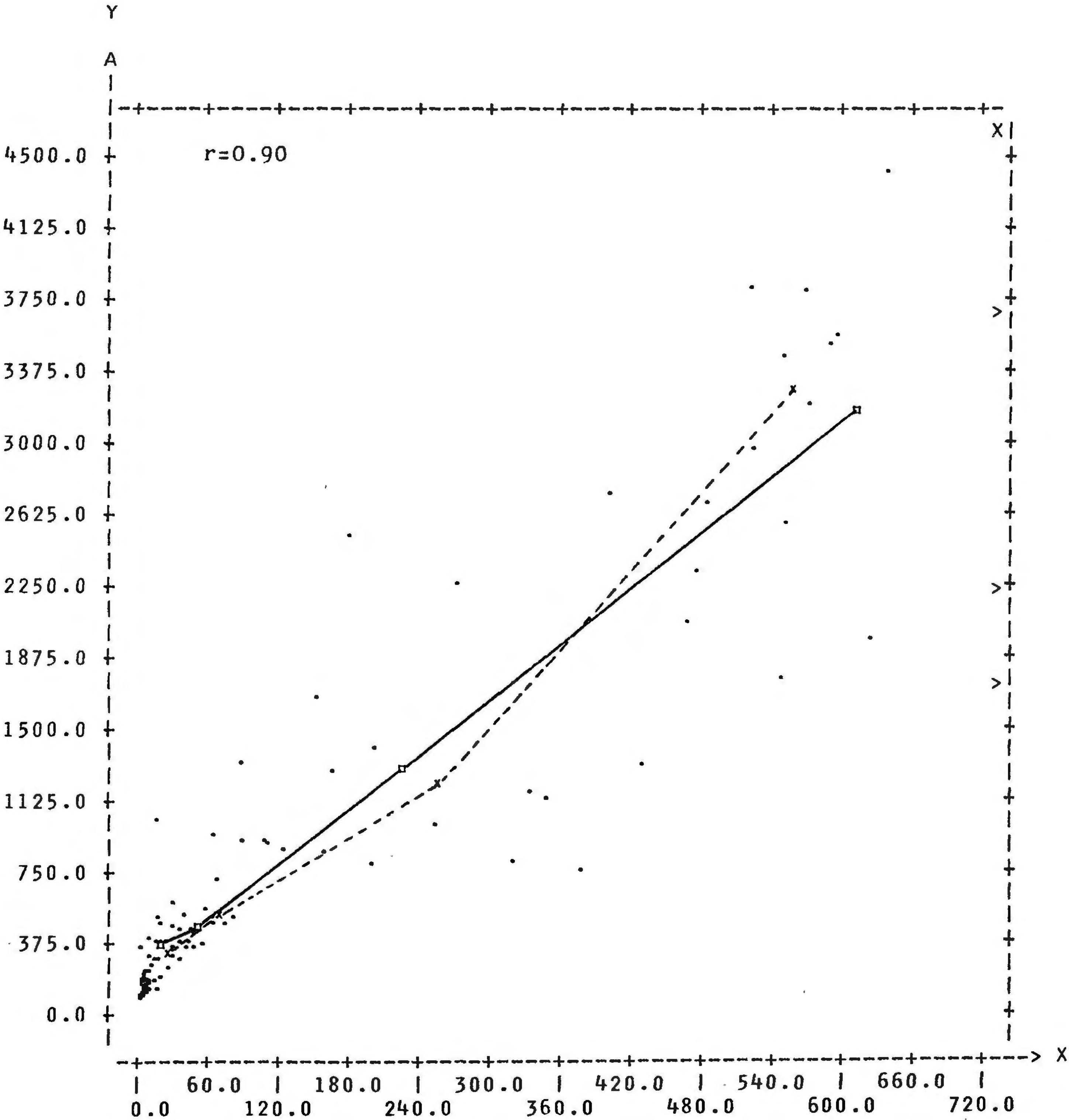
Figure 4.15

Steel Consumption per capita and Gross Domestic Product per capita

Values by ranked, equal-size groups

Ranked groups on X(¤)
Ranked groups on Y(x)

Y - G D P per capita



X - Steel Consumption per capita

UNIT X - 12.00 UNIT Y - 125.00

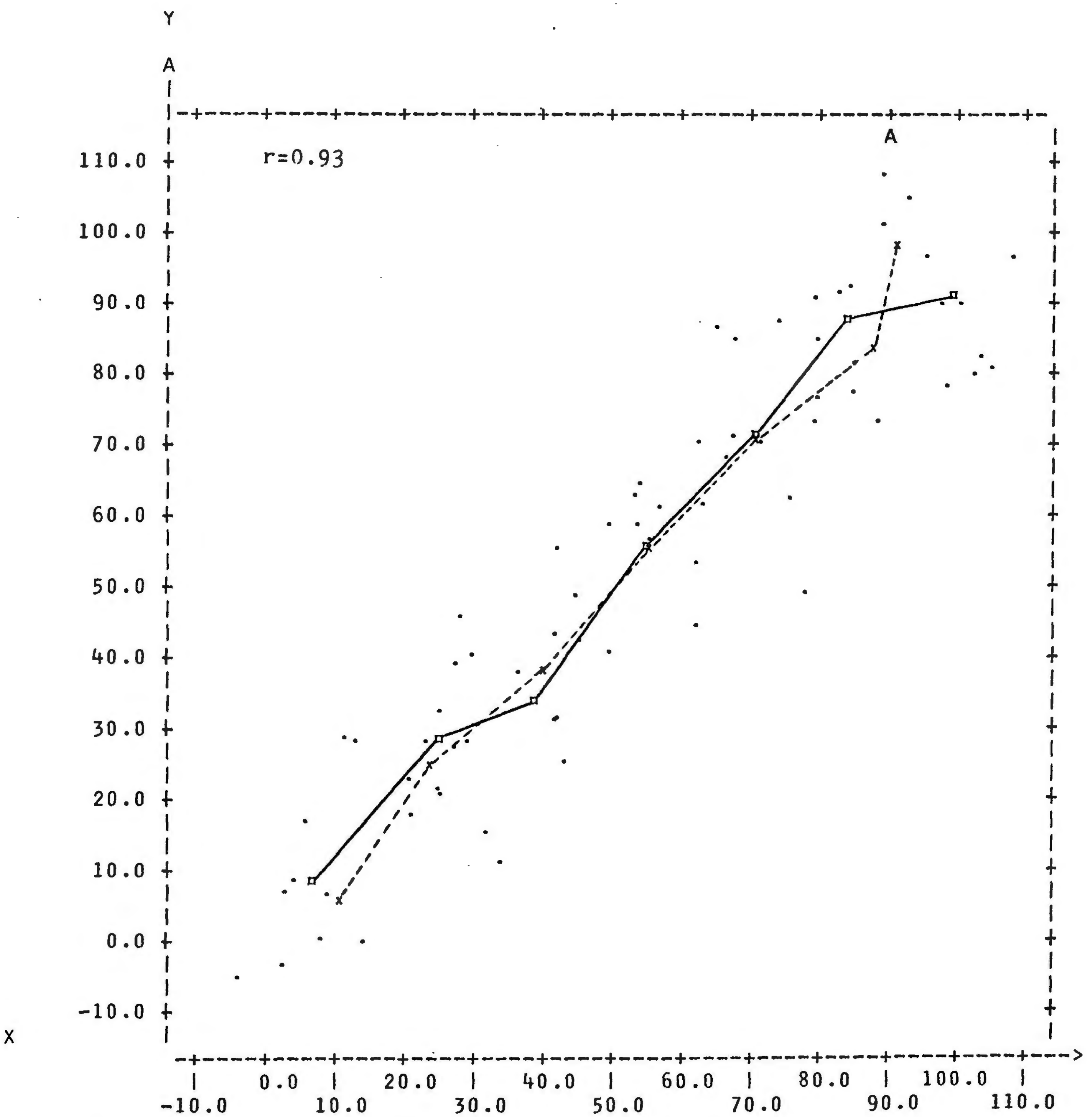
Figure 4.16

Combined Primary & Secondary Enrolment as % of 5-19 population and Expectation of life at birth

Transformed values by ranked, equal-size groups

Ranked groups on X(¤)
Ranked groups on Y(×)

Y - Life Expectation



X - Combined Primary and Secondary Enrolment

UNIT X - 2.00 UNIT Y - 3.33

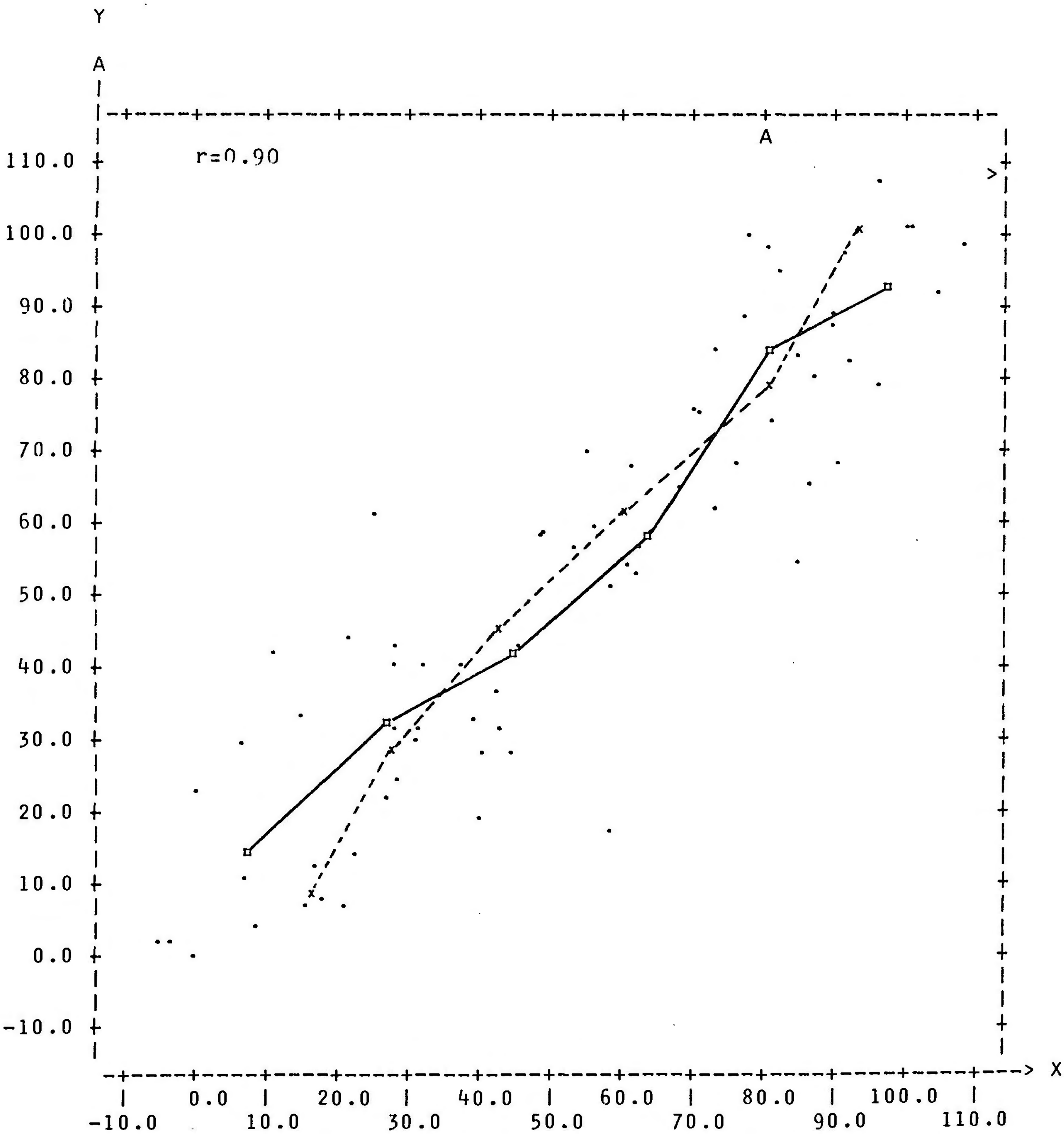
Figure 4.17

Expectation of life at birth and Gross Domestic Product per capita

Transformed values by ranked, equal-size groups

Ranked groups on X(¤)
Ranked groups on Y(×)

Y - G D P per capita



X - Life Expectation

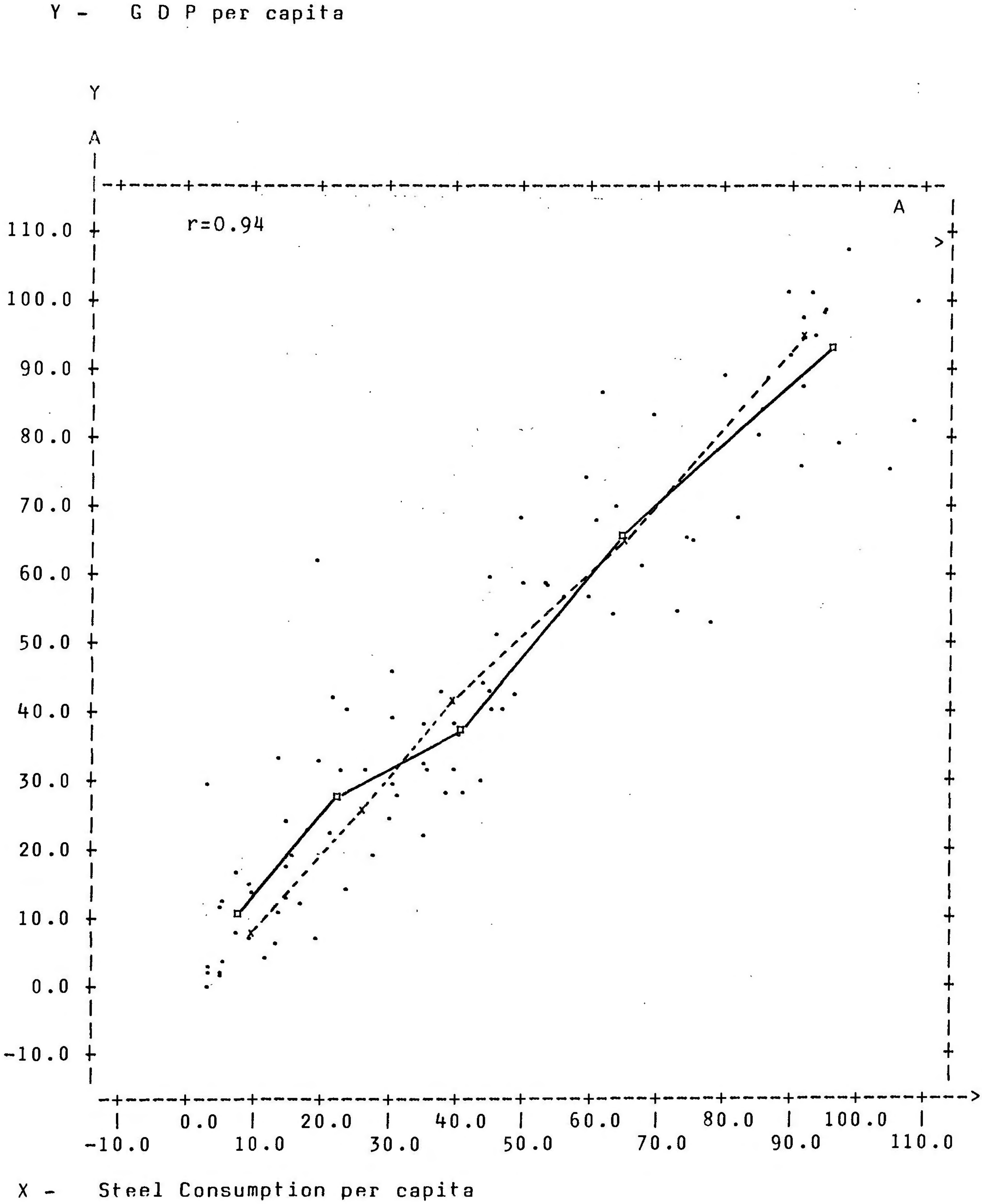
UNIT X - 2.00 UNIT Y - 3.33

Figure 4.18

Steel Consumption per capita and Gross Domestic Product per capita

Transformed values by ranked, equal-size groups

Ranked groups on X(¤)
Ranked groups on Y(×)



UNIT X - 2.00 UNIT Y - 3.33

Various conditions have been identified by mathematical statisticians and econometricians for the proper use of least squares regression. 1/ Most of these conditions are not met by development data, or there is no information as to whether they are met.

There are in fact many kinds of lines that can be drawn through a bivariate distribution. A strong case has been made for a line that minimizes, not the squares of the deviations from it, but the simple arithmetic or "absolute"

1/ E. Malinvaud, in Méthodes Statistiques de l'Econométrie (Paris: Dunod, 1969), gives six conditions that must be met for the proper use of regression in a linear model. If y is the dependent variable, x the independent variable and e represents the deviations, we can write:

$$y = ax + b + e$$

where a is the regression coefficient and b the intercept. The six conditions set forth by Malinvaud for the use of regression are then as follows:

1. Variables y_t and x_t $(t = 1......T)$ represent numerical values observed without error. Variable y_t is random and satisfies: $y_t = ax_t + b + e_t$ where a and b are numerical coefficients and e_t is a random variable, the expected value of which is nil, whatever the value of x.
2. Homoscedasticity.
   The error e_t follows a distribution independent of t and x_θ $(\theta = 1......T)$. It has a variance σ^2
3. Independence of the observations.
   The errors e_t and e_θ where θ, $t = (1......T)$ relative to two different observations t and θ are independent of one another.
4. Normality.
   The error e_t follows a normal law.
5. Exogeneous variables.
   When T increases indefinitely, the series of the exogeneous variable x_t $(t= 1, 2 ......)$ is such that its mean
   $$\bar{x} = \frac{1}{T} \Sigma x_t \text{ and its}$$
   quadratic deviation mean $\frac{1}{T} \Sigma (x_t - \bar{x})^2$ tend to approximate the finite limits of x_o and S^2 respectively.
6. There is no information on the two coefficients a and b except for the sample observed.

deviations. 1/ Such a line (called the minimum sum of absolute error or "MSAE line") does away with the excessive sensitivity of regression lines to points that lie outside the main stream of points ("outliers"). We shall come back to this line in the next chapter.

It is commonly assumed that the least squares regression line of Y on X is the proper line to use (the line that is the best estimator of the true line through a distribution and that will best predict future values) when Y is dependent on X and the deviations are on Y in a random normal distribution along the line. In practice, the dependency and the normal distribution of deviations are very often not known but "assumed". This is generally the case with applications to development data. The question of our knowledge of dependency or causality has been discussed above. As far as the normal distribution of deviations is concerned, examination of different sets of real points gives little reason to assume that it does exist in multivariate distributions of development data; 2/ and the curvilinearity of the bivariate or other multi-

---

1/ See John C. Wiginton, "MSAE estimation: An alternative approach to regression analysis for economic forecasting applications", Applied Economics, 1972, 4, 11-21. For use of MSAE in population analysis see Andrei Rogers, Matrix Analysis of International Growth and Distribution, University of California Press: Berkeley, 1968, Ch. 4. In an elaborate set of sampling experiments comparing MSAE and least squares estimators and following the Monte Carlo techniques, R. Blattberg and Thomas Sargent have shown that with symmetric stable function distributions defined by the log characteristic function

$$(\log \Phi_u (t) = i\delta t - |\sigma t|^{\alpha}).$$

the MSAE estimators of the slope "outperformed" the least squares estimators until (α) reached the value of 1.7, and the opposite was true for $1.7<\alpha\leq 2$ (for $\alpha=2$ the distribution is normal and for $\alpha=1$ it is the Cauchy distribution). See their article, "Regression with non-Gaussian stable disturbances: Some sampling results", Econometrica, Vol. 39, No.3, (May 1971).

2/ In view of the small number of countries that we can hope to get for a comparative analysis, it is not possible to take into account the "Central Limit Theorem" and assume with confidence that in a larger universe the deviations are going to follow a Gaussian law.

variate distributions will tend to submerge it if it does. 1/

There are, however, some cases where dependency would seem fairly clear (e.g., the dependency of literacy on schooling) and the possibility of random normal deviations cannot be excluded. Is it really true, however, that even when Y is clearly dependent on X and the deviations of Y on the Y regression line are random normal, then the Y regression line is necessarily the best line and the best estimator of the true line for the distribution? 2/

A Monte Carlo type test suggests that this may not be the case and that another formulation is called for. 3/ A straight line is taken with a series of points on it (say 45) in a coordinate system of X and Y and random deviations on the Y axis with given variance are generated by a computer and applied to

---

1/ A theoretical example will illustrate this statement. If 50 or 60 points are placed at equal distance on a straight line bisecting an X and Y coordinate system, the correlation will be perfect and the two regression lines will pass through all the points. If we transform this line into an exponential curve by a formula $Y = 2^x$, the regression line of Y on X will pass through the theoretical exponential curve in much the same way that the regression line of Y on X passes through the somewhat similar exponential curve in figure 10 (relating life expectation and GDP per capita). There are now, however, extensive deviations of the points, created by the curvilinearity alone. Furthermore, the distribution of these deviations created by curvilinearity shows no resemblance to a normal frequency distribution. (The greatest frequency is in deviations of relatively large size, below the line in this case, not of small deviations close to the line.)

2/ The analogy is sometimes made to a physical force A striking an object B - the greater A is, the greater the effect on B, but other random factors also affect B in a positive and negative manner, averaging out to zero or close to zero over a substantial number of observations.

3/ The test was carried out at UNRISD some years ago and is described in more detail in UNRISD Research Notes, No. 4, Geneva, 1974.

to these points, so that they are no longer on the straight line but distributed along it at random distances. If the regression lines of Y on X and of X on Y are now tested, it is found that the Y regression line approximates the original true line better (in slope and intercept) than does the X regression line in 95 out of 100 cases. This is consistent with the theory. But if horizontal cuts on the distribution are made by dropping the 10 points with the highest Y values and the 10 points with the lowest Y values, which line will now do better? The theory would require us to answer that the Y line will still do better, though not quite so well because of the reduced number of points. This answer is completely wrong. The X regression line now does better than the Y line more than 90 times out of 100.

A random distribution on Y can equally be a random distribution on X. The important point is not whether the deviations are generated on Y or X but how the cut-offs of the distribution are established, which variables are limited at the extremes. If random deviations are generated on Y, then the X values are held within limits at the two extremes and the Y regression line fits. But if top and bottom Y values are dropped, so that the distribution is cut horizontally, then the X values have the variance at the extremes and the X regression line fits. The bivariate distribution is reshaped and another line provides a better fit. In between the extremes, one can not tell in this case whether the deviations are deviations on Y or on X, unless the original points are known and it makes no difference to the results. (In practice, with development data, the deviations are likely to be on X and/or Y, varying from point to point.)

Distributions can thus be cut or shaped in different ways to fit different lines. (Addition of one or two "outliers" will make the MSAE lines do best.) It also follows that even where Y is clearly dependent on X, the regression line of X on Y (where X is supposed to be the "dependent" variable) can be the proper line to put through the distribution; this will be true in situations where the limits are placed on Y. 1/ The adequacy of lines to fit bivariate

---

1/ The point can be illustrated by an example of relationship that fulfils quite exceptionally the requirements of regression theory: the relationship between mental and chronological age of school children (from similar socio-economic and cultural background). Mental age by its very definition (the average level of performance on tests at given chronological ages) is dependent on chronological age, except for random influences which may be considered to cause deviations in the mental age variable, while chronological age may be considered to be given without error or deviation.

(Cont. on next page)

distributions such as those found in development data basically depends therefore not on what the causal relations are, but on the technical characteristics of the distributions at their extremes, which may or may not be related to causality.

In much analytic work using regression there is in fact a restriction on the range of one or other variable: for example, in a regression on time a specific range of years is taken as the X variable, while the "dependent" variable is allowed to vary at the extremes. But in the case of development data, the samples of countries on which analyses are carried out tend not to have a clear-cut and consistent restriction on the extremes of either variable. One takes the countries that are available, as a rule.

In the following chapter we shall be concerned with practical procedures of curve-fitting in relation to international development data. In this connection we shall be concerned also with a method of telling whether a point, which is out of line, is out of line primarily on X or on Y (whether, for example, a country that has a much higher level of education than would be expected for its level of health is out of line in education or in health). This means trying to identify which variable is the source of the deviation of each particular point. This is an essential part of a method used to draw analytic lines and establish a kind of analysis called "correspondence analysis".

(Cont. of footnote from last page)

Yet these facts are not really relevant to the correct analysis of a given set of data relating chronological and mental age. For children in a particular school system, the important thing is the kind of limitation placed on attendance; or otherwise stated, the kind of sample that the children in the particular school system represent. If the school system strictly follows a chronological age criterion (say 5-15 years) in admitting and graduating students, then the regression line whereby mental age is "dependent" on chronological age will be correct. But if it follows strictly a performance criterion (say mental ages 5-15), then the regression line whereby chronological age is the "dependent" variable will be correct - in spite of the fact that chronological age is not actually dependent on mental age and is known without error. (If both chronological and mental age criteria are applied to admittance and graduation of students, then for the resulting sample neither regression line will do very well in approximating the true one-to-one relation between chronological and mental age.)

Problems and limitations of regression analysis (as well as of correlation analysis) go beyond applications to development data. They have been widely recognized and various new methods of statistical analysis have been devised in recent years. Among the interesting new methods, for example, are "robust regression", applications of linear programming, path analysis and various "new" parametric techniques. A review of these methods suggests, however, that all of them, like simple regression, involve assumptions that are not tenable in relation to development data. Often this is because they are based on probabilistic models which make assumptions about distributions that are not evidently valid in relation to development data distributions. The problem is to work out a method or set of methods suitable to the nature of the data.